UNIVERSITI TEKNOLOGI MARA

ENHANCED CRYPTOCURRENCY PRICE PREDICTION USING INTEGRATED SENTIMENT AND PRICE FEATURES

NUR AZMINA BINTI MOHAMAD ZAMANI

PhD

JUNE 2026

UNIVERSITI TEKNOLOGI MARA

ENHANCED CRYPTOCURRENCY PRICE PREDICTION USING INTEGRATED SENTIMENT AND PRICE FEATURES

NUR AZMINA BINTI MOHAMAD ZAMANI

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science)

Faculty of Computer and Mathematical Sciences

JUNE 2026

CONFIRMATION BY PANEL OF EXAMINERS

I certify that a Panel of Examiners has met on 31 March 2026 to conduct the final examination of Nur Azmina Binti Mohamad Zamani on her Doctors of Philosophy thesis entitled "Enhanced Cryptocurrency Price Prediction Using Integrated Sentiment and Price Features" in accordance with Universiti Teknologi MARA Act 1976 (Akta 173). The Panel of Examiner recommends that the student be awarded the relevant degree. The Panel of Examiners was as follows:

Mohd Zaki Awang Chek, PhD
Associate Professor
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
(Chairman)

Suraya Masrom, PhD
Associate Professor
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
(Internal Examiner)

Asadullah Syah, PhD
Professor
Kulliyyah of Information and Communication Technology
IIUM Gombak Campus
(External Examiner)

PROFESSOR DR HJH ZURAEDA IBRAHIM
Dean
Institute of Postgraduates Studies
Universiti Teknologi MARA

Date: 3 June 2026

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Nur Azmina Binti Mohamad Zamani
Student ID. No.	:	2023906503
Programme	:	Doctor of Philosophy (Computer Science) – CDCS950
Faculty	:	Computer and Mathematical Sciences
Thesis Title	:	Enhanced Cryptocurrency Price Prediction Using Integrated Sentiment and Price Features

Signature of Student	:	……..
Date	:	June 2026

ABSTRACT

In the era of advanced technologies, digital currency such as cryptocurrency has attracted immense interests among traders and investors. The volatility and drastic shifts in cryptocurrency market make price predictions difficult. Thus, experiments were conducted to find the most optimised results. Majority findings have found that sentiment data improves cryptocurrency price prediction performance. But there exists a gap in research that discovers sentiment data for cryptocurrency price prediction for some languages (non-English) are not catered for. Therefore, the research aims to enhance the cryptocurrency price prediction performance and to extract sentiment features from online news and Twitter in English and Malay language, then builds a sentiment deep learning model integrating self-attention mechanism to investigate the effect of sentiment features on cryptocurrency price prediction. The methodology consists of three main phases, which are, data collection, feature extraction, and price prediction. Two corpora were constructed which are price and sentiment corpora in the first phase. During feature extraction phase, price features (open, high, low, close, and trading volume) were used directly, while text pre-processing was performed on the eight sentiment datasets. Sentiment data were then being annotated manually and automatically to be fed into Generalized Autoregressive Pre-training for Language Understanding (XLNet) model. Finally, the sentiment features are aggregated with the cryptocurrency price features and fed into the multi-layered bidirectional deep learning model with attention to predict the cryptocurrency closing price of the next day. The experimental results show that integrating sentiment features into deep learning models enhances cryptocurrency price prediction accuracy, with the Bi-GCL-A model outperforming traditional methods by leveraging self-attention mechanisms with an adjusted r^2 of between 0.907 and 0.918, with 2.4% improvement. Moreover, there is an improvement of 0.130 (Bitcoin) and 0.127 (Ethereum) in average adjusted r^2 for price prediction with price and sentiment features in comparison to price prediction with price features only. The research contributes to improving cryptocurrency price prediction performance while addressing the underrepresentation of the Malay language in Natural Language Processing (NLP) to make informed decision-making in optimising cryptocurrency trading profits.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my supervisor, Assoc. Prof. Ts. Dr Norhaslinda Kamaruddin, for the unwavering support, invaluable guidance, and encouragement throughout my PhD journey. Her expertise, patience, and insightful feedback have been instrumental in shaping this research, and I am truly grateful for the opportunity to have worked under her mentorship. I am also thankful to my co-supervisor, Dr. Sharifah Aliman who has supported and eased my PhD study.

I am also immensely thankful to my parents and my husband, whose unconditional love, sacrifices, and constant encouragement have been the foundation of my academic pursuits. Their belief in me has been a source of strength and motivation, pushing me to persevere even in the most challenging times.

To my family, thank you for the endless support, understanding, and patience. The words of encouragement and unwavering faith in me have been a source of comfort throughout this journey.

To my friends, who have been my pillars of strength, thank you for always being there whether it was to celebrate milestones, lend a listening ear, or provide much-needed laughter and distraction. Your companionship has made this journey all the more meaningful.

Finally, I would like to extend my gratitude to everyone who has contributed, directly or indirectly, to this research. The support and encouragement have played a crucial role in this achievement, and I am forever grateful.

TABLE OF CONTENTS

		Page
CONFIRMATION BY PANEL OF EXAMINERS		**ii**
AUTHOR'S DECLARATION		**iii**
ABSTRACT		**iv**
ACKNOWLEDGEMENT		**v**
TABLE OF CONTENTS		**vi**
LIST OF TABLES		**ix**
LIST OF FIGURES		**xii**
LIST OF SYMBOLS		**xiv**
LIST OF ABBREVIATIONS		**xv**
CHAPTER 1 INTRODUCTION		**1**
1.1	Introduction	1
1.2	Research Background	1
1.3	Problem Statement	6
1.4	Research Questions	8
1.5	Research Objectives	9
1.6	Scope and Limitation of Study	9
1.7	Significance of Study	10
1.8	Thesis Outline	11
CHAPTER 2 LITERATURE REVIEW		**13**
2.1	Introduction	13
2.2	Cryptocurrency	13
2.3	Cryptocurrency Market	15
2.4	Previous Works on Cryptocurrency Price Prediction	18
2.5	Types of Data Used in Cryptocurrency Price Prediction	23
	2.5.1 Price Data	23
	2.5.2 Sentiment Data	28
2.6	Data Annotation	30
2.7	Feature Extraction	34

2.7.1 NLP without Word embedding 36
2.7.2 NLP with Word embedding 43
2.7.3 Types of Sentiment Features Extracted 49
2.8 Classification for Cryptocurrency Price Prediction 52
2.8.1 Price Prediction Classification without Sentiment Features 53
2.8.2 Price Prediction Classification with Sentiment Features 56
2.9 Regression for Cryptocurrency Price Prediction 62
2.9.1 Price Prediction Regression without Sentiment Features 63
2.9.2 Price Prediction Regression with Sentiment Features 68
2.10 Research Gap 72
2.11 Summary 73

CHAPTER 3 RESEARCH METHODOLOGY 75
3.1 Introduction 75
3.2 Methodological Framework 75
3.3 Data Collection and Corpora Creation 78
3.4 Feature Extraction 84
3.4.1 Price Features 84
3.4.2 Sentiment Features 85
3.5 Cryptocurrency Price Prediction 100
3.6 Summary 103

CHAPTER 4 SENTIMENT DATA VALIDATION 105
4.1 Introduction 105
4.2 Sentiment Classification 105
4.2.1 Experimental Set-up 108
4.2.2 Results and Discussion 111
4.3 Sentiment Regression 113
4.3.1 Sentiment Regression using Traditional Machine Learning 113
4.3.2 Sentiment Regression using Language Model with Deep Learning 117
4.4 Summary 133

CHAPTER 5 CRYPTOCURRENCY PRICE PREDICTION **135**
5.1 Introduction 135
5.2 Cryptocurrency Price Prediction using Price Features only 135
5.3 Cryptocurrency Price Prediction using Sentiment Features 145
5.4 Summary 159

CHAPTER 6 CONCLUSION AND FUTURE WORK **161**
6.1 Introduction 161
6.2 Contributions 161
6.2.1 Corpora Creation for English and Malay 161
6.2.2 Data Annotation and Codebook 162
6.2.3 Validation of Sentiment Features using XLNet-GRU+ 164
6.2.4 Price Prediction model using Sentiment Feature 165
6.3 Summary of Thesis 167
6.4 Limitations and Future Work 168
6.5 Conclusion 169

REFERENCES **172**
APPENDICES **197**
AUTHOR'S PROFILE **209**

LIST OF TABLES

Tables	Title	Page
Table 2.1	Research Studies On Cryptocurrency From Year 2013 To 2023 (Bitcoin (B), Ethereum (E), Other coins (O))	20
Table 2.2	Summary Of The Cryptocurrency Price Prediction Results Achieved In Existing Research Studies	27
Table 2.3	Summary Of Datasets And Annotation Technique Used In Existing Research Studies	32
Table 2.4	Traditional Word Embedding Vs Contextual Word Embedding	43
Table 2.5	Contextual Word Embedding And Language Modelling	46
Table 2.6	Summary Of Classification For Cryptocurrency Price Prediction Without Sentiment Features Performance Results	60
Table 2.7	Summary Of Classification For Cryptocurrency Price Prediction With Sentiment Features Performance Results	61
Table 2.8	Summary Of Regression For Cryptocurrency Price Prediction Without Sentiment Features Performance Results	70
Table 2.9	Summary Of Regression For Cryptocurrency Price Prediction With Sentiment Features Performance Results	71
Table 3.1	Details Of Raw Data For News (Bitcoin And Ethereum)	82
Table 3.2	Details Of Raw Data For Tweets (Bitcoin And Ethereum)	82
Table 3.3	Number Of Documents (Posts) In Each Corpus	83
Table 3.4	Number Of Vocabulary In Each Corpus	88
Table 3.5	Rubrics For Codebook Creation: a) Positive And b) Negative Sentiment	91
Table 3.6	Krippendorf's Alpha Percentage Agreement For Each Corpus	93
Table 3.7	Annotated Scores Conversions	100
Table 4.1	Summary Of Sentiment Classification Results From Prior Studies	106
Table 4.2	Number Of Datasets After SMOTE-Tomek Application	108
Table 4.3	Sentiment Classification Experiment Results	111

Table 4.4	Summary Of Results On Sentiment Regression In Prior Studies	113
Table 4.5	Sentiment Regression Experiment Results	116
Table 4.6	Comparison Results Between VADER, BERT And XLNet Models	121
Table 4.7	Comparison Results Between BERT-GRU And XLNet-GRU Models	124
Table 4.8	Results Of Multi-Head Self-Attention Mechanism In XLNet-GRU Sentiment Model Performance (English And Malay News)	129
Table 4.9	Results Of Multi-Head Self-Attention Mechanism In XLNet-GRU Sentiment Model Performance (English And Malay Tweets)	130
Table 4.10	Comparison Results Between VADER, BERT-GRU And XLNet-GRU+ Models	131
Table 4.11	Results Of XLNet-GRU+ Sentiment Model Using Standard Labelled Datasets From Other Studies	133
Table 5.1	Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin)	141
Table 5.2	Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum)	142
Table 5.3	Price Prediction Error Difference Using Manual Annotation And Automated Annotation	144
Table 5.4	Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin English News)	146
Table 5.5	Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum English News)	148
Table 5.6	Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin Malay News)	150
Table 5.7	Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum Malay News)	151
Table 5.8	Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin English Tweets)	152

Table 5.9	Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum English Tweets)	154
Table 5.10	Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin Malay Tweets)	156
Table 5.11	Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum Malay Tweets)	157
Table 5.12	Average Error Differences	159

LIST OF FIGURES

Figures	Title	Page
Figure 1.1	Traders Reaction Towards Elon Musk's Tweet On Cryptocurrency Posted On March 14, 2022	4
Figure 1.2	Tweet Postings By A Malaysian Crypto Influencer, *MOSHED*	6
Figure 1.3	Thesis Outline	12
Figure 2.1	Statistic Of Cryptocurrency Price Prediction By Country Published Research	18
Figure 2.2	Statistics On Implemented Features For Cryptocurrency Price Prediction In Published Research	21
Figure 2.3	A Snippet Of Bitcoin Daily Price Data Based On Bitstamp Exchange (Bitstamp, 2025).	25
Figure 2.4	Generate Domain-Specific Sentiment Lexicon For Review Sentiment Analysis (Han et al., 2018)	37
Figure 3.1	Methodological Framework Overview	76
Figure 3.2	Sentiment Corpora Creation Framework	80
Figure 3.3	Examples Of Raw Online News Headlines: a) English And b) Malay	81
Figure 3.4	Examples Of Raw Twitter Posts: a) English And b) Malay	81
Figure 3.5	Emot Dictionary For a) English And b) Malay	86
Figure 3.6	Emoticon's Conversion Example For a) English And b) Malay tweets	87
Figure 3.7	Flowchart Of Annotation Process	94
Figure 3.8	Fine-tuning Process	96
Figure 3.9	XLNet Architecture (Yang et al., 2019): a) Content Stream Attention, b) Query Stream Attention, c) Overview Of The Permutation Language Modelling Training With Two-Stream Attention	98
Figure 4.1	GloVe + CNN-LSTM Model	109
Figure 4.2	BERT Model	110

Figure 4.3	Fine-tuned BERT Model	118
Figure 4.4	Fine-tuned XLNet Model	119
Figure 4.5	BERT-GRU Model	122
Figure 4.6	XLNet-GRU Model	123
Figure 4.7	Sentiment Regression Model Architecture With Multi-Head Self-Attention Mechanism	126
Figure 4.8	Visualization On A Negative Sentiment Word For a) English News And b) Malay News	128
Figure 4.9	Visualization On A Positive Sentiment Word For a) English Tweets And b) Malay Tweets	129
Figure 5.1	Cryptocurrency Price Prediction Model Architecture	136
Figure 5.2	LSTM Model Architecture	138
Figure 5.3	CNN Model Architecture	139
Figure 5.4	Comparison Of Price Prediction For Bitcoin Using Price Features Only	141
Figure 5.5	Comparison Of Price Prediction For Ethereum Using Price Features Only	142
Figure 5.6	Comparison Of Bitcoin Price Prediction Using English News Sentiment	146
Figure 5.7	Comparison Of Ethereum Price Prediction Using English News Sentiment	147
Figure 5.8	Comparison Of Bitcoin Price Prediction Using Malay News Sentiment	149
Figure 5.9	Comparison Of Ethereum Price Prediction Using Malay News Sentiment	151
Figure 5.10	Comparison Of Bitcoin Price Prediction Using English Tweets Sentiment	152
Figure 5.11	Comparison Of Ethereum Price Prediction Using English Tweets Sentiment	154
Figure 5.12	Comparison Of Bitcoin Price Prediction Using Malay Tweets Sentiment	156
Figure 5.13	Comparison Of Ethereum Price Prediction Using Malay Tweets Sentiment	157

LIST OF SYMBOLS

Symbols

α	Alpha (for Krippendorf's alpha measurement)
$\boldsymbol{D}_i^{Sentiment\ Score}$	Sentiment score for the document, D_i in a day
e	Margin of error
$\mathbf{E}_{ijk}$	Expected agreement between annotators, i,j,k
n	Sample size
N	Population size
$\mathbf{O}_{ijk}$	Observed agreement between annotators, i,j,k
p	Estimated proportion of population
r^2	R-squared
z	z-score

LIST OF ABBREVIATIONS

Abbreviations

Adagrad	Adaptive Gradient Descent
Adam	Adaptive Moment Estimation
Altcoin	Alternative coin
AMH	Adaptive Markets Hypothesis
API	Application Programming Interface
BERT	Bidirectional Encoder Representation from Transformers
Bi-GCL-A	Bidirectional GRU-Convolutional-LSTM with Attention
BNB	Binance coin
BOW	Bag-of-Words
BTC	Bitcoin
CNN	Convolutional Neural Network
COVID-19	2019 novel coronavirus
DNDF	Deep Neural Decision Forest
DNN	Deep Neural Network
ELMo	Embedding from Language Model
EMH	Efficient Market Hypothesis
ETH	Ethereum
FMH	Fractal Markets Hypothesis

GloVe	Global Vectors for Word Representation
GLS	Generalized Least Squares
GPT	Generative Pre-training Transformer
GRU	Gated Recurrent Unit
LLaMA	Large Language Model Meta AI
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MANE	Mean Absolute Normalized Error
MAPE	Mean Absolute Percentage Error
MCO	Movement Control Order
MSE	Mean Square Error
NB	Naïve Bayes
NLP	Natural Language Processing
NTUSD-Fin	A Market Sentiment Dictionary for Financial Social Media Data
OHLC	Open (O), High (H), Low (L), Close (C) prices
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
RMSProp	Root Mean Square Propagation
RNN	Recurrent Neural Network
RoBERTa	Robustly Optimised BERT Pre-training Approach
SANN	Stacked Artificial Neural Network

SGD	Stochastic Gradient Descent
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
T5	Text-to-Text Transfer Transformer
TF-IDF	Term Frequency – Inverse Document Frequency
ULMFIT	Universal Language Model Fine-Tuning
USDT	Tether
VADER	Valence Aware Dictionary and sEntiment Reasoner
XGA	XLNet-based Bi-GRU with Attention mechanism
XLNet	Generalized Auto Regressive
XLNet-GRU+	XLNet-GRU with Multi-head Self-attention mechanism
XRP	Ripple

CHAPTER 1
INTRODUCTION

1.1 Introduction

This chapter covers the introduction of the cryptocurrency and the way its price can be predicted. In this work, sentiment is hypothesised to be one of the factors that can influence the cryptocurrency price. Further detail on problem statement, research questions, research objectives, scope, limitation of study, and significance of the study are presented.

1.2 Research Background

Cryptocurrency is a virtual money utilising blockchain technology and cryptographic functions (Daskalakis & Georgitseas, 2020; Johnson, 2022; Pintelas et al., 2020). It is not in a physical form like fiat currency (cash money) such as Dollar, Yen, and Euro that holds value issued by the government of its respective country. Instead, cryptocurrency can be used to generate new coins and verify new transactions by any individuals using own computers or any electronic devices (Astuti et al., 2022; Goleman, 2018). Typically, the term *mine* in cryptocurrency is used to refer to a decentralized process where individuals use computational power to generate new coins and verify transactions, ensuring network security and transparency without central authority control.

There are several main characteristics of cryptocurrency; such as, transparency, decentralization, immutability, and volatility. Transparency is where information is readily available to be viewed by the users in an exchange (Awaysheh & Klassen, 2010), while decentralisation is the entrustment of power away from central authority (Hoffman et al., 2020). Once a transaction has been verified and recorded in the system, it cannot be altered which represents the immutability aspect of cryptocurrency (Politou et al., 2021). Moreover, volatility in cryptocurrency market refers to a sudden and rapid movement of price fluctuations (Bakas et al., 2022). The characteristics of cryptocurrency make it difficult to predict the future price of cryptocurrency.

Cryptocurrency price is determined by the relationship between demand and supply in the cryptocurrency market. Demand is when there are interests in purchasing the cryptocurrency asset, whereas supply refers to the ability of the cryptocurrency market to produce the assets (Giudici et al., 2020). In order for an individual to decide on whether to buy or sell cryptocurrency assets, prediction on future cryptocurrency values is essential by investigating various elements that may affect the cryptocurrency price directions (Kukacka & Kristoufek, 2023). This leads to the in-depth study on cryptocurrency price prediction. Cryptocurrency price prediction is important to assist cryptocurrency investors as an indicator for investment decisions of either to buy, sell, or hold cryptocurrency assets with the objective of to maximise earnings. Moreover, cryptocurrency price prediction is a solid research domain that can be used to analyse cryptocurrency market's behaviour (John et al., 2024).

The cryptocurrency price prediction was performed through various mathematical and computational methods by taking several factors that may influence the cryptocurrency price into account. The factors can be categorised into two types; which are, internal and external factors (İçellioğlu & Öner, 2019). The internal factor comprises of supply and demand, where the transaction cost, reward system, mining difficulty, coins circulation, and rule changes may influence the cryptocurrency traders' decision. On the other hand, external factors such as macro-financial issues involving rates based on stock market, gold price, interest rate, and political issues may affect the cryptocurrency market behaviour, thus, influencing the trader's judgement towards the cryptocurrency asset.

In addition, the popularity, market trend, and public's sentiments towards cryptocurrency asset can also contribute to the cryptocurrency price fluctuations (Hassan et al., 2022). The concept of sentiments that act as a manipulator or determiner in influencing the behaviour of the cryptocurrency investors and spontaneously affecting the cryptocurrency market are loosely discussed back in the 1990s (Deutscher et al., 1993; Dolan, 2002). However, Tetlock (2007) as reported by Karalevicius et al. (2018) lead the research direction to study the effect of sentiment as a predictor in examining the relationship between media reporting such as news and Twitter tweets, and the financial market movements. It is found that media content is mostly used as a proxy for investor sentiment, rather than a fresh fundamental information or market volatility.

Generally, the concept of sentiment can be understood as a form of trigger that sparks any forms of conflict. Higher crisis-related keywords found in the sentiment text, can result in higher risk of cryptocurrency price drop as confirmed by the Financial and Economic Attitudes Revealed by Search (FEARS) index (Anastasiou et al., 2021). The FEARS index is a sentiment indicator that tracks public attitudes toward economic and financial conditions by analysing the volume of online searches related to economic concerns. By aggregating search data on topics like market crashes, job security, and financial crises, the index aims to gauge the public's fear or optimism regarding the economy. Shifts in these search patterns can offer early signals of market trends or potential economic downturns, making the FEARS index a valuable tool for understanding how public sentiment may influence financial markets. Therefore, it is observed that the changing aspects of the cryptocurrency price market may be determined by the sentiment effect on the cryptocurrency investors.

Cryptocurrency prices are affected by both good and bad news, especially during bubble moments (Mnif et al., 2022; Rognone et al., 2020). A bubble moment refers to a phase in the market when the prices of assets, are artificially inflated well beyond their actual worth. This situation is driven by excessive speculation, excitement, or optimism. During these times, investors tend to rush in, propelling prices upward, often influenced more by market sentiment or news than by underlying fundamentals. However, when the market eventually recognizes the unsustainable nature of these inflated prices, a correction takes place, resulting in a significant drop in value, commonly known as the bursting of the bubble.

Typically, investors and traders gained insights from a cryptocurrency news online report (Hadan et al., 2024; Lamon et al., 2017). Then, they would start posting their perspectives, opinions, and comments through social media platforms such as Twitter, Reddit, Facebook, and other online forum platforms. Cerda (2021) and Hamza (2020) suggest that the opinions made by popular influencers such as celebrities, billionaires, or leaders, with a huge number of followers can cause a huge impact on any cryptocurrency condition. For instance, when Elon Musk tweets either a positive or a negative statement regarding Bitcoin, it highly influences the follower’s belief on the asset whether to buy, sell or hold. The impact may not be immediate, but the wave of change can be observed in the market. Figure 1.1 shows an example of a tweet posted

by one of the famous cryptocurrency influencers, Elon Musk stating about a cryptocurrency situation and several postings from the other traders regarding the post.

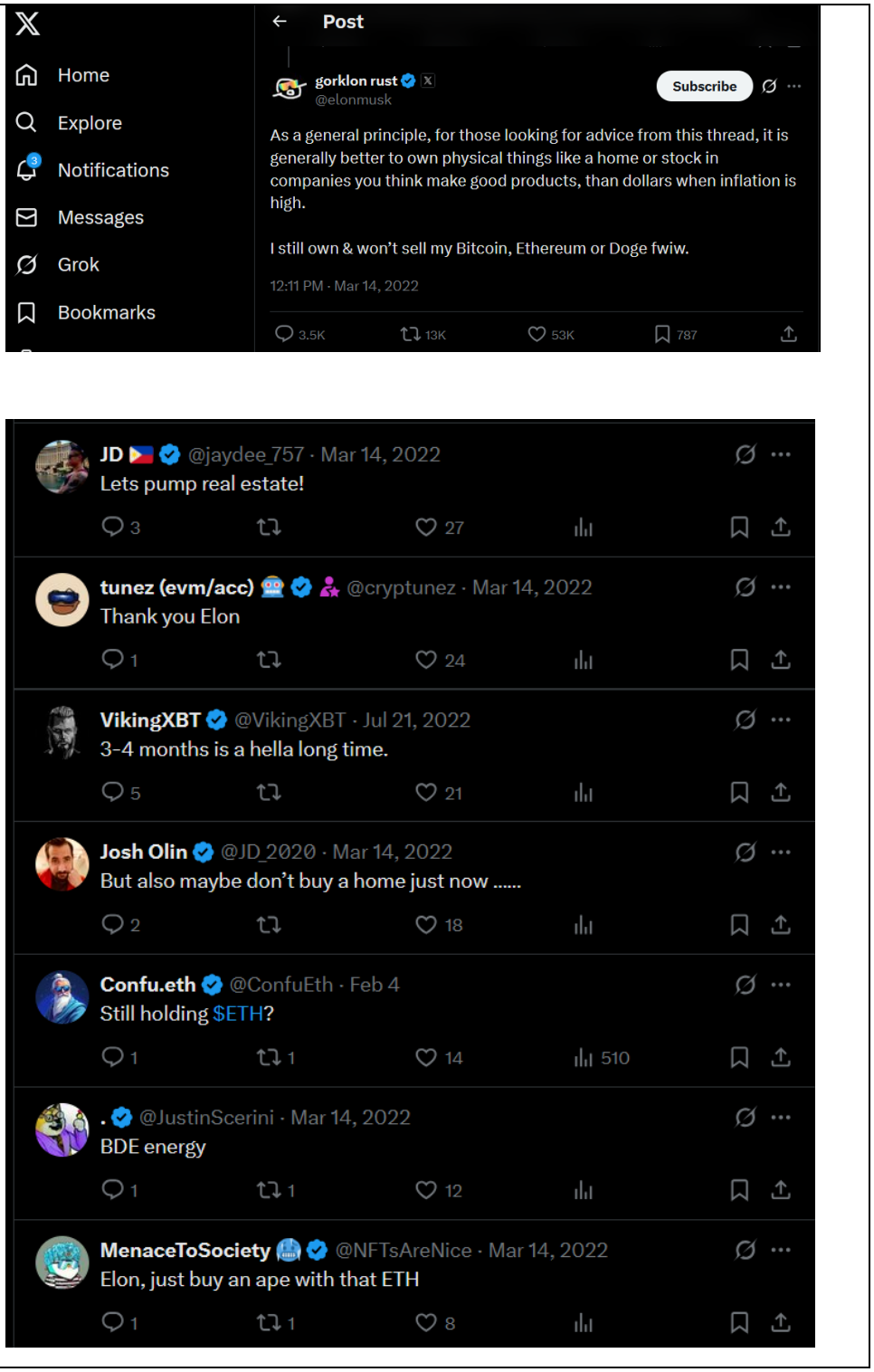


Figure 1.1 Traders Reaction Towards Elon Musk's Tweet On Cryptocurrency Posted On March 14, 2022

Based on Elon Musk’s tweet statement presented in Figure 1.1, it had likely sparked a ripple effect within the cryptocurrency market, influencing many followers to react. In particular, the market response would typically show fluctuations in the Bitcoin’s price as traders and investors react to Musk’s statements, whether positive or negative. For example, when Musk tweeted that he still owned Bitcoin and other cryptocurrencies, it potentially reaffirmed his support for these digital assets, leading to an initial surge in market confidence. The followers in the tweets, for example, user *JD* suggesting to "pump real estate" or others reacting with expressions like "Thank you Elon", reflected a mix of enthusiasm and consideration for his words.

Research by Ante (2023) highlights that Musk's tweets lead to substantial immediate price movements, especially for Dogecoin, which reacts with sharp price spikes. These effects typically last for a brief period before prices stabilize. However, the impact on Bitcoin is more complex, with Musk's mixed tone leading to varying market reactions, often cancelling out positive and negative effects. Pandey (2024) also confirms Musk’s influence during the COVID-19 pandemic, noting that his positive tweets generally caused price increases, particularly for Dogecoin, while his unfavourable remarks resulted in declines. Both studies emphasize that while Musk’s tweets have an immediate and significant impact on cryptocurrency markets, these effects are generally short-lived, with the market quickly correcting itself after the initial reaction. This highlights Musk's powerful but transient influence.

In Malaysia, many crypto influencers use their mother tongue, such as Malay, to connect with local audiences, making their content more relatable and accessible. By speaking in familiar languages, they break down language barriers and create a deeper emotional connection with followers, particularly when discussing complex or personal topics related to the cryptocurrency market. This approach not only fosters authenticity but also highlights the growing influence of local languages in global crypto discussions, ensuring that the space remains inclusive and engaging for a diverse audience. Therefore, it is vital for further study on sentiments and the cryptocurrency price movements. Figure 1.2 displays an example of a Malaysian crypto influencer named *MOSHED* posting about a cryptocurrency matter.

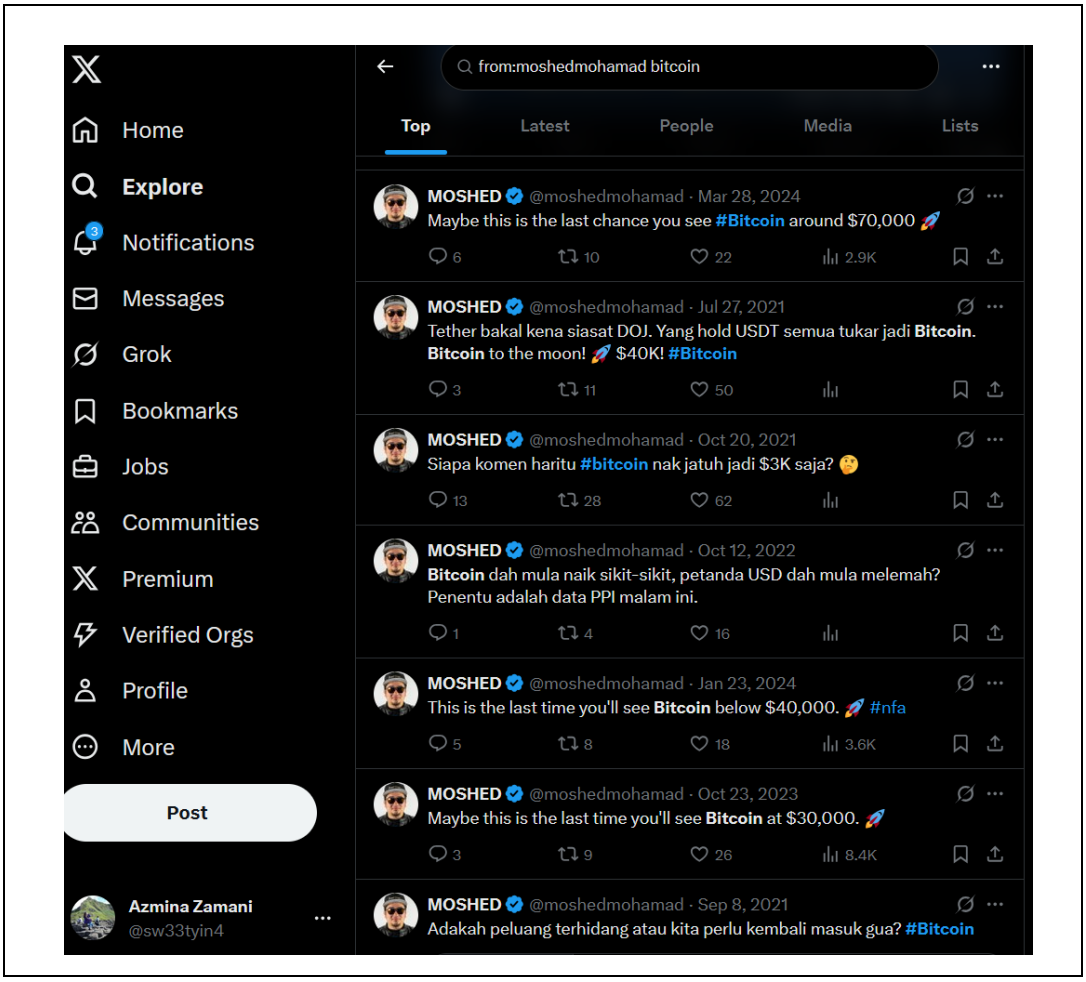


Figure 1.2 Tweet Postings By A Malaysian Crypto Influencer, *MOSHED*

From Figure 1.2, *MOSHED* used multiple languages such as English and Malay to express his opinion. From the tweets, it can be observed that Malay tweet governed more engagements as opposed to English tweets. Hence, considering Malay-based tweets to study sentiment in cryptocurrency price prediction in Malaysian perspective is a way forward to better understand the context.

1.3 Problem Statement

a) Inadequate Malay resources on cryptocurrency-related study

English is the world commonly used language, which makes its resources easier to obtain and more collections of data can be found to perform cryptocurrency-related study, particularly for sentiment analysis. Majority of

prior cryptocurrency-related research studies tend to focus mainly on English text to perform sentiment analysis (Arslan, 2025; Jung et al., 2025). Such experimentation task is a challenge when handling non-English texts, especially Malay since meticulous analysis are needed. The limited availability of resources further complicates matter (Sagnika et al., 2020). Different language has different rules that may not be similar to English (Brunato et al., 2022). Hence, the involvement of other language(s) in addition to English is essential as public users tend to be more expressive in their own mother-tongue language (Oostendorp et al., 2024).

b) Limitations in reliable cryptocurrency price prediction models and the role of sentiment analysis

Kuznetsov et al. (2025) states that modern computational methods, such as machine learning or deep learning, have a concern of insufficient accuracy and unreliability of prediction. The rapid growth of data has made the analysis more difficult due to its fluctuating and time-sensitive nature caused by various factors (Dunbar & Owusu-Amoako, 2023). The incorporation of sentiment features extracted from news and social media provides an additional layer of information that is not fully captured by the price data solely or by technical indicators. Therefore, the relationship between the sentiment features and cryptocurrency prices may contribute to the performance improvement of the existing machine learning or deep learning models for cryptocurrency price prediction (Shrivastava et al., 2025).

c) Complexity in performing sentiment prediction task

Classification and regression are the two types of prediction tasks to be addressed when performing sentiment analysis. Based on the existing cryptocurrency-related studies, classification task was selected by many researchers to discern the sentiment into either positive, negative, or neutral (Ortu et al., 2022; Passalis et al., 2022). Only few related works applied sentiment regression to perform the sentiment analysis (Haritha & Sahana, 2023; Sathiyapriya et al., 2023). Sentiment classification task predicts discrete class labels which is more generic in determining the level of positivity or negativity

in a document that leads to a less accurate prediction (Hartmann et al., 2023). Hence, a more precise sentiment polarity using continuous value (regression) is significant to achieve a better sentiment prediction performance.

Furthermore, many of the related works implemented lexicon-based approach and/or traditional word embeddings for sentiment analysis. These two techniques did not cater word polysemy problem unlike language models (Gomez-Perez et al., 2020). Although language models were introduced to address this problem, there are still very limited exploration on its application (Ortu et al., 2022; Passalis et al., 2022). Hence, an enhanced development of sentiment regression deep learning model incorporating language model is crucial in learning the semantic of texts to obtain a good quality sentiment feature to predict cryptocurrency prices.

Additionally, traditional deep learning models do not handle sequential reasoning, which leads to a lack of focus in the important words containing in a sentence. Thus, by having an attention mechanism, it allows a model to focus on the vital set of elements that steers the process of reasoning promoting to the enhancement of the sentiment analysis model (Kehinde et al., 2025), which can help to contribute to the performance gain of the cryptocurrency price prediction model.

1.4 Research Questions

The research is conducted to answer the following research questions:

a) How can cryptocurrency sentiment-related features be extracted from online news and tweets for Malay and English text for sentiment analysis?

b) How can the cryptocurrency price effectively be predicted based on deep learning model?

c) What is the effect of sentiment features on cryptocurrency price prediction using deep learning model?

1.5 Research Objectives

The research objectives are as follows:

a) To construct the cryptocurrencies corpora for English and Malay texts.

b) To develop the cryptocurrency price prediction system based on deep learning model.

c) To compare the performance of the cryptocurrency price prediction deep learning model for the constructed corpora with and without sentiment features.

1.6 Scope and Limitation of Study

The scope and limitation of study are as follows:

a) The study focused on extracting English and Malay online news headlines and tweets regarding Bitcoin and Ethereum cryptocurrency for a duration of one year (2021). It does not cover other languages. The corpora creation labelling is performed through manual human annotations which is time-consuming and may not be able to be completed in due time if more data need to be annotated.

b) Language model that is available for both English and Malay texts are implemented. Other language models without Malay text support are not applied.

c) Moreover, the study focused on the document-level or also known as post-level weightage for the sentiment polarity. The sentiments are labelled according to the overall sentence(s) contained in a news and tweet posts because it reflects the overall sentiment of a user. The language model and self-attention element learns the semantic of the whole sentence(s) to produce the weightage of vector representation. Since multiple sentences are involved, instead of only single sentence in determining the sentiment

value, other levels analysis such as sentence-level and word-level are not performed.

d) Subsequently, the scope of the study focused on the regression (continuous) values of sentiment scores ranging between -1 representing very negative sentiment to +1 representing very positive sentiment. The polarity range of sentiment score is selected because it is in line with the popularly used Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto & Gilbert, 2015) in existing studies and also to minimise the errors between annotator's agreement percentage during the annotation process, thus, other sentiment scores range are not used in this research.

1.7 Significance of Study

a) Construction of new cryptocurrency-related sentiment corpora in English and Malay

With the contribution of language resources by constructing new cryptocurrency-related sentiment datasets in English and Malay allows researchers to utilise the data in other related domains such as tourism, business, economics, or finances. Moreover, investors or traders of some organizations can use the datasets directly for the organization's research and development purposes. Due to the very limited availability of Malay corpora, this contributes to the addition of Malay resources, particularly in cryptocurrency domain.

b) Introduction of cryptocurrency word annotation rubrics

This research study contributes to designing an inter-annotator agreement codebook for manual annotation by providing rubrics. This is important because there is still lack of detailed process in designing the rubrics for the annotator's codebook in prior research studies. The rubrics can be used to ensure the standardisation of the annotation and minimize the error difference between the annotators. This approach can be emulated by other domains in producing a relevant codebook that can be tailored to specific domain and language.

c) Building a dynamic cryptocurrency price prediction model

This research study contributes to develop a dynamic cryptocurrency price prediction model by using the historical data for training, while the real-time data are used for testing to validate the model. It benefits the researchers to investigate the performance of the real-time data for the current day to predict the next-day cryptocurrency prices, while assisting traders to reduce the trading loss and maximising profit gains.

d) Integrating sentiment features to predict cryptocurrency prices

The methodological contribution of adding sentiment features to the cryptocurrency price prediction model produces an engine that generates a reliable sentiment scoring tool to be employed by any researchers who are involve in the language processing development and also offers developers from any organizations needing sentiment analysis for system enhancement. Furthermore, the incorporation of the sentiment features helps to produce a more reliable cryptocurrency price prediction model.

1.8 Thesis Outline

The thesis consists of six chapters. ***Chapter 1*** introduces the topic of the research by describing the research background, followed by the problem statement. There are three research questions and objectives being stated, followed by the scope, limitation and significance of the study. In ***Chapter 2***, literature review is conducted by reviewing existing related research articles on cryptocurrency and its market. The review on price prediction and sentiment analysis is also presented in this chapter with the research gap being highlighted. ***Chapter 3*** explains about the research methodology by illustrating the methodological framework and the main components of the research; which are: corpora creation, sentiment annotation, and feature extraction for cryptocurrency price prediction. This chapter has two sub-processes which are Corpora Creation, which involves the collection of data for analysis, and Annotation, which refers to the labelling of the data to provide meaningful insights. The chapter also discusses on the prediction of cryptocurrency prices, which forms part of the core analysis. This part of the process connects directly to the Results and Discussion section, which consists of two chapters,

Chapter 4, Sentiment Data Validation, and ***Chapter 5***, Cryptocurrency Price Prediction. These chapters focus on the analysis of the data and the discussion on the findings, including sentiment analysis and its impact on cryptocurrency price prediction. Finally, the thesis concludes with ***Chapter 6***, which wraps up the research findings and suggests potential directions for further research or improvements in the field. Figure 1.3 illustrates the general structure of the thesis.

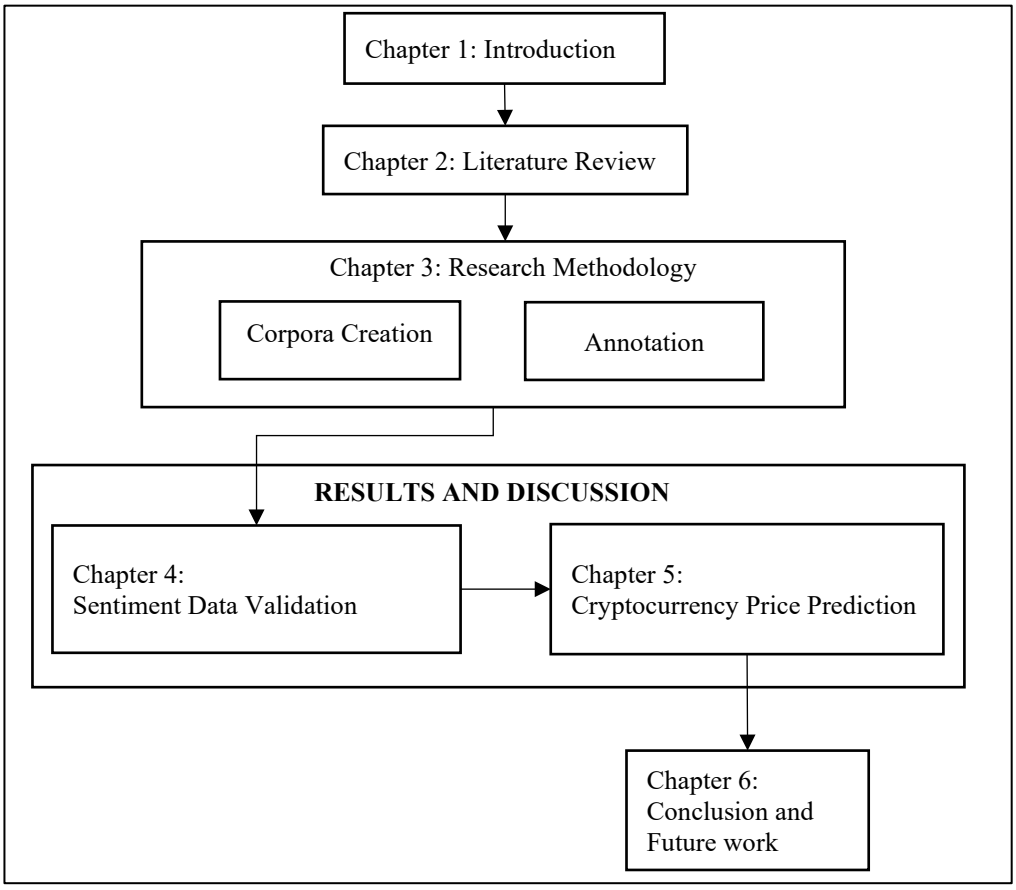


Figure 1.3 Thesis Outline

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

This chapter explains the cryptocurrency concept, cryptocurrency market including the market efficiency hypothesis and its predictability on cryptocurrency prices. Then, the review on the previous research works for cryptocurrency price prediction tasks is presented. Subsequently, the sentiment features being extracted as an added feature to the cryptocurrency price prediction model is also provided in this chapter.

2.2 Cryptocurrency

Cryptocurrency is a digital money that employs blockchain technology without going through central authority power (Protonotarios, 2019). Cryptocurrency transactions are recorded and displayed on the digital public ledger providing all details of transactions conducted by the users, making it highly transparent (Kshetri, 2018).

A centralised market is a financial market system in which all orders are routed through a single central exchange offering better management over the activities of the organization by ensuring consistency in operations and uniformity in decision-making. One of the most important features of centralised markets is that pricing is completely transparent and open to the public, such as the stock market (Niederle & Roth, 2009). On the contrary, decentralized market is emerging in tandem with the advancement of computer technology. It allows users to engage in online commerce without the use of a centralised market, for example, cryptocurrency market (Raval, 2016).

Decentralization of cryptocurrency via blockchain technology component operates by transferring from one party to another using a time-stamped and electronic signature by the cryptocurrency owner (sender). The cryptocurrency transaction is highly secured because a private key is used for authorization (Gowda & Chakravorty, 2021; Laurence, 2019). The blockchain technology element works as a middleman by becoming a trusted record keeping system in the operation of the cryptocurrency system

and it is not being controlled by any financial institutions or government, giving individuals a purchasing power (Daskalakis & Georgitseas, 2020; Goleman, 2018). Moreover, cryptocurrency runs on immutable transactions using encryptions (cryptographic function) that acts as a ciphering tool for secure online banking and commerce by using user authentication process; such as: obtaining usernames, passwords or public keys that requires a careful protection in giving access control decisions (Budaghyan, 2015; Tilborg, 2011). The counterfeiting activities such as manipulation, replacement, or fabrication of data stored on the network can be prevented (Laurence, 2019).

Observation by Liang et al. (2019) shows that the cryptocurrency market behaviour is similar to the stock market in terms of the central nodes performance because it is in a persistent transformation and expansion phase. However, cryptocurrency market behaviour differs from the traditional stock market behaviour in terms of its volatility and mutuality (Matkovskyy, 2019). Matkovskyy (2019) demonstrates that the volatility is higher in the cryptocurrency market (decentralized market) compared to the stock market (centralized market), making cryptocurrency market more fragile or unstable in terms of robustness (Liang et al., 2019). Additionally, investments in cryptocurrency market is known to be volatile during the rapid movement of the increment and decrement of prices daily due to certain factors influencing the market; such as: supply and demand, investor and user sentiments, government regulations, and media hype (Miraz et al., 2022).

In cryptocurrency market, there are various types of cryptocurrency assets, which is also called *cryptocurrency coins*. Initially, Bitcoin was created in the cryptocurrency system as the first cryptocurrency coin to be made public in 2008 under the pseudonym of Satoshi Nakamoto (Nakamoto, 2008). The idea of Bitcoin is formed to make the transfer of value between parties easier using its own value metric known as *BTC*. Such an approach is dissimilar to any traditional payment systems that transfer payments in sovereign currencies (Badev & Chen, 2014).

Besides Bitcoin, there are also other types of cryptocurrency coins known as *altcoins* (alternative coins) that use similar existing blockchain technology (Bashir, 2017; Kalodner et al., 2015). One of the earliest altcoin productions named *Namecoin*, was produced in April 2011. As time progresses, other altcoins are massively produced

to cater the desires and needs of the investors, traders, and cryptocurrency enthusiasts (Bashir, 2017). Altcoins are gaining popularity because of the improved functionality and conveniency over Bitcoin (Ramkumar, 2021). Based on the cryptocurrency coins market capitalisation reported in Forbes (Hooson & Pratt, 2023), Bitcoin and altcoins such as Ethereum (ETH), Tether (USDT), Binance coin (BNB), and Ripple (XRP) are the top five largest cryptocurrency asset being traded by the users up to this date implying the good reputation of altcoins.

Cryptocurrency trading is performed via various cryptocurrency exchange platforms worldwide such as Bitstamp, Binance and Coinbase for any transactions in buying or selling of cryptocurrencies (Podobas, 2019). Furthermore, cryptocurrency coins can also be used as a mean of payment for online purchasing (Cointelegraph, 2021; Fiorillo, 2018; Giudici et al., 2020).

2.3 Cryptocurrency Market

Cryptocurrency market is decentralised. The digital form of money is not being handled by a central authority such as the banks and government. It uses blockchain technology to perform the transactions, whereby the cryptocurrency assets can be bought or sold through an exchange platforms (Gowda & Chakravorty, 2021). Since there is no authority control over the cryptocurrency market, it is important to understand the behaviour of cryptocurrency market and to gain some knowledge about cryptocurrency. This is because there may be potential risks involving monetary policy of the standard banking system which has raised some concerns among the authorities. Hence, the standard monetary policies transmission mechanisms at the national and international levels may be weakening (Bibi, 2023).

Initially, cryptocurrency market price was assumed unpredictable based on the Efficient Market Hypothesis (EMH) theory (Fama, 1970). The EMH theory stated that cryptocurrency market is unpredictable because the market is efficient. The theory is outlined based on the assumption that share prices reflect all information that results in the stocks being traded at fair value on exchange. There is no reason for the investors to invest undervalued stocks or sell stocks for inflated prices (Fama, 1970). EMH follows the *Random Walk theory* which assumes that the market price trend cannot be

used to forecast future market movements because the technical and fundamental analysis are erratic (Van Horne & Parker, 1967).

Fama (1970) also outlined three forms of EMH; which are, 1) weak-form EMH, 2) semi-strong-form EMH, and 3) strong-form EMH. The weak-form EMH occurs when random price movements do not rely on past occasions (e.g., past prices, trading volume, etc) to predict future market's direction. The semi-strong-form EMH claims that significant market reactions are reflected by the public statement, inferring that neither fundamental nor technical analysis provide any significance in predicting future price movements. Conversely, a strong-form EMH is when a market is considered efficient by integrating the weak and semi-strong form EMH traits. Such a situation occurs when all information that are available to the public and not publicly known are completely accounted for in the current prices.

Kang et al. (2021) studied on the efficiency in the cryptocurrency market by conducting three Random-Walk tests for EMH. From the observation, it was found that from a total of 893 cryptocurrency units, 54 units satisfied the weak-form EMH, and 24 units fulfilled the semi-strong EMH. The rest of the units satisfied the strong EMH state inferring that the cryptocurrency market was more economically efficient initially, but decreasing over time. From the experimental results, it has shown that the larger-scaled cryptocurrency tends to satisfy the weak-form EMH especially those that were created before November 2017. After the end of 2017, the cryptocurrency market is more speculative causing more inefficient information being used and displaying the immaturity of the market. In this case, cautions are needed among investors and traders as the information that can be acquired from the market transaction data is not usually signalled in the transaction price of the cryptocurrency. This indicates that the profits and losses can only be gained alongside the obtainable information set. The results showed that it is possible to predict future market price movement, thus, giving a small advantage to investors in maximising profits and reducing the risk of loss.

However, Li et al. (2020) findings are in contrast. The researchers showed that EMH is rejected in the cryptocurrency market. This notion is supported by several studies demonstrated market inefficiency and the predictability of cryptocurrency (Cohen, 2020; Nadarajah & Chu, 2017; Urquhart, 2018; Zhang et al., 2018). Gurdgiev & O'Loughlin (2020) reported that the emergence of studies on behavioural finance,

sentiment analysis and cryptocurrencies implicitly has contributed to alternative theories such as the Fractal Markets Hypothesis (FMH) theory (Peters, 1994) and the Adaptive Markets Hypothesis (AMH) theory (Lo, 2004).

Whilst EMH hypothesize that the market is at equilibrium and always steady, FMH theory (Peters, 1994) proposed that markets are alleviated by coordinating the demand and supply of investors' investment horizons during market crises and stable market (Dar et al., 2017; Karp & Van Vuuren, 2019; Peters, 1994). Dar et al. (2017) found that a great number of investors are involved during crises period globally. Market instability tends to intensify during times of crisis, and investors often show risk aversion and tend toward safe-haven investments (Abou Tanos & Badr, 2024). A *safe haven* is a term that refers to an investment that is predicted to at least maintain in value. An example can be observed during the pandemic season of COVID-19, where higher volatility in cryptocurrency market is shown with a substantial rise in cashflow (Corbet et al., 2022). The involvement of many investors creates chaos, and it reflects the FMH theory. Hence, the price movement is not random and can be predicted (Gunay & Kaşkaloğlu, 2019). The validity of FMH is also supported by Çelik (2020) where the result demonstrates that the financial bubble pattern exists in the Bitcoin market.

Alternatively, the AMH theory (Lo, 2004) stated that markets are not always efficient and are highly context-dependent on certain market conditions. Lo (2004) combines EMH with the developmental behavioural aspects because the AMH theory claims that stable rational actors only happen during times of certainty. However, it is challenging to rationalise the actors' behaviour during ambiguous moments due to the emotion and instinct factors (Kraaijeveld & De Smedt, 2020). Consequently, Khursheed et al. (2020) shows that distinct market cycle has substantial control on price fluctuations of cryptocurrencies, which confirms the AMH practicality.

Based on the theories, this research study adopts the AMH theory in the assumption that sentiment factor of the public may influence the cryptocurrency market behaviour. Market is not perfect and therefore price can be predicted based on sentiment influence.

2.4 Previous Works on Cryptocurrency Price Prediction

Since cryptocurrency price can be predicted due to the market inefficiencies, cryptocurrency price prediction is gaining attention from researchers from various domains; such as, finance, economic, business, tourism, mathematics, and computing. One of the popular topics on cryptocurrency is cryptocurrency price prediction which is in line with the scope of this study. Figure 2.1 illustrates the density of research performed on cryptocurrency price prediction worldwide based on the paper published across 67 countries within the duration of 11 years (2013 to 2024), extracted from Scopus, Web of Science, and Google Scholar.

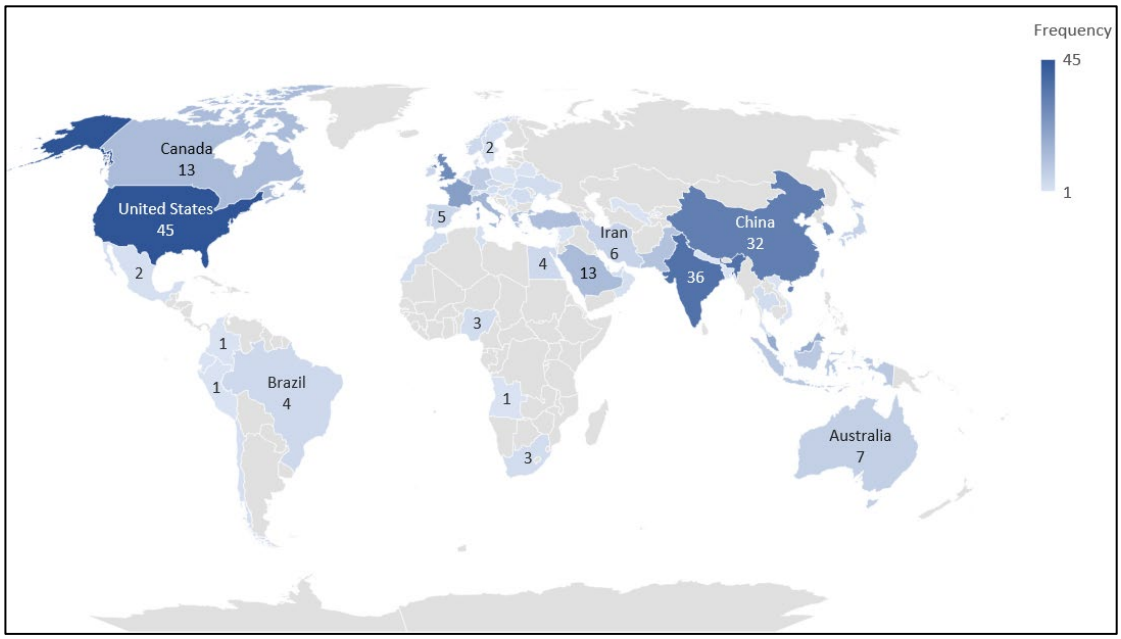


Figure 2.1 Statistic Of Cryptocurrency Price Prediction By Country Published Research

Publications on cryptocurrency price prediction topic started to gain traction in 2013. Bitcoin, although was introduced in 2008, was a relatively new and niche concept. There are limited data, infrastructure, or interest from the broader research community until the cryptocurrency market began to grow significantly around 2013. The expansion led to more attention and resources being allocated to studying price prediction and other related topics. As more researchers contribute to publish more cryptocurrency-related research studies, more resources are acquired to fill in the gaps

based on the domain expert. Continuous research on cryptocurrency is still expanding to date.

Based on Figure 2.1, United States of America (USA) has the most published work on cryptocurrency throughout the year 2013 to 2024 (10.4%), followed by other countries such as India (8.3%) and China (7.4%). This is because these countries are the top 10 countries with high cryptocurrency users up to this date (Parihar, 2025). USA has shown to be the lead in terms of research output due its active and high investment in cryptocurrencies. Moreover, it also contains a large number of research institutions and technology hubs with the presence of major blockchain-related startups and financial institutions. Nevertheless, publications on cryptocurrency price prediction work were done mostly by the authors from India during the year 2023 and 2024. With the substantial amount of cryptocurrency users, more ideas, resources, research, and experimentations are actively acknowledged. China, despite its regulatory challenges regarding cryptocurrency, remains a key player in cryptocurrency research. This is due to its advanced technological infrastructure and significant involvement in the blockchain sector. On the other hand, regions such as Africa and other smaller countries display lower research density, potentially due to limited access to resources, lower technological infrastructure, and regulatory barriers that inhibit research or adoption of cryptocurrencies.

There are various experiments on different types of cryptocurrency coins. Table 2.1 depicts some of the existing research studies performed on different types of cryptocurrency coins from 2013 to 2024. From Table 2.1, it has been observed that 91% of the existing studies regarding cryptocurrency price prediction has selected Bitcoin as the case study due to its popularity among the other cryptocurrencies with the highest market capitalization up until today, followed by Ethereum (27%) and other altcoins (26%) (Yahoo! Finance, 2025). Ethereum and other altcoins have started to gain more attention starting from 2018 onwards, inferring that multiple cryptocurrency coins are being investigated in order to have a wider understanding. In addition, the types of features utilised in prior studies are also being explored.

Table 2.1
Research Studies On Cryptocurrency From Year 2013 To 2023 (Bitcoin (B), Ethereum (E), Other coins (O))

Authors	2013			2014			2015			2016			2017			2018			2019			2020			2021			2022			2023			2024		
	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O	B	E	O
Kristoufek (2013)	■																																			
Shah & Zhang (2014)				■																																
Madan et al. (2015)							■																													
Greaves & Au (2015)							■																													
Kim et al. (2016)										■	■																									
Attila (2017)													■	■																						
Lamon et al. (2017)													■	■	■																					
Stenqvist & Lönnö (2017)													■																							
Yenidoğan et al. (2018)																■		■																		
Mohanty et al. (2018)																■																				
Steinert & Herff (2018)																■	■	■																		
Shintate & Pichl (2019)																			■																	
Valencia et al. (2019)																			■	■	■															
Chen et al. (2019)																					■															
Salač (2019)																			■																	
Vo et al. (2019)																				■																
Jahjah & Rajab (2020)																						■														
Akyildirim et al. (2020)																						■	■	■												
Rognone et al. (2020)																						■		■												
Shahzad et al. (2021)																									■											
Hitam et al. (2021)																									■	■	■									
Akba et al. (2021)																									■											
Erfanian et al. (2022)																												■								
Alghamdi et al. (2022)																												■	■							
Edgari et al. (2022)																												■								
El Haddaoui et al. (2023)																															■					
Sathiyapriya et al. (2023)																																■				
Acosta & Arana (2023)																															■					
Girsang & Stanley (2023)																																■	■			
Zubair et al. (2024)																																		■	■	■
Perera et al. (2024)																																		■		
Arslan (2024)																																		■		

The research works on cryptocurrency price prediction using sentiment features were starting to gain traction from 2019. This may be due to the sentiment chaos caused by the users during the pandemic of COVID-19 (Vacca et al., 2021). The cryptocurrency influencers, for example Elon Musk, tend to speculate a cryptocurrency coins, which gives major impact on the cryptocurrency market (Ante, 2023). A study on the features involved in the cryptocurrency price prediction need to be investigated. The understanding of feasible features may lead to better cryptocurrency price prediction performance. Figure 2.2 shows the statistics on implemented features for cryptocurrency price prediction in published research.

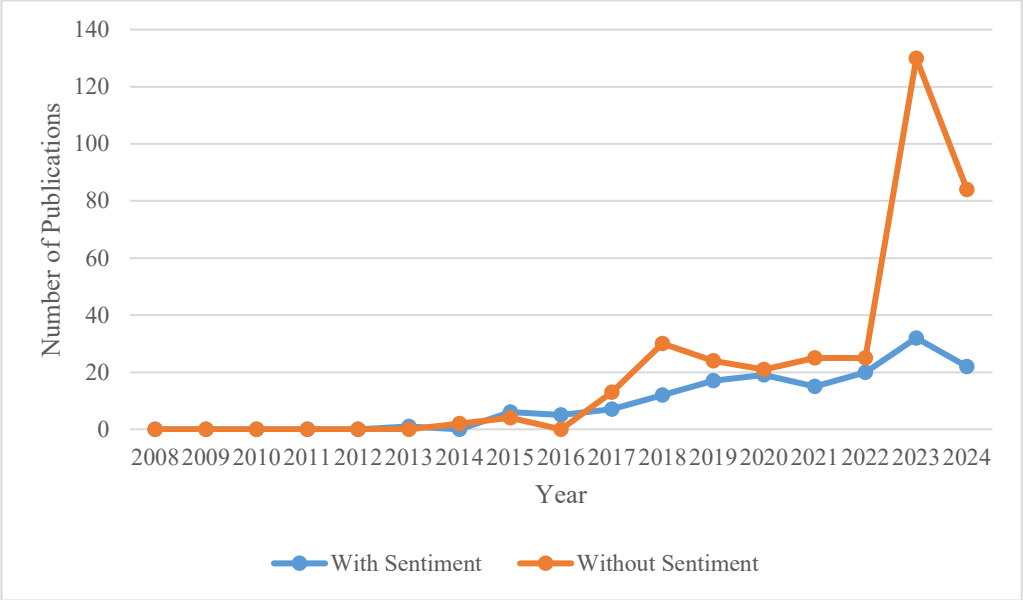


Figure 2.2 Statistics On Implemented Features For Cryptocurrency Price Prediction In Published Research

Figure 2.2 presents the trend of studies on the features for cryptocurrency price prediction from 2008 to 2024. The trend was manually plotted from the total count of the collected published research articles. Based on Figure 2.2, research on cryptocurrency price prediction gain momentum in 2013 and is increasing throughout the years. However, there was a 2% slight plunging from 2019 to 2021 that is hypothesize to be the effect of COVID-19 pandemic (Kruger et al., 2022). Once the pandemic is over, the pattern shows upward trend.

In 2023, there is about 23% increase in the research on cryptocurrency price prediction with sentiment whereas for the research on cryptocurrency price prediction

without sentiment shows a soar of about 67%. This infers that the research on cryptocurrency captures a peak attention among researchers in 2023 but has a slight reduction in the year 2024. Furthermore, it is noticeable that less explorations were performed on cryptocurrency price prediction using sentiment features. The limited explorations may be due to the chaotic and unstructured characteristics of sentiment data. Therefore, this study focuses on sentiment features for better understanding on the gap.

There are various statistical and computational methods being implemented in the past research studies for cryptocurrency price prediction. Statistical methods are usually used to determine the correlation (relationship) and causation between sentiment data and the cryptocurrency prices, whereas computational methods (i.e., machine learning or deep learning) are typically used to predict the cryptocurrency prices such as the future price direction of either going up or down, price returns, and/or estimated future prices (Maqsood et al., 2022; Ortu et al., 2022; Ye et al., 2022). Hence, this study focuses on the computational methods in predicting the cryptocurrency prices. This is because machine learning, especially ensemble and deep learning, is a more reliable technology since traditional statistical methods are difficult to use due to the lack of seasonal effects and the need to meet a number of unrealistic requirements (Bouteska et al., 2024; Kehinde et al., 2025).

Previous research studies have shown that combining multiple features, such as price and sentiment data, can significantly enhance the accuracy of cryptocurrency price predictions (Arslan, 2025; Perera et al., 2024). Price data, often represented by the Open, High, Low, and Close (OHLC) prices along with trading volume, provides essential information on market movements and trends (Acosta & Arana, 2023; Arslan, 2025; Zubair et al., 2024). On the other hand, sentiment data, often derived from sources like news articles and social media platforms, capturing the emotional tone and public perception on cryptocurrency can influence price fluctuations (Passalis et al., 2022; Schulte & Eggert, 2021).

Moreover, studies have explored both price and sentiment features for two types of prediction tasks; which are, classification and regression. In classification tasks, machine learning and deep learning models are typically used to predict the direction of price movements, such as whether the price will go up or down (Loginova et al., 2021; Tandon et al., 2021). In contrast, the regression task focuses on predicting the actual price or price returns of cryptocurrencies (Mohagheghzadeh et al., 2024; Perera

et al., 2024). Research has shown that incorporating sentiment data alongside traditional price features in regression tasks improves the model's predictive performance, as sentiment provides additional context that may explain underlying price trends (Arslan, 2025; Mohagheghzadeh et al., 2024; Park & Seo, 2023). By combining these two features, previous studies have demonstrated that machine learning and deep learning models can produce more accurate and reliable predictions for both the direction (classification) and exact values (regression) of cryptocurrency prices (Jung et al., 2023; Zubair et al., 2024). The detailed discussion on data collection, data annotation, feature extraction, classification and regression for cryptocurrency price prediction is provided in the next subsection.

2.5 Types of Data Used in Cryptocurrency Price Prediction

For price data, information is gathered from various reliable sources, such as exchanges and market tracking platforms, which provide both real-time and historical data on prices, trading volumes, and other relevant metrics. The data is collected at different intervals to ensure accuracy and relevance. Sentiment data is obtained from online platforms, including social media, news, and forums, where discussions and opinions about cryptocurrencies are prevalent. Sentiment analysis tools are used to categorize the collected data into positive, negative, or neutral sentiments. The collected sentiment data helps identify public perceptions that may influence market behaviour and, in turn, impact cryptocurrency prices. Combining price and sentiment data provides a comprehensive view of the factors that drive price movements, offering valuable insights for further analysis.

2.5.1 Price Data

Cryptocurrency price prediction often involves analysing key market features such as open (O), high (H), low (L), close (C), and trading volume, which are fundamental indicators for understanding market trends and making predictions (Huang et al., 2024). Based on prior research studies, the price data can be obtained from various financial websites such as CoinMarketCap.com, CoinDesk, and Yahoo! Finance (Prashanth et al., 2023; Kumar & Rath, 2020; Struga & Qirici, 2018). The extraction of price data can be implemented using APIs to allow users to access the data. The API

key needs to be registered initially, then requests can be submitted to retrieve historical price data in structured formats such as JSON or CSV. Another method that can be performed is by manually download the price data in CSV format directly from the cryptocurrency exchange website. Additionally, automated data gathering can be done through web scraping, using tools such as BeautifulSoup or Selenium to harvest data from websites in the absence of APIs.

A comparison between manual and automated data collection highlights distinct advantages and drawbacks. Manual downloading is relatively straightforward, requires minimal technical elements, and ensures that data is obtained directly from the official platform. This method is suitable for small-scale studies or one-off analyses. However, it is less efficient for large datasets since it is prone to human error, and impractical for research requiring frequent or continuous updates. On the contrary, automated methods such as API extraction and web scraping provide efficiency, scalability, and the ability to capture real-time or high-frequency data, making them more suitable for large-scale or longitudinal studies. Nevertheless, collecting data automatically requires technical proficiency. In addition, the automated data collection is subject to constraints such as API rate limits and may be disrupted by frequent changes to website structures during web scraping (PropTrack, 2025). Consequently, the choice of data collection method is determined by the research objectives, dataset size, and the need for periodic or real-time data updates.

Typically, there are few price data that are used by the researchers; such as, open price, high price, low price, close price, and trading volume. Other non-sentiment features that are also being utilised in several prior research studies are technical indicators and Google Trends (OSL News, 2025).

- ***Open price*:** the price of a cryptocurrency at the beginning of a given time period, such as daily and hourly. It provides insight into the starting market sentiment for that particular time frame. The opening price can indicate whether there is a bullish or bearish trend at the start of a new trading period.

- ***High price*:** the highest price reached during a specific time period, representing the peak of market demand within that frame. The high price can reveal the maximum value that buyers were willing to pay before a reversal or decline occurred, showing potential resistance levels.

- ***Low price***: the lowest point reached during that time period. This represents the most significant price drop, indicating support levels where buyers started to step in and halt further decline, often suggesting areas of potential price stability.

- ***Close price*****:** the final trading price within the specified time period. It is a crucial reference point for analysing market sentiment and determining the overall direction, either bullish or bearish of the market.

- ***Trading Volume*****:** the total number of units of a cryptocurrency traded during a specific period. A high trading volume indicates strong investor interest and can amplify price movements, while low trading volume suggests weak market participation, often leading to price stagnation or increased volatility.

Collectively, these features are often used in technical analysis to create predictive models for cryptocurrency prices. Understanding how the open, high, low, close, and trading volume data interacted helps traders forecast potential price movements and make informed decisions (John et al., 2024). Figure 2.3 displays an example of the extracted Bitcoin daily price data with the five common features of open, high, low, close prices and trading volume. The key price indicators; the open, high, low, and close prices, which represent the price at the start, peak, lowest point, and end of each trading day, respectively. Additionally, the volume BTC column indicates the total amount of Bitcoin traded during that day, reflecting market activity and liquidity.

date	open	high	low	close	Volume BTC
1/1/2021 0:00	28999.63	29700	28720	29402.64	8781.465555
2/1/2021 0:00	29410.77	33333	29050	32216.53	17867.8789
3/1/2021 0:00	32216.51	34800	31977.45	33097.83	14160.58128
4/1/2021 0:00	33068.83	33669.76	27734	32005.88	22446.94001
5/1/2021 0:00	32054.23	34500	29936.26	34039.05	14806.49651
6/1/2021 0:00	34081.24	37050.83	33348.27	36849.94	19219.28789
7/1/2021 0:00	36875.96	40402.46	36327.5	39510.49	23102.26121

Figure 2.3 A Snippet Of Bitcoin Daily Price Data Based On Bitstamp Exchange (Bitstamp, 2025).

From Figure 2.3, the date represents daily data captured at 12.00 am. The four different prices shown, for example, looking at the date 2/1/2021 at time 0.00 (12.00

am), the open price value was taken at the start point of time, which is at 12.00 am on 2/1/2021 and not at the close price value of 11.59 pm on 1/1/2021 because the cryptocurrency market is live 24/7. The closing price is the price at the last minute of the current day, which is at 11.59 pm. The high price is when the value of price is at its highest point of the current day, while the low price is when the value of price is at its lowest point of the current day from 12.00 am until 11.59 pm. The Volume BTC is the trading volume for the day, showing the number of traders who traded.

The technical indicators such as moving averages, Relative Strength Index (RSI), and Bollinger Bands are derived from the OHLCV data. For example, moving averages use the closing prices over specific time periods, and RSI uses the closing prices along with the high and low values to determine momentum. The RSI enables the traders to spot potential entry or exit points and also to confirm market trends.

In addition, Google Trends captures the amount of search volumes by the publics' interest in a cryptocurrency. While OHLCV data reflects actual market transactions, Google Trends provides additional context by showing the level of public curiosity or search activity, which often precedes market movements.

The summarized cryptocurrency price prediction results using several types of price features with the implemented models are shown in Table 2.2. Based on the findings, although various price features were used to predict the cryptocurrency prices, not all price features produced significant results. Thus, new features are needed to enhance the cryptocurrency price prediction performance. From the observations in prior studies, the integration of sentiment features may be needed to enhance the reliability of the cryptocurrency price prediction model. By having mixed results in the cryptocurrency price prediction, experimentation on sentiment features is required to address any insufficiencies that exist in the prediction as suggested in several studies to include sentiment features particularly those extracted from social media and news to improve prediction results (Vo et al., 2019; Yao et al., 2019).

Table 2.2
Summary Of The Cryptocurrency Price Prediction Results Achieved In Existing Research Studies

Author	Additional Price Features			Price prediction model	Performance Result
	Trading Volume	Technical Indicators	Google Trends		
Madan et al. (2015)		✓		Random Forest	Accuracy = 94.98%
Greaves & Au (2015)		✓		Neural Network (Classification), Linear Regression (Regression)	Accuracy = 55%, MSE = 1.94
Sin & Wang (2017)		✓		MLP	Accuracy = 60%
McNally et al. (2018)		✓		LSTM	Accuracy = 52%
Saxena et al. (2018)	✓			LSTM	RMSE = 456.78 (min = 0, max = 20,000)
Wu et al. (2018)	✓			LSTM with AR(2)	RMSE = 247.33 (std deviation = 2269.18)
Phaladisailoed & Numnonda (2018)	✓			GRU	r^2 = 0.992, MSE = 0.00002
Dutta et al. (2019)		✓	✓	GRU	RMSE = 0.016 (15-days lookback period - normalized)
Ji et al. (2019)		✓		DNN (Classification), LSTM (Regression)	Accuracy = 53.6%, MAPE = 3.52%
Mallqui & Fernandes (2019)		✓		ML Ensemble Algorithm (Classification), SVM (Regression)	Accuracy = 62.91%, MAPE = 1.28% - 1.91% (normalized)
Shintate & Pichl (2019)		✓		Random Sampling Method	Accuracy = 53.53%
Akyildirim et al. (2020)	✓	✓		SVM	Accuracy = 69%
Borges & Neves (2020)		✓		Ensemble Voting	Accuracy = 55.61%
Chowdhury et al. (2020)	✓			Ensemble Learning Method	Accuracy = 92.4%
Kumar & Rath (2020)	✓			LSTM	RMSE = 20.53 (Daily - normalized)
Livieris et al. (2020)	✓			CNN-BiLSTM	Accuracy = 54.62%
Mudassir et al. (2020)		✓		SANN (Classification), SVM (Regression)	Accuracy = 65%, MAPE = 1.44% (normalized)
Patel et al. (2020)	✓			LSTM & GRU-based hybrid	RMSE = 2.2986 (Litecoin), RMSE = 3.2715 (Monero)
Sun et al. (2020)		✓		LightGBM	Accuracy = 60.7% (2-week), Accuracy = 47.6% (2-day)
Jaquart et al. (2021)		✓		LSTM	Accuracy = 56% (1-hour)
Schulte & Eggert (2021)	✓	✓	✓	LSTM	MAPE = 1.18%
Erfanian et al. (2022)		✓		SVR	Pearson's r = 0.89, RMSE = 8.68 (std deviation = 2.48)
Kanat (2023)		✓		Chaid Algorithm	Profit = 80%
Jung et al. (2023)	✓	✓		XGBoost	Accuracy = 90.58%
Perera et al. (2024)	✓			GRU	RMSE = 659 (normalized)

2.5.2 Sentiment Data

A range of online platforms where users actively express their thoughts and feelings are usually the source of sentiment data used to forecast Bitcoin prices. Social networking sites like Twitter, Reddit, and Telegram, financial news sources, forums like Bitcointalk, blogs, and even market reports or analysis postings are examples of these sources. Large volumes of this data are frequently collected using tools like web scraping and APIs, and sentiment indicators are extracted from the text using Natural Language Processing (NLP) techniques (Perera et al., 2024; Mohagheghzadeh et al., 2024; Loginova et al., 2021).

From straightforward positive or negative words to more intricate statements or even emoticons, people can communicate their emotions in a variety of ways. Because tone, sarcasm, urgency, and exaggeration can all be used to express sentiment, NLP approaches are required for accurate parsing and interpretation. For example, someone may use hashtags like #Bullish or #BearMarket, or emojis like 🚀 or 💰, to show their enthusiasm for a cryptocurrency asset. Similar to this, negative attitude can be expressed by using terms like "sell" or "dump," as well as by making more critical remarks about the asset or its performance in the market. The difficulty of using the sentiment lies in accurately interpreting context, identifying irony or sarcasm, and gleaning emotion from colloquial language or acronyms that are frequently used in social media conversations (Durmuş Şenyapar, 2024).

Predicting cryptocurrency prices requires sentiment data to be accurate and relevant. Predictions may be distorted if the data is extremely generic, derived from untrustworthy sources, or does not accurately reflect the sentiment of the target market. Inaccurate models may result from the influence of noise, such as sentiment collected from spammy, low-quality posts. Furthermore, it is critical to think about the appropriate time range for sentiment analysis; for instance, sentiment from a few hours or days earlier might not be as applicable to forecasting short-term price swings (Rani et al., 2024; Sun et al., 2020; Kumar & Rath, 2020).

According to previous studies, sentiment data can be collected from various sources, such as Twitter, Reddit, and cryptocurrency-specific forums (Ortu et al., 2022; Loginova et al., 2021). Researchers typically focus on sentiment scores, consisting of positive, negative, and neutral sentiments, and sentiment volatility, which can provide

crucial insights into market sentiment shifts. Additionally, sentiment features are often combined with traditional market indicators like price data and technical indicators to improve prediction accuracy in cryptocurrency price forecasting. The various sentiment features and its functions are unique and can be used for different kind of analysis. It involves features such as positive counts, negative counts, and sentiment scores.

The ***positive-negative count*** refers to the number of positive and negative words in a given text. It is typically determined using sentiment lexicons or models. In the context of cryptocurrency price prediction, these counts serve as numerical indicators of market sentiment. A higher positive count suggests optimism, potentially indicating upward price movement. A higher negative count signals pessimism, possibly predicting a price drop. These counts are often used as input features in machine learning models or statistical analyses to correlate sentiment with price trends, enabling more informed predictions by capturing the emotional tone of market-related discussions. Positive and negative counts denote the quantity of words in a specific text categorized as positive or negative according to a sentiment lexicon. These counts present a clear, quantitative assessment of the ratio of positive and negative sentiment conveyed in the text. Tools such as VADER, TextBlob, or alternative sentiment lexicons classify words inside a text, then computing the counts according to the number of terms linked to positive or negative sentiment. These counts are significant attributes utilized in machine/deep learning models, especially for predicting cryptocurrency market fluctuations. These sentiment features offer valuable input for diverse classification and regression applications.

The ***sentiment score*** is a numerical value that quantifies the overall sentiment of a text, often calculated by aggregating the polarities (positive or negative) of words or phrases in the text. It typically ranges from negative (indicating pessimism) to positive (indicating optimism). In cryptocurrency price prediction, sentiment scores are used to capture market sentiment from news, social media, or other textual data. Positive sentiment scores may suggest bullish market trends, while negative scores indicate bearish trends. These scores are integrated as input features in predictive models, helping to identify correlations between market sentiment and price fluctuations, thus enhancing the accuracy of price forecasts. Sentiment scores are obtained by text analysis to assess the overall sentiment. Scores are generally produced by sentiment analysis tools like VADER, TextBlob, or BERT, and are frequently categorized into positive,

negative, neutral, and compound scores. The compound score consolidates all individual emotion scores into a singular value, spanning from -1 (strongly negative) to 1 (strongly positive). The overall sentiment score is crucial for comprehending the emotional tone of the text and is often utilized as a feature in predictive models. By incorporating sentiment scores, these models can better capture market psychology, which is often a driving force behind price movements. However, challenges like noisy data, contextual ambiguity, and the risk of overfitting need to be carefully managed.

The positive-negative counts quantify the occurrence of sentiment-laden words, while the sentiment score aggregates these into a single value reflecting overall sentiment polarity. This combination provides a comprehensive sentiment analysis feature by capturing both the frequency and intensity of sentiments in text data. In cryptocurrency price prediction, the combination allows for a nuanced understanding of market sentiment by leveraging both granular word-level insights and an overall sentiment trend. Once the data are collected, data annotation phase is conducted.

In this study, sentiment scores are chosen as the primary sentiment analysis feature instead of using positive–negative word counts. Although positive–negative counts are useful for capturing the frequency of sentiment-laden terms, the features treat all sentiment words equally and do not account for variations in intensity or contextual weight (Alabdulkarim et al., 2024; Taboada et al., 2011). In contrast, the sentiment scores combined sentiment polarity and intensity into a single continuous value. The sentiment scores also provide a more expressive and compact feature, allowing predictive models to capture the sentiment holistically, thus, many prior related studies applied sentiment scores as the sentiment features (Feizian & Amiri, 2023; Frohmann et al., 2023).

2.6 Data Annotation

In order to train machine learning models, data annotation is the process of labelling or categorizing the data. In this work, sentences in text data are categorized, to segregate the positive, negative, or neutral sentiment as implemented by Tan et al. (2024). Machine learning models learn from labelled instances. Without precise and consistent annotations, models are unable to recognize patterns or predict outcomes.

The model's effectiveness are dependent on accurate annotation, based on pertinent and high-quality data (Klie et al., 2024).

Price data annotation is the process of adding context to historical cryptocurrency prices by tagging them with relevant information. This could include marking certain price movements with labels like positive news, market crash, or high/low volatility. This annotation in some studies can be an indicator of what might have caused the price shift. It can be whether it is a news story, a regulatory announcement, or a change in investor sentiment. The annotated data is valuable for building models that can predict future price changes, as it helps to show how different factors have influenced the market in the past.

However, it is not always necessary for the price data to be annotated. The open, high, low, close, and trading volume (OHLCV) price features can be directly used without the need for annotation because the price features inherently provide valuable information about market behaviour and serve as sufficient input for predictive analysis. By analysing the raw price data, machine learning models can detect patterns and trends autonomously (Chauhan et al., 2025).

In addition, sentiment data such as the sentiment score is obtained from the text of some posts or messages. One of the annotation types is to categorize the sentiment into positive, negative, or neutral classes, whereas the sentiment strength can be measured on a scale of -1 to 1. For instance, "Bitcoin to the moon" might be categorized as a positive tweet, but "Bitcoin crashing, get out now!" might be categorized as a negative sentiment (Frohmann et al., 2023; Girsang & Stanley, 2023; Zubair et al., 2024). The strength of the sentiment can be indicated by very positive (+1) to very negative (-1). Table 2.3 presented the dataset sizes and annotation technique used in existing research studies for comparison purposes.

There are two ways to annotate the data with sentiment values, which are manual and automatic annotation. The practice of human annotators manually labelling data is known as manual annotation. This method assigns appropriate tags, labels, or categories after human experts examine the data. In the text sentiment analysis, an annotator mark a statement as neutral, negative, or positive (Akba et al., 2021). The annotators can also indicate the intensity of the sentiment.

Table 2.3
Summary Of Datasets And Annotation Technique Used In Existing Research Studies

Author(s)	Sentiment Datasets	Collection Period	Number of Data	Annotation Technique (Manual/Auto)	Results
Lamon et al. (2017)	Online news, Twitter	1/1/2017 – 30/11/2017	Online news: 3,600, Twitter: 10,000	Manual	Accuracy = 24.2% to 75.8%
Stenqvist & Lönnö (2017)	Twitter	11/5/2017 - 11/6/2017	2,271,815	Auto	Accuracy = 83.3%
Galeshchuk et al. (2018)	Twitter	Jan 2014 – Sept 2017	2.5 million	Auto	Accuracy = 68.6%
Jain et al. (2018)	Twitter	1/3/2018 – 11/3/2018	1,492,987	Auto	Adj $r^2 = 0.426$
Pant et al. (2018)	Twitter	1/1/2015 – 31/12/2017	7,454	Manual	Accuracy = 77.62%
Valencia et al. (2019)	Twitter	16/2/2018 – 21/4/2018	13,096,598	Auto	Accuracy = 44% to 72%
Sattarov et al. (2020)	Twitter	12/3/2018 – 12/5/2018	92,550	Auto	Accuracy = 62.48%
Chin & Omar (2020)	Online news	Sep2017 - Aug 2019	15,295	Auto	RMSE = 0.0471
Steinert & Herff (2018)	Twitter	21/3/2017 - 4/6/2017	426,520	Auto	$r^2 = 0.025$
Jahjah & Rajab (2020)	Twitter	1/1/2018 – 31/5/2018	25,746	Auto	Accuracy = 78%
Agarwal et al. (2021)	Twitter	1/1/2016 – 15/5/2021	22,147	Auto	RMSE = 0.02
Cerda (2021)	Twitter	Jul 2018 - Jan 2019	9,146	Manual	RMSE = 1.78
Shahzad et al. (2021)	Twitter	1/1/2015 – 1/1/2019	25,000	Auto	RMSE = 6.21
Tandon et al. (2021)	Twitter	Jan 2012 to Mar 2021	12,562	Auto	Accuracy = 96%
Hitam et al. (2021)	Twitter	Jan 2017 - Jan 2020	1,126	Auto	RMSE = 8.517
Alghamdi et al. (2022)	Twitter	1/1/2022 – 9/5/2022	75,995	Auto	RMSE = 0.503 to 0.684
Hasan et al. (2022)	Twitter	27/1/2015 – 20/2/2020	10,260	Auto	Accuracy = 98.75%
Critien et al. (2022)	Twitter	1/1/2012 – 31/12/2020	over 20 million	Auto	Accuracy = 63%
Edgari et al. (2022)	Twitter	5/2/2021 – 17/12/2021	1,970,666	Auto	Accuracy = 88%
Passalis et al. (2022)	Twitter, Reddit, Online news	2015 - 2020	570,865 (Twitter), 154,481 (Reddit + Online news)	Auto	Accuracy = 92%
Ye et al. (2022)	Twitter	Sep 2017 - Jan 2021	more than 7 million	Auto	Accuracy = 57.51%
Zhou et al. (2023)	Twitter	1/1/2021 – 30/6/2021	28,799,774 tweets	Auto	Accuracy = 61%
Sekioka et al. (2023)	Twitter	29/6/2021 – 6/1/2022	534,155 tweets	Auto	F1-score = 29.6%

Manual annotation is very accurate and able to capture intricate details because human annotators are able to thoroughly examine and classify data. As the dataset gets larger, human annotation gets more expensive and time-consuming. It is also labour-intensive and prone to human mistake while handling huge amount of data (Mikulová et al., 2023). Hence, an automated data annotation is needed.

Automated annotation is the process of automatically labelling data according to preset criteria or patterns using algorithms or machine learning models. Because it can handle big datasets with minimal human involvement, this method is quicker and more scalable than manual annotation (Manaf & Nordin, 2009). For example, based on learnt patterns, a pre-trained natural language processing model can automatically categorize text with sentiment. However, human monitoring is frequently necessary for automated annotation to guarantee the correctness and quality of the labels (Wu et al., 2023). The likelihood of using automated annotation increases with the size of the dataset. Hence, it can be summarized that automatic annotation is suitable for large dataset whereas manual annotation is suitable for small dataset.

Based on the observation of Table 2.3, the automated methods are mostly used because it is useful in sentiment analysis tasks involving vast amounts of data gathered from Twitter posts, Reddit discussions, and online news articles. Studies such as Stenqvist & Lönnö (2017) and Passalis et al. (2022) used automated annotation on datasets containing millions of instances. The ability to process such large volumes of data is one of the primary reasons why automated techniques are preferred in sentiment analysis, due to its time-saving and economical with labour.

On the other hand, although manual annotation is more resource-intensive, it offers certain advantages especially in terms of accuracy. Many of the studies that employed manual annotation achieved relatively higher accuracy rates despite working with smaller datasets. For example, Pant et al. (2018) used manual annotation on 7,454 data points and achieved an accuracy of 77.62%. The study by Lamon et al. (2017) also achieved an accuracy of 75.8% using manually annotated data of 3,600 online news and 10,000 tweets. The experimental results suggest that manual annotation is more precise in capturing the subtleties of sentiment, making it ideal for smaller-scale analyses where the quality of the annotation outweighs the need for processing large volumes of data. Manual techniques are particularly useful when the sentiment or context is more complex or nuanced, which can sometimes be difficult for automated systems to detect.

In conclusion, both annotation techniques are useful for sentiment analysis. Manual annotation is suited to smaller datasets where accuracy and nuance are paramount. It tends to offer higher precision and reliability in capturing sentiment, but it is limited by its ability to scale. Automated annotation, while potentially less accurate in terms of capturing subtleties, excels at handling large datasets and providing consistent results. Based on efficiency and scalability, the quantity of data gathered influences the annotation technique selection.

Therefore, this study performs manual annotation on online news headlines and tweets in English and Malay texts. Since this study involves a non-English text, manual annotation is more suitable because the linguistic and cultural subtleties that automated systems would find difficult to account for can be taken into consideration by human annotators. Many machine learning models and NLP technologies are mostly trained on English datasets. Therefore, it is a struggle to correctly understand the syntax, grammar, and meaning of non-English languages. Additional difficulties may arise from non-English texts because of distinctive sentence patterns, colloquial idioms, regional dialects, and cultural allusions that automated systems could misunderstand. Manual annotation ensures that the annotations catch these subtleties by allowing native or competent speakers to use their knowledge of the nuances of the language.

Additionally, there are limited pre-trained models or resources available for non-English languages, necessitating human interaction to guarantee correctness and dependability in the annotation process. In this study, an inter-annotator agreement is carried out for validity similar to the annotation process in Gadi & Sicilia (2024) but using Krippendorf's alpha measurement (Krippendorff, 2004b) with the sentiment range of -1 to 1. Then, feature extraction from the annotated datasets is performed using several approaches to learn the sentiment texts for prediction tasks.

2.7 Feature Extraction

In cryptocurrency price prediction, it is not always necessary to perform feature extraction on price data, especially when working with raw price data. Price data, such as open, high, low, close (OHLC), and trading volume, can often be directly used in models without the need to extract additional features (Kang & Kim, 2025). This is because the raw price data itself contains significant information about market trends,

volatility, and price behaviour. In some cases, especially when using advanced machine learning models like deep learning, the raw price data may be processed directly by the model, allowing the algorithm to learn relevant patterns and relationships without the need for traditional feature extraction. However, while feature extraction can enhance model performance by providing additional context, it is not always a requirement when the raw data is sufficient.

Moreover, sentiment features can be extracted using either traditional Natural Language Processing (NLP) techniques that does not require word embedding or modern NLP techniques that involves word embedding. In recent experiments, a variety of machine learning and deep learning models, including SVM, Random Forest, GloVe + CNN-LSTM, and BERT, have been employed to evaluate the effectiveness in sentiment analysis (Alghamdi et al., 2022; Passalis et al., 2022; Jahjah & Rajab, 2020; Pant et al., 2018). Traditional models like SVM and Random Forest rely on feature extraction techniques and are known for their simplicity and interpretability. However, deep learning approaches, such as GloVe + CNN-LSTM, leverage on pre-trained word embeddings and sequential pattern learning to capture complex linguistic relationships. Meanwhile, a state-of-the-art transformer-based model, BERT has revolutionized sentiment classification by understanding contextual nuances within text. Comparing these models across datasets provides insights into their strengths, limitations, and suitability for handling varying levels of text complexity and domain-specific challenges.

Convolutional Neural Network (CNN) is a class of deep neural network consisting of convolutional layer, pooling layers, and fully connected layers (Wu, 2017), while Long Short-Term Memory (LSTM) is a type of recurrent neural network that captures long-term dependencies in sequential data (Hochreiter & Schmidhuber, 1997). In addition to CNN-LSTM model, Global Vectors for Word Representation (GloVe) embedding was implemented as the text representation. GloVe is an unsupervised learning algorithm for generating vector representations (embeddings) of words based on global word co-occurrence statistics in large text corpora. GloVe aimed to capture semantic relationships between words by considering the global context of word co-occurrence across the entire corpus (Pennington et al., 2014).

To deal with class imbalance in datasets for classification tasks, Synthetic Minority Over-sampling Technique (SMOTE) with Tomek Links (SMOTE-Tomek)

can be used. By balancing the dataset and removing noise, the hybrid approach produces a higher-quality dataset that produces models that are more reliable and accurate (Sasada et al., 2020). Further descriptions are provided in the next subsection.

2.7.1 NLP without Word embedding

Conventional NLP approach without word embedding, such as lexicon-based approach (Osgood et al., 1957), Bag-of-Words (BoW) (Harris, 1954), Term Frequency-Inverse Document Frequency (TF-IDF) (Robertson & Jones, 1976), and dictionary-based tools, regard text as a simple collection of individual words or tokens, neglecting the semantic interconnections among them. These techniques depend on quantifying word frequencies or aligning words with established lexicons to execute functions such as sentiment analysis or document classification. Although the techniques are computationally straightforward and interpretable, the meaning of the words, synonyms, or word associations are inadequately captured, making the rendering less efficient in addressing intricate linguistic subtleties such as polysemy or nuanced variations in meaning contingent upon context (Khoo & Johnkhan, 2018).

2.7.1.1 Lexicon-based approach

The lexicon-based approach was commonly used in related research studies, where it classifies the words into either positive, negative, or neutral by producing the sentiment scores of each word (Chin & Omar, 2020; Hasan et al., 2022). Lexicon-based sentiment analysis uses a preset dictionary of terms (sentiment lexicon) to classify sentiment in raw text. Each word in the lexicon has a label or sentiment score of either neutral, negative, or positive. The text is split into discrete words or tokens by the algorithm. Then, it compares these tokens to the dictionary. The sentiment score of each matching word is added to the text overall sentiment score. Once all the words have been processed, the overall score is used to identify the text sentiment. A score close to zero indicates neutral sentiment, a score that is positive shows positive sentiment, and a score that is negative indicates negative sentiment (Han et al., 2018).

An example provided in Figure 2.4 illustrates a sentiment analysis process based on a domain-specific sentiment dictionary and mutual information scores (Han et al.,

2018). The flow begins with testing documents, where text preprocessing is applied to clean and prepare the data. For each term in the review, mutual information scores are extracted based on the part of speech (POS), which are then summed across all remaining terms in the document. A weight factor (α) is applied to adjust the term scores. The final step involves calculating a final score for the document. If the score is negative, the review is labelled negative. The same situation applies to positive and neutral score. This approach combines sentiment dictionaries with information theory to classify reviews based on their sentiment.

There are two types of lexicon-based approach that can be applied: 1) ***Lexicon-based approach with built-in library*** such as, Valence Aware Dictionary for Sentiment Reasoning (VADER) (Maqsood et al., 2022; Zhou et al., 2023), and Textblob (Hitam et al., 2021; Prajapati, 2020), and 2) ***Lexicon-based approach with sentiment dictionary*** such as, NTUSD-Fin dictionary (Chin & Omar, 2020), Loughran-McDonald Master Dictionary (Gurdgiev & O'Loughlin, 2020), and SentiWordNet (Cerda, 2021).

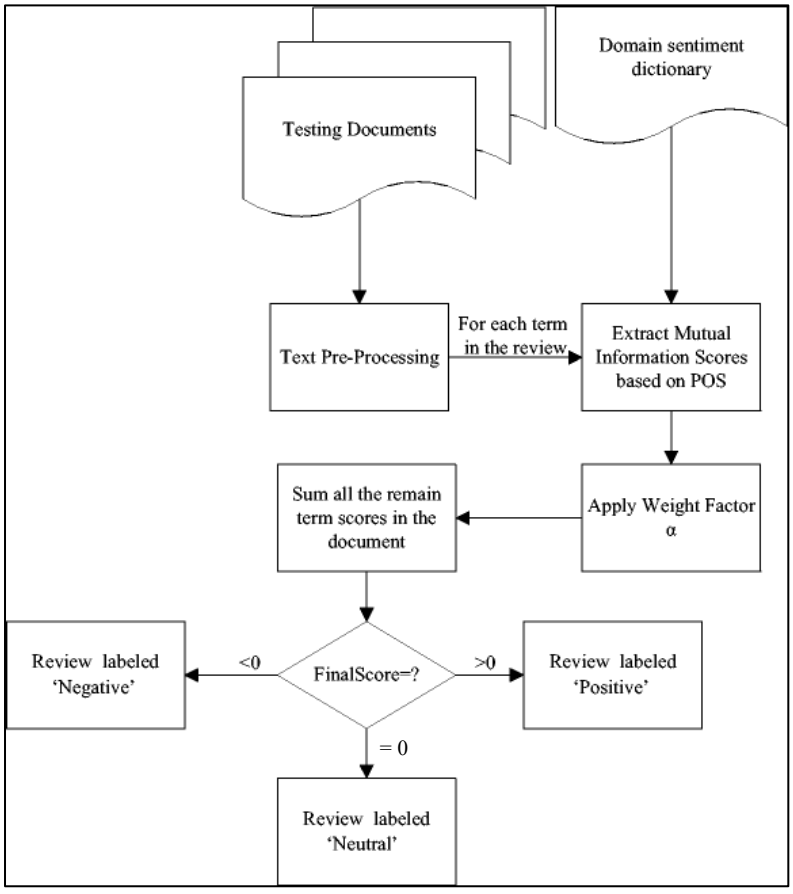


Figure 2.4 Generate Domain-Specific Sentiment Lexicon For Review Sentiment Analysis (Han et al., 2018)

VADER (Hutto & Gilbert, 2015) is one of the most popularly used automated techniques in the existing studies for sentiment scoring of English text. It is popular due to its ease of use, functionality, and suitability for different use cases. VADER is perfect for real-time sentiment tracking because it is very good at assessing short, informal texts, such as posts on social media (Mountstephens et al., 2023). It also employs a lexicon-based approach with heuristics to account for slang, emoticons, and intensity modifiers.

VADER consists of four types of scores which are positive, negative, neutral, and compound score. The range of VADER score is in between -4 to +4, ranging from very negative to very positive respectively with 0 indicates neutral sentiment according to predetermined threshold. The entire sentiment of a text is summarized by the VADER compound score, a numerical metric that goes from -1 (most negative) to 1 (most positive). It is computed as a normalized sum of sentiment scores at the word level, with context-specific corrections made for negations, capitalization, and punctuation.

Edgari et al. (2022) illustrated the positive effect of sentiment score for cryptocurrency price prediction by applying VADER compound score in computing the sentiment index. The study reported the performance of the cryptocurrency price prediction model using the sentiment feature gained 82% accuracy. However, after applying sentiment feature, an accuracy of 88% was achieved with 6% improvement. This indicates that VADER compound score can improve prediction outcomes for Bitcoin price movements. The compound score reflects the market's current perception of Bitcoin and encapsulates sentiments such as fear, greed, and speculation that frequently drive the unpredictability of Bitcoin value.

Contradictory, a study by Salač (2019) utilised VADER compound scores to predict cryptocurrency prices show a conflicting result. An r^2 of 0.095 and 0.312 was produced by the cryptocurrency price prediction model using Reddit and Twitter data respectively. However, the result of Reddit and Twitter combination data yielded r^2 = 0.337 which is higher than using the sentiment data separately. The experimental results implied the low precision in predicting the cryptocurrency prices due to the limitation exists in VADER because VADER is incapable to evaluate negations in phrases, especially when dealing with informal words in Twitter. There are text noises in the Twitter data making it harder to produce very accurate predictions.

The research studies by Zhou et al. (2023) implemented VADER scores to predict Bitcoin and Ethereum price movements using SVM model. The accuracy of 61% for both Bitcoin and Ethereum price prediction was obtained. No sentiment evaluation was conducted in this research study. The result without sentiment shows low accuracy that implies the sentiment features are needed to boost the performance in predicting the cryptocurrency price movements.

Another approach for NLP without word embedding is TextBlob (Loria, 2020), a general-purpose NLP package that manages tasks like tokenization and part-of-speech tagging in addition to sentiment analysis that focuses on polarity and subjectivity of a sentence or an overall statement. The polarity values range from -1 to +1, with -1 representing a negative sentiment and +1 representing a positive sentiment. Moreover, the subjectivity value ranges between 0 and 1, with 0 denoting very objective and 1 denoting very subjective. Subjectivity is the degree to which a text conveys subjective feelings, judgments, or views rather than objective facts. Low subjectivity signifies factual information, such as news reports or scientific articles, but high subjectivity denotes opinionated content, such as reviews, opinions, or editorials. This measure helps distinguish between texts that are meant to inform and are meant to persuade or express feelings. The understanding of subjectivity is helpful in applications such as sentiment analysis, content categorization, and understanding the tone of communication. The characteristics make TextBlob more comprehensive tool for formal text processing.

A study by Hitam et al. (2021) applied TextBlob to generate sentiment scores. The average of all the scores of each tweet in a day was taken as the sentiment feature for cryptocurrency price prediction. An accuracy of 77% was achieved for the sentiment classification model performance. To prove the significant impact of the extracted sentiment feature, further study was carried out to evaluate the cryptocurrency price prediction. The experimental result shows the errors of RMSE = 0.0367 (Bitcoin) and RMSE = 8.5171 (Ethereum). The result implicates usefulness of sentiment score in predicting the cryptocurrency prices. However, VADER and TextBlob lexicon-based approaches are not readily available for Malay text.

Some research studies also utilized the application of dictionary to generate the sentiment features. The NTUSD-Fin dictionary (Chen et al., 2018) provides various scoring methods such as frequency, Term Frequency - Inverse Document Frequency

(TF-IDF), squared value, market sentiment score and word vector for the tokens based on StockTwits financial dataset. The chi-squared is used for categorical data, while the r-squared (r^2) is used for continuous data. A study by Chin & Omar (2020) generated positive and negative counts from the NTUSD-Fin dictionary to perform the cryptocurrency price prediction. From the sentiment evaluation result, an accuracy of 65.83% was achieved, implying the potential effect of the sentiment feature in predicting cryptocurrency prices. In addition, the combination of positive and negative counts with cryptocurrency historical prices produced an RMSE of 0.0471, confirming the usefulness of the positive and negative counts as one of the features for cryptocurrency price prediction. These counts, derived using a dictionary-based method, quantify the frequency of positive and negative sentiment words in relevant text data (e.g., news, social media, forums). The result demonstrates that sentiment, particularly the combination of positive and negative in market discourse, has a consequential effect on cryptocurrency prices, which are extremely sensitive to trader and market perceptions. The model improves price prediction accuracy and complements existing indicators by gaining a deeper knowledge of the emotional elements influencing market patterns through the capture of these sentiment traits.

A study by Gurdgiev & O'Loughlin (2020) applied the Loughran-McDonald Master Dictionary (Loughran & Mcdonald, 2011) dictionary with Generalized Least Squares (GLS) model to investigate how investor decision-making behaviour and publicly available data flows correspond to affect volatility in cryptocurrency prices. The Loughran-McDonald Master Dictionary created a sentiment lexicon specifically for business and financial communications. It covers the English language and context of financial papers, including annual reports, earnings calls, and regulatory filings, in contrast to general-purpose sentiment dictionaries. In order to capture the tone and behaviour in financial circumstances, the dictionary divides terms into categories such as positive, negative, uncertain, litigious, modal, and constraining. This is to clarify the way the language used in finance can convey different attitudes, risks, or influences on decision-making and actions. It is a useful tool for domain-specific analysis and is frequently used in financial sentiment research to assess how language in financial communication may affect investor decisions or market activity.

From the experiment of quantitative analysis of the relationship between investor sentiment and cryptocurrency price movements, the findings demonstrate that

the price direction of cryptocurrencies can be predicted. The analysis shows that anchoring and herding biases directly affect the sentiment data collected from the Bitcointalk.org forum, with a coefficient value of 0.0389. This value is statistically significant at the 1% level because the probability of the result occurring by chance is very low. In statistical terms, the finding indicates the likelihood of observing such a strong relationship between anchoring and herding biases in the data without any real effect is less than 1%. The result also suggests that the relationship is real and not due to random fluctuations in the data. The experiment signifies the usefulness of the sentiment dictionary in determining the relationship between public's sentiment and the cryptocurrency market behaviour.

In more recent study, Cerda (2021) applied SentiWordNet to generate the sentiment scores between -1 to 1 from Twitter data. The researcher combines the OHLC prices, market cap and sentiment score using XGBoost to predict Bitcoin price. The SentiWordNet is a lexical resource that extends the WordNet database by assigning sentiment scores to sets of synonymous words (synsets). Each synset is assigned three scores of positivity, negativity, and objectivity, which add up to one. The results indicate whether the word conveyed is positive, negative, or neutral. SentiWordNet works by mapping words to the WordNet synsets and uses the resulting scores to identify sentiment orientation. It captures the intricacies of words with numerous meanings by delivering sentiment data at the word sense level, allowing for more detailed and accurate sentiment analysis. This makes it especially beneficial for activities such as opinion mining and social media analysis, which require recognizing the sentiment of individual words in context (Baccianella et al., 2010). Based on the result, an RMSE = 1.78 was achieved, inferring that this technique is greatly aided by the sentiment score that SentiWordNet provided since the model's prediction accuracy, with a lower RMSE generally signifying better performance (Cerda, 2021).

Even though dictionary-based method produces a good result based on some prior research study, it still incapable to capture domain-specific language, context, and subtleties. It produces skewed findings because it is unable to recognise irony, sarcasm, and conflicting emotions in complicated words (Frankel et al., 2022). It is less flexible when it comes to new terms, slang, or dynamic areas because it relies on a predefined lexicon and considers all words equally, disregarding differences in intensity. It is also

language-specific, necessitating the use of distinct dictionaries for multilingual analysis. Because of these drawbacks, it is less reliable and accurate (Frankel et al., 2022).

2.7.1.2 Bag-of-Words (BOW)

Bag-of-Words (BoW) is a text representation, which transforms text into a structured format. Each dimension corresponds to a distinct word in the vocabulary, and it depicts text as a vector of word frequencies or binary occurrences. Word order and context are ignored, yet machine learning models can interpret the text thanks to this organized representation (Malik et al., 2024). A study by Shahzad et al. (2021) applied BoW and labelling the sentiment as positive, negative, and neutral for sentiment feature. The sentiment feature is then combined with the Bitcoin close price to predict several few days of Bitcoin future price using Deep Neural Network (DNN) regressor and obtained an RMSE of 6.21. This shows a significant result of applying the text representation technique compared to the study by Inamdar et al. (2019) that only directly utilised deep learning model to perform the sentiment analysis, and yielded higher RMSE of 13.70.

2.7.1.3 Term Frequency-Inverse Document Frequency (TF-IDF)

A statistical metric called TF-IDF is another form of text representation (Robertson & Jones, 1976). It is used to assess a word's significance in a document in relation to a corpus, or group of documents. Inverse Document Frequency (IDF), which lessens the significance of common terms that occur in several documents, and Term Frequency (TF), which gauges how frequently a word occurs in a document, are its two main components (Ramos, 2003). Words that appear frequently in a particular document but infrequently throughout the corpus receive better scores according to the algorithm, which increases TF by IDF. TF-IDF is frequently used for tasks like document ranking or classification in text mining and information retrieval to represent text as numerical vectors, highlighting keywords that are more specific to a document (Fakharchian, 2022).

In a study by Rateb et al. (2024), TF-IDF was performed on Twitter data for sentiment classification with the deployment of SVM model. Based on the sentiment

evaluation, it gained an accuracy of 79%. Such result is the lowest in comparison to other sentiment feature extraction techniques applied with accuracy ranging from 82% to 92% using CNN-LSTM, SVM+GloVe, and *pysentimento*. From the results, it can be inferred that *pysentimento*, based on RoBERTa (a contextual word embedding) provides the most reliable sentiment classification performance.

2.7.2 NLP with Word embedding

Words can have their meanings captured via dynamic vector representations called contextual word embeddings, which are dependent on the context of a sentence. Contrary to traditional embeddings, which give each word a fixed vector, contextual embeddings adjust to various word usages, addressing ambiguity and polysemy. These embeddings are produced by models such as BERT, XLNet, and GPT, which analyse complete sentences or passages to record complex, context-sensitive relationships, for example, "bank" in "river bank" versus "financial bank". These embeddings offer a richer comprehension of language semantics, making it extremely useful for tasks like sentiment analysis, translation, and question answering (Asudani et al., 2023). Table 2.4 shows the summarized difference between traditional word embedding and contextual word embedding.

Table 2.4
Traditional Word Embedding Vs Contextual Word Embedding

Traditional Word Embedding	**Contextual Word Embedding**
Cannot capture polysemy (treat the same word as same meaning in different context).	Capture polysemy (same word different meaning).
A shallow model that integrates previous knowledge in the first layer of the model in which the rest of the network still needs to be trained from scratch for a new target task (only taking vectors as output).	Taking the trained model and vectors as the output.
Fail to capture higher-level information that might be more significant.	Able to understand higher-level concepts such as anaphora, long-term dependencies, agreement, and negation.
Example: Word2Vec, GloVe	Example: BERT, XLNet, GPT

Contextual word embeddings encapsulate semantic linkages and contextual significance, positioning analogous words in proximity within the vector space (Kaur

& Sharma, 2023). In contrast to the traditional NLP approaches that regard words as discrete tokens, word embeddings are more adaptable and efficient for nuanced sentiment analysis because it can handle synonyms, polysemy, and context-dependent meanings. The approach has significantly enhanced the efficacy of tasks such as machine translation, text production, and sentiment analysis, by allowing models to attain a substantially elevated level of language comprehension.

Word embeddings can be categorized into two types; which are, traditional word embedding and contextual word embedding. Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) are popular examples of traditional word embeddings, which are static vector representations of words in a dense vector space. Each word is mapped to a fixed vector according to how it appears in a huge text corpus. The words are context-independent because each word in these embeddings has the same vector regardless of its context, which restricts their capacity to handle polysemy or complex meanings. The traditional word embedding can capture the syntactic and semantic links between words quite well, but unable to adjust to phrases or context. The approach offers simple and effective implementation as opposed to the traditional embeddings that lacks the dynamic, context-aware features of more modern embeddings such as BERT (Asudani et al., 2023). Further descriptions are presented in the next subsection.

2.7.2.1 Word2Vec

Word2Vec emphasizes local co-occurrence patterns in text by using shallow neural networks to train embeddings by either predicting the context given a word Continuous Bag-of-Word (CBOW) or predicting a word given its context (Skip-Gram) (Mikolov et al., 2013). Mohanty et al. (2018) used Word2Vec word embedding with LSTM for the sentiment model on Twitter data to generate the sentiment scores. Then, the sentiment and price features of open and close prices, trading volume, market cap, and blockchain data are fed into Bi-LSTM for price prediction. The accuracy of 50% is achieved, denoting a low accuracy with only 60% precision. These findings have demonstrated that the model is susceptible to improvement, in which additional useful input data may contribute to the prediction performance.

2.7.2.2 GloVe

In contrast, Global Vectors for Word Representation (GloVe) uses both local context and global statistics to produce word vectors by factorizing a word co-occurrence matrix (Pennington et al., 2014). GloVe explicitly models word co-occurrence, whereas Word2Vec concentrates on word prediction. While both are effective, different approaches take different techniques to capture word associations (Asudani et al., 2023).

GloVe was used in the studies by Attila (2017) and Jahjah & Rajab (2020). From the sentiment model evaluation, an MSE of 0.927 was achieved in Attila (2017) with GloVe and Convolutional Neural Network (CNN) on 100 tweets per hour. The sentiment feature was fed into the cryptocurrency price prediction model along with the cryptocurrency historical prices and obtained an MSE of 0.043. The low error signifies the good performance of the cryptocurrency price prediction. Similar application of GloVe combined with CNN-LSTM has shown significant sentiment classification result with an accuracy of 88.7% in generating the sentiment scores from Twitter (Jahjah & Rajab, 2020). The sentiment feature was then combined with cryptocurrency historical OHLC prices and gained an accuracy of 78% for the cryptocurrency price prediction model (Jahjah & Rajab, 2020). The result shows that sentiment score can help in achieving better accuracy in predicting the cryptocurrency prices.

Since the research studies by Attila (2017) and Jahjah & Rajab (2020) used almost the same number of input data as Mohanty et al. (2018), it can be observed that the number of input features does not necessarily improve the performance of the prediction model. Therefore, enhancing feature quality can have a favourable effect on model performance, and selecting the right combination of input features can both enhance model performance and lessen the detrimental effects of the curse of dimensionality (Ji et al., 2022).

2.7.2.3 Contextual word embedding

The traditional word embedding models are lacking in out-of-vocabulary words coverage and contributes to inadequate modelling of vague words because it does not support word polysemy issue (Gomez-Perez et al., 2020). Hence, contextual word

embedding model such as Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) and Generalized Autoregressive Pretraining for Language Understanding (XLNet) (Yang et al., 2019), a Transformer-based language model are recommended to overcome the deficiency in traditional word embeddings.

Word representations that consider the context in which they occur are known as contextual word embeddings. Contextual embeddings such as; BERT, XLNet or GPT modify the vector according to the surrounding words, in contrast to static (traditional) embeddings (e.g., Word2Vec and GloVe), which give each word a fixed vector. These embeddings are a significant advancement for managing the subtleties of language in tasks like sentiment analysis, question answering, or summarization. The embeddings are created dynamically during text processing (Miaschi & Dell'Orletta, 2020). The scope of a language model extends beyond contextual word embeddings. In order to comprehend and depict the links between words in context, it internally employs contextual embeddings. A language model is a system that can comprehend, produce, and forecast text to learn grammar, meaning, context, and patterns. A language model is trained on massive text datasets (Asudani et al., 2023). Table 2.5 summarizes the important features in contextual word embedding and language modelling.

Table 2.5
Contextual Word Embedding And Language Modelling

Features	Contextual Word Embedding	Language Modelling
Definition	A representation of words as dynamic vectors based on their context in a sentence.	A model trained to understand, predict, and generate text by learning language patterns.
Purpose	Encodes semantic meaning of words in context.	Performs tasks like text generation, translation, summarization, and question answering.
Scope	Focuses on representing individual words or tokens.	Operates at the level of sentences, paragraphs, or entire documents.
Dependency	Often used as part of a language model.	Uses contextual embeddings internally for understanding text.
Examples of Models	BERT (embedding layer), ELMo, GPT embeddings.	GPT, BERT, T5, LLaMA (full models)
Training Objective	Learns to represent words based on their context.	Learns language patterns by predicting the next word (causal) or masked words (masked language modelling).
Output	Provides numerical vectors for words based on context.	Provides predictions, probabilities, or generates coherent text.

Use Case Examples	Semantic similarity, clustering, word sense disambiguation.	Text generation, translation, summarization, sentiment analysis.

Table 2.5 compares contextual word embedding with language modelling, emphasising their unique functions in natural language processing. Contextual word embedding highlights the representation of individual words as dynamic vectors influenced by circumstances, functioning mainly as a component inside more extensive models. Conversely, language modelling functions at the phrase or document level, acquiring the ability to comprehend and produce cohesive content. Embeddings facilitate tasks such as semantic similarity and word sense disambiguation, whereas language models address more intricate tasks including text production, translation, and summarisation. Models such as BERT and GPT can operate in dual capacities, utilising the embeddings in one context and the full representations in another. Contextual embeddings establish the fundamental comprehension of word meaning within context, which language models utilise to execute advanced language tasks.

The contextual word embeddings or language modelling have begun to be applied in recent research studies concerning sentiment-based cryptocurrency price prediction (Haritha & Sahana, 2023; Sekioka et al., 2023; Passalis et al., 2022). One of the popular language models is Bidirectional Encoder Representations from Transformers (BERT). Cerda (2021) performed two types of experiments on sentiment classification to analyse sentiment and detect stance. Three labels are used such as *Support*, *Against*, and *None*. *Support* is referred to as positive, where the content shows bullish trend in the trading market, whereas *Against* represents negative, showing bearish trend in the trading market. *None* refers to neutral sentiment that shows factual content that does not express positive nor negative sentiment. For sentiment analysis, SentiWordNet was used to generate sentiment score ranging from -1 to 1 representing negative to positive sentiment. For stance detection, the pre-trained Universal Language Model Fine-tuning (ULMFiT) and BERT models were fine-tuned individually as the feature to perform the classification. A precision of 48% was achieved for sentiment analysis and 72% of precision was yielded for stance analysis. The result shows that the language models used for stance detection performed better than the non-language model applied for sentiment analysis. By using the generated sentiment feature, cryptocurrency price prediction was performed and produce an RMSE of 1.78 in the

prediction errors. The small error produced implies the potential of implementing language model for sentiment prediction and helps to improve the cryptocurrency price prediction model's performance.

Another study to support the good effect of sentiment scores generated by a language model in increasing the accuracy performance of cryptocurrency price prediction model is presented by Passalis et al. (2022). Passalis et al. (2022) employed BERT language model to generate the sentiment scores. The sentiment scores are classified and yielded 68% accuracy. Passalis et al. (2022) then feed the sentiment score and cryptocurrency past closing price as the features into the cryptocurrency price prediction model, and manage to obtain an accuracy of 92%.

In addition, Sekioka et al. (2023) performed classification task on cryptocurrency price prediction to predict the cryptocurrency market trend by applying LightGBM model and used BERT to generate the sentiment features automatically. However, the price prediction results only achieved an F1-score of 29.6%, that suggested an improvement is needed to gain higher accuracy.

Subsequently, this study employs the XLNet language model for sentiment analysis. The XLNet model is fine-tuned by training the manually labelled sentiment data. Then, the data is fed into the GRU deep learning model for sentiment analysis. This study implements XLNet and GRU layers for sentiment feature extraction due to its ability to handle the complexity and nuances of language. The XLNet captures contextual relationships bidirectionally while preserving the autoregressive nature of training, making it highly effective in understanding sentiment nuances, such as sarcasm, negations, or polysemy. By fine-tuning XLNet on manually annotated sentiment data, the model learns domain-specific sentiment patterns, ensuring tailored and accurate predictions. The integration of GRU layers further enhances the analysis by efficiently processing sequential information from the embeddings, reducing computational overhead compared to other recurrent architectures while maintaining strong performance. The implementation leverages on the contextual power of XLNet and the sequential modelling efficiency of GRU, providing a robust and scalable solution for sentiment analysis.

2.7.3 Types of Sentiment Features Extracted

From the extraction of sentiment features using various NLP techniques, different sentiment features are extracted to capture the emotional tone and polarity of the text. These features are then used as an input to the models, which categorize the sentiment of either positive, negative, or neutral or generate the sentiment polarity into a range of values (e.g., from -1 to +1) to infer market trends. By associating the sentiment data with price movements, the approach enables the prediction of crypto prices based on market sentiment and behavioural patterns. Typically, positive and negative counts and sentiment scores are extracted for the cryptocurrency price prediction.

Sentiment scores has shown to have a positive effect as a feature in addition to historical prices such as the typical open-high-low-close (OHLC) prices (Hitam et al., 2021; Pillai et al., 2021; Schulte & Eggert, 2021; Jahjah & Rajab, 2020; Kilimci, 2020; Pathak & Kakkar, 2020; Wang & Chen, 2020; Vo et al., 2019; Lamon et al., 2017), only close price (Critien et al., 2022; Passalis et al., 2022; Sattarov et al., 2020; Yao et al., 2019), cryptocurrency exchange rate (Galeshchuk et al., 2018; Stenqvist & Lönnö, 2017; Wołk, 2019), market cap (Cerda, 2021; Pathak & Kakkar, 2020; Wooley et al., 2019), price return (Nasekin & Chen, 2020), gold prices (Aggarwal et al., 2019), technical indicators (Ortu et al., 2022; Ye et al., 2022) and trading volume (Edgari et al., 2022; Hasan et al., 2022; Oikonomopoulos et al., 2022; Zamani et al., 2022; Loginova et al., 2021; Poongodi et al., 2021; Luo, 2020; Li et al., 2019) in performing cryptocurrency price prediction.

However, some experiments achieved only partial correlation between sentiment score and the cryptocurrency price (Alghamdi et al., 2022; Glenski et al., 2019). Alghamdi et al. (2022) reported that the sentiment feature is useful for Bitcoin but not so much for Ethereum. Research by Cerda (2021) discovered that using the historical prices only as the feature or using both historical prices and sentiment score in the experiment produced similar predictive performance. The lack of information resources had been claimed to be the cause of the outcome (Cerda, 2021). The limitation also exists in Lamon et al. (2017) in which the experiment did not combine both news and social media data when deriving the sentiment score, thus causing inaccurate predictions that were not coherent with the general trend.

Inamdar et al. (2019) argued that there is no correlation between the sentiment score and cryptocurrency price due to the huge number of neutral sentiments. Attila (2017) also expressed that the sentiment score cannot be the only feature to enhance the cryptocurrency price prediction model.

However, some studies also highlight the limitations of sentiment features. For example, studies like Mohanty et al. (2018) which used sentiment scores with BiLSTM, only achieved 50% accuracy, indicating that sentiment alone may not be enough to drive highly accurate predictions in certain settings. This suggests that while sentiment is an important additional signal, it may require careful handling and model tuning to fully capture its predictive power. Furthermore, simple models like Linear Regression or Decision Trees often show weaker performance compared to more complex models (e.g., LSTM or XGBoost), which suggests that more advanced methods are better suited to integrating sentiment data.

Tandon et al. (2021), Balfagih & Keselj (2019), and Mai et al. (2018) discovered a correlation between the positive and negative sentiment counts, and the historical open-high-low-close (OHLC) prices as well as the trading volume (Akba et al., 2021). The reported findings consistently reflected that selected sentiment features are statistically significant and useful to enhance cryptocurrency price prediction. The significant effect of positive and negative counts combined with historical prices in cryptocurrency price prediction is evident from the experimental result (Chahooki et al., 2023).

Other research studies have also found that both positive count or negative count is useful when combined with historical price features such as closing price (Shahzad et al., 2021), average price (Jain et al., 2018), percentage of price change (Kim et al., 2016), and market cap (Matta et al., 2015). According to Jain et al. (2018) and Matta et al. (2015), the positive count feature is more prominent in contrast to the negative count feature. However, Kaminski (2016) claimed that negative sentiment count has fairly good correlation with the price inclination. It is observed that the relationship between positive sentiment count and the upswing of cryptocurrency price is weak. Such finding is in line with Kim et al. (2016) study that positive sentiment count feature affected Bitcoin's price fluctuation more significantly, while negative sentiment count feature has more influence on Ethereum's and Ripple's price change.

In addition, Zhou et al. (2023) and Wooley et al. (2019) reported solid accuracy with VADER-derived sentiment, achieving 61% and 74% accuracy for Bitcoin and Ethereum price prediction respectively. The result suggests that VADER's ability to capture both positive and negative sentiments in textual data can be a powerful predictor when used alongside traditional market features, like trading volume and price data.

On the contrary, there are studies that indicate low correlation between both positive and negative sentiment counts for cryptocurrency price prediction. Steinert & Herff (2018) combined past cryptocurrency exchange rate and trading volume features with positive and negative sentiment counts features to predict cryptocurrency price. From the result, it was observed that even though sentiment features have significant effect on the price returns, not all cryptocurrencies are affected by social media sentiment. Despite having sentiment correlated with cryptocurrency prices, the combination of the positive and negative counts with historical closing price (Mohapatra et al., 2019; Yao et al., 2019) or price return (Mai et al., 2015) did not help improve price prediction performance.

Another notable insight derives from studies that combine sentiment with additional factors, such as user affinity, blockchain data, or technical indicators. Edgari et al. (2022) achieved an impressive 88% accuracy using sentiment features in conjunction with market cap and trading volume, showcasing how sentiment can act as a valuable augmenting factor when combined with broader market context. Additionally, Ye et al. (2022) shows that integrating VADER sentiment with technical indicators using an ensemble model (LSTM and GRU) leads to improved performance, indicating that sentiment data may provide valuable insights when paired with other factors influencing price movements.

Furthermore, sentiment analysis tools like TextBlob and Word2Vec are also used and show mixed performance. TextBlob contributes to improve model performance in some cases but results in higher Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE) when used in isolation or without proper data integration (Pathak & Kakkar, 2020; Prajapati, 2020). The finding suggests that while TextBlob is effective for basic sentiment scoring, more sophisticated models like LSTM or XGBoost are needed to fully leverage sentiment data for price prediction.

Limited previous researches utilised the combination of the three sentiment features of positive counts, negative counts and sentiment score to perform

cryptocurrency price prediction. Maqsood et al. (2022), Chen et al. (2019), and Valencia et al. (2019) applied all three sentiment features with the past OHLC prices and transaction volume and reported that the proposed model performed well. Historical price return was employed with the three sentiment features, and it was found that the features were more robust in predicting cryptocurrency prices during the non-bubble period in comparison to the bubble period. The related study showing negative results is yet to be found.

Overall, these findings underscore the potential of sentiment features to boost predictive accuracy in cryptocurrency price forecasting, but the degree of success depends heavily on the combination of sentiment analysis tools, the type of machine learning models used, and the integration of other relevant market data.

2.8 Classification for Cryptocurrency Price Prediction

Classification is the process of distinguishing data into multiple categorical classes in discrete values and has the advantage of flexibility in utilising different models in machine learning (Chen et al., 2019). In cryptocurrency price prediction, classification tasks aim to predict discrete outcomes, such as whether a cryptocurrency price will rise, fall, or remain the same, based on historical data and market indicators. The finding can assist to either buy, sell or retain the cryptocurrency assets.

In order to classify price movements into discrete classes, such as buy, sell, or hold based on historical data and market indications, traditional machine learning algorithms are used in cryptocurrency price prediction. Through the evaluation of features such as trading volume, sentiment in the markets, and technical indicators classification allows analysts to discern patterns and trends in price fluctuations and help traders and investors make better decisions, optimize trading strategies, and manage risk by pointing out possible future price movements. In the end, the method can lead to more successful investment strategies in the erratic cryptocurrency market.

In terms of model selection, deep learning techniques, particularly LSTM networks, are increasingly preferred for their ability to capture temporal dependencies in time-series data. These models excel at recognizing patterns over time and have shown strong performance in cryptocurrency forecasting. For example, LSTM-based models used in Pant et al. (2018) and (Vo et al., 2019) demonstrate that deep learning

can offer more accurate predictions when trained on large volumes of market data. The features used can be traditional data such as historical prices and trading volume combined with sentiment data.

2.8.1 Price Prediction Classification without Sentiment Features

Akyildirim et al. (2020) used Support Vector Machine (SVM) to predict the direction of cryptocurrency return and obtained 69% accuracy using 11 main features; which are, the open, close, high, low prices, high-low range, number of trades in the given close to open time intervals, US dollar denominated volume, number of cryptocurrencies traded (cryptocurrency volume) for all trades and similarly number of trades, US dollar trade volume, and cryptocurrency volume for buyer initiated trades. It was also reported that by only using the past Open (O), High (H), Low (L), and Close (C) prices, the price return of cryptocurrency cannot be predicted efficiently. Additional price features such as technical indicators, trading volume, and/or market capitalization are also used in several prior research studies (Jung et al., 2023; Shintate & Pichl, 2019; Sin & Wang, 2017) to investigate for any performance gain in predicting the cryptocurrency price movements. The findings show accuracy of between 53.53% to 90.58%, which shows better performance than using the OHLC prices only.

Another study using the OHLC prices and technical indicators was conducted by Sun et al. (2020) and forecasted cryptocurrency price trend for binary class of falling or not falling with the application of Light Gradient Boosting Machine (LightGBM). The proposed approach obtained an accuracy of 60.7% for 2-weeks prediction and 47.6% for 2-days prediction. It can be observed that the 2-day prediction achieved low accuracy reflecting the inefficiency of the prediction task. This was because more fluctuation patterns and the strength of the investment cycle can be seen in 2-weeks prediction compared to the 2-day prediction, which allows the model to learn the trend more adequately to perform the prediction.

Studies by Mallqui & Fernandes (2019) and Greaves & Au (2015) applied the same price features of OHLC with additional technical indicators to predict Bitcoin price movements. Machine Learning Ensemble algorithm, which combined Recurrent Neural Network (RNN) and tree classifiers are utilised by Mallqui & Fernandes (2019) with an accuracy of 62.91%, while Greaves & Au (2015) applied neural network and

achieved 55% of accuracy. The observation suggests that the more complex ensemble approach, which integrates multiple machine learning models, provides a better performance in terms of predictive accuracy compared to the neural network.

On the contrary, Borges & Neves (2020) utilised the OHLC prices as the features and fed into an ensemble learning model consisting of four different classifiers combination; which are: Logistic Regression, Random Forest, Gradient Tree Boosting, and Support Vector Machine (SVM). The accuracy of 59.26% was obtained. The result shows that the relevant reliable learning models combination may also contribute to the enhancement of the price prediction performance.

Subsequently, Madan et al. (2015) predicted the sign of daily cryptocurrency price change using 16 features from a total of 26 features relating to Bitcoin trading and the Bitcoin network. The features are fed into Random Forest classifier and achieved a high accuracy of 94.98%. Therefore, the prior experimental works indicated that the number and type of features used to perform classification task on cryptocurrency price prediction is crucial for machine learning implementation to improve the predictability performance. Besides the implementation of machine learning models, deep learning models are also explored in previous related research studies.

A study of analysing the Bitcoin market predictability for four different prediction horizons which are 1-minute, 5-minutes, 15-minutes, and 60-minutes intervals was conducted by Jaquart et al. (2021) by applying open, high, low, and close (OHLC) prices and technical indicators into Long Short-Term Memory (LSTM) deep learning model to analyse the predictability of Bitcoin market across prediction horizons ranging from 1 to 60 min. From the experiment conducted, the accuracy ranging from 51% to 56% were yielded for the two classes. In a binary classification problem, random guessing would yield an accuracy of 50%. Therefore, the accuracy range of 51% to 56% is slightly better by 1% to 6%. The result suggests that the model is providing a modest improvement over random predictions, but it also indicates that the model's performance could be enhanced further.

Detailed results of the study are as follows: 51.93% (1-minute), 52.49% (5-minutes), 53.2% (15-minutes), and 56% (60-minutes) for binary classification. This study shows that short-term fluctuations in the Bitcoin market can be predicted by machine learning algorithms. It is evident from the forecasting accuracy of just over 50% that there is some degree of predictability in the Bitcoin market. This could be

caused by a variety of factors, such as an instantaneous market response to the features that were used or a potentially substantial amount of information that affects the Bitcoin market in addition to these features. The findings confirmed that the implementation of solely OHLC prices and technical indicators features may not result in reliable cryptocurrency price prediction performance.

Similar results were also recorded in Shintate & Pichl (2019) work in predicting the trend of cryptocurrency that are non-stationary using the same price features with the implementation of Random Sampling Method deep learning model (Jung et al., 2023). The accuracy of 53.53% was obtained for three classes of up, down, and static. However, the technical indicators still did not show any significance in improving the performance of predicting the cryptocurrency price movements.

A breakthrough in the experimental result is observed with promising outcomes of an accuracy of 92.4% in predicting the trend of cryptocurrency price with an additional feature of market capitalization (market caps) besides the open, high, low, close prices, and trading volume (Chowdhury et al., 2020). The market cap shows the popularity of the cryptocurrency since higher market caps often have greater liquidity and higher trading volume (Leirvik, 2022).

Moreover, Sin & Wang (2017) and McNally et al. (2018) predicted the direction of Bitcoin prices using historical prices with technical indicators and market cap, and then implemented Multilayer Perceptron (MLP) and LSTM, achieving 60% and 52% accuracy respectively. Sin & Wang (2017) used features from Blockchain.info and Bitcoinity.org such as the market price, transaction volume, hash rate, market capitalisation, revenue, and rank of each exchange. The experimental result indicates that the cryptocurrency features consist of information that is useful to predict the next-day price level of Bitcoin (Sin & Wang, 2017).

On the other hand, McNally et al. (2018) used open, high, low, close prices from CoinDesk, with additional features of difficulty and hash rate taken from the Blockchain site. McNally et al. (2018) stated that from the findings, deep learning models are apparently effective to predict Bitcoin price change, especially the ability of LSTM model in understanding longer-term dependencies, but there is still some issue of overfitting a model, leading to an insufficient learning which requires a more suitable optimization implementation. Moreover, a larger amount of data or longer duration of

data was also suggested by McNally et al. (2018) to improve the learning of the deep learning model.

In addition, a multiple combination of deep learning models in Ensemble Learning Method were used to perform the prediction task (Livieris et al., 2020). The work aims to classify the direction of the price for the following hour using the historical OHLC price data. Then, ensemble learning combining CNN-LSTM and CNN-BiLSTM were implemented. The achieved highest accuracy of 54.52% (CNN-LSTM) and 54.62% (CNN-BiLSTM) for Bitcoin prediction are presented. The finding indicates that while the ensemble deep learning models slightly improve prediction accuracy over random guessing, which would be about 50% for a binary classification task, the accuracy remains relatively modest, suggesting that further refinement of the models or the use of additional features might be needed to enhance prediction performance.

2.8.2 Price Prediction Classification with Sentiment Features

The research results from the various studies provide valuable insights into the integration of sentiment analysis and machine learning models for cryptocurrency price prediction. The most notable takeaway is the significant role that sentiment plays in improving the accuracy of predictions.

The findings underscore the growing trend of combining sentiment analysis with technical indicators such as OHLC prices and trading volume to enhance predictive models. This integration enables models to capture both market sentiment and technical trends, which are often interrelated in the volatile world of cryptocurrencies.

A research study by Jung et al. (2023) utilised sentiment features in addition to the technical indicators to predict cryptocurrency price. The added sentiment feature was included since several research studies have suggested a correlation between sentiment and cryptocurrency price changes, in which the public's behaviour in trading cryptocurrency usually influenced by news or social media (Jung et al., 2023; Farimani et al., 2022). Based on the findings in Jung et al. (2023), it shows an accuracy of 90.58% using the Extreme Gradient Boosting (XGBoost) to predict Bitcoin price trends. This infers that additional features are significant for classification task in predicting cryptocurrency market trends.

A study by Ibrahim (2021) highlights the effectiveness of ensemble models in which XGBoost is combined with multiple decision trees to improve predictive performance. The finding shows that combining sentiment features with XGBoost model yields strong results, achieving F1-score of 88.8%. The model outperformed simpler models like Logistic Regression (54.2%), SVM (59.8%), Naïve Bayes (66%), and XGBoost (79.9%). The result achieved was also supported by Edgari et al. (2022), who also used XGBoost and obtained an accuracy of 88%. In conclusion, advanced ensemble models like XGBoost are preferable for sentiment-driven prediction tasks due to their improved performance and ability to handle sentiment data complexity.

Studies by Jahjah & Rajab (2020) and Lamon et al. (2017) utilised sentiment scores with OHLC prices into Multiple Linear Regression and Logistic Regression respectively. Based on the results, Jahjah & Rajab (2020) achieved an accuracy of 78%, while Lamon et al. (2017) obtained accuracy of between 24.2% to 75.8%. The observation indicates that Multiple Linear Regression can better capture the relationship between sentiment and OHLC prices. On the other hand, Logistic Regression may struggle with the complexities of integrating sentiment data. Consequently, the findings can be deduced that Multiple Linear Regression may provide more consistent performance when incorporating sentiment scores, while logistic regression may need additional optimization or may lack reliability across various circumstances in sentiment-based predictions.

Another study by Stenqvist & Lönnö (2017) also used sentiment scores with OHLC price as features to predict the cryptocurrency market. Naïve Bayes classifier was applied and achieved an accuracy of 83.3%. This shows that it can handle the probabilistic relationships inherent in sentiment analysis well and offers a more stable and effective solution for sentiment-driven predictions.

Similarly, a study by Luo (2020), used sentiment scores aggregating with OHLC prices and trading volume, then fed into Decision Tree classifier. An accuracy of 76% was achieved to observe the cryptocurrency price movement. Although the Decision Tree model demonstrates commendable performance, there remains potential for enhancement. An accuracy of 76% indicates that sentiment significantly influences market behaviour, and its incorporation with standard market indicators such as OHLC prices and trade volume provides essential context for predictive models.

Using the same sentiment and price features as Luo (2020), the study by Zhou et al. (2023) implemented SVM classifier to perform the cryptocurrency price prediction and obtained an accuracy of 61%. This suggests that while SVM can be effective, it may not perform as well as some other models in handling sentiment data alongside market features, especially when compared to models like Decision Trees or XGBoost. SVM might not be the best model for integrating sentiment features with market data in this specific task, or that it might require more hyperparameter tuning or data preprocessing to achieve better performance.

A study by Sattarov et al. (2020) used sentiment scores and close price only for price prediction using Random Forest classifier. The result is not very outstanding with an accuracy of 62.48%. This result implies the same as the study by Luo (2020). The classifiers may require further tuning or additional features to improve their performance. This emphasizes the need for more complex models, which can be better suited for handling sentiment and market data in this context. Moreover, sentiment analysis serves as a crucial component for market forecasting, but the selection of the machine learning model profoundly impacts the reliability and precision of the outcomes.

The commonly used deep learning model, LSTM was implemented in the studies by Pillai et al. (2021) and Pant et al. (2018). Both studies utilised sentiment scores and historical price data as features. Pillai et al. (2021) used solely LSTM, while Pant et al. (2018) applied RNN-LSTM-GRU hybrid model to perform cryptocurrency price prediction. The findings show an accuracy of 82.34% and 77.62% respectively. The performance suggests that LSTM is a strong model for capturing the sequential dependencies in cryptocurrency market data, where both historical prices and sentiment play a role. In addition, although hybrid models such as RNN-LSTM-GRU can include many architectures, simpler models like LSTM may occasionally produce better results. The research highlights the significance of sentiment analysis in predicting cryptocurrency prices.

However, a study by Pathak & Kakkar (2020) shows contradicting result when utilising LSTM to predict the price with accuracy of 77.89%. Such a different result may be the result of the usage of user affinity features. User affinity includes number of likes, number of followers, and/or number of reposts. Moreover, the study also employed market capitalization as an added feature. The lower accuracy suggest that

incorporating excessive features, such as user affinity and market capitalization does not inherently enhance model performance. The finding indicates that meticulous feature selection is crucial. The inclusion of irrelevant features could impede the performance and cause overfitting.

Studies by Critien et al. (2022) and Mohanty et al. (2018) both used BiLSTM model to predict the cryptocurrency prices with sentiment score. However, the price data taken by each study is different. Critien et al. (2022) used closing price only, while Mohanty et al. (2018) utilised open and close price, with trading volume, market capitalization, and blockchain data. The results achieved by the studies are 63% and 50% in accuracy respectively. Adding more features such as market capitalization or blockchain data might have made the model more complex than necessary, leading to poorer generalization on unseen data. This highlights the importance of carefully selecting relevant features and avoiding adding too many unnecessary variables that may not contribute positively to model performance.

Valencia et al. (2019) experimented on sentiment scores, positive and negative counts as the sentiment features incorporated with the OHLC prices with trading volume as the price features. The features are fed into MLP model for price prediction and obtained accuracy = 72% for Bitcoin prediction and accuracy = 44% for Ethereum. The result demonstrates that MLP serves as an effective model for combining sentiment and market data to forecast cryptocurrency prices, exhibiting superior performance for Bitcoin that yielded 72% accuracy as opposed to Ethereum 44% accuracy. The findings indicate that the distinct attributes of cryptocurrency, particularly Bitcoin is more consistent to price fluctuations might be more susceptible to forecasting using sentiment data. The subpar performance of Ethereum suggests that various cryptocurrencies may necessitate distinct feature sets or models to attain high accuracy in price prediction.

Table 2.6 and Table 2.7 present the summary of classification for cryptocurrency price prediction using various features performed in prior research studies.

Table 2.6
Summary Of Classification For Cryptocurrency Price Prediction Without Sentiment Features Performance Results

CLASSIFICATION FOR CRYPTOCURRENCY PRICE PREDICTION (USING PRICE FEATURES)			
Authors	**Price Prediction**		**Cryptocurrency Price Prediction Results**
	Features	**Model**	
Madan et al. (2015)	16 features relating to Bitcoin trading and the Bitcoin network	Random Forest	Accuracy = 94.98%
Greaves & Au (2015)	OHLC prices + technical indicator	Neural Network (Classification), Linear Regression (Regression)	Accuracy = 55%, MSE = 1.94
Sin & Wang (2017)	OHLC prices + blockchain	MLP	Accuracy = 60%
McNally et al. (2018)	OHLC prices + blockchain	LSTM	Accuracy = 52%
Mallqui & Fernandes (2019)	OHLC prices + technical indicator	ML Ensemble Algorithm (Classification)	Accuracy = 62.91%
Shintate & Pichl (2019)	OHLC prices + technical indicator	Random Sampling Method	Accuracy = 53.53%
Akyildirim et al. (2020)	OHLC prices + trading volume	SVM	Accuracy = 69%
Borges & Neves (2020)	OHLC prices	Ensemble Voting	Accuracy = 55.61%
Chowdhury et al. (2020)	OHLC prices + trading volume + market cap	Ensemble Learning Method	Accuracy = 92.4%
Livieris et al. (2020)	OHLC prices	CNN-BiLSTM	Accuracy = 54.62%
Sun et al. (2020)	OHLC prices + technical indicator	LightGBM	Accuracy = 60.7% (2-week), Accuracy = 47.6% (2-day)
Jaquart et al. (2021)	OHLC prices + technical indicator	LSTM	Accuracy = 56% (1-hour)
Jung et al. (2023)	OHLC prices + trading volume + technical indicator + market cap	XGBoost	Accuracy = 90.58%

From Table 2.6, it is observed that the performance measurements vary significantly across studies, with accuracy ranging from low accuracy of 61% for SVM to exceptionally high accuracy of 94.98% for Random Forest as reported by Madan et al. (2015). The results indicate that more sophisticated models, particularly those utilizing deep learning techniques such as LSTM and XGBoost, generally yield superior predictive accuracy, especially when integrated with extensive variables, including sentiment analysis and various market indicators.

Table 2.7
Summary Of Classification For Cryptocurrency Price Prediction With Sentiment Features Performance Results

CLASSIFICATION FOR CRYPTOCURRENCY PRICE PREDICTION (USING SENTIMENT AND PRICE FEATURES)					
Authors	**Sentiment Analysis**		**Price Prediction**		**Cryptocurrency Price Prediction Results**
	Features	**Model**	**Features**	**Model**	
Lamon et al. (2017)	Sentiment scores	Logistic regression	OHLC prices	Logistic regression, Bernoulli Naive Bayes	Bitcoin: Accuracy = 43.9% - 56.1%, Ethereum: Accuracy =75.8% - 24.2%
Stenqvist & Lönnö (2017)	Sentiment scores	VADER compound score	OHLC prices	Naïve Bayes classifier	Accuracy = 83.3%
Mohanty et al. (2018)	Sentiment scores	Word2Vec (word embedding) + LSTM	OC prices + market cap, trading volume, blockchain data	BiLSTM	Accuracy = 50%
Pant et al. (2018)	Sentiment scores	Bag-of-words + Random Forest classifier	historical prices	RNN-LSTM+GRU	Accuracy = 81.39% (Sentiment evaluation), Accuracy = 77.62% (Price prediction)
Valencia et al. (2019)	Pos & Neg counts, Sentiment scores	VADER positive, negative, neutral, and compound score	OHLC prices + trading volume	MLP	Accuracy = 72% (Bitcoin), Accuracy = 44% (Ethereum)
Jahjah & Rajab (2020)	Sentiment scores	GloVe (word embedding) + CNN-LSTM	OHLC prices	Multiple Linear Regression	Accuracy = 78%
Luo (2020)	Sentiment scores	VADER compound score	OHLC prices + trading volume	Decision Tree	RMSE = 0.0214 (log return), Accuracy = 76% (price movement)
Pathak & Kakkar (2020)	Sentiment scores + user affinity	Textblob	OHLC prices + market cap , trading volume	LSTM	Accuracy = 77.89%
Sattarov et al. (2020)	Sentiment scores	VADER compound scores	Close price	Random Forest Classifier	Accuracy = 62.48%
Ibrahim (2021)	Sentiment scores	VADER compound score	Close price + technical indicator	XGBoost-Composite ensemble model (CEPM)	F1-score = 88.8%
Pillai et al. (2021)	Sentiment scores	(VADER positive, negative, neutral and compound score) + Naïve Bayes	historical prices	LSTM	Accuracy = 82.34%, RMSE = 462.39
Critien et al. (2022)	Sentiment scores	VADER positive, negative scores	Close price	BiLSTM	Accuracy = 63%
Zhou et al. (2023)	Sentiment scores	VADER positive, negative scores	OHLC prices + trading volume	SVM	Accuracy = 61%

Table 2.7 highlights that integrating sentiment features such as VADER scores, TextBlob polarity, and word embeddings with market features such as the OHLC prices, trading volume, and market capitalization consistently improves cryptocurrency price prediction performance. Studies using deep learning models such as LSTM, BiLSTM, and CNN-LSTM achieved notably higher accuracies than traditional classifiers. The results demonstrate that sentiment information provides imperative insights into market behaviour, allowing models to better capture the sentiment influencing price movements. Hence, the effectiveness of these models depends on the type of features used, the amount of data, and the complexity of the cryptocurrency market being predicted.

2.9 Regression for Cryptocurrency Price Prediction

Regression prediction task refers to the process of predicting a numerical value. It is used to predict the future price of cryptocurrency in numerical form instead of in categorical values. By using regression, it can identify which of the independent variables included are most strongly associated with changes in the cryptocurrency prices (Biedscheid, 2023). While classification is useful for trend prediction and decision-making (e.g., buy or sell), regression provides more precise price forecasts. Instead of using the recognised metric of accuracy in classification task, regression results are reported in error metrics for predictions (Chen et al., 2019). The performance of different classifiers and regressors in predicting cryptocurrency prices varies significantly, with the key metrics of accuracy, RMSE and r^2 (Coefficient of Determination) providing insight into how well these models can forecast price movements.

Traditional regression machine learning algorithms for cryptocurrency price prediction generally encompass methods such as Linear Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM). It is employed to examine historical price data and various market indicators, including trading volume and technical indicators. Linear Regression models presume a linear correlation between characteristics and price, rendering them straightforward yet constrained in their ability to capture intricate patterns. Decision Trees and Random Forests exhibit greater flexibility, as they may represent non-linear connections and accommodate interactions

among features. Support Vector Machines for regression problems seek to identify the optimal hyperplane for forecasting continuous values.

Deep learning regression models, including Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Deep Neural Networks (DNN), have garnered considerable interest in cryptocurrency price prediction owing to their capacity to identify intricate, non-linear relationships within extensive datasets. These models are proficient in analysing sequential data, including previous prices and trade volumes, and can incorporate variables such as sentiment scores from social media, news, or other sources. LSTM models are especially proficient in managing time-series data, enabling them to consider the temporal dependencies in price fluctuations. Deep learning methods utilize numerous layers of neural networks to discern complex patterns and trends that conventional models, such as linear regression or decision trees, may overlook. Consequently, these models frequently surpass basic machine learning methods, providing enhanced forecast accuracy in volatile and dynamic markets like cryptocurrency. Nevertheless, deep learning models necessitate substantial training data and computational resources, rendering their implementation more intricate yet more potent when properly calibrated.

2.9.1 Price Prediction Regression without Sentiment Features

Erfanian et al. (2022) used technical indicators as an added price feature in predicting cryptocurrency prices and achieved a significant result of Pearson's r = 0.89, with a Root Mean Square Error (RMSE) of 8.68 (from a standard deviation of 2.48). From the results, it shows that macroeconomic and Blockchain information features such as the technical indicators are a significant long-term predictor.

Mudassir et al. (2020) and Mallqui & Fernandes (2019) performed the regression task on cryptocurrency price prediction using SVM model with OHLC prices and technical indicators features. With the normalized data, a Mean Absolute Percentage Error (MAPE) of between 1.28% and 1.91% (Mallqui & Fernandes, 2019), and MAPE = 1.44% (Mudassir et al., 2020) were achieved. The experimental results indicate a good result since there are minimal errors in the prediction using the stated price features. Furthermore, the study by Greaves & Au (2015) performed Bitcoin price prediction in 1-hour using linear regression model with OHLC prices and technical

indicators features. The reported result was in MSE metrics of MSE = 1.94 showing a good prediction result with minimal errors.

The positive results from using machine learning models like SVM and linear regression for cryptocurrency price prediction, with minimal errors such as low MAPE and MSE, demonstrate the effectiveness of traditional models in this domain. However, as the field progresses, there is growing interest in exploring deep learning techniques, which have the potential to further improve prediction accuracy by capturing more complex patterns in the data, leading to a more sophisticated approach in forecasting cryptocurrency prices.

As a result, technical indicators can be used to examine prior performances by taking into account both the indicators' performance and the derived margins of error. This scenario demonstrates how various indications and analysis conducted at various times can be used. It is crucial that experts give the topic the attention it deserves and carry out fresh research on various cryptocurrency coins, indicators, and time periods. Therefore, it is not recommended to solely use technical indicators to predict cryptocurrency prices. Various other predictors should be considered to improve the performance of the prediction models.

There are research studies that applied trading volume as an added feature to the OHLC prices in performing the regression task and produced a slightly better result than using the added technical indicators feature alone. Saxena et al. (2018) and Wu et al. (2018) used the closing price and trading volume as features to predict the Bitcoin daily prices by applying LSTM deep learning model. An RMSE of 456.78 (with minimum price = 0 and maximum price = 20,000) was achieved by Saxena et al. (2018), while an RMSE of 247.33 (with standard deviation = 2269.18 and mean closing price = 8840.64) was obtained by Wu et al. (2018). The result achieved by Wu et al. (2018) indicates a better predictive power in predicting Bitcoin price compared to Saxena et al. (2018)’s prediction performance. This is because Wu et al. (2018) implemented a hybrid LSTM, where it combines the conventional LSTM with AR(2) model, while Saxena et al. (2018) only used LSTM model to perform the prediction. This shows that by using hybrid model for price prediction, it can help improve its performance. For example, Wu et al. (2018) has shown that hybrid LSTM has an improved RMSE of 9.08 compared to the conventional LSTM. In addition, Saxena et al. (2018) also suggested that other

features besides using the closing price only should be applied to perform the cryptocurrency price prediction.

A research study by Kumar & Rath (2020) predict the price trends of Ethereum using LSTM and MLP, whereas Patel et al. (2020) used both LSTM and Gated Recurrent Unit (GRU)-based hybrid model to predict Litecoin and Monero prices. Both studies utilised the OHLC prices with trading volume, then performed normalization of (0,1) range on the price data. The normalization of the data to a (0,1) range is typically performed to ensure that the data is scaled uniformly, improving the efficiency and performance of the deep learning models. Normalization helps prevent issues related to large value ranges, making the training process more stable and allowing the models to better capture complex patterns in the data (Rodrigues & Machado, 2025).

Kumar & Rath (2020) performed experiments on daily, hourly and per minute price prediction comparing MLP and LSTM model performance. Based on the findings, for all three experiments, LSTM has found to be the best model with an RMSE of 20.53 (daily), 7.12 (hourly), and 18.16 (per minute). For daily experiment, only a slight difference of 0.77 in the RMSE when comparing between MLP and LSTM, but for hourly and per minute experiment, the loss difference between MLP and LSTM are obvious with the difference of 10.17 and 71.9 in the RMSE respectively. The hourly prediction has shown to be better among the three experiments and could be attributed to the fact that it strikes a balance between capturing meaningful price movements and managing the noise that becomes more prevalent in per-minute predictions. In shorter intervals, such as per minute, there is likely a lot of noise and rapid fluctuation that makes accurate predictions much harder, resulting in a higher RMSE. Therefore, LSTM's ability to manage and process these sequences of data proves most effective in the hourly time frame, where price patterns are still consistent enough to be predicted with lower error. This infers the importance of sample dimensions in machine learning algorithms (Ranjan et al., 2023; Kumar & Rath, 2020).

Patel et al. (2020) also experimented on three different sample dimensions of 1-day, 3-days, and 7-days price prediction using LSTM and hybrid model of the combination of LSTM and GRU. Based on the results obtained, the hybrid model has shown to have better prediction compared to using only LSTM model. RMSE results of 2.30 (Litecoin) and 3.27 (Monero) were achieved for daily prediction, as contrast to RMSE results of 2.03 (Litecoin) and 5.50 (Monero) 3-days prediction. Moreover, the

RMSE results of 4.55 (Litecoin) and 20.24 (Monero) were gained for 7-days prediction. The result indicated that the 3-days prediction produced the smallest error. Such a result aligns with the work by Wu et al. (2018) showing the improvement of prediction performance when combining more than one learning models and also supports the claim by Ranjan et al. (2023) that the sample duration taken plays a significant role in determining the quality of the predictive model. An acceptable RMSE result of 20.53 was obtained for daily data (Kumar & Rath, 2020) and RMSE = 2.299 (Litecoin) and RMSE = 3.272 (Monero) was achieved in Patel et al. (2020).

Moreover, a study by Phaladisailoed & Numnonda (2018) using the same price features model of OHLC and trading volume using GRU deep learning to predict Bitcoin prices using 1-min interval data. A very high precision of $r^2 = 0.992$, with a Mean Square Error (MSE) of 0.00002 was achieved. This may be due to the large dataset used for the 1-min interval (from 1 January 2012 to 8 January 2018). Phaladisailoed & Numnonda (2018) also mentioned that the past OHLC prices are useful but insufficient, hence more input features are needed.

Dutta et al. (2019) added more features to the OHLC prices, which are the technical indicators and Google Trends data to predict daily Bitcoin prices. With these features, GRU deep learning model was implemented and achieved a normalized RMSE of 0.016 for 15-days look back period experiment. The very low prediction error confirms the significance of the input features used in performing cryptocurrency price prediction for the regression task.

Another study by Schulte & Eggert (2021) using OHLC prices, trading volume, technical indicators, Google Trends and also additional sentiment features from news headlines and social media platforms of Twitter, Reddit, Facebook and GitHub. Four predictive models were implemented for the experiment which are LSTM, Bi-LSTM, Vector autoregression (VAR), and Seasonal Autoregressive Integrated Moving Average with Exogenous (SARIMAX). Among the four models, LSTM has shown to be the best with the least error prediction of 8.93% in average with Mean Absolute Percentage Error (MAPE) = 1.18%, while the other three gained more than 10% of error prediction. This shows that the variety of features is crucial to obtain a good performing price prediction model.

Yet, a study by Ji et al. (2019) using the same OHLC prices with technical indicator features but with the implementation of LSTM deep learning model to predict

the Bitcoin price achieved a slightly higher error in the prediction result (MAPE = 3.52%). This shows that although the same price features were utilised, the different prediction model used can produce different prediction results.

The studies highlight that incorporating diverse features like technical indicators, Google Trends data, and sentiment from social media can significantly enhance cryptocurrency price prediction accuracy. Dutta et al. (2019) and Schulte & Eggert (2021) demonstrated that adding these features improved prediction results, with LSTM performing the best. However, Ji et al. (2019) showed that even with similar features, different models such as LSTM can yield varying levels of accuracy. Overall, the research studies suggest that a comprehensive feature set, combined with the suitable model, is essential for accurate cryptocurrency price prediction.

Based on the results, most of the work that applied regression to perform cryptocurrency price prediction obtained a more precise prediction of the cryptocurrency prices with minimal errors and achieving up to r^2 result of 0.992 when using a larger dataset. Meanwhile, in majority of the existing research studies, the accuracy of cryptocurrency price prediction classification was usually between 55% to 65%. Hence, the focus of this study is on a regression task to predict cryptocurrency prices.

Even though various price features were used in addition to the historical OHLC prices, not all price features produce significant result in predicting the cryptocurrency prices. Thus, more new features are needed to explore in enhancing the cryptocurrency price prediction performance. From the observations in prior studies, the integration of sentiment features may be needed to enhance the reliability of the cryptocurrency price prediction model. By having mixed results in the consistency of the cryptocurrency price prediction classification and regression tasks, experimentation on sentiment features is required to address any insufficiencies that exist in the prediction as suggested in several studies to include sentiment features particularly those extracted from social media and news to improve prediction results (Vo et al., 2019; Yao et al., 2019).

2.9.2 Price Prediction Regression with Sentiment Features

A study by Inamdar et al. (2019) used LSTM model to directly perform multi-class classification with ranging sentiment scores from -2 to 2 on Twitter and news data, then mapped the sentiment features to Bitcoin price features to predict the price using Random Forest. RMSE of 13.70 was obtained from the experiment, and it was reported that the sentiment data do not have much impact on the Bitcoin price. Hence, a better-quality feature is needed to improve the evaluation.

A study by Jain et al. (2018) used Multiple Linear Regression model to predict cryptocurrency price with the utilisation of positive and negative counts as the sentiment features, and average historical price of Bitcoin. The findings show an adjusted r^2 of 0.426, indicating a moderate level of predictive power. In addition, the inclusion of positive and negative counts as sentiment features, along with historical price data, contributed to the model's performance. The relatively low adjusted r^2 suggests that other factors not captured by the model, such as market volatility, external news events, or advanced sentiment analysis techniques, could significantly influence the price of Bitcoin. This result implies that while Multiple Linear Regression can provide some insight into price movements, it may not fully capture the complexities of cryptocurrency markets.

A very low performance was shown in a study by Steinert & Herff (2018) and Salač (2019). Linear Regression was used to predict the price with sentiment scores as the sentiment feature. However, for the price features, exchange rate and trading volume were utilised by Steinert & Herff (2018), while historical prices was used by Salač (2019). An r^2 of 0.025 was achieved by Steinert & Herff (2018). In Salač (2019)'s work, very low r^2 of 0.095 was achieved if only using sentiment score from Reddit source, but when the sources of Reddit and Twitter were combined, the findings show r^2 result of 0.337. The results indicate that Linear Regression is not an optimal approach for forecasting Bitcoin values, particularly when utilizing sentiment data. The low r^2 values suggest that simpler models fail to capture the intricacies of Bitcoin market. Integrating other sentiment sources, as demonstrated in Salač (2019), can enhance the model's efficacy; nonetheless, more advanced methodologies may be necessary for achieving greater predictive accuracy.

Studies by Alghamdi et al. (2022) and Chin & Omar (2020) used LSTM model to perform price prediction with the application of sentiment scores and positive-negative counts as sentiment features respectively. Both studies aggregated the sentiment feature with OHLC prices and trading volume. In Alghamdi et al. (2022), RMSE of 0.503 for Bitcoin, and RMSE of 0.684 for Ethereum. The higher RMSE for Ethereum compared to Bitcoin suggests that the model may struggle more with Ethereum's price movements, which could be more volatile or influenced by factors not fully captured by the model. While Chin & Omar (2020) experiment obtained RMSE of 0.0471, indicating a better performance. This indicates that although LSTM algorithms proficiently capture the correlation between sentiment variables and market data, the accuracy of predictions can be significantly influenced by data quality, feature selection, and model optimization. The disparities in RMSE between the two studies indicate that, although employing analogous models, variances in data and methodology might result in varying performance levels.

Another good performance result of using LSTM to predict cryptocurrency price was shown in the studies by Agarwal et al. (2021) and Poongodi et al. (2021). Sentiment scores and OHLC prices with trading volumes are used as features by both studies, with RMSE of 0.02 and RMSE of 0.029 respectively, indicating highly accurate predictions. The low RMSE indicates that LSTM effectively captures the intricate, non-linear interactions between sentiment data and market indicators in the volatile cryptocurrency market. Poongodi et al. (2021) reported an r^2 value of 0.446, indicating that approximately 44.6% of the volatility in cryptocurrency prices can be analysed by the model, hence reinforcing its predictive efficacy. These findings highlight the efficacy of LSTM models in utilizing sentiment research alongside conventional market attributes to generate precise price forecasts, indicating that meticulous feature selection and model optimization are critical components for attaining robust performance.

Table 2.8 and Table 2.9 presents the summary of regression for cryptocurrency price prediction using price features and using the combination of sentiment and price features performed in existing research studies.

Table 2.8
Summary Of Regression For Cryptocurrency Price Prediction Without Sentiment Features Performance Results

REGRESSION FOR CRYPTOCURRENCY PRICE PREDICTION (USING PRICE FEATURES)			
Authors	**Price Prediction**		**Cryptocurrency Price Prediction Results**
	Features	**Model**	
Greaves & Au (2015)	OHLC prices + technical indicator	Linear Regression	MSE = 1.94
Saxena et al. (2018)	Close price + trading volume	LSTM	RMSE = 456.78 (min = 0, max = 20,000)
Wu et al. (2018)	Close price + trading volume	LSTM with AR(2)	RMSE = 247.33 (std deviation = 2269.18)
Phaladisailoed & Numnonda (2018)	OHLC prices + trading volume	GRU	r^2 = 0.992, MSE = 0.00002
Dutta et al. (2019)	OHLC prices + technical indicator + Google Trend	GRU	RMSE = 0.016 (15-days lookback period - normalized)
Ji et al. (2019)	OHLC prices + technical indicator	LSTM	MAPE = 3.52%
Mallqui & Fernandes (2019)	OHLC prices + technical indicator	SVM	MAPE = 1.28% - 1.91% (normalized)
Kumar & Rath (2020)	OHLC prices + trading volume	LSTM	RMSE = 20.53 (Daily - normalized)
Mudassir et al. (2020)	OHLC prices + technical indicator	SVM (Regression)	MAPE = 1.44% (normalized)
Patel et al. (2020)	Historical prices	LSTM & GRU-based hybrid	RMSE = 2.2986 (Litecoin), RMSE = 3.2715 (Monero)
Schulte & Eggert (2021)	OHLC prices + trading volume + technical indicator + Google Trend	LSTM	MAPE = 1.18%
Erfanian et al. (2022)	OHLC prices + technical indicator	SVR	Pearson's r = 0.89, RMSE = 8.68 (std deviation = 2.48)

From Table 2.8, the experiments that used deep learning models like LSTM and GRU can achieve strong predictive performance. The deep learning models outperform traditional regression models such as Linear Regression and SVM due to the ability to capture temporal dependencies and nonlinear relationships in time-series data. Based on the observation, the prior research studies that utilised additional features such as technical indicators or external factors like Google Trends further enhance accuracy, suggesting that incorporating diverse yet price-based inputs enhance prediction reliability. In addition, it can be seen that deep learning regression models are more effective and robust for cryptocurrency forecasting compared to conventional approaches when using only market-derived features. Phaladisailoed & Numnonda (2018), for example, used GRU to get an outstanding r^2 = 0.992, indicating how well these models capture intricate patterns in cryptocurrency values.

As demonstrated by the work of Kumar & Rath (2020), models that incorporate historical price data and market indicators, such as OHLC prices and trade volume

produce better results. The finding indicates that feature selection has a substantial impact on performance.

Table 2.9
Summary Of Regression For Cryptocurrency Price Prediction With Sentiment Features Performance Results

REGRESSION FOR CRYPTOCURRENCY PRICE PREDICTION (USING SENTIMENT AND PRICE FEATURES)					
Authors	**Sentiment Analysis**		**Price Prediction**		**Cryptocurrency Price Prediction Results**
	Features	**Model**	**Features**	**Model**	
Jain et al. (2018)	Pos & Neg counts	Textblob	average price	Multiple Linear Regression	Adj $r^2 = 0.426$
Steinert & Herff (2018)	Sentiment scores	VADER positive, negative, neutral, and compound score	exchange rate + trading volume	Linear regression	$r^2 = 0.025$
Aggarwal et al. (2019)	Sentiment scores	VADER	historical prices + Gold prices	LSTM	RMSE = 47.91
Inamdar et al. (2019)	Sentiment scores	RNN-LSTM	historical prices	Random Forest regression	MAE = 2.7526, RMSE = 13.7033 (1st-day), MAE = 3.1885, RMSE = 15.1686 (2nd-day)
Chin & Omar (2020)	Pos & Neg counts	NTUSD-Fin dictionary	OHL prices + trading volume	LSTM	RMSE = 0.0471
Poongodi et al. (2021)	Sentiment scores	Latent Dirichlet Allocation (LDA)	OHLC prices + trading volume	LSTM	RMSE = 0.029, r^2 = 0.446
Alghamdi et al. (2022)	Sentiment scores	SVM	OHLC prices + trading volume	LSTM	MAE = 0.255, RMSE = 0.503 (Bitcoin), MAE = 0.384, RMSE = 0.684 (Ethereum)

Based on Table 2.9, traditional models like Linear Regression exhibit weak predictive power, as seen by a $r^2 = 0.025$ (Steinert & Herff, 2018), indicating a discernible variation in model performance across the investigations. Conversely, more sophisticated models with lower RMSE and higher r^2 values, such as LSTM, GRU, and RNN-LSTM, typically perform better.

Additionally, the study by Chin & Omar (2020) have shown that sentiment analysis is a useful tool for enhancing price forecasts, merging sentiment features with market data to achieve RMSE = 0.0471. It consistently demonstrates great performance by deep learning models, especially LSTM, as demonstrated by Poongodi et al. (2021). An RMSE = 0.029 and $r^2 = 0.446$ are obtained, demonstrating the model's ability to

grasp complex patterns. By integrating the advantages of several approaches, hybrid and ensemble models further increase accuracy.

Overall, the findings imply that while conventional models like Linear Regression and SVM typically perform poorly for predicting cryptocurrency prices, while LSTM and other deep learning models, particularly when combined with sentiment analysis and extensive market data are mostly gained better outcomes.

2.10 Research Gap

From the literature review, there are several gaps that can be explored. The existing studies predominantly extract sentiment features in English language and very limited explorations were performed in other languages for cryptocurrency price prediction experimentation. Malay language, which is the mother tongue of Malaysian, is always underutilised. Since impactful sentiment tend to be expressed in each country's mother-tongue language, it is significant to include Malay in addition to English for creating new datasets on cryptocurrency-related sentiment data. Manual annotations were selected for corpora annotation. Although, such approach is less preferred by the previous researchers. However, experimental results reported that manual annotation especially in a small and medium-sized dataset, can provide better labelling and consequently result in higher performance. Hence, in order to construct a cryptocurrency-specific sentiment-labelled datasets, this study performs manual annotation on the collected data from online news site and Twitter.

Additionally, there were also limited experimentations on language models to cater polysemy problem in a sentence as part of the sentiment feature extraction process for cryptocurrency price prediction. Moreover, the focus on the impactful sentiment words at sentence-level is also beneficial to produce a better-quality sentiment prediction model. This leads to the implementation of self-attention mechanism (Ji et al., 2019; Vaswani et al., 2017). To date, the application of self-attention mechanism in the sentiment model for cryptocurrency-related study is still has not been thoroughly explored. Therefore, this study explores on the XLNet language model coupled with the multi-head self-attention mechanism adopted from Vaswani et al. (2017) in the GRU layer to develop a more robust sentiment regression model for sentiment feature generation to predict cryptocurrency prices.

Subsequently, the aggregations of sentiment features and historical cryptocurrency price features will then be fed into a multi-layered Bi-GRU deep learning model for cryptocurrency price prediction as GRU deep learning model has not been thoroughly explored in the sentiment-based cryptocurrency price prediction related research studies. The proposed cryptocurrency price prediction model development has the advantage of taking the current day real-time data to predict the next-day cryptocurrency prices, however, the accuracy performance of using very small data is yet to be tested. Hence, this study focuses on using past data instead of the real-time data to perform the training and testing of the prediction models since this study aims to build a dependable sentiment model for sentiment feature extraction and will be used as the added features to the cryptocurrency price prediction model. The purpose of adding additional features from previous collected data is to investigate the effect of using sentiment features on the performance of the price prediction model.

2.11 Summary

The literature review section presents the concept of cryptocurrency, and its market based on the financial theories. Prior studies on cryptocurrency price prediction have shown that it is possible to predict future prices and have thus rejected the random-walk theory of EMH. Two types of prediction tasks, namely, classification and regression on cryptocurrency price prediction have been studied and it was found that regression task is more reliable in predicting the cryptocurrency prices. In the effort of enhancing the existing cryptocurrency price prediction model, extra features such as the sentiment features were implemented. This study focuses on constructing new datasets in English and Malay that is manually labelled with sentiment scores ranging between -1 to 1 (regression) based on the cryptocurrency topic. The continuous sentiment values are preferred over categorical polarity in cryptocurrency forecasting because the intensity and variations of market emotions can be preserved, enabling more precise modelling of price movement. Then, the implementation of sentiment regression model will be performed to generate the sentiment features by using XLNet language model with GRU and self-attention mechanism. XLNet is selected for sentiment analysis due to its bidirectional modelling and permutation-based text prediction that allows better understanding of complex linguistic context. The sentiment feature will then be

integrated into the cryptocurrency price prediction model to investigate for any performance gain.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

This chapter provides an overview of the methodology applied in this research study. Firstly, the chapter explains the methodological framework to highlight the phases involved in addressing the research questions. There is a total of three phases involved and each of the phases are described further in the following sections. Part of this chapter are published in (Mohamad Zamani et al., 2025; Zamani & Kamaruddin, 2024).

3.2 Methodological Framework

The methodological framework for this work consists of three phases; which are, corpora creation, feature extraction, and cryptocurrency price prediction. Phase 1 involves data scraping from cryptocurrency price exchange, online news and Twitter, and text pre-processing to perform data annotation. Phase 2 uses the pre-processed datasets for feature extraction. The sentiment data then undergo the sentiment classification and regression to evaluate the quality of the features. Subsequently, Phase 3 completed the main focus of the study on price prediction using price and sentiment features. The performance of cryptocurrency price prediction activities is evaluated. The overall methodological framework is illustrated in Figure 3.1.

Referring to Figure 3.1, in Phase 1, two types of data are collected; namely price data and sentiment data. In 2021, Bitcoin and Ethereum experienced significant price volatility, driven by heightened adoption, market enthusiasm, and macroeconomic trends. Bitcoin reached an all-time high of over $68,000 in November, spurred by growing institutional investments, widespread adoption as a digital asset, and increasing acceptance as *digital gold* (Taskinsoy, 2021). Ethereum also soared, hitting a record high of over $4,800 in the same month, fuelled by the explosion of decentralized finance (DeFi) projects, non-fungible tokens (NFTs), and anticipation of Ethereum 2.0 upgrades (Thomas, 2025).

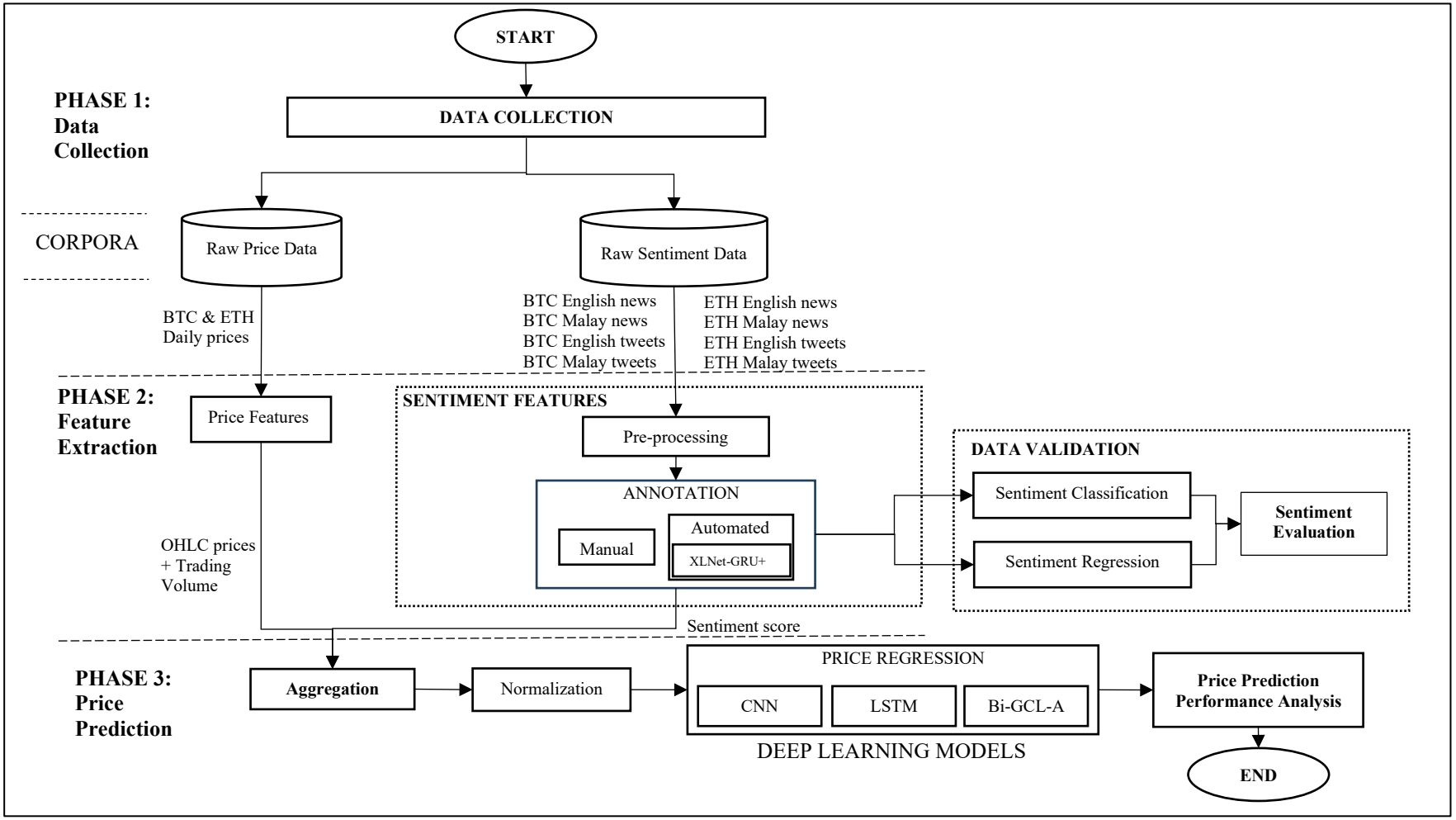


Figure 3.1 Methodological Framework Overview

Both cryptocurrencies faced periods of sharp corrections due to regulatory concerns, environmental debates about mining, and global economic shifts, but 2021 marked a landmark year for their mainstream recognition and growth (Hajric & Greifeld, 2021). Therefore, the historical price features for both Bitcoin and Ethereum are obtained throughout the year 2021 from Bitstamp exchange consisting of price features such as date, open, high, low, and close prices (OHLC), and the trading volume (CryptoDataDownload, 2025). In addition, sentiment raw data are collected from online news and Twitter in English and Malay languages to be pre-processed using Natural Language Processing (NLP) techniques.

Phase 2 focuses on feature extraction for price and sentiment data. Price data can be used directly without the necessity to be processed, as it already encompasses significant time-series information that depicts market patterns and price fluctuations, potentially negating the need for further feature extraction in certain analyses or models. Sentiment data need to undergo several processes before it is cleaned. First, the data must be annotated to classify it as positive, negative, or neutral sentiments using a range of sentiment scores (-1 to +1), facilitating a more precise analysis of its influence on cryptocurrency price fluctuations. Annotating sentiment text is essential for teaching computers to recognize human emotions in language. It offers precise and consistent labelled data, for more accurate and perceptive analysis (Majlinda, 2024). Both manual and automated annotation is conducted. Manual annotation was performed by three human annotators using a designated codebook as the guide. Several trainings were done to finalise the sentiment scores for all instances. On the other hand, automated annotation was generated by fine-tuning the XLNet model (Yang et al., 2019) with additional GRU and attention layer. From the sentiment feature extraction, sentiment classification and sentiment regression were performed.

Sentiment classification and sentiment regression are performed as part of Phase 2 by encoding the data from the created labelled corpora to be fine-tuned using the XLNet pre-trained language model. XLNet is a sophisticated language model that captures bidirectional context by taking into account every conceivable word order permutation. XLNet is especially good at sentiment classification and regression tasks because of its improved comprehension of language, paired with its capacity to manage intricate linguistic phenomena and make use of previously learned information. It can reliably identify long-range dependencies and details in text, resulting in more accurate

and sophisticated sentiment predictions (Yang et al., 2019). Fine-tuning is carried out for transfer learning in which the weights of a pre-trained language model are trained based on the new features. The fine-tuned model is then utilised for both sentiment classification and regression.

The work is carried out to predict price using both price and sentiment data in Phase 3. The price features of OHLC prices and trading volume are combined with the sentiment scores. Normalization for the prices are done to ensure all the features are in a common scale. Then, price regression is performed. This study focuses on price regression instead of price classification because regression models yield continuous outputs, facilitating more accurate predictions of specific price values, whereas classification merely categorises data into distinct price ranges, which may be less informative. The results achieved is to observe the performance of the price prediction model. Detailed description of each phase is presented in the next subsection.

3.3 Data Collection and Corpora Creation

The cryptocurrency price raw data was downloaded from *CryptoDataDownload* website (CryptoDataDownload, 2025). The price raw data is based on the daily date which is taken at 12.00 UTC each day based on Bitstamp exchange for Bitcoin and Ethereum. The price raw data consists of features; such as: open, high, low, close (OHLC) prices, and trading volume for the current day. The price raw features are taken as is without any feature extraction because the price directly reflects the market behaviour and applying feature extraction could risk distorting the original price signals or losing valuable temporal dependencies that are crucial for accurate forecasting. The study by Chowdhury et al. (2020) shows a high accuracy of 92.4% in predicting cryptocurrency prices with the direct usage of the OHLC prices. In addition, El Haddaoui et al. (2023) also utilised the OHLC prices and trading volume directly to predict cryptocurrency prices and obtained an r^2 of 0.974, indicating a good result.

The cryptocurrencies selected for the study are Bitcoin (BTC) and Ethereum (ETH) as the current top two highest value in market capitalization (Gupta & Chaudhary, 2022). Two corpora are generated for each of the cryptocurrency, which are: price raw data and sentiment raw data. From the price corpora, two datasets are obtained, which are: BTC daily price and ETH daily price. Subsequently, the sentiment

corpora produced eight datasets, which are: BTC English news, BTC Malay news, BTC English tweets, BTC Malay tweets, ETH English news, ETH Malay news, ETH English tweets, and ETH Malay tweets. Both corpora were collected for a duration of one year from 1 January 2021 until 31 December 2021. Pandemic season of COVID-19 happened critically with the Movement Control Order (MCO) worldwide making the social media and other online platforms are actively used during the year 2021, thus, the duration selected for the data collection is sufficient to observe the effect of the sentiment during cryptocurrency market instability.

Sentiment corpora are collected from Twitter and online news. The online news is extracted using NewsAPI to extract English news, whereas Parsehub tool is used to extracted Malay news. It can be divided into two, which are news headlines and Twitter posts. The news headlines were included in the analysis because news headlines capture most of readers' attention (Hossain et al., 2021). English news headlines are extracted using News API[1] Python library that consists of various online news sources such as CNN[2], Bloomberg[3], Business Insider[4], and New York Times[5]). The Malay news headlines are extracted using the Parsehub[6] tool from several different websites, namely Intraday[7], Utusan Malaysia[8], Berita Harian[9], Harian Metro[10], Astro Awani[11], and Malaysiakini[12]. English and Malay tweets are extracted using the TwitterSearchScrapper[13] Python package. The keywords "bitcoin," "btc," "ethereum," and "eth" are used to collect information about Bitcoin and Ethereum. The collected

[1] News API: https://newsapi.org/docs/client-libraries/python

[2] CNN news: https://edition.cnn.com

[3] Bloomberg: https://www.bloomberg.com/asia

[4] Business Insider: https://www.businessinsider.com

[5] New York Times: https://www.businessinsider.com

[6] Parsehub: https://www.parsehub.com

[7] Intraday: https://intraday.my

[8] Utusan Malaysia: https://www.utusan.com.my

[9] Berita Harian: https://www.bharian.com.my

[10] Harian Metro: https://www.hmetro.com.my

[11] Astro Awani: https://www.astroawani.com

[12] MalaysiaKini: https://www.malaysiakini.com/

[13] Twitter Scraper: https://github.com/JustAnotherArchivist/snscrape

sentiment data are stored into English and Malay corpora for Bitcoin and Ethereum. The process of data collection is illustrated in Figure 3.2.

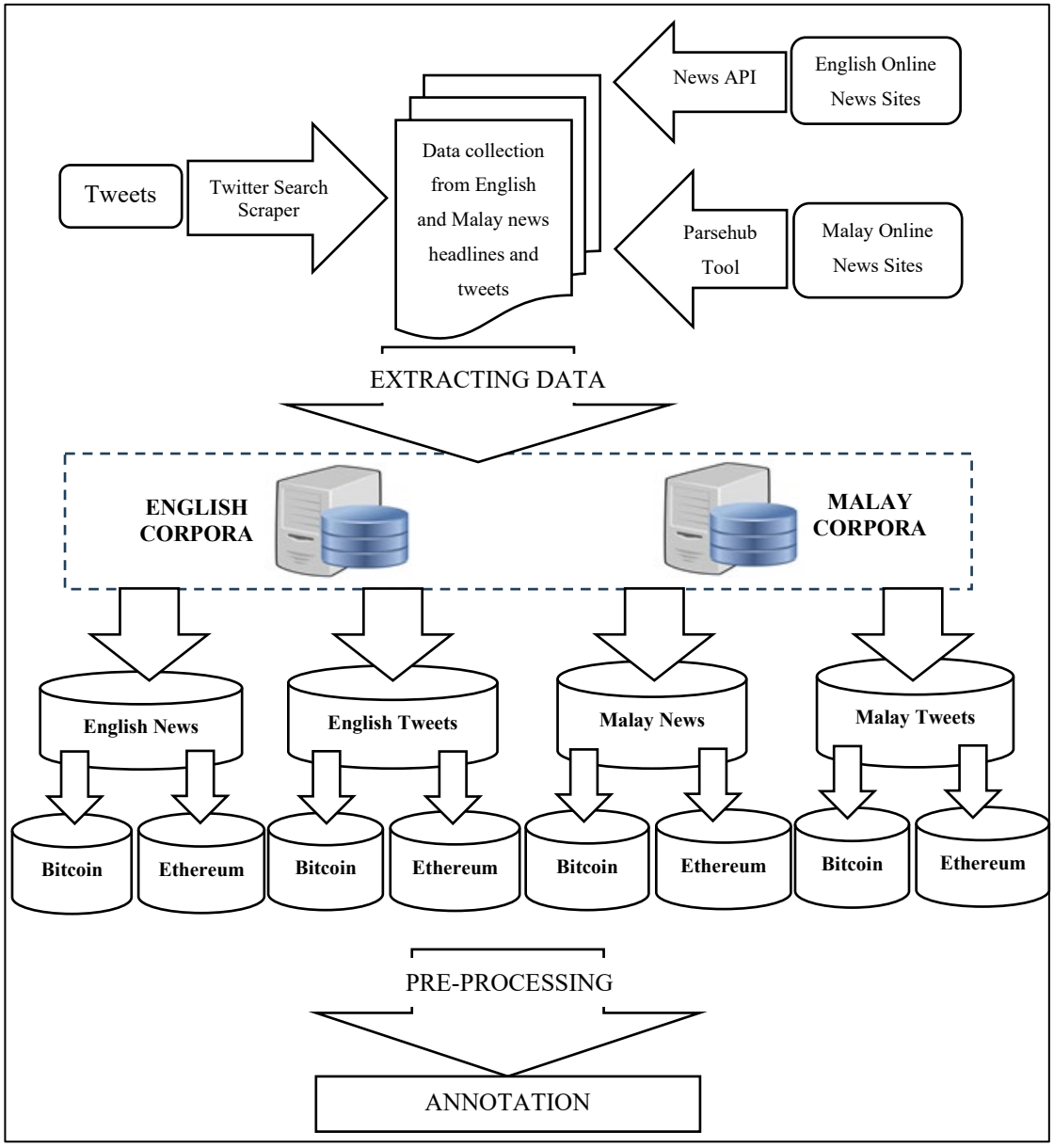


Figure 3.2 Sentiment Corpora Creation Framework

Based on Figure 3.2, after the data extraction phase, the collected data are organised into two main corpora based on language, namely English and Malay corpora. Each corpus is further divided into news and Twitter datasets, which are then categorised according to the cryptocurrency which are Bitcoin and Ethereum. The structured organisation allows for more focused sentiment analysis across different data sources and languages. Subsequently, the datasets undergo a pre-processing stage to

clean and standardize the text, followed by an annotation process where sentiment labels are assigned for further analysis.

From the sentiment corpora, the raw data need to undergo text pre-processing before being annotated. Some of the examples of raw news headlines collected are shown in Figure 3.3 a) for English and b) for Malay, while raw tweets examples are displayed in Figure 3.4 a) for English and b) for Malay, respectively.

Date	Source	Title
01/01/2021 03:43	Reuters	Bitcoin touches record above $29,000 extending 2020 rally ��� Reuters India
02/01/2021 12:59	Seeking Alpha	Bitcoin blasts through $30K Seeking Alpha
02/01/2021 15:17	Reuters	Bitcoin rallies above $30,000 for first time ��� Reuters
02/01/2021 19:14	CNN	Bitcoin soars past $33,000 its highest ever
02/01/2021 19:46	Business Insider	Bitcoin crosses $30,000 for the first time as it charges into 2021
02/01/2021 20:16	Entrepreneur	Record in a matter of hours! Bitcoin price exceeds $ 31,000
03/01/2021 03:03	Gizmodo.com	Bitcoin Keeps Moving On Up Reaching $33,000 in Value
03/01/2021 18:45	Slashdot.org	Bitcoin Surges 25% In One Week. Warren Buffett Still Won't Buy It
04/01/2021 01:44	Bloomberg	Bitcoin Holds Near Record Ether Surges Amid Crypto Rally - Bloomberg

(a)

Date/Time	Source	Title
03/01/2021 16:48	Harian Metro	Sindiket Macau Scam beli hartanah RM336 juta hasil penipuan
04/01/2021 11:00	Intraday.my	Ethereum Tembus $1,000, Bitcoin Capai Tahap Tertinggi $34,700!
04/01/2021 15:30	Intraday.my	Litecoin Atasi XRP, #3 Terbesar Selepas Bitcoin Dan Ethereum
04/01/2021 17:00	Intraday.my	Sehari Je Ethereum Catat Rekod Tertinggi Lebih $1,100!
04/01/2021 17:30	Intraday.my	Puluhan Altcoins Catat Lonjakan Tertinggi! Adakah Alt-Season Telah Kembali?
07/01/2021 10:40	Intraday.my	Permodalan Pasaran Kripto Kini Dah Cecah $1 Trilion!
07/01/2021 11:20	Intraday.my	1,620 Ladang Perlombongan Kripto Di Iran Ditutup Serta-Merta
07/01/2021 13:09	Berita Harian	Nilai pasaran kripto cecah AS$1 trilion, Bitcoin catat rekod tertinggi

(b)

Figure 3.3 Examples Of Raw Online News Headlines: a) English And b) Malay

created	tweet	name
14/4/2021 0:00	@JessicaVaugn The 100k #bitcoin party is Jessica and 10,000 nerds trying not to look or talk to you.	realfuckingnews
14/4/2021 0:00	The Fall of Bitcoin Prices is a Good ThingÃ°Å¸Â¤â€"!	BitcoinLifeClub
14/4/2021 0:00	#Bitcoin	valueofcrypto
14/4/2021 0:00	Market Id : BTC-AUD	RoyotPi
14/4/2021 0:00	Ã°Å¸Å½â„¢ Today on The Decrypt Daily:	decryptmedia
14/4/2021 0:00	The price of	BitcoinBellyB
14/4/2021 0:00	Bitcoin: $63558.48	HourlyBTCUpdate
14/4/2021 0:00	The first ever 10k bitcoin candle is upon us! Ã°Å¸â€™Â¥Ã°Å¸Å¡â¬Ã°Å¸Å'â„¢ LFGO!!!	AdamWelsh128

(a)

created	tweet	name
01/04/2021 00:00	Mar 31, 2021 23:59:00 UTC | 58,771.40$ | 50,120.90��| 42,640.50| #Bitcoin #BTC $BTC #crypto #financial #news	NowBitcoinPrice
01/04/2021 00:00	$BTC: $58918.833 (0.02%)$ETH: $1918.362 (3.97%)$BNB: $302.139 (-2.55%)$USDT: $1.0 (-0.09%)$ADA: $1.194 (-1.75%)$D(	BastionExchange
01/04/2021 00:01	#Bitcoin hari ini...Rp 850,770,00024h : ���0.78%7d : ���8.43%#btc	BotFinansial
01/04/2021 00:01	DASH/BTC yeni sinyal:buy @ fiyat:0.003785 #DASHBTC #CryptoCat #KriptoKedisi #KriptoSinyalleri #TeknikAnaliz #Trend	cryptocatvisor
01/04/2021 00:02	Strong #Downtrends in 30/60/120 mins (Binance $BTC pairs):$CHZ $KSM $REN $RVN $SUSHI $UNI #crypto #coin #RSIhttps:/	yekturus
01/04/2021 00:03	@1MillionTokens @300Token Amazing. @Martins44160536 @PhanNhuNguyen @xyfhy @Bendafit #bitcoin $BTC $ETH $DO	SubamosN
01/04/2021 00:03	������$2,500,000 #bitcoin LONGED @$59,169.7334 [01/04/21 00:03:39]����BitMEX | $XBTUSD����S	WhaleTrades
01/04/2021 00:05	$Bidao $Bid $10 #EOY$ETH $TRON $ONT $LINK DGBBTC $BNB $HUSD $TELLOR $NEM $DASH $IOST $BCD $QTUM $KMD	DDivs3

(b)

Figure 3.4 Examples Of Raw Twitter Posts: a) English And b) Malay

A total of 2,722 English news, 2,972 Malay news, 49,093,840 English tweets, and 653,551 Malay tweets are captured. From the details of the datasets in Table 3.1 and Table 3.2, English language sources from news and tweets are more extensive than the Malay language.

Table 3.1
Details Of Raw Data For News (Bitcoin And Ethereum)

	ENGLISH				MALAY			
	BITCOIN		ETHEREUM		BITCOIN		ETHEREUM	
Month	No. of Posts	No. of Words	No. of Posts	No. of Words	No. of Posts	No. of Words	No. of Posts	No. of Words
January	132	491	103	259	125	377	112	229
February	193	574	85	261	143	424	114	292
March	113	430	61	199	133	421	113	304
April	110	465	86	339	125	414	118	321
May	155	567	154	490	122	403	142	395
June	145	464	97	339	139	425	103	277
July	85	347	63	243	130	428	122	370
August	86	382	78	291	123	395	131	342
September	136	504	84	332	109	378	108	288
October	149	516	123	390	134	414	138	383
November	118	495	145	479	117	409	138	317
December	96	391	125	433	120	383	113	319
Total	**1,518**	**5,626**	**1,204**	**4,055**	**1,520**	**4,871**	**1,452**	**3,837**

The observation shows that most cryptocurrency users use English as the language medium to post the articles and opinions. Moreover, there are more tweets posts compared to the news headlines because people tend to tweet multiple times per day, while news online usually being published once a day.

Table 3.2
Details Of Raw Data For Tweets (Bitcoin And Ethereum)

	ENGLISH				MALAY			
	BITCOIN		ETHEREUM		BITCOIN		ETHEREUM	
Month	No. of Posts	No. of Words	No. of Posts	No. of Words	No. of Posts	No. of Words	No. of Posts	No. of Words
Jan	2,376,474	2,380,471	650,777	698,426	34,879	41,226	9,504	15,412
Feb	2,394,282	2,399,533	764,213	780,555	38,745	46,082	10,218	17.032
Mar	2,523,767	2,544,636	909,851	979,557	16,535	25,237	2,434	6,240
Apr	2,547,110	2,557,175	1,118,968	1,504,227	39,512	48,546	11,635	20,270
May	3,177,834	3,183,219	1,638,043	1,730,769	55,373	65,774	18,769	27,425
Jun	3,093,244	3,099,512	1,163,861	1,299,163	32,784	42,367	9,314	17,049
Jul	2,571,812	2,575,654	1,287,381	1,487,313	35,183	41,962	10,245	19,556
Aug	2,359,768	2,364,955	1,824,287	1.898.425	31,031	40,271	15,421	23,092
Sep	2,729,877	2,751,870	2,086,821	2,270,893	43,775	47,448	17,960	27,643
Oct	3,037,281	3,044,010	1,968,094	2,390,988	51,929	63,709	19,778	30,367
Nov	2,684,860	2,692,397	1,943,826	2,165,190	59,715	65,921	21,548	33,356
Dec	2,345,512	2,398,862	1,895,897	1,998,678	40,363	45,108	26,901	27,716
Total	**31,841,821**	**31,992,294**	**17,252,019**	**17,305,759**	**479,824**	**573,651**	**173,727**	**248,143**

From Table 3.1, it is noted that news headlines are cleaner than Twitter data because Twitter data contains noise such as non-English and non-Malay texts, insignificant sentences, misspelt words and special characters including hyperlinks. From the overall monthly analysis, it can be observed that news and tweets are being posted actively in the month of May 2021. Such a situation may result from Elon Musk's announcement regarding Tesla and Bitcoin transaction which caused havoc among the cryptocurrency users, leading to the highest data extracted in May throughout the year 2021 as reported by Rahman (2021).

The minimum sampling size for Twitter data is calculated based on Equation 3.1 to obtain the minimum recommended sample size (Cochran, 1977; Lohr, 2009).

$$n = \frac{N \times z^2 \times p \times (1-p)}{(N-1) \times e^2 + z^2 \times p \times (1-p)} \tag{3.1}$$

where:

n: sample size,
N: population size,
z: z-score (which corresponds to the desired level of confidence)
p: estimated proportion of the population with the attribute of interest
e: margin of error

Based on Equation 3.1, the recommended minimum sample size is 2,401 and 2,393 for English and Malay tweets respectively with the settings of 95% confidence level, 2% margin of error, and the population proportion of 50%. These settings was commonly used by prior studies to calculate the recommended minimum sample size in the domain (Fitts, 2022; Hazra, 2017). Therefore, 6,000 tweets data are selected to comply to the minimum recommended sample size.

Table 3.3
Number Of Documents (Posts) In Each Corpus

Corpus	Cryptocurrency	English	Malay
News headlines	Bitcoin	1,518	1,520
	Ethereum	1,204	1,452
Tweets	Bitcoin	6,000	6,000
	Ethereum	6,000	6,000

As for news headline, the number selected lies in between the number of data presented by previous work (Cerda, 2021; Rognone et al., 2020; Pant et al., 2018). Typically, researchers manually annotate between 3,000 to 10,000 data samples. For instance, Rognone et al. (2020) performed manual annotation on 3,108 English news data, while Cerda (2021) and Pant et al. (2018) manually annotated 9,146 and 7,454 English tweets, respectively. This shows that this study contributes to a larger number of manually labelled data with the aim to produce a better quality corpora in order to reduce the tendency for over-fitting if the dataset is too small (Althnian et al., 2021). The collected raw data are then utilised for feature extraction process in the next phase (Phase 2).

3.4 Feature Extraction

Feature extraction incorporates the extraction of price data and sentiment data. The price data are based on the daily price data of Bitcoin (BTC) and Ethereum (ETH), focusing on the OHLC prices and trading volume. Concurrently, sentiment features are derived from sentiment data, encompassing news headlines and tweets in both English and Malay. The sentiment data is subjected to pre-processing and annotation. This extensive feature extraction technique establishes the basis for constructing a resilient model to predict cryptocurrency prices by amalgamating market data and public opinion.

3.4.1 Price Features

The price features taken for this study is the OHLC prices and trading volume of Bitcoin and Ethereum cryptocurrencies. The price data did not undergo any feature extraction and utilised the data as it is. The extracted price features are essential as the market dynamics of cryptocurrency and function are reflected as a crucial input to be combined with the sentiment features in Phase 3 to perform the cryptocurrency price prediction.

3.4.2 Sentiment Features

The sentiment feature extraction in Phase 2 plays a crucial role in understanding the market sentiment and its influence on cryptocurrency price movements. The collected sentiment data undergo pre-processing, which involves cleaning and preparing the data for analysis. Following this, annotation is performed on the cleaned sentiment data, which can be carried out either manually or automatically. Once annotated, the sentiment is classified or regressed into sentiment scores.

The sentiment scores feature consists of three different sentiment score, which are: average overall sentiment score for the current day, average positive sentiment score for the current day, and average negative sentiment score for the current day. The generated average sentiment scores are shown in Equation 3.2. It is applied to the three types of sentiment score feature.

$$Average_{sentimentscore} = \frac{1}{n}\sum_{i=1}^{n} D_i^{sentimentscore} \tag{3.2}$$

where

n: Total number of documents in a day,

$D_i^{sentimentscore}$: *The sentiment score for the document* D_i *in a day.*

The sentiment classification and regression are also evaluated to measure the performance of the extracted features. These sentiment features are essential in capturing the public's perception of the market, helping to predict how sentiment can influence the price of cryptocurrencies.

3.4.2.1 Pre-processing

The pre-processing of the data starts with the manual removal of non-English and non-Malay texts that are produced after raw data extraction. Then, the raw sentiment data must go through text pre-processing stage where unknown symbols, insignificant special characters, hyperlinks, and hashtags are removed. Besides performing the elimination of unwanted instances and objects, emoticons conversion is also conducted.

Emoticons add emotional context to text, which can help in understanding the sentiment or tone behind a message. Converting emoticons to words such as ":)" becomes *happy* or ":(" becomes *sad* ensures that these emotional cues are captured in sentiment analysis models. Without this conversion, the model might miss out on important context, leading to inaccurate or incomplete sentiment classification or regression. Thus, emoticon conversion helps improve the accuracy and understanding of the emotional tone in the text.

The emoticons consist in tweets are handled using a Python library called Emot[14]. Emot translates an emoticon into text that expresses the emotion and was originally developed for English text. Nevertheless, to cater the emoticons contain in Malay text, the English words in Emot dictionary (Shah, 2021) are translated to Malay words for this research study. The Emot dictionary for English and Malay are shown in Figure 3.5 a) for English and b) for Malay, respectively.

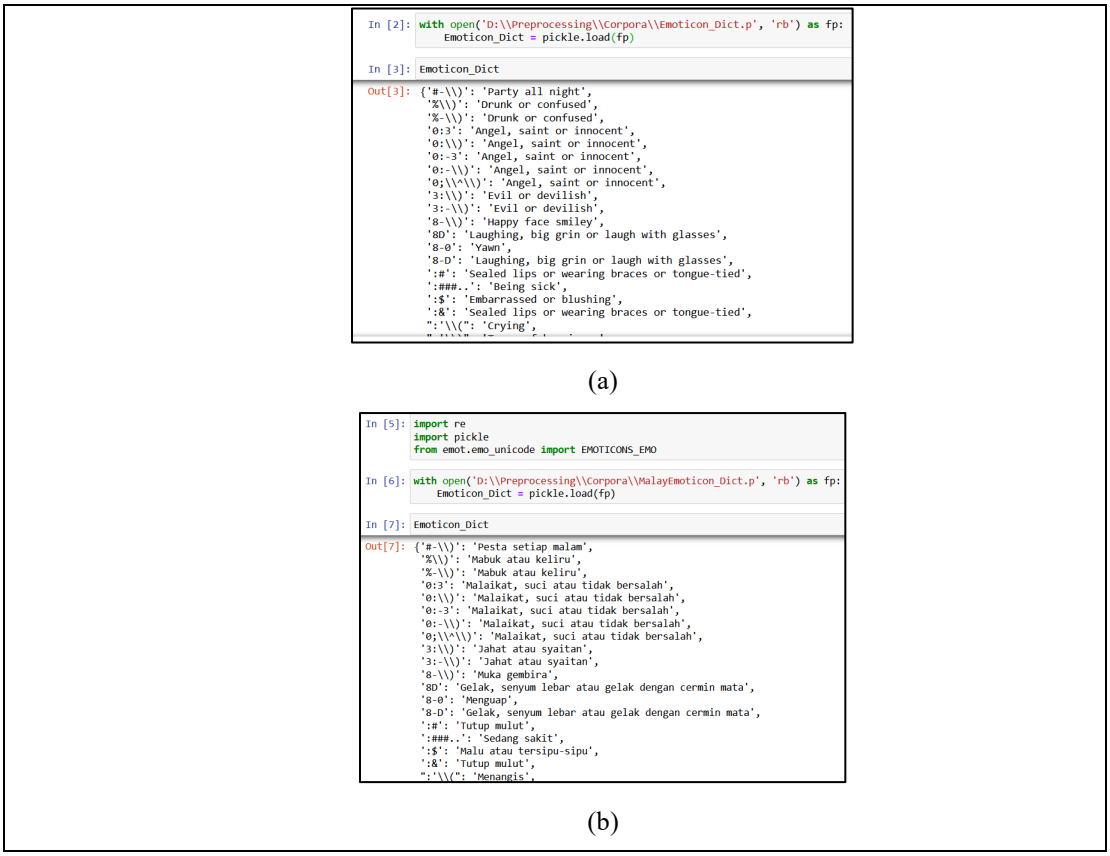


Figure 3.5 Emot Dictionary For a) English And b) Malay

[14] Emot: https://github.com/NeelShah18/emot

Figure 3.6 shows the example of Emot dictionary conversion for a) English and b) Malay documents. Once the corpora have been created by undergoing the necessary processes, sentiment analysis will be performed in the next phase (Phase 2).

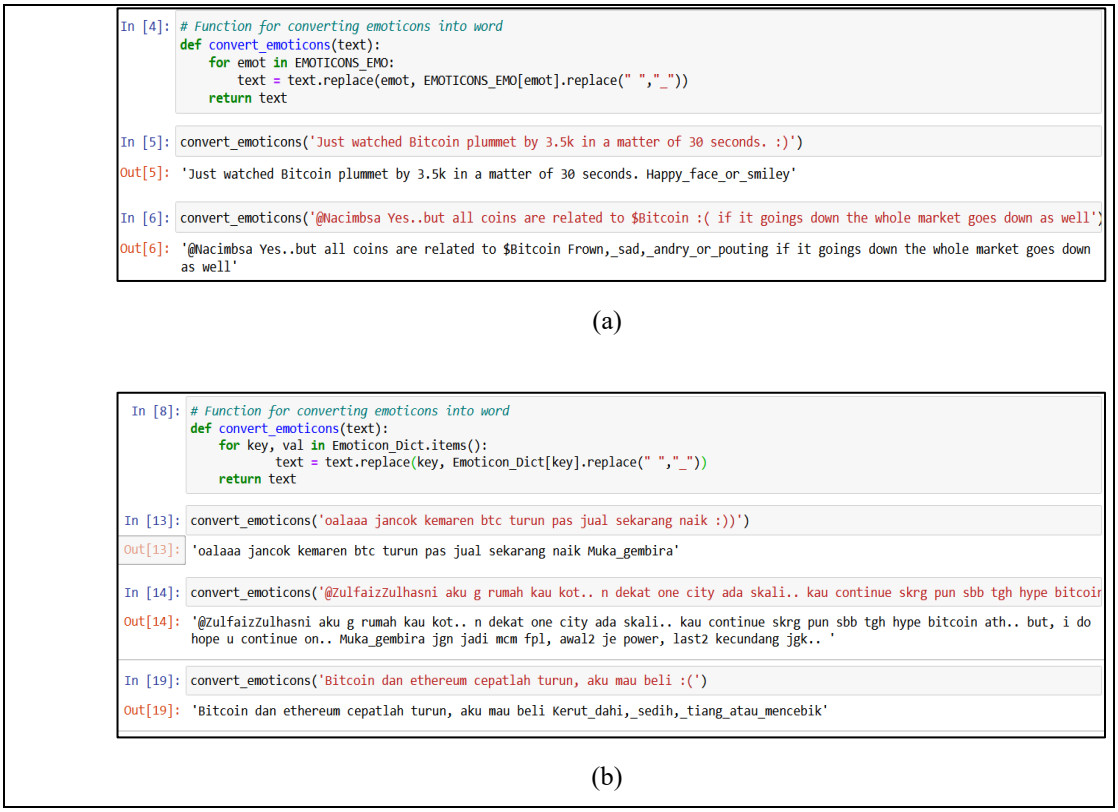


Figure 3.6 Emoticon's Conversion Example For a) English And b) Malay tweets

This step produces cleaner texts for annotation process in order for the annotators to effectively and precisely identify the text by concentrating on meaningful information, standardizing the data, and minimizing noise. By eliminating unwanted artefacts, the result of sentiment analysis is more dependable and perceptive (Jim et al., 2024).

Table 3.4 shows the total vocabulary (number of unique words) after the pre-processing of words including unwanted symbols and hyperlinks removal, and also the corrected misspelt words for each corpus to be used in computing the word frequency ranking and weightage.

Table 3.4
Number Of Vocabulary In Each Corpus

Corpus	Cryptocurrency	English	Malay
News headlines	Bitcoin	2,376	1,870
	Ethereum	1,703	1,499
Tweets	Bitcoin	10,484	7,285
	Ethereum	10,769	8,366

3.4.2.2 Annotation

The next process is data annotation. Both manual and automated data annotation are explored to give different perspectives. Manual annotation exploits the human annotation intervention by applying the inter-annotator percentage agreement of Krippendorf's technique (Krippendorff, 2004a; Krippendorff, 2004b) to determine the final sentiment value for each document. Three annotators with an addition of one domain expert are employed to verify the sentiment labels. In addition, three sentiment model set-ups which are XLNet model, XLNet-GRU model, and XLNet-GRU with self-attention models are employed for automated data annotation.

A. Manual Sentiment Data Annotation

Three annotators are selected based on their expertise and fluency in English and Malay languages. The first annotator, denoted as Annotator A, is a PhD holder and a senior lecturer of language studies at Universiti Teknologi MARA with six years working experience. Annotator A labels the data based on the language knowledge, whereby the validation of sentiment labelling is according to the expert understanding on generic sentiment value in a document containing either one sentence or multiple sentences. The second annotator, denoted as Annotator B, has been involved in the cryptocurrency trading for about six years and currently working as a P2P dispute specialist for three years which job scopes includes handling issues on cryptocurrency transactions between users. The annotated data are labelled based on the cryptocurrency terms containing in the sentences signifying the positivity and negativity of each document. The positive and negative sentiment values in cryptocurrency term is when a sentence consists of a statement referring to the rising or dropping of the price. The

document is labelled as positive using continuous value within the specified range if the statement is referring to a rising of price, containing words such as *buy* or any positive comments about the cryptocurrency coin. On the contrary, the document is labelled as negative if the statement is referring to a dropping of price, containing words such as *sell* or any negative comments about the cryptocurrency coin. Similar annotation understandings and validation of data are conducted by the third annotator denoted as Annotator C, who is involved in cryptocurrency and forex market with six years' experience.

The annotation procedure was validated by Expert A, a remisier with 26 years of experience at Kenanga Investment Bank, Malaysia, specializing in financial markets like stocks, forex, and cryptocurrency, though not directly involved in the annotation. The annotators' demographics are provided in Appendix 1. The annotators are trained on standard labelled dataset on financial-related news and media to ensure consistency of interpretation and annotation (John & Vechtomova, 2017a). Whilst, Malay news and tweets datasets on economics were obtained from Malaya Documentation (Zolkepli, 2022).

All annotators are provided with a codebook as a guidance to determine the sentiment score for each document. It serves as a reference guide, ensuring that all annotators interpret and label data in a consistent manner. The uniform set of guidelines of the codebook is essential to obtain an inter-annotator agreement between the annotators (Ritchie et al., 2022). The preparation of the codebook is based on the applicable guide for manual annotation and determining the validity of the codebook for annotation activity (Krippendorff, 2004a; Mohammad, 2016; Van Atteveldt et al., 2021). The designated codebook for the annotators in this study can be referred at Appendix 2. Based on the codebook, every daily headline and tweet posted by each user is represented by the term *document* in this study, where the overall sentences in each document are annotated with sentiment score range between -1 to +1, representing from very negative to very positive sentiment (refer to Table 3.5).

In order to create a codebook for the annotators to perform the annotation task, a rubric was designed to assign the sentiment scores on each of the document. The rubrics consists of category, the description of the category (statement), and the weight (sentiment score). Four relevant categories are determined; namely, level of influential person, country, organization, and event and news. The categories are identified with

the help of a word cloud. The weightage of the sub-categories is assigned based on the TF-IDF vectorizer.

TF-IDF works by ranking the term frequency within a document relative to its frequency across all documents. Terms that appear frequently within a document are given a higher weight, indicating their significance within the document. By analysing the most significant terms using TF-IDF, categories such as level of influential person, country, organization, and event/news are identified.

Once the categories are identified, the rubric for the codebook is created. Each category is associated with a description that explains the nature of the category and its relevance to the sentiment analysis. The sentiment score, which represents the strength of the sentiment is assigned based on the importance of the category as determined by its TF-IDF weight. For example, a high TF-IDF score for a term related to an influential person or a significant event would result in a higher sentiment score, reflecting the stronger impact of these factors on the overall sentiment.

The annotators utilised the codebook by referring to the rubric and examine the document for relevant information. The annotators then check for terms and statements that match the identified categories (influential person, country, organization, event/news) and assign a sentiment score accordingly. The rubric for designing a codebook is provided in Figure 3.7.

Krippendorff's alpha is then used to calculate the percentage agreement between all annotators due to its reliability in measuring continuous data with multiple coders (Krippendorff, 2004b). An experiment on measuring its reliability has been conducted by Antoine et al. (2014) that involved similar number of coders (i.e, three coders/annotators) on similar type of data that contains the intensity of sentiment. Thus, an estimate of seven rounds of training with same sets of data are provided for the annotators until an inter-annotator agreement of more than 60% based on Krippendorff's alpha measurement (Krippendorff, 2004a) has been achieved between the annotators.

The benchmark of 60% agreement percentage is considered a reasonable value due to the complicated sentiment scoring using continuous value task of determining the degree of positivity or negativity in the document (Artstein & Poesio, 2008). After the percentage agreement is fulfilled, the average sentiment scores across all annotators are taken as the final sentiment value for each document. The inter-annotator agreement was validated by Expert A.

Table 3.5
Rubrics For Codebook Creation: a) Positive And b) Negative Sentiment

CATEGORY	STATEMENT	SCORE	CHECK	TOTAL SCORE
Influential person	Crypto influencer who made a statement/action			
	Has more than 1 million followers (Mega-influencer)	0.4		
	Has between 100,000 to 1 million followers (Macro-influencer)	0.3		
	Has between 10,000 to 100,000 followers (Micro-influencer)	0.2		
	Has below than 10,000 followers (Nano-influencer)	0.1		
	Billionaire	0.3		
	Other non-crypto influencer or authority who made a statement/action			
	Leader/government	0.4		
	International celebrity (well-known)	0.3		
	Local celebrity	0.2		
	High post in a company	0.1		
	Action			
	Buy, purchase, invest, hold or HODL	0.2		
	Launching	0.1		
	Planning	0.1		
	Opinion or feelings			
	Extreme positive	0.2		
	Moderate positive	0.1		
Country	Country status based on GDP per capita[15]			
	High-income (developed) country	0.4		
	Newly emerging economies (emerging) country	0.3		
	Low-income (developing) country	0.2		
	Action			
	Buy, purchase, invest, hold or HODL	0.2		
	Launching	0.1		
	Planning	0.1		
	Opinion			
	Extreme positive	0.2		
	Moderate positive	0.1		
Organization	Bank			
	Top world bank (developed country – high market cap)	0.3		
	Local bank (from other countries)	0.2		
	Company			
	International (well-known – Top 100 based on market cap)	0.4		
	Local	0.2		
	Action			
	Buy, purchase, invest, hold or HODL	0.2		
	Launching	0.1		
	Planning	0.1		
	Opinion			
	Extreme positive	0.2		
	Moderate positive	0.1		
Event/News	Positive announcement (gaining profits or anything positive)	0.3		
	Promote crypto trading/investment	0.2		
	Planning to launch a promotional event	0.1		
	Free giveaways during sign-up in a crypto coin investment	0.1		
	Price is going up			
	A stronger emphasis on price going up (e.g., “to the moon”, “rocket”, etc)	0.4		
	General statement of price going up (with no strong emphasis)	0.3		
	Pumping	0.2		
	Stating price value	0.1		
			FINAL SCORE (positive)	

a) Positive Sentiment Rubrics

[15] https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD

CATEGORY	STATEMENT	SCORE	CHECK	TOTAL SCORE
Influential person	Crypto influencer who made a statement/action			
	Has more than 1 million followers (Mega-influencer)	-0.4		
	Has between 100,000 to 1 million followers (Macro-influencer)	-0.3		
	Has between 10,000 to 100,000 followers (Micro-influencer)	-0.2		
	Has below than 10,000 followers (Nano-influencer)	-0.1		
	Billionaire	-0.3		
	Other non-crypto influencer or authority who made a statement/action			
	Leader/government	-0.4		
	International celebrity (well-known)	-0.3		
	Local celebrity	-0.2		
	High post in a company	-0.1		
	Action			
	Sell crypto asset	-0.2		
	Launching	-0.1		
	Planning	-0.1		
	Opinion or feelings			
	Extreme negative	-0.2		
	Moderate negative	-0.1		
Country	Country status based on GDP per capita[16]			
	High-income (developed) country	-0.4		
	Newly emerging economies (emerging) country	-0.3		
	Low-income (developing) country	-0.2		
	Action			
	Against cryptocurrency, boycott, or banning	-0.3		
	Planning	-0.1		
	Opinion			
	Extreme negative	0.2		
	Moderate negative	0.1		
Organization	Bank			
	Top world bank (developed country – high market cap)	-0.3		
	Local bank (from other countries)	-0.2		
	Company			
	International (well-known – Top 100 based on market cap)	-0.4		
	Local	-0.2		
	Action			
	Against cryptocurrency, boycott, or banning	-0.3		
	Sell crypto asset	-0.2		
	Planning	-0.1		
	Opinion			
	Extreme negative	-0.2		
	Moderate negative	-0.1		
Event/News	Negative announcement (having loss or losing something)	-0.3		
	Against cryptocurrency (boycotting or banning	-0.2		
	Planning to boycott or ban cryptocurrency	-0.1		
	Price is going down			
	A stronger emphasis on price going down (e.g., “slide”, “rock bottom”, etc)	-0.4		
	General statement of price going down (with no strong emphasis)	-0.3		
	Dumping	-0.2		
	Stating price value	-0.1		
			FINAL SCORE (negative)	

b) Negative Sentiment Rubrics

[16] https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD

For inter-annotator agreement, Krippendorf's alpha (Krippendorff, 2018) for three annotators was applied and the calculation is shown in Equation 3.3.

$$\alpha = 1 - \frac{\sum_i \sum_j \sum_k error(O_{ijk})}{\sum_i \sum_j \sum_k error(E_{ijk})} \tag{3.3}$$

where:

i, j, and k: indexes for the three annotators,

O_{ijk}*:* observed agreement between coders i, j, and k for a specific item (unit of analysis),

E_{ijk}*:* expected agreement between coders i, j, and k for the same item,

Error function: the error function is typically 0 when O_{ijk} equals E_{ijk} and 1 otherwise.

The final Krippendorf's alpha percentage agreement for each corpus is presented in Table 3.6. The results infers that the annotators were more used to the English financial terms than Malay financial terms, which makes it a bit more complex to understand certain statements in Malay for cryptocurrency-related issue. Figure 3.7 summarizes the annotation process by depicting the annotation process encompassing both English and Malay texts. The procedure commences with cleaned English and Malay texts, accompanied by a codebook that provides annotation guidelines. The inputs are utilised during the sentiment scoring phase, where the annotators determine sentiment values to the texts.

Table 3.6
Krippendorf's Alpha Percentage Agreement For Each Corpus

Corpus	Cryptocurrency	English	Malay
News headlines	Bitcoin	64%	61%
	Ethereum		
Tweets	Bitcoin	63%	60%
	Ethereum		

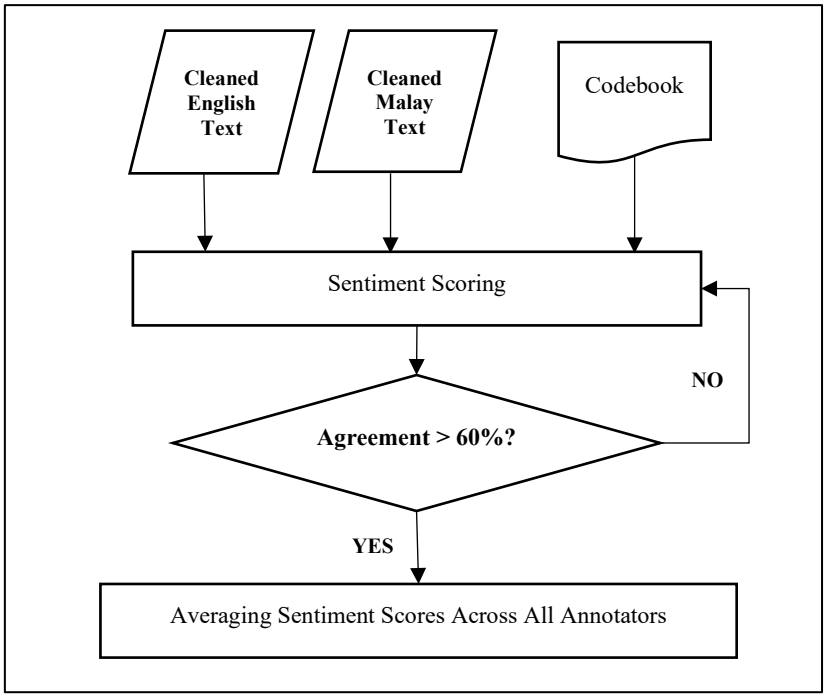


Figure 3.7 Flowchart Of Annotation Process

The procedure subsequently assesses if the annotators' agreement based on Krippendorf's alpha surpasses 60%. Should the agreement fall below this criterion, the texts are reassessed utilising the codebook for re-scoring guidance. However, if the agreement remains below 60% even after the rescoring process, those samples are excluded from the dataset to maintain the reliability of the annotated data. Upon exceeding 60% agreement, the sentiment scores from all annotators are summed to generate the final sentiment label. The training process for the annotators guarantees dependability and uniformity in the sentiment annotation procedure. For the news dataset, all samples are included for annotation due to its relatively small size, ensuring comprehensive coverage. In contrast, for the Twitter dataset, certain samples are excluded if they contain unmeaningful, null, or repetitive sentences, as these do not contribute meaningful sentiment information.

Many of the existing studies do not precisely describe manual annotation or agreement thresholds explicitly; instead, the studies rely on automated sentiment extraction (from lexicons or language models) or hybrid features. In the study by Mochtak et al. (2022), the sentiment dataset called *ParlaSent-BCS*, used a six-level sentiment schema for parliamentary debates and result is reported to be useful based on the inter-annotator agreement (measured with Krippendorff's alpha) of around 0.60. A

study on cryptocurrency by Gadi & Sicilia (2025), performed manual annotation on the corpus called *CryptoLin* with either positive or negative label. The study used multiple inter-annotator agreement metrics including Krippendorff's alpha and based on the finding, it achieved an average of 0.90 coefficient. The study achieved a high Krippendorff's alpha of 0.9 for simple positive-negative labels, reflecting strong agreement among annotators. For the continuous sentiment scale of -1 to +1, the suitable coefficient is 0.6. This is due to the increased subjectivity in fine-grained labelling and the additional challenges of interpreting non-English text, including informal expressions. Despite the lower value, $\alpha = 0.6$ represents moderate and acceptable reliability, ensuring that the dataset remains both consistent and sufficiently large for effective cryptocurrency price prediction (Hughes, 2021).

B. Automated Sentiment Data Annotation

Another way to label the data is by using pre-defined automated annotation. XLNet (Yang et al., 2019), BERT (Devlin et al., 2019), and GPT (Radford et al., 2018) are all transformer-based language models. Although BERT has started to be used commonly in current research studies on cryptocurrency, XLNet is selected in this research study because it considers all possible permutations of the factorisation order instead of the fixed forward-backward factorisation order as applied in BERT (Devlin et al., 2019; Yang et al., 2019), and also does not undergo pretrain-finetune discrepancy as BERT does (Yang et al., 2019).

The XLNet pre-trained language model is directly implemented using Hugging Face[17]. Hugging Face is an artificial intelligence (AI) platform that offers pre-trained machine learning models for natural language processing (NLP) applications such as sentiment analysis, translation, and text classification. It is well-known for its Transformers library. Developers and academics can easily create and implement cutting-edge NLP applications with its model sharing (Hugging Face Hub), dataset, and fine-tuning tools. Mainly, the XLNet language model is used to tokenize and encode the texts to produce vector representations. The process is illustrated in Figure 3.9.

[17] Huggingface: https://huggingface.co/

In the XLNet model, tokenization, encoding, and the attention mask are key pre-processing steps that prepare textual input for the transformer model to produce context vectors. Based on Figure 3.9, the sentences containing in each document from the corpora are taken as the input to the XLNet pre-trained language model. Then, each sentence will be tokenized and encoded with the attention masks to produce context vector. Figure 3.8 highlights the steps in fine-tuning process.

```
1. Read CSV files (corpus).
2. Create a list of texts and labels.
3. Find the maximum sentence length.
4. Split the samples into the train-test set (80%-20%) with random_state.
5. Tokenization of each sentence using 'xlnet-base-cased' (for English
   corpus) or 'xlnet-base-bahasa-cased' (for Malay corpus).
6. Encode the data into input_ids, attention_mask, and label.
7. Encapsulate the three parameters into data loaders.
8. Initialize batch size and optimiser = AdamW with learning rate = 2e-5.
9. Feed data encodings into XLNetForSequenceClassification function from
   HuggingFace with num_labels = 1 #for regression.
10. Set the number of epochs.
11. Training and evaluation.
12. Save fine-tuned XLNet pre-trained language model.
```

Figure 3.8 Fine-tuning Process

Based on Figure 3.8, in order to create the vector representations, each sentence from the corpus is tokenized using the XLNetTokenizer called *xlnet-base-cased* for English and *xlnet-base-bahasa-cased* for Malay. The *xlnet-base-cased* for English and *xlnet-base-bahasa-cased* for Malay are referred to a specific pre-trained version of the XLNet transformer model that are used during tokenization (Gong et al., 2019).

The XLNet-base-cased models are selected because it provides a balance between performance and computational efficiency, making it suitable for tasks with limited resources while still leveraging the power of the XLNet architecture. Its cased configuration is important for tasks where capitalization conveys meaning, such as distinguishing proper nouns (e.g., *Apple* the company vs. *apple* the fruit). Additionally, it serves as a good starting point for fine-tuning on specific NLP tasks, offering robust contextual understanding without the need to train from scratch.

The hyperparameter setting is based on the XLNet architecture from (Yang et al., 2019) and illustrated in Figure 3.9. Figure 3.9(a) represents the content stream attention, which is similar to the standard self-attention, while Figure 3.9(b) shows the query stream attention, which does not have access information about the content.

Subsequently, Figure 3.9(c) represents the overview of the permutation language modelling training with two-stream attention (Yang et al., 2019).

The tokens are encoded with attention mask to provide the context for each prediction. The encoding works by applying the permutation technique allowing bidirectional contexts to be captured for the positional encoding based on factorization order instead of the sequence order by relying on the attention mask (Yang et al., 2019).

A SentencePiece (Kudo & Richardson, 2018) is an unsupervised text tokenizer and detokenizer specifically designed for natural language processing (NLP) models. By employing SentencePiece's subword tokenization, the input text is divided into smaller parts called tokens that can be mapped to numerical representations. By doing so, uncommon or unfamiliar words can be broken up into smaller, more understandable chunks because originally, XLNet cannot interpret text directly. With this, the encoding maps tokens to numerical IDs and adds positional information. This is crucial to process tokens since the XLNet needs numerical IDs to assist the model in comprehending the order of words in a sequence.

Then, the encoding converts tokens to integers and adds positional embeddings. The attention mask is a binary tensor (0s and 1s) used to differentiate meaningful tokens from padding tokens during attention computation, where 1 indicates the tokens, the model should attend to and 0 indicates the tokens the model should ignore which is known as padding. Hence, the attention mask is essential to ensure that padding tokens are ignored in self-attention computations.

Once the input is tokenized, encoded, and passed through the attention mechanism, the XLNet model processes the sequence using its transformer architecture. During this process, the self-attention computes the relationships between all tokens in the input sequence, incorporating the attention mask to focus on meaningful tokens, then each token is represented as context vector that captures its meaning in relation to other tokens in the sequence.

GridSearch Cross Validation (Beheshti, 2022) is also used to find the optimised hyperparameter settings for the XLNet model. GridSearch is a brute-force technique to find the best combination of hyperparameters by exhaustively testing all possible values from a predefined parameter grid and it is beneficial in reducing the risk of overfitting by evaluating performance on multiple train-test splits (Shah, 2024), while the 5-fold

Cross Validation is a resampling method where the dataset is divided into 5 equal parts (folds).

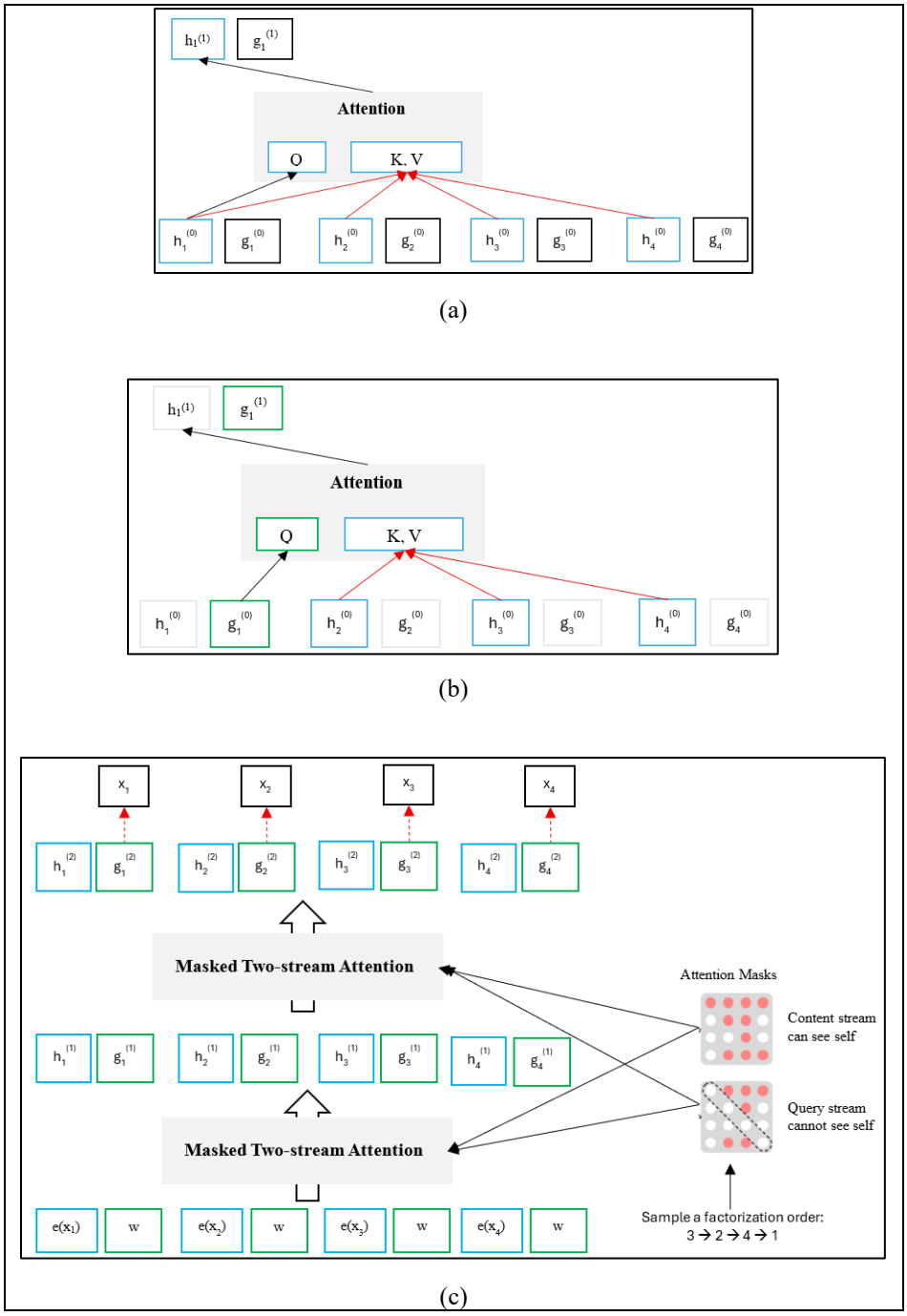


Figure 3.9 XLNet Architecture (Yang et al., 2019): a) Content Stream Attention, b) Query Stream Attention, c) Overview Of The Permutation Language Modelling Training With Two-Stream Attention

Based on the current studies on Transformer-based language modelling approach in sentiment analysis, Adam optimizer (Kingma & Ba, 2017) is commonly used to perform the optimization. Adam optimizer is computationally efficient with small memory usage. In addition, it is suitable for dealing with large data or parameters and handles noisy problems (Jais et al., 2019). Compared to other optimizers, AdamW is suitable for sentiment regression when utilizing XLNet since it successfully handles the difficulties associated with training big transformer models (Ali et al., 2024).

AdamW decouples weight decay from learning rate, preventing over-regularization and ensuring better generalization. It adapts learning rates dynamically, handles sparse gradients efficiently, and is specifically optimized for transformer architectures, leading to stable and efficient training. Whilst other optimizers such as SGD, RMSProp, and Adagrad lack these combined advantages, making them less effective for XLNet (Liu et al., 2023).

This process is extended by fine-tuning the model to learn new weights of sentiment score that are optimum for the task while maintaining the pre-trained model's overall language comprehension and modifying it to learn the particular patterns and correlations in the created corpus. This method minimizes the need to train a model from scratch and is computationally efficient. Fine-tuning is performed for each of the corpus.

3.4.2.3 Sentiment Evaluation

To investigate the quality of the dataset using several machine learning and deep learning models, sentiment evaluation is performed. Sentiment classification and sentiment regression are conducted.

Sentiment classification is a task that categorizes text into predefined sentiment labels, such as positive, negative, or neutral. The weights in the datasets are converted into categorical classes. The purpose of the experiment is to compare the performance of the dataset on sentiment classification. In sentiment classification, eliminating neutral sentiment frequently enhances model performance by streamlining the process and lowering ambiguity. Neutral sentiment tends to be more subjective and less defined, which might cause noise and confound the model (Talaat, 2023). Therefore, the neutral sentiment is removed throughout the sentiment classification experiments.

In order to cater class imbalanced in datasets for classification task, Synthetic Minority Over-sampling Technique (SMOTE) with Tomek Links (SMOTE-Tomek) is used in this experiment. SMOTE improves dataset balance and improves model performance without duplicating data by interpolating between existing instances to create synthetic examples for the minority class (Chawla et al., 2004). Conversely, SMOTE-Tomek incorporates SMOTE with Tomek Links, which detects and eliminates overlapping or unclear samples close to the decision boundary. Although SMOTE is the best option for expanding the representation of minority classes, SMOTE-Tomek is the better choice when both data cleansing and balancing are required (Sasada et al., 2020).

The classification problem is made binary by emphasizing only on positive and negative classes, which produces more precise predictions and clearly defined decision limits (Jain et al., 2021). The conversions of the scores are shown in Table 3.7. The corpora are trained and evaluated using Support Vector Machine (Vapnik & Lerner, 1963), Random Forest (Tin, 1995), CNN-LSTM (Fukushima, 1980; Hochreiter & Schmidhuber, 1997) and BERT (Devlin et al., 2019).

Table 3.7
Annotated Scores Conversions

Annotated score	Class
> 0	1 (positive)
< 0	-1 (negative)

On the contrary, sentiment regression is a natural language processing task aimed at predicting a continuous sentiment score, rather than categorizing text into discrete labels like positive, negative, or neutral. This approach provides finer-grained insights into the intensity of sentiment, making it useful for applications such as financial market predictions or customer satisfaction analysis. Sentiment regression is scalable and consistent, often using techniques like regression-based algorithms, pre-trained embeddings, or transformer models.

3.5 Cryptocurrency Price Prediction

In Phase 3, the extracted sentiment features of average sentiment scores are aggregated with the cryptocurrency historical price features of open, high, low, close, and trading volume to predict the closing price of the next day. Feature aggregation is

the process of combining multiple input features into a single dataset to enhance the predictive power of a machine learning model. In this context, sentiment scores are combined with cryptocurrency price features, to provide a more comprehensive view of the variables affecting cryptocurrency prices. The purpose of this aggregation is to capture both market-driven and sentiment-driven effects on the price changes. Different combinations of these features are tested in order to determine the most impactful and informative set of features that best capture the variables influencing cryptocurrency prices, increasing the prediction model's accuracy and stability.

Then, data normalization or feature scaling is performed on the cryptocurrency prices since the scale range for the sentiment scores feature is radically differed from the price features. The range for BTC is over 10,000 whereas the range for ETH is over 100. Two types of feature scaling techniques which are data normalization (MinMaxScaler() function) and data standardization (StandardScaler() function) are tested to select the most optimum scale for cryptocurrency price prediction model.

Data normalisation transformed the values in a feature to a distribution between 0 to 1 or between -1 to +1, while data standardisation rescaled values in a feature to have a mean value of 0 and standard deviation value of 1 (Ali et al., 2014). Feature scaling, through normalization or standardization, ensures all input features in a cryptocurrency price prediction model have comparable scales, preventing features with larger ranges from dominating the model. Testing both methods helps determine the most effective scaling technique for improving the model's accuracy, stability, and optimization performance. By utilising these features, several combinations from the features are experimented to extract the best features for cryptocurrency price prediction.

Bahdanau attention (Bahdanau et al., 2016) is employed in the price prediction model to focus on the most pertinent historical time steps while anticipating future prices. The attention mechanism prioritises significant historical data events, such as abrupt spikes, declines, or volume surges, to enhance the prediction of the subsequent price point. The attention mechanism allows the model to dynamically assess significant aspects, enhancing precision in time-series forecasting, particularly when patterns are non-linear or affected by intricate market dynamics (Bahdanau et al., 2016).

Finally, the best performed feature combinations are reported to compare the cryptocurrency price prediction result between the manual human-annotated corpora

and the automatic-annotated corpora. The training and evaluation are performed for each of the eight datasets.

There are four classification metrics were used to evaluate sentiment classification models, which are: accuracy, precision, recall and F1-score. Accuracy is the ratio of correctly predicted instances to the total number of instances in the dataset. It provides an overall measure of how well a model performs across all classes. Precision is the ratio of correctly predicted positive observations to the total predicted positives. It measures the accuracy of the positive predictions. Recall is the ratio of correctly predicted positive observations to the total actual positives. It measures the ability of the model to capture all the positive instances. F1-score is a metric that combines precision and recall into a single value and is particularly useful in evaluating the performance of classification models. It provides a balance between precision and recall, offering a more comprehensive assessment of a model's effectiveness, especially in situations where there is an imbalance between the classes (Powers, 2011).

On the other hand, there are three metrics are used to evaluate the regression models which are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and adjusted r^2. RMSE and MAE metrics are used to measure the errors in the prediction for regression task, whereas the adjusted r^2 is to evaluate the precision and accuracy of the model (Chicco et al., 2021; Karch, 2019).

The RMSE metric is good for getting unbiased forecasts but in some cases RMSE is bound to penalize significant errors. On the other hand, MAE metric is used to protect outliers and is more robust than RMSE, making it a more appropriate metric for error assessment (Robeson & Willmott, 2023).

The adjusted r^2 is selected as the performance metric because it is more accurate and informative (Chicco et al., 2021). Adjusted r^2 differs from r^2 by testing many independent variables against the model, which is not being performed in r^2. In addition, adjusted r^2 measurement is favoured by most investors as it provides a more accurate assessment of the correlation between one variable and another (Karch, 2019), which aligns with this research study’s focus on cryptocurrency that falls under similar financial domain.

3.6 Summary

The chapter describes the research methodology to construct a prediction model for cryptocurrency price prediction. It emphasised the three-phased methodological framework consisting of data collection, feature extraction, and cryptocurrency price prediction.

The first phase constructs cryptocurrency-related sentiment corpora containing English and Malay texts, producing a total of 10 datasets. The two datasets include the daily price data of Bitcoin (BTC) and Ethereum (ETH), along with the OHLC prices and trading volumes features. The eight datasets consist of sentiment data from various sources, each for BTC and ETH: English and Malay news, as well as English and Malay tweets. Manual and automated annotation tasks were also performed.

The second phase involves extracting relevant features from the datasets. Price features, consisting of the OHLC price and trading volume data, are utilised directly. On the contrary, sentiment features were processed by undergoing text pre-processing to remove unwanted artefacts. Then, the cleaned datasets were used for data annotations. The data is then annotated using both manual and automated methods. Although, manual data annotation is effort extensive and time consuming, the accuracy is hypothesized to be good. Sentiment classification and sentiment regression are applied to classify the sentiment as positive, negative, or neutral and to quantify the sentiment strength, respectively. The evaluation of these sentiment scores is crucial to observe the extracted features' performance.

In the third phase, the sentiment and price features are used to build a price prediction model. The data aggregation was performed by combining price features and sentiment features. Next, normalization of price data was conducted. A regression model was then applied to predict the price of BTC and ETH. The performance of the price prediction model is assessed to evaluate its accuracy and effectiveness.

In this study, two different attention mechanisms are applied in separate components of the pipeline, each selected to suit the specific task. For sentiment analysis, multi-head self-attention is employed within the XLNet model to capture contextual dependencies across the entire input text. This mechanism allows the model to attend simultaneously to different positions in the sequence with multiple attention heads, thereby learning nuanced semantic and syntactic relationships that are critical for

predicting fine-grained continuous sentiment scores. In contrast, for cryptocurrency price prediction, the model incorporates Bahdanau attention within a sequential forecasting framework to dynamically weight historical price and feature sequences. Unlike self-attention, Bahdanau attention computes a context vector based on the relevance of each previous time step to the current prediction, which is particularly effective for modelling temporal dependencies in time-series data.

CHAPTER 4
SENTIMENT DATA VALIDATION

4.1 Introduction

This chapter describes about the sentiment classification and regression model by reporting the evaluation results for XLNet-GRU experimentation with different number of heads using the multi-head self-attention mechanism. Then, the function of the attention mechanism is expressed through visualization. In addition, the performance of own sentiment model is compared with BERT as the state-of-the-art (SOTA) baseline model. The sentiment classification and regression are conducted to show the performance of the extracted features from the collected data. The sentiment data validation is being done to ensure data accuracy, consistency, reliability, and proper alignment between sentiment and market data. Moreover, issues such as noise, mislabelled sentiment, and overfitting can be catered with data validation. Through these checks, models can generalize better and produce better-quality cryptocurrency price predictions. Part of this chapter have been published in Zamani & Kamaruddin (2024).

4.2 Sentiment Classification

The performance of several models on the Bitcoin (BTC) and Ethereum (ETH) datasets from prior research studies are highlighted in Table 4.1, which compares the sentiment classification results. Performance metrics comprise of accuracy and precision are presented. Based on Table 4.1, most existing research studies used Bitcoin tweets dataset and mainly in English language. From the overall findings, it can be deduced that since Malay language resources are very limited, English data tend to give a higher performance than Malay. Even so, Malay data still managed to give comparable results to English data. The comparison further validates the robustness of the utilised approaches by highlighting the improvements in sentiment analysis made possible by utilizing contemporary techniques.

Table 4.1
Summary Of Sentiment Classification Results From Prior Studies

Author	Dataset	Model	Result
Alghamdi et al. (2022)	17.608 English tweets (BTC), 17,373 English tweets (ETH)	SVM	Accuracy = 93.95% (BTC), Accuracy = 95.59% (ETH)
Pant et al. (2018)	7,454 English tweets (BTC)	BOW + Random Forest	Accuracy = 78.49%
Jahjah & Rajab (2020)	25,746 English tweets (BTC)	GloVe + CNN-LSTM	Accuracy = 88.7%
Passalis et al. (2022)	119,286 English news (BTC)	BERT	Accuracy = 68%
Cerda (2021)	9,186 English tweets (BTC)	BERT	Precision = 72%

The experiment on SVM classifier was conducted in the study by Alghamdi et al. (2022) that used a dataset with 17,608 tweets for BTC and 17,373 tweets for ETH using categorical sentiment labels of -1 (negative), 1 (positive), and 0 (neutral). The same approach was implemented using this study's dataset containing 6,000 BTC and 6,000 ETH English tweets. Alghamdi et al. (2022) dataset is considerably larger than this study's dataset, thus, achieved higher accuracy results of 93.95% (BTC dataset) and 95.59% (ETH dataset). Although this study applied SVM and trained on 6,000 tweets for each BTC and ETH, the model produced a significant accuracy of 75.5% and 76.3% respectively. Therefore, the findings demonstrate that, even though larger datasets can result in greater performance, smaller datasets can still offer valuable insights, particularly in situations where resources or data collecting are scarce.

A study by Pant et al. (2018) manually annotated 7,454 Bitcoin tweets in English and reported an accuracy result of 78.49% using random forest, thus, this study's English Bitcoin tweets dataset obtained an accuracy of 74.3% with 6,000 number of instances in the corpus. Hence, a difference of 4.19% can be observed. The difference in the performance may be due to the corpus size impact or model consistency. A larger dataset could potentially lead to a better accuracy, but it is not always the case. The pre-processing of data, the contents, or feature extraction can also play a role in achieving a good accuracy. The study's technique may still be enhanced by optimising the model, expanding the dataset, or improving the feature set.

In comparison with the Bitcoin English tweets corpora and the same GloVe + CNN-LSTM model implemented in Jahjah & Rajab (2020), this study achieved an accuracy of 64.4% with 6,000 instances, while Jahjah & Rajab (2020) obtained an

accuracy of 88.7% with 25,746 number of tweets data. The study utilisation of smaller dataset indicates that limited datasets might yield significant insights, albeit they may not perform as well as models trained on bigger datasets. The gap of 24.3% between the two experiments suggest room for improvement in this study. The model's performance could be enhanced with the incorporation of additional data or alternative methodologies.

The study by Passalis et al. (2022) and Cerda (2021) applied BERT language model to evaluate the sentiment model. Passalis et al. (2022) used 119,286 Bitcoin news (English) to be trained on BERT model and achieved an accuracy of 68%, while Cerda (2021) used BERT to learn 9,146 English Bitcoin tweets and reported a precision of 72% as the result. This study achieved 16.5% higher in accuracy than Passalis et al. (2022) for English Bitcoin news with only 1,518 instances. Meanwhile, with lesser number of English Bitcoin tweets than Cerda (2021)'s dataset, a precision of 63.6% was attained, 8.4% lower than reported in (Cerda, 2021). In spite of this, the datasets in this study still managed to produce results at par with the experiments conducted in the stated existing research studies.

Moreover, BERT model's performance was noteworthy to be used as the baseline for this study because it surpassed the accuracy of Passalis et al. (2022) despite using a significantly smaller dataset (1,518 instances vs. 119,286 instances). This suggests that BERT model can be efficient with relatively smaller datasets, with a well-annotated datasets and the model is properly optimized.

From the observations of the experiments conducted, the four machine learning models are employed, which are: SVM, Random Forest, GloVe combined CNN and LSTM, and BERT. The models were selected for its ability to capture both simple and complex patterns in text data. SVM and Random Forest provides robust traditional machine learning approaches, while GloVe+CNN-LSTM and BERT offer powerful deep learning techniques that can effectively capture contextual and semantic nuances in the text.

Additionally, one of the factors that can be enhanced to achieve better sentiment classification accuracy is catering class imbalance issue. Sentiment datasets often suffer from imbalance, where certain classes (e.g., neutral) dominate others (e.g., positive or negative). This imbalance can caused bias in the model's predictions and concede the

reliability of the results. SMOTE (Chawla et al., 2002) generates synthetic samples for the underrepresented classes, while TOMEK (Tomek, 1976) links help remove borderline instances, improving the quality and reliability of the dataset. Therefore, the study applies SMOTE-TOMEK technique. The following section outlines the experimental setup used to implement the methods and validate the results.

4.2.1 Experimental Set-up

The daily average sentiment score is taken as the sentiment feature for sentiment classification utilised across all the eight datasets. For sentiment classification, all instances with neutral sentiment scores are removed. The classes that were taken are positive denoted by 1 and negative denoted by -1. Each of the dataset have imbalanced instances for the positive and negative classes. Hence, the SMOTE-Tomek technique was applied to all the sentiment datasets with the aim to enhance the efficiency of the sentiment classification model. Table 4.2 shows the number of instances in each class for each sentiment datasets.

Table 4.2
Number Of Datasets After SMOTE-Tomek Application

Dataset	Before SMOTE-Tomek		After SMOTE-Tomek	
	1	-1	1	-1
BTC English news	756	353	714	714
ETH English news	588	200	571	571
BTC Malay news	801	492	721	721
ETH Malay news	924	300	886	886
BTC English tweets	2,464	993	2,353	2,353
ETH English tweets	2439	650	2,367	2,367
BTC Malay tweets	2,863	1,679	2,632	2,632
ETH Malay tweets	3,276	1,150	3,137	3,137

Based on Table 4.2, it can be observed that the positive class tend to have more instances than the negative class. Major imbalanced can be seen on tweets datasets. Following the SMOTE-Tomek application with random_state setting of 130, the datasets were balanced, as evidenced by rising numbers for the minority classes, which brought the dataset sizes closer to the majority classes. Then, the balanced datasets are used to assess the performance of four sentiment classification models, which are: SVM classifier, Random Forest classifier, CNN-LSTM deep learning model, and BERT model.

For sentiment classification using SVM classifier, TFIDF vectorizer with max_features = 10000 and ngram_range(1,2) was utilised initially. Then, the model is fit with linear kernel and a 5-fold cross validation was used for training. Lastly, the macro-averaged parameter was used for the model's evaluation.

Subsequently, the Random Forest classification model applied CountVectorizer with max_features = 10000 and ngram_range(1,2). Similarly, 5-fold cross validation macro-averaged parameter were used for the model's training and evaluation.

For the experiment on CNN-LSTM model, GloVe was implemented into the CNN-LSTM architecture for binary sentiment classification, built with the Keras Sequential API. The CNN-LSTM model architecture is presented in Figure 4.1.

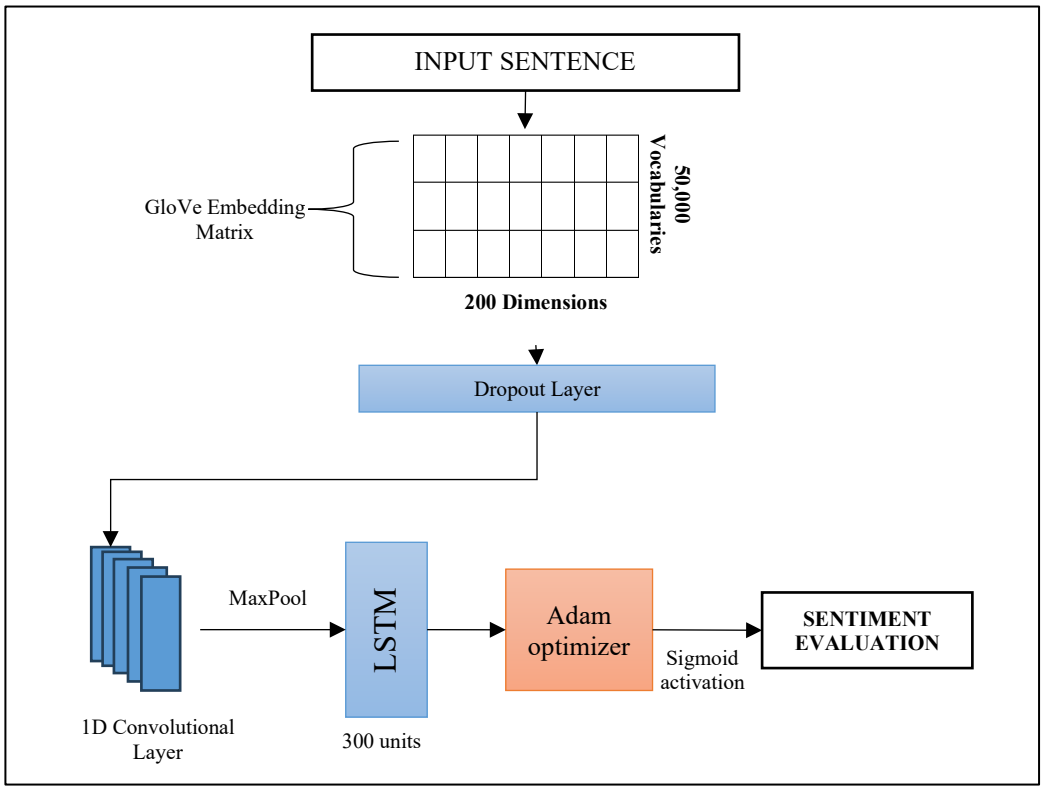


Figure 4.1 GloVe + CNN-LSTM Model

The embedding layer is initialized with a GloVe pre-trained embedding matrix of size 50,000 vocabulary and 200 dimensions, set as non-trainable to leverage pre-trained word vectors. A dropout layer of 0.2 is applied to prevent overfitting, followed by a 1D Convolutional layer (64 filters, kernel size = 5, and ReLU activation) to extract local patterns in the input. A MaxPooling1D layer reduces dimensionality and

highlights key features, which are then passed to an LSTM layer (300 units) to capture sequential dependencies. The final dense layer uses a sigmoid activation function for binary classification, and the model is compiled with the Adam optimizer, binary cross entropy loss, and accuracy as the evaluation metric. The model is trained for 10 epochs with a 20% validation split.

For BERT experiment, pre-trained English and Malay BERT language model were used accordingly. Hence, this study conducted an experiment on BERT using the annotated corpora. The BERT-based model is configured using *BertForSequenceClassification*. It is set to handle a single output label (num_labels=1) for binary classification, with attention outputs and hidden states disabled (output_attentions=False, output_hidden_states=False). Data is loaded into DataLoaders with a batch size of 16, using a RandomSampler for the training set and a SequentialSampler for validation and test sets, ensuring consistent batching. The optimizer is configured as AdamW with a learning rate of 1e-5 and epsilon of 1e-8 to ensure stable training. The model is trained for 5 epochs using a linear learning rate scheduler (get_linear_schedule_with_warmup) with no warmup steps (num_warmup_steps=0), and the number of training steps is dynamically set based on the size of the training DataLoader and the number of epochs. Figure 4.2 illustrates the overview of BERT model for the experiment. The performance results of each classifier are described in the next subsection.

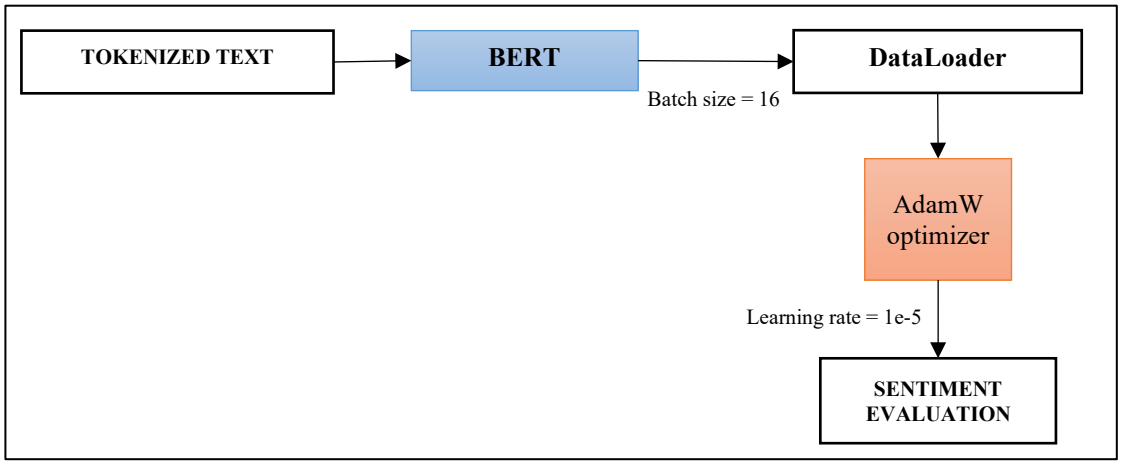


Figure 4.2 BERT Model

4.2.2 Results and Discussion

Table 4.3 presents a performance comparison of sentiment classification using various machine learning and deep learning models for Bitcoin and Ethereum using the eight sentiment datasets. The models used for experimentation are SVM, Random Forest, GloVe + CNN-LSTM, and BERT, with the evaluation metrics: accuracy, precision, recall, and F1-score.

Table 4.3
Sentiment Classification Experiment Results

Dataset	Model	BITCOIN				ETHEREUM			
		Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
English news	SVM	90.0	90.3	89.8	84.9	95.5	94.9	88.2	82.8
	Random Forest	79.0	77.1	55.0	60.6	85.1	72.0	66.5	66.2
	GloVe + CNN-LSTM	89.0	86.0	81.2	85.3	90.0	93.5	85.35	87.0
	BERT	84.5	84.9	83.4	83.8	76.9	83.3	76.9	69.8
English tweets	SVM	95.5	92.6	86.1	87.9	96.3	93.9	86.5	87.5
	Random Forest	74.3	58.2	54.5	55.2	70.5	63.2	55.6	56.3
	GloVe + CNN-LSTM	84.4	82.4	83.7	81.8	87.6	86.7	83.9	84.5
	BERT	78.3	73.6	78.7	78.0	71.5	71.3	61.5	58.6
Malay news	SVM	86.2	89.4	87.7	80.6	89.0	81.3	83.3	86.1
	Random Forest	70.5	54.2	49.6	59.2	72.5	61.2	49.0	59.0
	GloVe + CNN-LSTM	81.5	81.7	76.5	75.2	86.0	83.4	81.5	84.3
	BERT	85.7	85.9	85.7	84.8	87.0	86.7	85.1	83.0
Malay tweets	SVM	85.1	82.5	83.3	83.8	88.4	89.7	81.5	82.9
	Random Forest	62.2	62.3	59.7	50.2	66.9	65.0	56.0	51.7
	GloVe + CNN-LSTM	80.8	86.7	84.6	75.2	85.5	87.7	88.2	89.3
	BERT	78.3	74.8	78.3	75.5	77.8	71.0	70.5	62.7

Based on Table 4.3, across both cryptocurrencies, English news and tweets datasets generally yield better classification results than the Malay news and tweets datasets. From the observation of the English tweets dataset for Bitcoin, SVM achieves a high accuracy of 95.5%, while the same model on Malay tweets drops 10.4% in accuracy. Similarly, for Ethereum, the SVM accuracy on English tweets is higher than

Malay tweets by 7.9%. The difference in the result is because English language resource is well-used in majority of the existing experiments, making the learning of English text more accurate than Malay text.

Furthermore, tweets consistently outperform news articles across most models and metrics. In the case of Ethereum using SVM, English news shows 95.5% accuracy while, 96.3% accuracy on English tweets. The performance gap is wider for Malay datasets, where GloVe + CNN-LSTM achieves an F1-score of 89.3% on tweets versus 75.2% on news. The observation shows that the dataset size can caused an impact on the accuracy. On the contrary, the findings for the experiment of GloVe + CNN-LSTM shows that English news on Bitcoin and Ethereum achieved 70% in accuracy, the highest among other datasets. Malay dataset obtained lower accuracy since there is still no pre-trained GloVe embedding designed specifically for Malay text up to this date. Nevertheless, the results achieved for the Malay datasets are comparably good. The results also show that the BERT model performs better on structured and formal text, such as news articles, compared to informal text like tweets. It is evident from the higher accuracy and F1-scores for both Bitcoin and Ethereum on news datasets. The model's lower performance on tweets may infer that it struggles with the informal language, abbreviations, and noise commonly found in social media data.

From the performance across Bitcoin and Ethereum, Ethereum generally shows slightly better results in Accuracy and Precision across both news and tweets datasets. For example, SVM on English news achieves 5.5% higher accuracy for Ethereum, than for Bitcoin. The GloVe + CNN-LSTM model also presents similar outcome achieving 90.0% accuracy for Ethereum and 86.0% for Bitcoin on English tweets. For Bitcoin, Random Forest performs relatively poorly across all datasets, with the lowest scores among all models on English news. In contrast, Ethereum tends to produce better F1-scores across all models. In Malay tweets, Ethereum again shows better results across models. The findings infer that Ethereum datasets contain more consistent sentiment compared to Bitcoin.

From the overall observation, the English tweet datasets consistently yield the best performance across all models, with Ethereum sentiment generally being easier to classify than Bitcoin. Tweets, regardless of language, tend to offer better results than news articles, likely due to the informal nature of tweets providing clearer sentiment

signals. Finally, English datasets perform better than Malay datasets, potentially due to better pre-training of models on English data or the inherent complexities of Malay language sentiment expression.

4.3 Sentiment Regression

Based on prior research studies, there are very limited experiments on sentiment analysis for cryptocurrency price prediction that reported the evaluation results of the sentiment regression. Since this research study focus on sentiment regression, the findings are compared with the most similar datasets as the curated sentiment-based cryptocurrency corpora. In this section, two main results are discussed; which are, 1) Sentiment Regression using Traditional Machine Learning, and 2) Sentiment Regression using Language Models with Deep Learning.

4.3.1 Sentiment Regression using Traditional Machine Learning

John & Vechtomova (2017b) applied regression models such as Linear Regression and Support Vector Regression on financial news headlines in English. Another similar work by Symeonidis et al. (2017) employed Linear Regression, Boosted Decision Tree Regression and Decision/Random Forest Regression on English financial microblogging (data from Stocktwits and Twitter) and news articles with a total number of 2,333 microblogging data and 1633 news articles data. To date, there is no available related work on Malay financial dataset that can be compared with this research study. Table 4.4 shows the results achieved in the existing similar studies.

Table 4.4
Summary Of Results On Sentiment Regression In Prior Studies

Author	Dataset	Model	Result
John & Vechtomova (2017b)	1,143 English financial news	TF-IDF + SVM	$r^2 = 0.38$
Symeonidis et al. (2017)	2,333 English financial tweets	Linear Regression	MAE = 0.446, RMSE = 0.568

Based on Table 4.4, Symeonidis et al. (2017) reported the errors (MAE and RMSE) as the evaluation metrics for tweets dataset only instead of r^2 metric as applied in John & Vechtomova (2017b). The result shows that Linear Regression model is also

suitable to be experimented with this study's English and Malay news datasets. The result achieved by John & Vechtomova (2017b) indicates that the model could explain 38% of the variance in the sentiment scores derived from the news data, suggesting a moderate level of predictive capability. The result reflects that while TF-IDF with SVM can capture some meaningful patterns in textual financial data, there remains substantial room for improvement in predictive accuracy. On average, the predicted sentiment scores obtained by Symeonidis et al. (2017) deviated from the actual values by about 0.45, and the overall model error which is sensitive to larger deviations was about 0.57. The finding suggests that while linear regression captured some underlying trends in the tweet data, the model's performance is still not optimal.

Both studies demonstrate that traditional machine learning models TF-IDF + SVM for financial news and Linear Regression for financial tweets can provide a baseline level of performance for sentiment regression tasks in finance. However, the moderate r^2 value and relatively high error metrics imply limitations in capturing nuanced financial sentiment, highlighting potential benefits of exploring more advanced models such as deep learning or transformer-based approaches to achieve higher predictive accuracy. The experimental set-up of implementing the methods is explained in the following section.

4.3.1.1 Experimental Set-up

From the eight sentiment datasets, the following experimental set-ups are implemented for each of the datasets. A random state (seed) of 52 was set to ensure reproducibility during the resampling process, and the dataset was split into a training set and a test set using a test size of 0.2. For Linear Regression, SVM, Decision Tree, and Random Forest, text pre-processing are performed by tokenizing the texts, removing stop words, and conducting lemmatization. Next, feature extraction is performed using TfidfVectorizer() for all four models with max_features = 10000, ngram_range = (1,2) as the hyperparameter settings. Then, the features are fed into each of the machine learning model.

The Linear Regression is fit with n_job = 1, while for SVM regressor, the fitting is performed with linear kernel and gamma kernel coefficient. Epsilon = 0.1 is set for

the model training. On the contrary, the training of Decision Tree using DecisionTreeRegressor() utilises the squared_error criterion. Moreover, the strategy used to choose the split at each node is splitter = "best" with the default min_samples_split = 2 and min_samples_leaf = 1. The Random Forest regressor is trained with n_estimators = 100 and a random_state = 42, criterion = "squared_error" as the hyperparameter settings. Random Forest also applied the same minimum samples split for the internal node and leaf node as the Decision Tree regressor. Lastly, a 5-fold cross validation is applied for training all four of the machine learning models.

4.3.1.2 Results and Discussion

The sentiment regression experiment using Linear Regression, SVM and Random Forest are conducted using the sentiment datasets. The results are presented in Table 4.5. The findings show an average r^2 of 0.393 (English news dataset) and 0.411 (Malay news dataset) better than John & Vechtomova (2017b) financial news headlines dataset John & Vechtomova (2017a).

Moreover, the results for SVM are comparable with the findings by John & Vechtomova (2017b), obtaining an r^2 of 0.38 with the implementation of TF-IDF and SVM model on English financial news headlines dataset. Thus, the Bitcoin English news dataset gained 0.028 more of the r^2 value, whereas a slightly lower r^2 of 0.279 was attained by the Ethereum English news dataset with a difference of 0.101 from John & Vechtomova (2017b) dataset.

In spite of this, SVM model did not perform very well on English tweets corpora. Meanwhile, the MAE = 0.349 and RMSE = 0.404 for microblogging datasets reported in Symeonidis et al. (2017) portraying the weak performance of the boosted decision tree model. Even though the performance for the dataset is adequate for the Random Forest model, it managed to produce lesser errors of MAE and RMSE compared to the findings by Symeonidis et al. (2017) using the microblogging dataset. Symeonidis et al. (2017) reported to have achieved an MAE of 0.384 and RMSE of 0.472. The finding shows that Random Forest regression model can trained the manually annotated corpora sufficiently.

Table 4.5
Sentiment Regression Experiment Results

Dataset	Model	BITCOIN				ETHEREUM			
		MAE	RMSE	r^2	Adjusted r^2	MAE	RMSE	r^2	Adjusted r^2
English news	Linear Regression	0.274	0.378	0.439	0.436	0.198	0.293	0.347	0.344
	SVM	0.303	0.388	0.408	0.405	0.250	0.308	0.279	0.276
	Decision Tree	0.322	0.497	0.029	0.024	0.242	0.398	-0.204	-0.210
	Random Forest	0.305	0.432	0.267	0.264	0.236	0.332	0.165	0.161
English tweets	Linear Regression	0.333	0.453	-1.414	-1.416	0.269	0.356	-1.353	-1.355
	SVM	0.200	0.262	0.191	0.191	0.168	0.216	0.143	0.143
	Decision Tree	0.236	0.342	-0.376	-0.377	0.191	0.288	-0.521	-0.522
	Random Forest	0.200	0.272	0.128	0.127	0.159	0.226	0.063	0.063
Malay news	Linear Regression	0.287	0.373	0.365	0.361	0.227	0.316	0.456	0.453
	SVM	0.314	0.377	0.351	0.348	0.253	0.316	0.457	0.454
	Decision Tree	0.292	0.463	0.022	0.017	0.218	0.390	0.173	0.169
	Random Forest	0.264	0.372	0.369	0.365	0.197	0.307	0.487	0.484
Malay tweets	Linear Regression	0.346	0.470	-0.525	-0.526	0.293	0.400	-0.752	-0.753
	SVM	0.228	0.288	0.430	0.430	0.194	0.241	0.353	0.353
	Decision Tree	0.238	0.373	0.042	0.042	0.210	0.320	-0.143	-0.144
	Random Forest	0.200	0.290	0.421	0.420	0.178	0.255	0.275	0.275

Based on Table 4.5, Linear Regression obtained negative values of r^2 and adjusted r^2 shown for English and Malay tweets (Bitcoin and Ethereum) is treated as 0 and denotes a very low accuracy signifying that the regression performed poorly (Chicco et al., 2021) on the tweets datasets. The poor accuracy may be due to the amount of noise in tweets text that is more than in news text. Subsequently, it can be observed that SVM model performed well on the Malay dataset with an average adjusted r^2 of 0.401 (Malay news dataset) and average adjusted r^2 of 0.392 (Malay tweets dataset). In comparison with English dataset, Bitcoin English news achieved a substantial result with an adjusted r^2 of 0.405.

For the Decision Tree experiment, the negative adjusted r^2 values acquired for Bitcoin and Ethereum English tweets datasets, Ethereum English news dataset as well as Malay tweets dataset signifies the poor performance of the Decision Tree in learning the data for sentiment regression. Although the other datasets obtained a positive adjusted r^2 values, the performance is still not good enough in which the values are

much closer to 0. Since decision tree model is known to best fit for a decision-related problem, it may be the caused for the poor results achieved (Song & Lu, 2015). According to the findings shown in Table 4.5, the Random Forest model fits well for Ethereum Malay news data with an adjusted r^2 of 0.484, while it performed moderately for the other datasets.

4.3.2 Sentiment Regression using Language Model with Deep Learning

Zubair et al. (2024) used a hybrid model of VADER and BERT to perform sentiment analysis to predict the price of Bitcoin, Ethereum, and Dogecoin. The study reported a good RMSE result of 0.0241 (Bitcoin), 0.0645 (Ethereum), and 0.0978 (Dogecoin). The findings motivate this study to experiment on VADER and BERT as the baseline to compare with the XLNet model to explore the performance between a lexicon-based approach and the language models. VADER calculates sentiment polarity scores based on a lexicon and heuristics, commonly used for social media and short text, whereas BERT and XLNet utilize transformer-based architecture, with the ability to understand contextual and nuanced language.

Then, the study further explores the performance between the two language models, which are BERT and XLNet with the implementation of deep learning. The experimental set-up and results and discussion are described in the following subsections.

4.3.2.1 Lexicon-based VS Language models Experimental Set-up

VADER implemented for English news and tweets datasets used the *SentimentIntensityAnalyzer()* function directly to obtain the positive, negative, neutral, and compound score. However, for the Malay news and tweets datasets, the *nltk* library of *omw* and *wordnet* are downloaded to handle non-English texts. Then, the *synset()* function with the Malay language setting (lang = “zsm”) is used during translation process from Malay to English. Lastly, the *sentiwordnet* is implemented to obtain the positive, negative, neutral, and compound score. For all the datasets, only the compound score is taken to perform the experiment.

BERT architecture consists of three embeddings which are token embeddings to capture the meaning of each word, segment embeddings that differentiate between distinct sentences in a pair, and position embeddings to provide the order of the tokens. Special tokens such as *[CLS]* are added at the start for classification tasks, and *[SEP]* is used to separate sentences. The combined embeddings are then passed through multiple layers of bidirectional Transformer encoders, which use self-attention to learn contextual relationships between words in both directions (Devlin et al., 2019). The experimental set-up for fine-tuned BERT is shown in Figure 4.3.

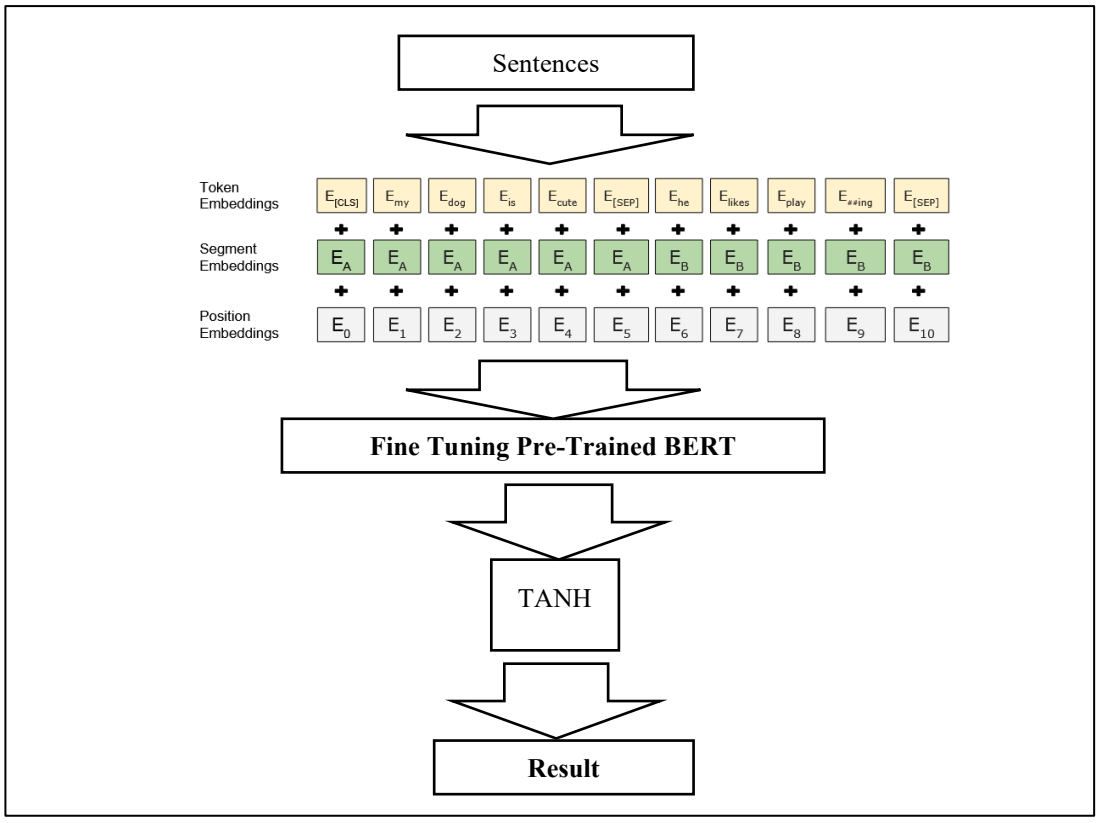


Figure 4.3 Fine-tuned BERT Model

Based on Figure 4.3, the sentence is fed into the three embedding layers. The dataset is loaded into *DataLoaders* with a batch size of 8, where the training set uses a *RandomSampler* to shuffle the data, while the validation and test sets use a *SequentialSampler* to maintain the order. The pre-trained BERT model is initialized with a single output node (num_labels=1) for regression tasks. For English datasets, *BertForSequenceClassification.from_pretrained("bert-base-cased")* pre-trained BERT

model is utilised, whereas for Malay datasets, *BertForSequenceClassification.from_pretrained("malay-huggingface/bert-base-bahasa-cased")* is applied. The optimizer is set to *AdamW* with a learning rate of 2e-5, as recommended in the original BERT paper (Devlin et al., 2019), and the scheduler is configured with a linear learning rate decay over the total number of training steps. The training is conducted for 50 epochs with no warmup steps (num_warmup_steps=0) to ensure gradual adjustment of the learning rate throughout the training. The hyperparameter settings are the same throughout the experiment for all eight datasets.

The XLNet model has 12 layers of transformer blocks, 768 hidden layers (dimensions), and 12 self-attention heads. Then, the *AdamW* optimiser with *MSELoss()* function is used as the criterion for training and evaluation. A learning rate of 2e-5 works best with *AdamW* for XLNet because it ensures stable updates, preserves pre-trained weights, prevents overfitting, and effectively refines task-specific features without disrupting the model's general knowledge. Figure 4.4 shows the fine-tuned XLNet Language Model architecture.

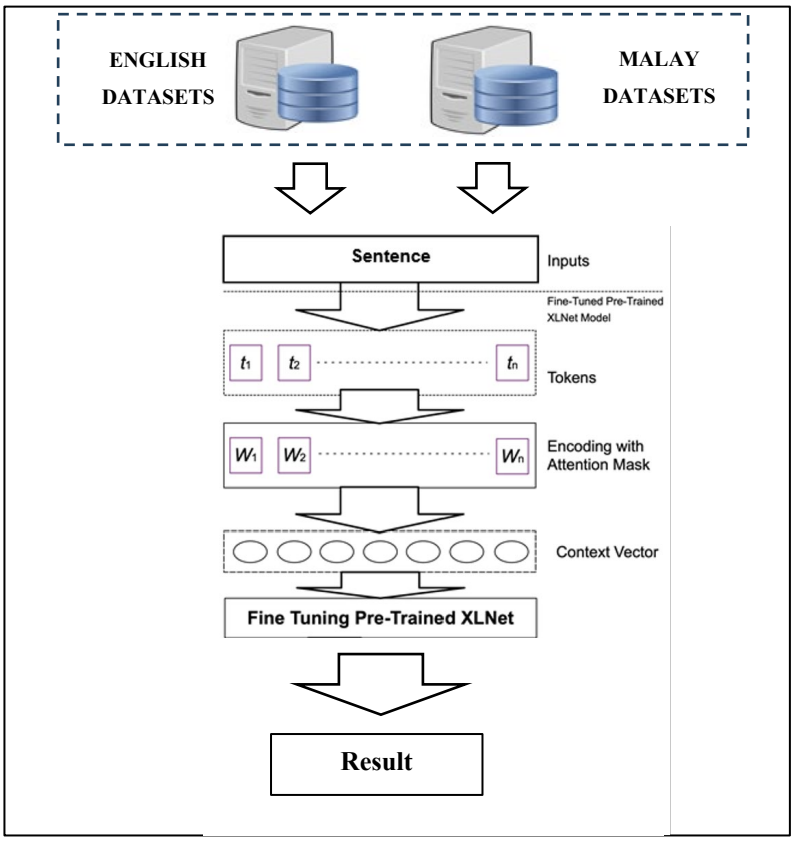


Figure 4.4 Fine-tuned XLNet Model

The architecture shown represents a pipeline that utilizes XLNet for text processing and fine-tuning. The process begins by gathering text data from the English and Malay corpora. These corpora are then split into eight sentiment datasets. For each of the dataset, the sentences in the respective languages are utilised for text pre-processing.

After the cleaned text is passed through the XLNet model, two key processes, which are: tokenization and encoding are executed. In the tokenization phase, the sentences are divided into smaller, manageable pieces known as tokens. These tokens can represent words or sub-words. Then, the tokens are encoded into numerical representations, allowing the model to interpret and process the text data. The encoding is crucial for transforming human-readable language into a format that machine learning algorithms can understand.

The encoded data is then subjected to fine-tuning, a process in which the pre-trained XLNet model is adjusted and optimized for the specific task at hand. Fine-tuning allows the model to better understand the nuances and structure of the dataset. Finally, the model's performance is reported as the result.

4.3.2.2 Results and Discussion on Lexicon-based VS Language models

The eight datasets are used for the experiments on VADER, BERT and XLNet. The three models are compared and shown in Table 4.6. From the findings on both Bitcoin and Ethereum, the results show a clear performance gap between the lexicon-based methods and the language models. Overall, the comparison highlights the clear advantage of language models over lexicon-based approach in sentiment analysis for cryptocurrency-related texts.

Based on Table 4.6, VADER has shown a poor performance in handling sentiment-based cryptocurrency datasets. The adjusted r^2 for all datasets except English Bitcoin news corpus achieved negative values that denotes poor fit of the model. Although English Bitcoin news dataset obtained a positive value, it is still a very low adjusted r^2 leaning towards the zero value. It can be perceived that VADER, a lexicon-based approach is not suitable to be used for the research study especially in catering

Malay texts since VADER was only designed for general English sentiment text and very weak at learning the sentiment-based cryptocurrency English and Malay texts.

Table 4.6
Comparison Results Between VADER, BERT And XLNet Models

		VADER			BERT			XLNet		
		MAE	RMSE	Adj r^2	MAE	RMSE	Adj r^2	MAE	RMSE	Adj r^2
Bitcoin	English News	0.388	0.438	0.081	0.215	0.366	0.482	0.186	0.295	0.681
	Malay News	0.471	0.584	-0.590	0.228	0.352	0.432	0.195	0.312	0.572
	English Tweets	0.362	0.457	-1.457	0.178	0.247	0.390	0.162	0.249	0.428
	Malay Tweets	0.450	0.551	-1.090	0.235	0.312	0.375	0.205	0.305	0.411
Ethereum	English News	0.324	0.423	-0.139	0.176	0.308	0.321	0.099	0.196	0.763
	Malay News	0.485	0.598	-0.956	0.198	0.325	0.451	0.162	0.278	0.602
	English Tweets	0.327	0.423	-2.292	0.185	0.266	0.299	0.143	0.216	0.348
	Malay Tweets	0.424	0.532	-2.161	0.207	0.283	0.340	0.184	0.256	0.389

In contrast, BERT and XLNet demonstrate the strength in understanding contextual and semantic nuances, leading to more accurate sentiment predictions. The superior performance of the language models across both English and Malay suggests that the pre-training on large datasets generalize well even in multilingual and less structured text environments. Additionally, the better results for news datasets compared to tweets datasets imply that formal, well-structured language is easier for models to interpret accurately.

The consistent better performance of BERT and XLNet over VADER indicates VADER's limitations in handling linguistic complexity, whereas the language models appear more resilient to noise and language variation. Therefore, using a more advanced models can lead to more reliable sentiment analysis.

4.3.2.3 BERT-GRU VS XLNet-GRU Experimental Set-up

The development is expanded with the additional of GRU layer on each fine-tuned BERT and XLNet language models. The samples from the corpus are split into 80%-20% train-test set. The 80-20 split offers sufficient data for training while preserving enough for trustworthy assessment, which is an adequate balance. For

machine learning tasks, the ratio is frequently employed as a criterion to prevent overfitting or underfitting. The BERT-GRU model's architecture is presented in Figure 4.5.

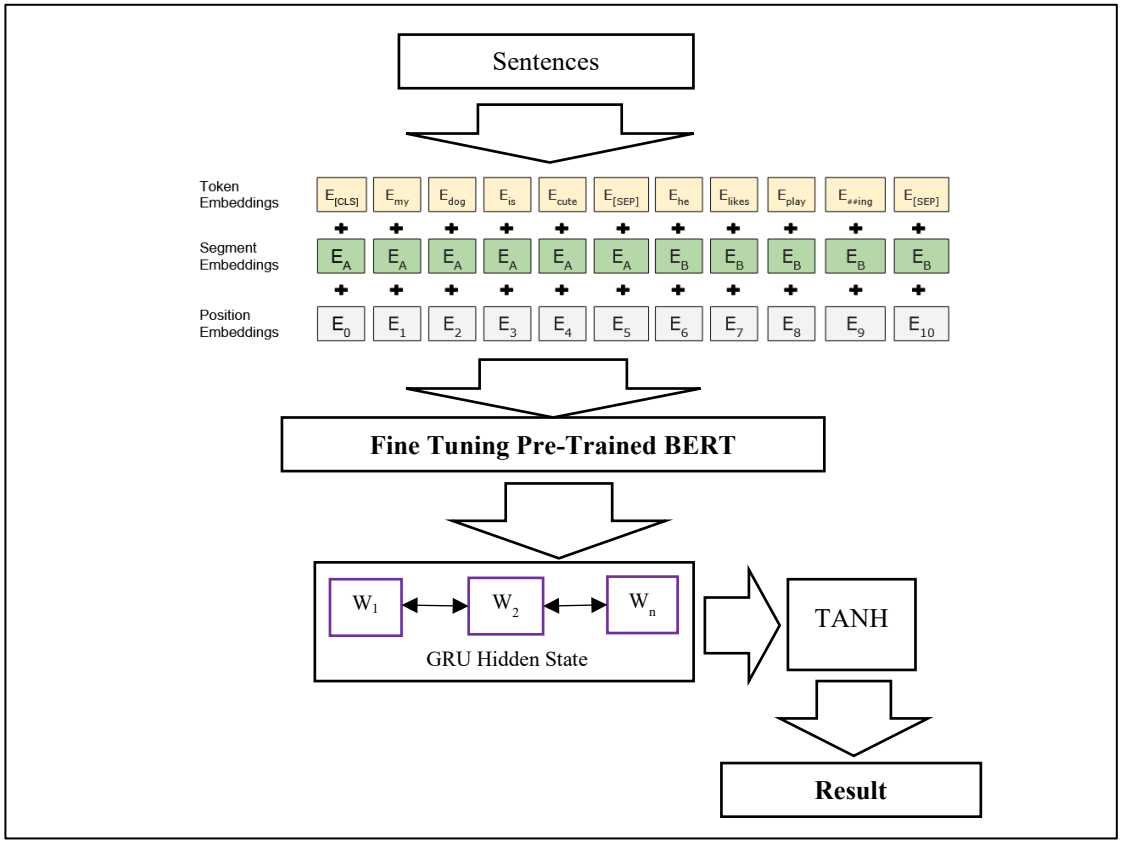


Figure 4.5 BERT-GRU Model

Based on Figure 4.5, the experimental setup for the BERT-GRU model involves combining a fine-tuned pre-trained BERT encoder with a GRU layer. First, input sentences are tokenized and embedded using BERT's token, segment, and positional embeddings. The output contextual embeddings from BERT are then passed through a bidirectional GRU layer, allowing the model to capture temporal dependencies and contextual flow within the sequence. The GRU hidden states are pooled into a single representation, followed by a dropout layer to reduce overfitting. This representation is then fed into fully connected linear layers, followed by a *Tanh* activation function to produce the final output. The architecture uses a hidden dimension of 16, with dropout regularization (dropout = 0.5), and linear layers mapping from 32 to 16, then to a single output unit. The set-up leverages BERT for rich language understanding and GRU for efficient sequence modelling, making it well-suited for sentiment prediction or other

text classification tasks. Similar experimental set-up is also used to implement the XLNet-GRU model, and the architecture is shown in Figure 4.6.

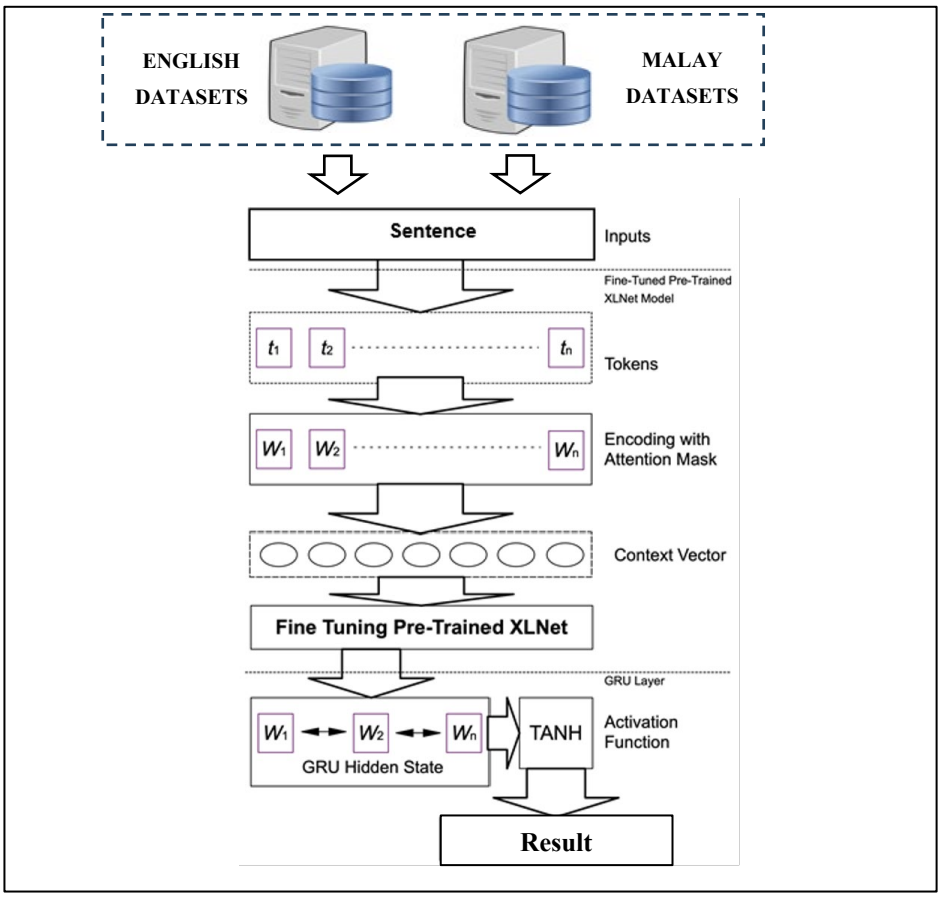


Figure 4.6 XLNet-GRU Model

From Figure 4.6, the fine-tuned XLNet language model is fed to the GRU layer (bidirectional) encapsulating 16 hidden dimensions, one number of layers, dropout value of 0.5, and one output value with Tanh activation function. Tanh activation function is chosen as it supports the range between -1 to +1 and also gives better training performance compared to other activation functions such as *Sigmoid* or *ReLU* (Ali, 2024; Nwankpa et al., 2018). Additionally, GridSearch of 5-fold cross validation was be used for hyperparameter tuning to find the most optimised values.

The model is trained on 4 folds and tested on the remaining fold. This process is repeated 5 times, with each fold used once as the test set. The purpose of this is to evaluate the model's performance more constantly by ensuring it is tested on multiple data splits (Shakya et al., 2024). It is to ensure the model does not see or learn about the

testing data fold so that generalization can be achieved. Since XLNet language model and GRU deep learning layer are utilised, the study's sentiment regression model is denoted as XLNet-*GRU* model. The comparison of the performance results between BERT-GRU and XLNet-GRU model are explained in the next subsection.

4.3.2.4 Results and Discussion on BERT-GRU VS XLNet Language models

Table 4.7 presents a comparative analysis between the BERT-GRU and XLNet-GRU models. The evaluation focuses on Bitcoin and Ethereum and examines four different data sources, which are English news, Malay news, English tweets, and Malay tweets.

Table 4.7
Comparison Results Between BERT-GRU And XLNet-GRU Models

		BERT-GRU			XLNet-GRU		
		MAE	RMSE	Adj r^2	MAE	RMSE	Adj r^2
Bitcoin	English News	0.196	0.344	0.534	0.175	0.280	0.702
	Malay News	0.205	0.337	0.478	0.185	0.301	0.593
	English Tweets	0.150	0.221	0.422	0.150	0.231	0.445
	Malay Tweets	0.214	0.291	0.415	0.194	0.295	0.430
Ethereum	English News	0.154	0.291	0.353	0.085	0.178	0.790
	Malay News	0.179	0.302	0.484	0.151	0.263	0.624
	English Tweets	0.164	0.239	0.325	0.132	0.201	0.364
	Malay Tweets	0.189	0.257	0.366	0.174	0.242	0.404

The results in Table 4.7 demonstrate that the XLNet-GRU model consistently achieves lower MAE and RMSE values and higher adjusted r^2 scores compared to the BERT-GRU model across most datasets. The finding suggests that the XLNet-GRU model enables more effective extraction of semantic and contextual dependencies, particularly in structured text such as news articles. The superior performance in these domains indicates XLNet-GRU's strength in modelling coherent and contextually rich language data.

Nonetheless, the performance gap narrows in Twitter data, where informal language, abbreviations, and noise are prevalent. The observation implies that while XLNet-GRU effectively captures sequential information, it may not optimally differentiate between relevant and irrelevant tokens in less structured contexts. To address this issue, integrating an attention mechanism could allow the model to

selectively emphasize important linguistic features while suppressing noisy or redundant information. Such enhancement is hypothesized to improve the model generalization and predictive performance, particularly for short, noisy, and sentiment-rich text such as tweets to achieve a more balanced and robust performance across different data sources.

The idea in building this approach is based on the analogous result achieved in Chen et al. (2020)'s work on Cantonese tweet rumour detection. The study's adoption of the XLNet-based Bi-GRU with Attention mechanism (XGA) model demonstrates its ability to outperform other models in tasks requiring semantic and sentiment feature extraction, such as rumour detection on Twitter.

The success of integrating XLNet for contextualized embeddings, BiGRU for sequential dependencies, and attention mechanisms for prioritizing critical words highlights the effectiveness of combining advanced NLP techniques. Using a similar hybrid approach for sentiment analysis-based cryptocurrency price prediction is feasible because XLNet is able to extract insightful contextual details from sentiment and financial data. Bi-GRU efficiently simulates time-series price data's sequential dependencies, and the prediction accuracy is improved due to the implementation of attention mechanisms that give priority to important textual and sentimental indications. This approach appears to adapt well to problems involving sentiment-driven decision-making, based on the study's good results across measures (i.e., accuracy, precision, and F-score). This makes it an ideal basis to employ a similar algorithm to predict cryptocurrencies with sentiment.

4.3.2.5 XLNet-GRU+ Experimental Set-up

The same architecture of the XLNet-GRU deep learning model is implemented with the additional component of multi-head self-attention mechanism being integrated into the Bi-GRU deep learning model. The experiment is to observe for any potential performance enhancement on the sentiment regression model. The architecture is illustrated in Figure 4.7.

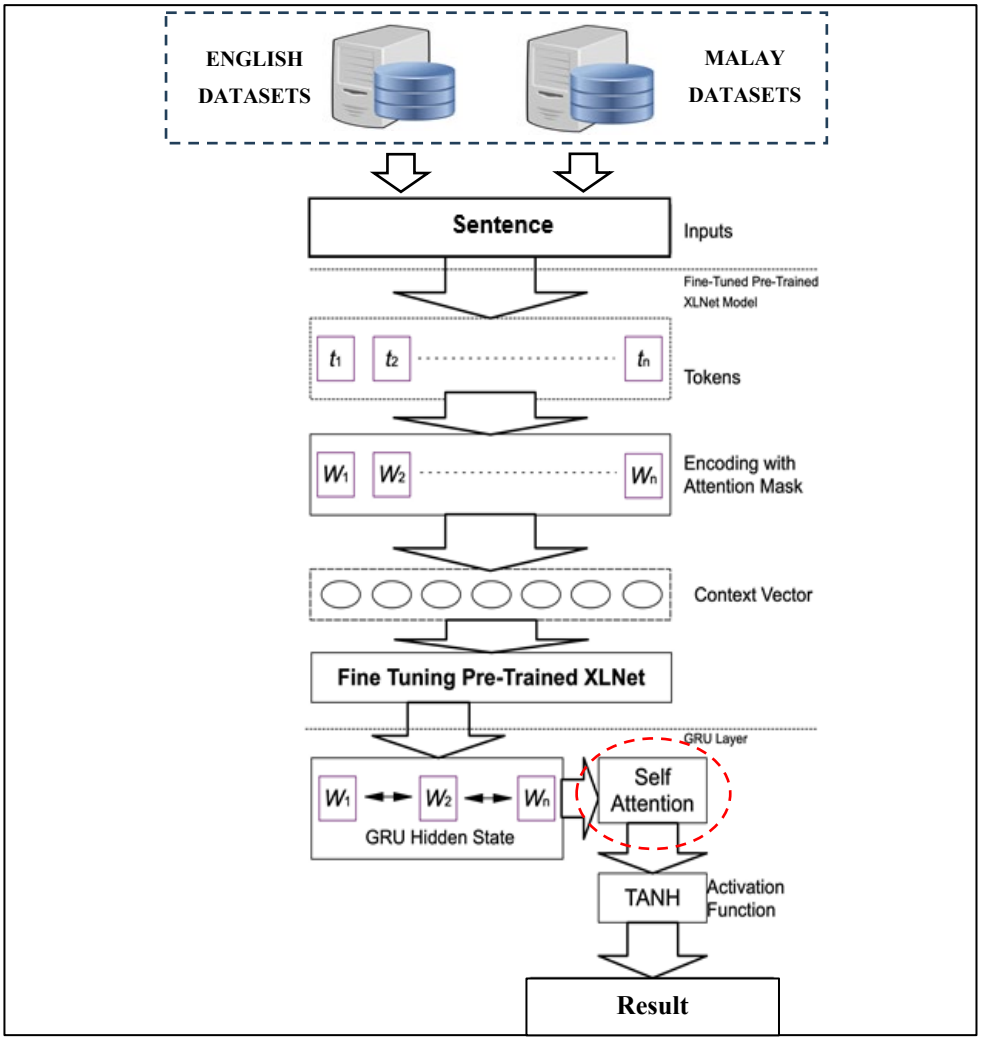


Figure 4.7 Sentiment Regression Model Architecture With Multi-Head Self-Attention Mechanism

The multi-head self-attention mechanism by Vaswani et al. (2017) are appended to the GRU deep learning layer model architecture with the aim to learn the semantic information of the sentences by capturing the significant words associated with sentiment towards the cryptocurrency issue and helps preventing the sentiment regression model from overfitting.

There are four different number of attention heads applied, which are single-head, 2-head, 4-head, and 8-head self-attention. These attention head's values are selected in accordance to the findings by Li et al. (2022), stating that having more than eight attention heads will cause the model to learn the attention weight repetitively, instigating noise in the sentiment regression. In addition, it is important to note that

increasing the number of attention heads also raises computational cost and memory usage. Each attention head has its own set of weight matrices for query, key, and value, and the model must calculate attention separately for each head before combining the results. Therefore, adding more heads increases the amount of data the model needs to process and store during training. For example, with an input sequence of 128 words and a model with 768 hidden dimensions, increasing the number of heads from 1 to 8 led to about 7–8 times more intermediate data that the model must hold in memory and 25–30% longer training time per batch. While more heads allow the model to learn from multiple perspectives in the text, beyond 8 heads, the additional computational cost outweighs the performance benefits, as the model starts learning redundant information. This is also supported by Meng & Wang (2024), stating that an 8-head attention model acts as an efficient, high-performance benchmark compared to models with a higher number of attention heads. The experiment conducted by Meng & Wang (2024) shows that a smaller number of attention heads reduces redundancy and noise computations.

The multi-head self-attention mechanism will produce a weightage for each word in a sentence. Such an approach is done at sentence-level, while maintaining its contextual aspect. Thus, from a given whole sentence, the sentiment words will have a higher weight value than the other words in the sentence, showing the strength or focus on the impactful word that determines its overall sentiment score.

The experimental setup for the XLNet-GRU+ model integrates the XLNet's autoregressive pre-training with GRU's sequential modelling and also a self-attention mechanism. The dataset is loaded into *DataLoaders* with a batch size of 8, where the training set uses a *RandomSampler* for shuffling, while the validation and test sets use a *SequentialSampler* to maintain the order. The pre-trained XLNet model is initialized with a single output node (num_labels=1) for regression tasks. A custom neural network class incorporates a bidirectional GRU with a hidden dimension of 384, doubling the hidden state output size due to bidirectionality. The GRU output is passed through a multi-head self-attention mechanism, capturing the global dependencies in the text.

The final layers include a linear layer of (32, 16) dimension, followed by an output layer of (16, 1) dimension for sentiment regression. Dropout rate of 0.5 is applied to prevent overfitting, and a *Tanh* activation function is used for non-linearity. The optimizer is set to *AdamW* with a learning rate of 2e-5, and a linear learning rate

scheduler is employed over 30 epochs. This architecture is designed to capture contextual, sequential, and long-range relationships in the text, ensuring effective sentiment regression.

The purpose of multi-head self-attention is to focus in learning the significant words of a sentence to determine the sentiment score. Each word in a sentence was assigned an attention weight representing the strength of the word in the sentence. The visualizations of the attention weights examples were taken from the test sets are illustrated in Figure 4.8 and Figure 4.9 for English and Malay news and tweets, respectively.

The aim is to enhance the XLNet-GRU sentiment model performance in predicting the sentiment scores within the range of -1 to +1 based on the scope of study. For this experiment, single-head, 2-head, 4-head, and 8-head self-attention were selected and the results achieved are presented in Table 4.8 for news datasets and Table 4.9 for tweets datasets. Darker shades of colour indicate the stronger impact of the word on the sentence than the lighter-shaded words.

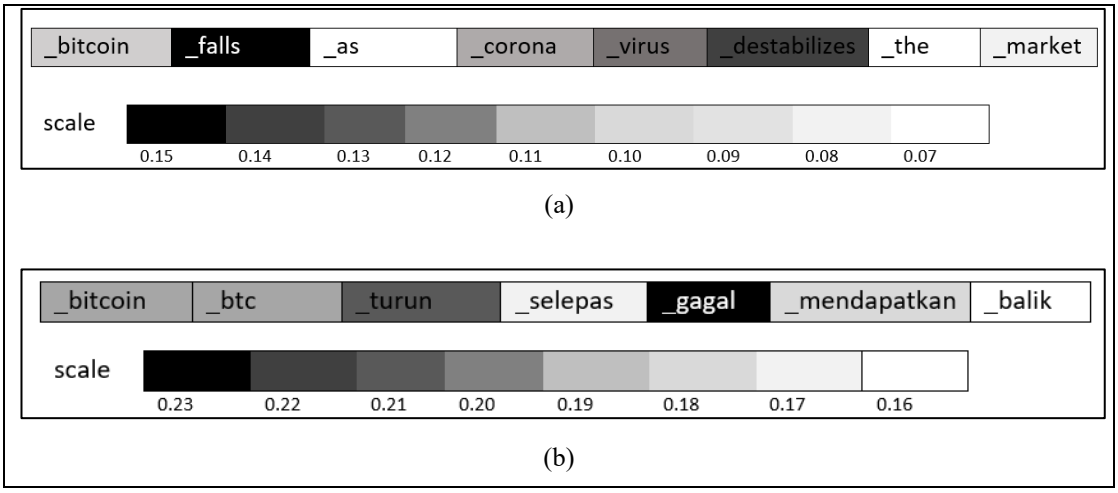


Figure 4.8 Visualization On A Negative Sentiment Word For a) English News And b) Malay News

From the observation in Figure 4.8, the English words “fall” and “destabilizes”, and similarly the Malay words “turun” (translated as “fall” in English) and “gagal” (translated as “fail” in English) in the sentences show stronger attention weight. This implies the significant function of the attention mechanism in the sentiment model to

emphasize more words that represent sentiment, thus producing the overall negative sentiment.

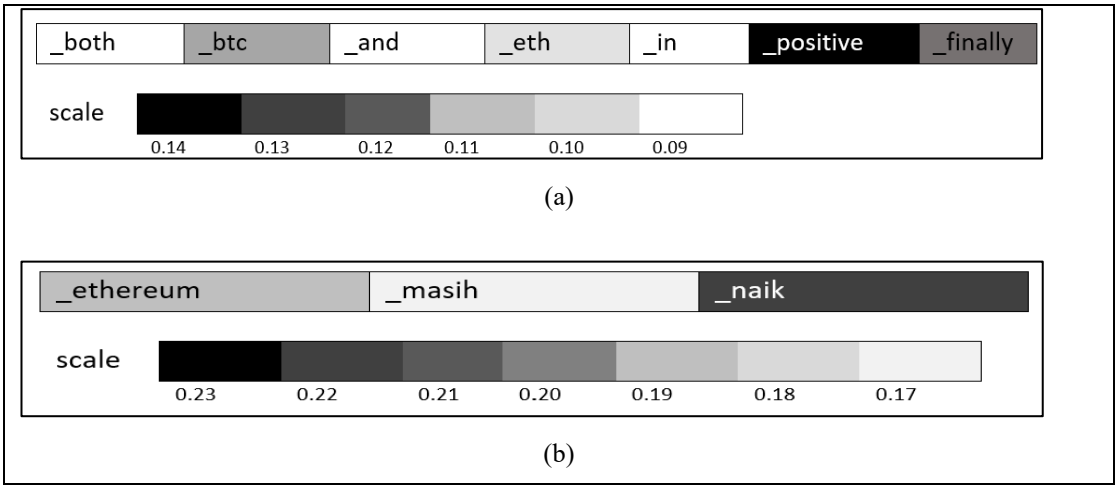


Figure 4.9 Visualization On A Positive Sentiment Word For a) English Tweets And b) Malay Tweets

Similarly, the English word "positive" and the Malay word "naik" (translated as "up" in English) has shown to have the darkest shade with highest attention weight in the sentences. This also infers the usefulness of the attention mechanism to predict the overall positive sentiment of the sentence. The results for the experimented attention heads are shown in Table 4.8 and Table 4.9.

Table 4.8
Results Of Multi-Head Self-Attention Mechanism In XLNet-GRU Sentiment Model Performance (English And Malay News)

ENGLISH NEWS						
Number of attention heads	Bitcoin			Ethereum		
	MAE	RMSE	Adjusted R^2	MAE	RMSE	Adjusted R^2
No attention head	0.175	0.272	0.701	0.084	0.168	0.795
Single head	0.164	0.271	0.711	0.077	0.157	0.812
2-head	0.163	0.268	0.717	0.070	0.163	0.798
4-head	0.162	0.268	0.717	0.080	0.156	0.799
8-head	**0.165**	**0.265**	**0.722**	**0.074**	**0.156**	**0.815**
MALAY NEWS						
Number of attention heads	Bitcoin			Ethereum		
	MAE	RMSE	Adjusted R^2	MAE	RMSE	Adjusted R^2
No attention head	0.179	0.292	0.608	0.144	0.255	0.631
Single head	0.176	0.290	0.615	0.140	0.252	0.641
2-head	**0.186**	**0.285**	**0.628**	0.145	0.255	0.632
4-head	0.175	0.293	0.605	0.139	0.250	0.647
8-head	0.186	0.287	0.621	**0.139**	**0.248**	**0.653**

From Table 4.8, there are substantial outcomes obtained from the self-attention mechanism. There is a small improvement of 0.021 and 0.020 in using the 8-head self-attention compared to without the self-attention head for Bitcoin and Ethereum English news respectively. Comparable result can be observed for Ethereum Malay news with an adjusted r^2 of 0.653. On the contrary, the 2-head self-attention is better for Bitcoin Malay news with 0.020 improvement than the no-head self-attention. The findings show the advantage of applying the self-attention mechanism to improve the XLNet-GRU sentiment model's performance.

The 8-head self-attention also more suited for Bitcoin and Ethereum English news, and also for Ethereum Malay tweets with an adjusted r^2 of 0.464, 0.379, and 0.418, hence, in comparison with the no attention head performance, the 8-head self-attention improved by 0.029, 0.031, and 0.037 in the adjusted r^2, whilst the 2-head self-attention improved the Bitcoin Malay tweets performance by 0.031. The findings verified that the multi-head self-attention mechanism does provide a slight improvement gain in the XLNet-GRU sentiment model's performance with a designated number of attention head(s).

Table 4.9
Results Of Multi-Head Self-Attention Mechanism In XLNet-GRU Sentiment Model Performance (English And Malay Tweets)

ENGLISH TWEETS						
Number of attention heads	Bitcoin			Ethereum		
	MAE	RMSE	Adjusted R^2	MAE	RMSE	Adjusted R^2
No attention head	0.150	0.223	0.435	0.132	0.194	0.348
Single head	0.148	0.229	0.383	0.127	0.197	0.287
2-head	0.142	0.219	0.433	0.119	0.191	0.327
4-head	0.143	0.219	0.437	0.121	0.192	0.323
8-head	**0.139**	**0.213**	**0.464**	**0.122**	**0.184**	**0.379**
MALAY TWEETS						
Number of attention heads	Bitcoin			Ethereum		
	MAE	RMSE	Adjusted R^2	MAE	RMSE	Adjusted R^2
No attention head	0.189	0.289	0.426	0.159	0.236	0.381
Single head	0.208	0.282	0.453	0.177	0.238	0.366
2-head	**0.183**	**0.281**	**0.457**	0.161	0.235	0.385
4-head	0.207	0.287	0.434	0.162	0.235	0.384
8-head	0.205	0.283	0.444	**0.165**	**0.229**	**0.418**

Based on the overall findings in the implementation of multi-head self-attention mechanism into the sentiment model, English datasets gained better performance compared to Malay datasets due to the very limited availability of Malay language

resources in structuring the language modelling. However, the study's results still managed to produce comparable results with similar related works (Chen et al., 2020). With the comparable results produced in using Malay text, it can be comprehended that there is a potential in adding Malay text to the English text using the XLNet-GRU sentiment model for further exploration.

4.3.2.6 Results and Discussion on XLNet-GRU+ model

The XLNet-GRU with self-attention further improves performance by assigning higher weights to critical sentiment-related words, ensuring a more refined focus on impactful inputs, hence, this approach is imperative in cryptocurrency price prediction because price movements are sentiment-driven, with key terms or phrases such as "crash" or "bullish", carrying significant weight.

The self-attention mechanism identifies and prioritizes these terms, aligning sentiment with market trends. While the models can be used solely, integrating additional financial features such as historical prices and trading volume together with sentiment feature is recommended for better prediction accuracy, as cryptocurrency prices are influenced by both sentiment and market data. The results of VADER, BERT-GRU and XLNet-GRU+ are displayed in Table 4.10.

Table 4.10
Comparison Results Between VADER, BERT-GRU And XLNet-GRU+ Models

		VADER			BERT-GRU			XLNet-GRU+		
		MAE	RMSE	Adj r^2	MAE	RMSE	Adj r^2	MAE	RMSE	Adj r^2
Bitcoin	English News	0.388	0.438	0.081	0.196	0.344	0.534	0.175	0.265	0.722
	Malay News	0.471	0.584	-0.590	0.205	0.337	0.478	0.185	0.285	0.628
	English Tweets	0.362	0.457	-1.457	0.150	0.221	0.422	0.150	0.213	0.464
	Malay Tweets	0.450	0.551	-1.090	0.214	0.291	0.415	0.194	0.281	0.457
Ethereum	English News	0.324	0.423	-0.139	0.154	0.291	0.353	0.085	0.156	0.815
	Malay News	0.485	0.598	-0.956	0.179	0.302	0.484	0.151	0.248	0.653
	English Tweets	0.327	0.423	-2.292	0.164	0.239	0.325	0.132	0.184	0.379
	Malay Tweets	0.424	0.532	-2.161	0.189	0.257	0.366	0.174	0.229	0.418

From Table 4.10, the performance gap between VADER, a lexicon-based approach and the language models is substantial. VADER shows limited ability to capture the complex linguistic nuances and context-dependent semantics present in both news and tweet datasets, resulting in high errors and negative or near-zero adjusted r^2 values. The finding indicates that lexicon-based approach is inadequate for nuanced financial sentiment prediction tasks, especially when dealing with informal or multilingual text.

Next, the shift from BERT-GRU to XLNet-GRU demonstrates a notable improvement across all datasets, which reflects the combined strength of XLNet's autoregressive pretraining and deep learning architecture. By integrating GRU layers on top of the contextual embeddings, the model can capture sequential dependencies more effectively, allowing it to process both global and local contextual cues. The hybrid structure enhances the model's ability to interpret complex sentence structures and long-range dependencies. As a result, XLNet-GRU proves to be more robust and accurate, particularly for structured data sources such as news, where linguistic patterns are more complex and informative.

Lastly, the integration of attention mechanisms in XLNet-GRU (XLNet-GRU+) further enhances the model performance, particularly for noisy and informal text sources such as tweets. The improvement suggests that the attention layer enables the model to focus selectively on salient features, filtering out less relevant information leading to a better predictive accuracy.

The results infer that the experimental progression from a simple lexicon model of VADER to contextual models of BERT-GRU and XLNet-GRU, and finally to an attention-enhanced hybrid (XLNet-GRU+) reflects a systematic enhancement of linguistic understanding and predictive power. The performance also validates the importance of attention mechanisms in improving the robustness of language models for sentiment analysis.

Additional experiments on other publicly available datasets that are associated with cryptocurrency and other financial domains were conducted using the XLNet-GRU+ sentiment model. For English datasets, only the sentiment data (with sentiment scores ranging between -1 to +1) on Bitcoin are obtained. There is still no publicly available Malay cryptocurrency dataset that can be tested on the XLNet-GRU+

sentiment model; thus, financial and economic news and tweets datasets are attained, containing several issues on Bitcoin and Ethereum. Since the acquired Malay datasets (news and tweets) were unlabelled, the XLNet-GRU+ sentiment model is applied to score the sentiment between -1 to +1 (aligning with the scope of our research study) for each document in the Malay dataset. By using these public datasets, the performance results achieved by the XLNet-GRU+ sentiment model are displayed in Table 4.11.

Table 4.11
Results Of XLNet-GRU+ Sentiment Model Using Standard Labelled Datasets From Other Studies

Study by	Dataset	MAE	RMSE	Adjusted r^2
John & Vechtomova (2017b)	English Bitcoin News (1143)	0.17	0.243	0.594
Al Mercado (2023)	English Bitcoin Tweets (44400)	0.018	0.049	0.959
Zolkepli (2022)	Malay Financial News (23360)	0.124	0.172	0.624
Zolkepli (2022)	Malay Economic Tweets (39405)	0.043	0.060	0.570

According to Table 4.11, the number of documents for each public dataset fed into the XLNet-GRU+ sentiment model is larger than own datasets. A total of 1,143 English Bitcoin news, 44,400 English Bitcoin tweets, 23,360 Malay financial news, and 39,405 Malay economic tweets are gathered from other sources such as Kaggle and Github. Based on the results for English news and tweets corpora, it can be observed that the larger number of datasets contributes to the enhancement of the XLNet-GRU sentiment model's performance. Even though it is incapable to compare the Malay datasets with the same domain from other datasets, similar topic is taken instead, and the findings signify that the XLNet-GRU+ sentiment model manage to perform well in terms of predicting the unlabelled Malay data, achieving results at par with the labelled Malay data. Therefore, it can be implied that the XLNet-GRU+ sentiment model is sufficient to be applied for other financial or economic news and tweets data in addition to cryptocurrency sentiment data.

4.4 Summary

This chapter explores the experimental setups for sentiment analysis, focusing on both sentiment classification and sentiment regression. Sentiment classification was

conducted with models like SVM, Random Forest, GloVe + CNN-LSTM, and BERT, showcasing the evolution from traditional to state-of-the-art methods for categorizing sentiments into discrete labels. The integration of advanced preprocessing techniques, TF-IDF vectorization and resampling with SMOTE-Tomek, further enhanced the performance of these models. For sentiment regression, VADER, BERT, XLNet, BERT-GRU, XLNet-GRU and XLNet-GRU with attention mechanisms (XLNet-GRU+) are compared since the models offered scalable solutions for predicting fine-grained sentiment intensity. The insights and tools discussed provide a comprehensive basis for developing a cryptocurrency price prediction system based on sentiment features derived from English and Malay texts.

CHAPTER 5
CRYPTOCURRENCY PRICE PREDICTION

5.1 Introduction

Cryptocurrency price prediction involves forecasting future prices of digital assets like Bitcoin and Ethereum using historical data. This often includes analysing factors such as historical price trends, trading volumes, market volatility, and external influences like news or social sentiment. Two types of experiments are conducted in this chapter. Firstly, implementing only price features, which rely solely on historical price data and market metrics, and secondly, incorporating sentiment features, which analyses public sentiment from news and Twitter sources. Sentiment features are added because cryptocurrency markets are highly speculative and often influenced by public opinion and emotional reactions. Studies show that inclusion of sentiment analysis can improve accuracy, as positive or negative market sentiment can drive price movements beyond traditional technical indicators. For cryptocurrency price prediction, only regression task is performed and not price classification because it provides a continuous output, offering precise price prediction of the target day. Since prices can take any value within a range, regression models can capture the nuances of these variations, which classification models cannot.

5.2 Cryptocurrency Price Prediction using Price Features only

Cryptocurrency price prediction using the price features solely relies on historical price data (OHLC) and trading volume to predict the prices. This method focuses purely on quantitative metrics and applies techniques like time series analysis using deep learning regression models such as the LSTM and CNN models to detect patterns in price movements. The LSTM and CNN deep learning models are chosen to compare with this study's model called Bi-GCL-A because CNN and LSTM are used in most of the existing related studies, and also due to the strength of each in handling financial data (Jain et al., 2021; Mohagheghzadeh et al., 2024; Ye et al., 2022). CNN is suitable for capturing short-term price movements and feature interactions, while LSTM is good at handling long-term temporal relationships, which is crucial for time

series forecasting. The comparison allows leveraging both spatial feature extraction and temporal learning capabilities, thereby improving predictive performance in the highly volatile cryptocurrency market. Hence, this section describes the experiments conducted to predict Bitcoin and Ethereum prices using price features only.

The cryptocurrency price features consist of open, high, low, close prices (OHLC) and trading volume for the current day. The cryptocurrency data is based on the daily date which is taken at 12.00 UTC each day based on Bitstamp exchange for Bitcoin and Ethereum. Figure 5.1 illustrates the architecture of the cryptocurrency price regression model taking only the price features as the input. The study's cryptocurrency price regression model is called "Bi-GCL-A" model, which is applied to all the experiments on price prediction in this chapter.

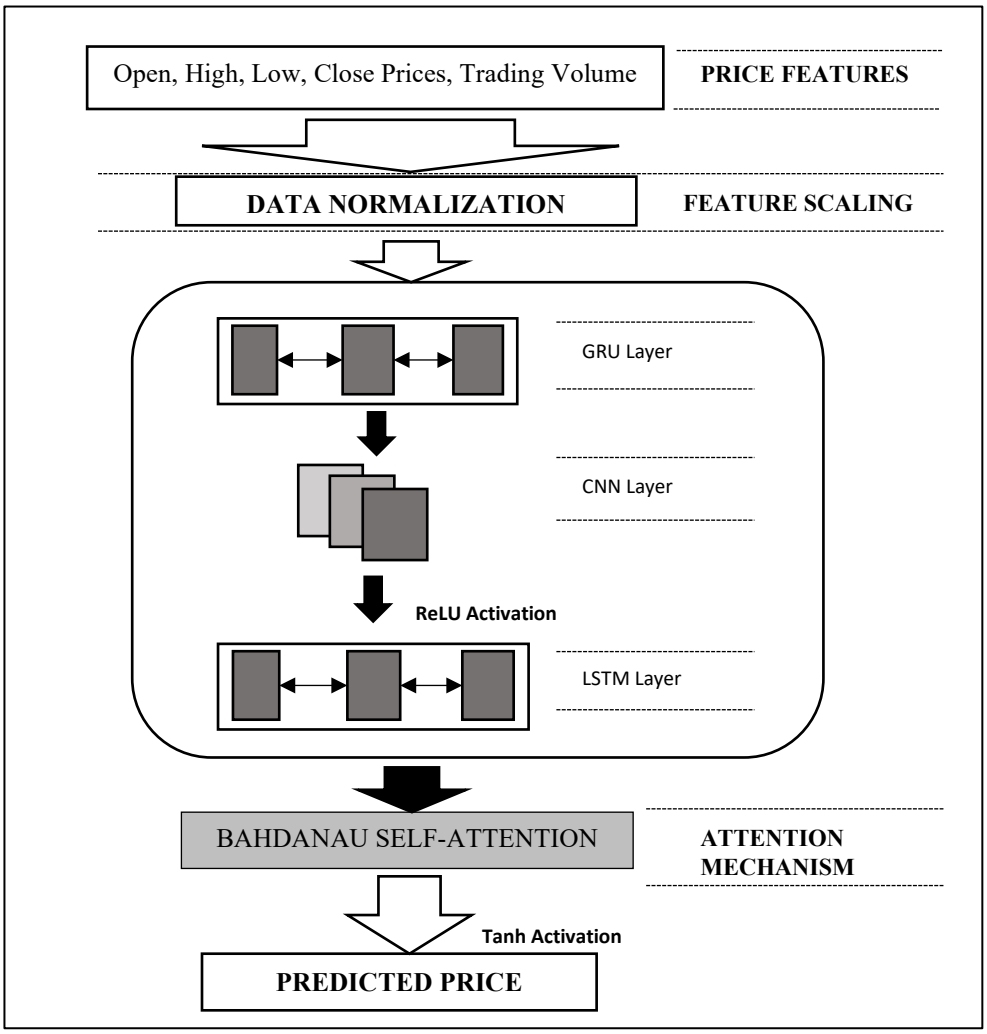


Figure 5.1 Cryptocurrency Price Prediction Model Architecture

Based on Figure 5.1, the experimental setup prepares the data for a cryptocurrency price prediction model by scaling the input features ('open', 'high', 'low', 'close', and 'volume') and the target variable ('TOMORROW_CLOSE'). The *MinMaxScaler()* is used to normalize both the features and the target to a range of 0 to 1, ensuring consistent scaling and improving the model's performance by preventing large feature values from dominating. The scaling is applied separately to the training and testing datasets to avoid data leakage, maintaining the integrity of the testing process. This feature scaling process is applied in all three of the models (Bi-GCL-A, LSTM, and CNN) to ensures the models can effectively learn from normalized data while preserving the independent distribution of the test set for evaluation.

The experiment on Bi-GCL-A model defines a custom deep learning model for cryptocurrency price prediction that integrates multiple neural network architectures and attention mechanisms. The model takes the price data as input. A Conv1D layer with 64 filters and a kernel size of 1 extract features from the input sequence, followed by a Bidirectional GRU layer (64 units). Dropout with a rate of 0.2 is also applied for regularization. The output of Conv1D is then passed to a Bidirectional LSTM layer (64 units) for additional sequential learning. Bahdanau attention mechanism is applied into an attention layer with 64 units to focus on the most relevant features from the LSTM outputs. The final output is passed through a dense layer to predict the target value (closing price). The model is compiled with Adam optimizer and mean squared error (MSE) loss function. It is trained for 50 epochs with a batch size of 32 using. Hence, this setup combines convolutional, recurrent, and attention mechanisms for robust sequential data modelling.

The Bi-GCL-A is compared with the commonly used price prediction models in earlier studies, which are LSTM and CNN models to observe the performance since LSTM and CNN models are most used in past related works (Sathiyapriya et al., 2023; Passalis et al., 2022) to predict cryptocurrency prices. Figure 5.2 illustrates the LSTM architecture and Figure 5.3 shows the CNN architecture.

For LSTM model as shown in Figure 5.2, the input price features have been normalized. Then, the scaled price features are fed into a single LSTM layer with 64 units and process the sequential input data to capture temporal dependencies. The output from the LSTM layer is passed through a Tanh activation function, which scales the outputs to a range of -1 to 1, aiding in capturing non-linear patterns in the data. A dense

layer with a single unit is used as the final output layer to predict the closing price. The model is compiled using the Adam optimizer for adaptive learning and the mean squared error (MSE) loss function, which is suitable for regression tasks. It is trained for 50 epochs with a batch size of 32. This setup focuses on leveraging the sequential learning capabilities of LSTM for time-series forecasting.

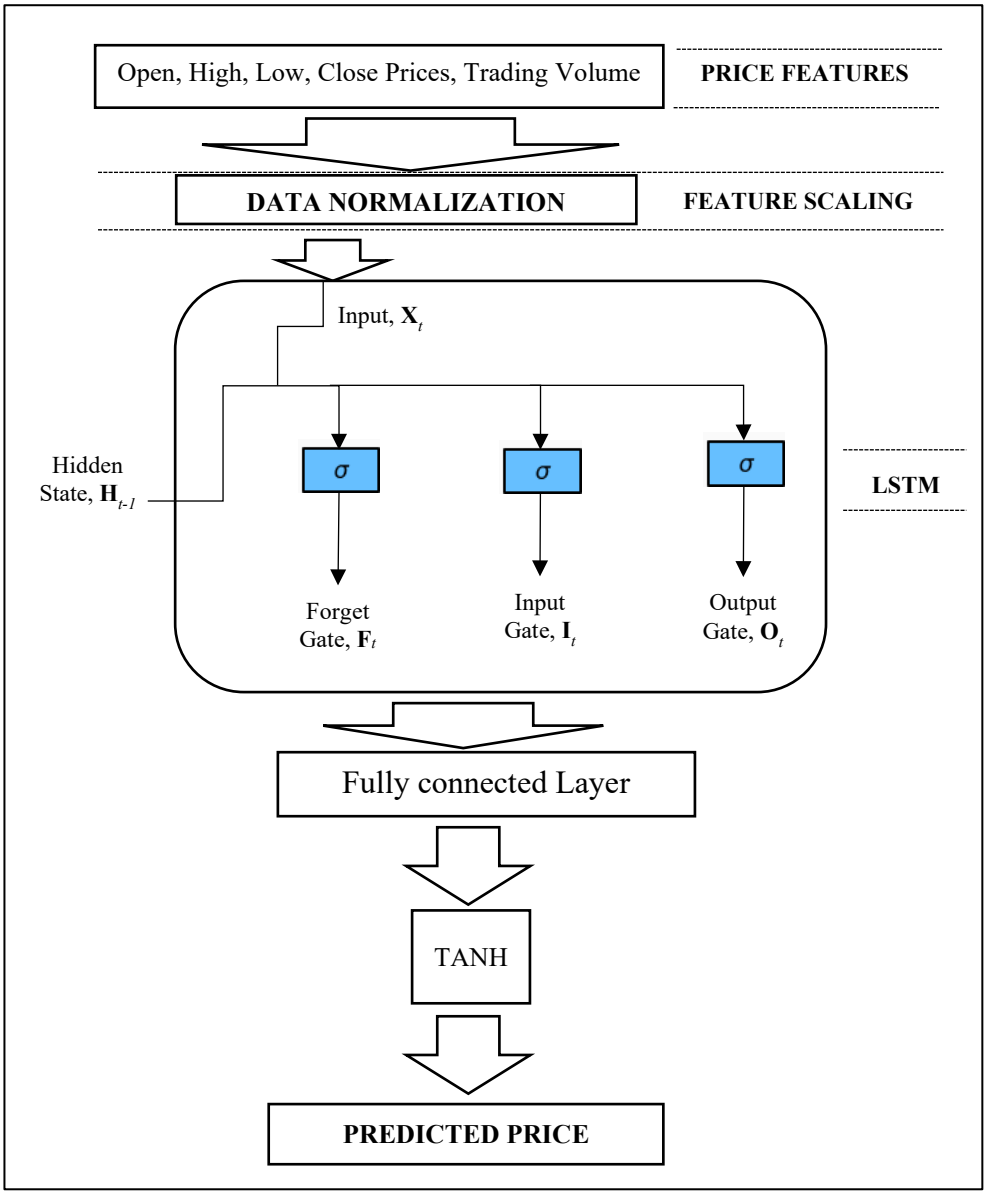


Figure 5.2 LSTM Model Architecture

Subsequently, the CNN model shown in Figure 5.3 for cryptocurrency price prediction processes the input price features into a Conv1D layer with 64 filters, a kernel size of 1, and *ReLU* activation to extract local patterns in the input data. The output of the convolutional layer is passed through a *Tanh* activation function. The final layer is

a dense layer with a single unit, producing the predicted closing price. The model is also compiled using the *Adam* optimizer for adaptive learning and the mean squared error (MSE) loss function. The model is trained for 50 epochs with a batch size of 32. This setup utilizes the strength of CNN in identifying spatial patterns in time-series data.

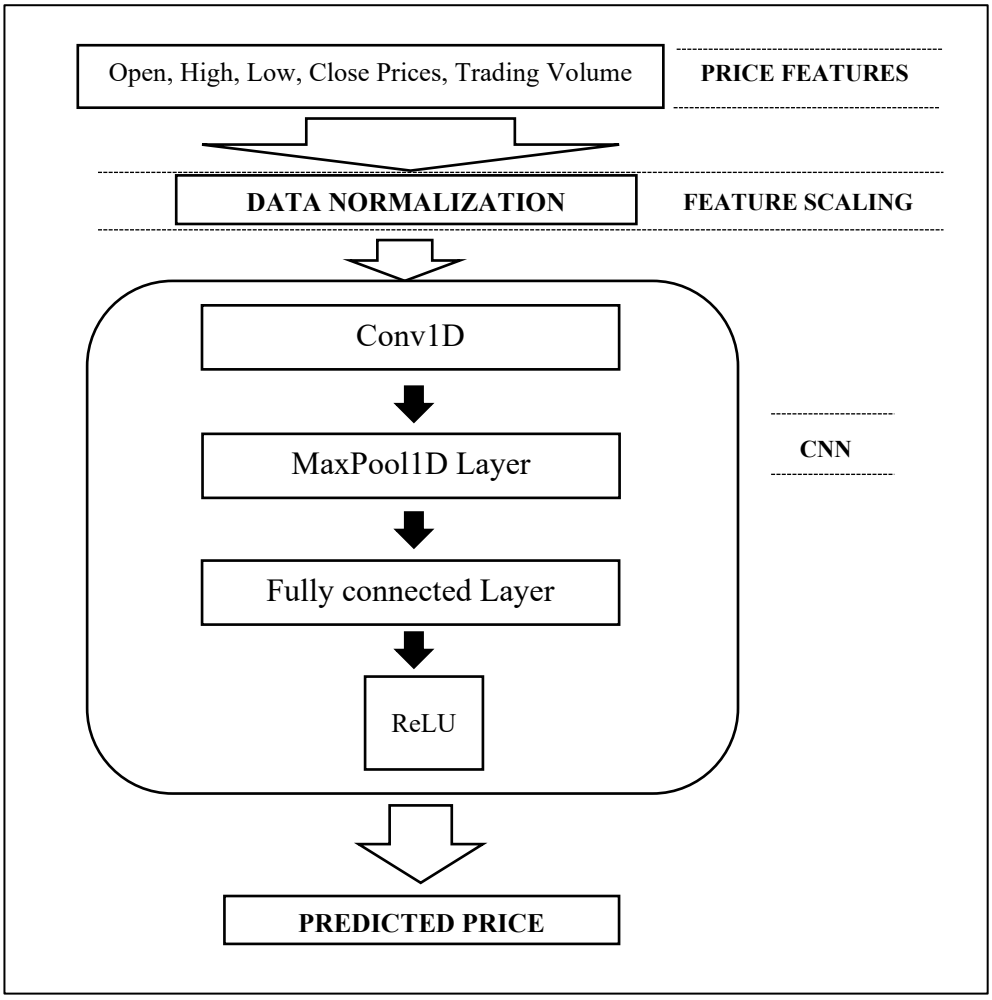


Figure 5.3 CNN Model Architecture

Based on the three models, the experiments follow a similar pipeline by firstly conducting price feature extraction (open, high, low, close, and trading volume), followed by data normalization for feature scaling, then passing through the respective architectures before applying an activation function to generate the predicted price.

Each model differs in terms of the ability to capture temporal and spatial patterns. The Bi-GCL-A architecture combines GRU, CNN, and LSTM layers with a Bahdanau self-attention mechanism. GRU captures sequential dependencies, CNN

extracts local patterns, and LSTM refines long-term dependencies, while attention enhances relevant feature weighting, making it the most complex and hybrid model. In contrast, the LSTM architecture relies solely on LSTM's gating mechanisms which are *forget*, *input*, and *output* gates, to learn temporal dependencies, followed by a fully connected layer to produce predictions. The CNN architecture uses 1D convolution and max-pooling layers to capture spatial patterns and short-term trends efficiently, followed by *ReLU* activation and a dense layer for price prediction.

The comparison of actual closing prices against predicted closing prices using Bi-GCL-A, LSTM, and CNN models for Bitcoin and Ethereum are presented respectively. The blue line represents the actual closing prices, while the red dashed line shows the predictions from the custom model (pred_close_Bi-GCL-A). The LSTM model predictions are represented by the dotted green line (pred_close_LSTM), and the predictions from the CNN model are shown with the purple dashed line (pred_close_CNN). The line representations are applied to all the experiments. The graph highlights the variations between the actual prices and the different model predictions, showcasing the performance of each model in predicting future closing prices over the observed period. Figure 5.4 and Figure 5.5 illustrates the comparisons for Bitcoin and Ethereum respectively. While Table 5.1 and Table 5.2 shows the average error differences generated from the percentage deviation calculations between the predicted prices and the actual prices and the performance results for the three models on Bitcoin and Ethereum respectively. The average error difference in percentage is calculated by comparing the predicted prices with the actual prices over a given number of days.

Based on Figure 5.4, the Bi-GCL-A model predicted price follows the actual price closely with minor deviations of -0.02%, in which it captures both trend and detailed fluctuations more accurately than the other models. On the other hand, the predicted price generated by LSTM model captures the overall trend and trails the actual price quite closely but with noticeable deviations, whereby it misses some of the more detailed fluctuations. Moreover, a visible difference can be seen in the prediction where the peak and lows seem slightly delayed compared to the actual prices.

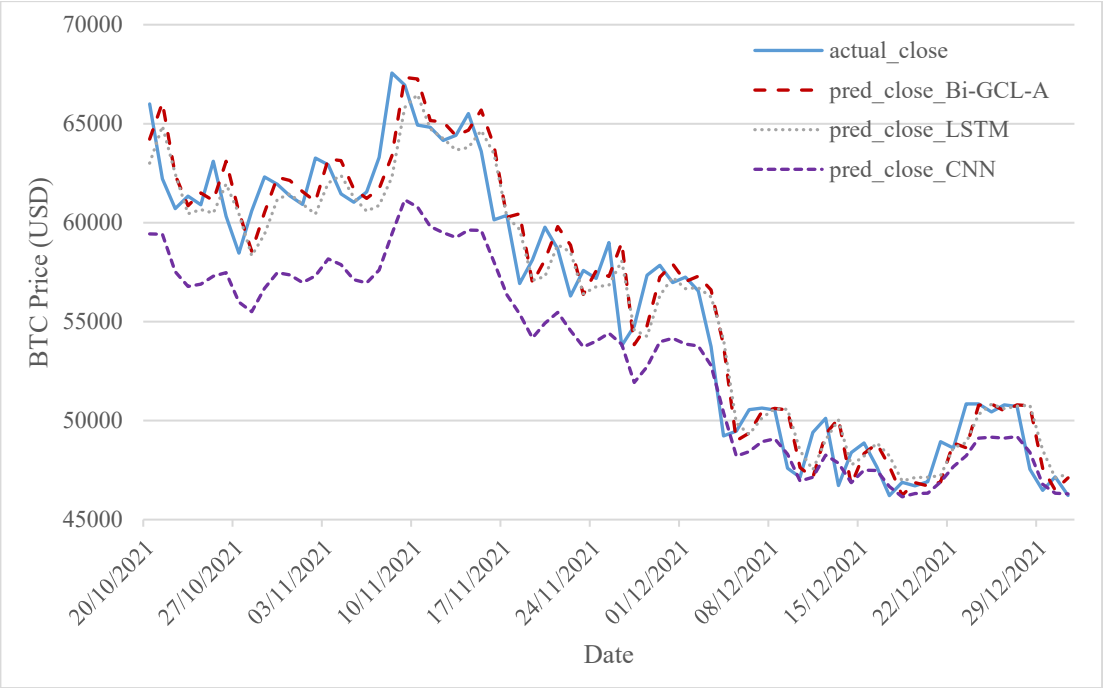


Figure 5.4 Comparison Of Price Prediction For Bitcoin Using Price Features Only

Table 5.1
Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin)

Model	Avg. Error Difference	RMSE	MAE	R^2	Adjusted R^2
Bi-GCL-A	-0.02%	0.058	0.060	0.785	0.787
LSTM	-0.48%	0.063	0.065	0.770	0.772
CNN	5.02%	0.066	0.068	0.760	0.762

The observation in Figure 5.4 displays lower trend for the predicted price than the actual price throughout the period. The CNN model underestimated the price, especially during the price is at its peak. Therefore, CNN model manages to predict the general downward trend but unable to associate intently with the actual price fluctuations.

From the overall findings on using Bitcoin price features only, the Bi-GCL-A model displays the best accuracy in following the actual price trend and fluctuations with minimal prediction error, whereas LSTM and CNN models capture the general trend but with more substantial abnormalities and delays. The CNN model is seen to struggle with underprediction, leading to a less accurate prediction of trend.

In terms of the models' performance, the Bi-GCL-A model has the lowest RMSE and MAE, and the highest r^2 and adjusted r^2, showing the best overall

performance for Bitcoin predictions among the three models. The LSTM model shows an equal MAE result to the Bi-GCL-A model but has the highest RMSE result among the three models, showing the inefficiency in predicting the Bitcoin price accurately. However, CNN model obtained the highest MAE result than the other two models, with the lowest r^2 and adjusted r^2 values, indicating the least effective regression model for Bitcoin price prediction and perform poorer than the Bi-GCL-A and LSTM models.

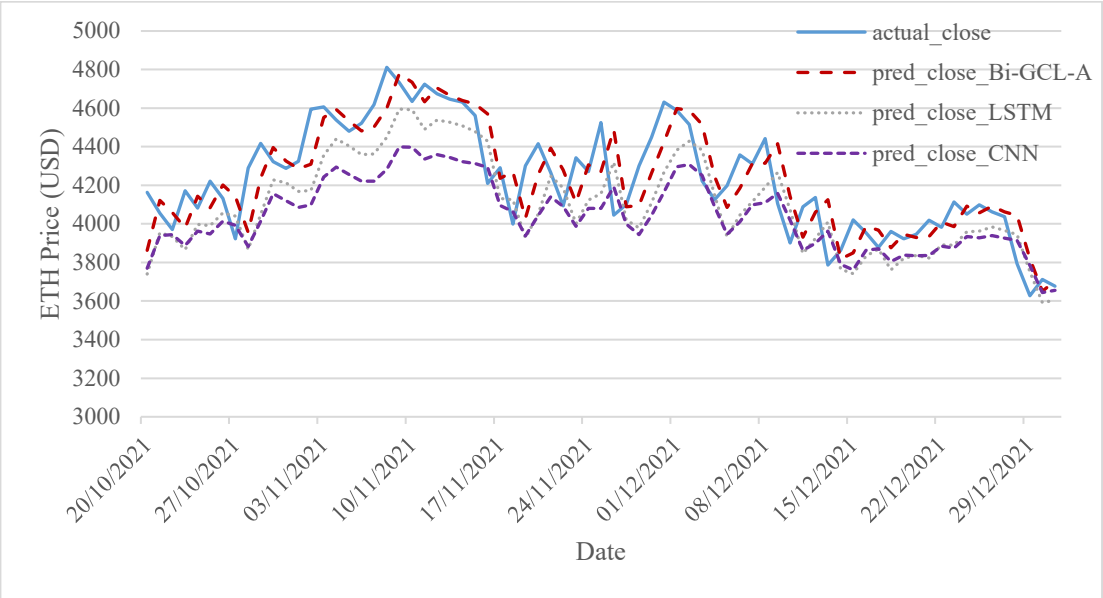


Figure 5.5 Comparison Of Price Prediction For Ethereum Using Price Features Only

Table 5.2
Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum)

Model	Avg. Error Difference	RMSE	MAE	R^2	Adjusted R^2
Bi-GCL-A	0.16%	0.048	0.044	0.780	0.782
LSTM	3.28%	0.070	0.063	0.769	0.771
CNN	4.64%	0.072	0.062	0.761	0.763

Based on the graphs shown in Figure 5.5 for Ethereum, the predicted price produced by Bi-GCL-A model follows the actual price trend very closely with minor deviations of 0.16%. This depicts that the model manages to capture the overall trend and short-term fluctuations accurately. In addition, the alignment between the actual and predicted prices is resilient, inferring the model's good performance.

The LSTM model also shows similar trend between the predicted price and the actual price, whereby close trails can be seen with slight deviances. LSTM also captures

the overall trend and short-term fluctuations well despite the slight difference in prediction, indicating a good model performance. On the contrary, the predicted price from CNN model displays much lower trend than the actual price, whereby it captures the general upward and downward trends but failed to follow the short-term fluctuations precisely. Furthermore, a noticeable gap between the actual and predicted prices can be observed, especially when the price is moving upwards.

In a nutshell, the Bi-GCL-A model shows a slightly better accuracy by showing less error and high alignment with actual prices than the LSTM model in following the actual price trend. The LSTM model also captures the detailed fluctuations with minimal error and good alignment with actual price movements. The CNN model in contrast exhibits a consistent bias towards underprediction and shows a significant gap from the actual prices. Therefore, the Bi-GCL-A and LSTM models perform well, with close alignment to actual Ethereum prices and accurate capture of both trend and fluctuations, while the CNN model consistently underestimates the prices and does not align well with short-term movements.

Based on the models' performance results, Bi-GCL-A model displays the best performance with the least RMSE and MAE results, implying that the model is effective in predicting the Ethereum prices as it explains the most variance in the data. Conversely, LSTM has shown to be the worst performed model in predicting Ethereum prices among the three models, obtaining the highest RMSE and MAE results, indicating the inability to predict the Ethereum price accurately compared to the Bi-GCL-A and CNN models. Based on the observation on the CNN model, it performs better than LSTM but not as good as the Bi-GCL-A in terms of RMSE and MAE. Its r^2 and adjusted r^2 are lower than Bi-GCL-A model but higher than LSTM, indicating a moderate performance model.

From the observations between Bitcoin and Ethereum on using the price features only, it can be deduced that the Bi-GCL-A model has the smallest error difference for both Bitcoin and Ethereum, inferring the highly accurate prediction of Bitcoin price with small digression from the actual prices. The Bi-GCL-A model has shown minimal prediction bias for both cryptocurrencies. LSTM on the other hand has a higher error difference than the Bi-GCL-A model but still showing a good accuracy in predicting the Bitcoin and Ethereum prices. The LSTM model has a small bias for Bitcoin and a larger underprediction bias for Ethereum. On the contrary, the CNN model

displays the lowest accuracy in predicting both Bitcoin and Ethereum prices with error differences of 5.02% and 4.64% respectively. This indicates that CNN has a significant underprediction bias for both Bitcoin and Ethereum. Hence, this confirms the reliability of the Bi-GCL-A model in predicting the daily Bitcoin and Ethereum prices.

Subsequently, the Bi-GCL-A Model outperforms both LSTM and CNN in all metrics for both Bitcoin and Ethereum, indicating it is the most accurate and reliable model for predicting cryptocurrency prices in this analysis. Though, the LSTM model performs less than Bi-GCL-A, but it is comparable to CNN for Bitcoin, but worse for Ethereum. The CNN model performs better than LSTM for Ethereum but worse for Bitcoin in terms of MAE.

The experimental results aligns with the work by Gudavalli & Kancherla (2023) and Phaladisailoed & Numnonda (2018) who used LSTM to predict Bitcoin prices achieving an $r^2 = 0.806$ and $r^2 = 0.992$ respectively. Both works used open, high, low, and close prices as the price features. In addition, Wen & Ling (2023) applied CNN model to predict Bitcoin prices and obtained an $r^2 = 0.988$ as the result. Thus, the Bi-GCL-A model demonstrates a comparable result with the existing related research studies besides its superior performance in this study's experimentation, suggesting it has the best capability in capturing the patterns and predicting the prices for both cryptocurrencies (Bitcoin and Ethereum).

In order to select the type of annotated datasets to be utilised for cryptocurrency price prediction with sentiment features, an experiment to compare the error differences achieved in using manual and automated annotation was conducted. Based on the percentage deviation measurement to calculate the average error differences for the test set of each corpus, the results between using manual annotated and automatic annotated datasets are shown in Table 5.3.

Table 5.3
Price Prediction Error Difference Using Manual Annotation And Automated Annotation

	BITCOIN		**ETHEREUM**	
Corpus	**Manual**	**Auto**	**Manual**	**Auto**
English News	-0.20%	-0.43%	-0.12%	0.77%
English Tweets	-0.39%	-0.20%	0.77%	0.59%
Malay News	-0.45%	-0.52%	-0.54%	0.91%
Malay Tweets	-0.29%	-0.19%	0.85%	0.86%

From Table 5.3, the results show variations in error differences for both methods. For Bitcoin, the error differences are generally closer to zero for manual annotations compared to automated annotations, indicating less deviation. Similarly, for Ethereum, while automated annotations sometimes show higher deviations (e.g., 0.77% for English News), manual annotations tend to be more stable and consistent across all corpora. Moreover, for Bitcoin, average error difference for the manual annotation (-0.33%) is slightly smaller than the automatic annotation (-0.34%). Similar with Ethereum, where the average error difference for manual annotation is 0.24%, while for automatic annotation is 0.78%. It can be observed a large difference between the manual and automatic annotation error difference for Ethereum. Therefore, manual annotation is selected because it typically provides more accurate and reliable sentiment labelling, as it captures nuanced meanings and context that automated systems might miss. This can improve the quality of the data used in the prediction model, leading to better accuracy and reducing noise from incorrect annotations. The visualization for each model comparisons for Bitcoin and Ethereum are shown in the following subsections according to the English news, Malay news, English tweets, and Malay tweets corpora.

5.3 Cryptocurrency Price Prediction using Sentiment Features

Further experiments on the same three models (Bi-GCL-A, LSTM, and CNN) are conducted with added sentiment features using equivalent parameters in the experimental set-up for the price features only experiment.

Figure 5.6 illustrates the comparison between the actual closing prices of Bitcoin and the predicted closing prices generated by three different models which are Bi-GCL-A, LSTM, and CNN models. Table 5.4 shows the error difference between the actual and closing price when using the Bitcoin English news sentiment feature, and the performance results of the three models.

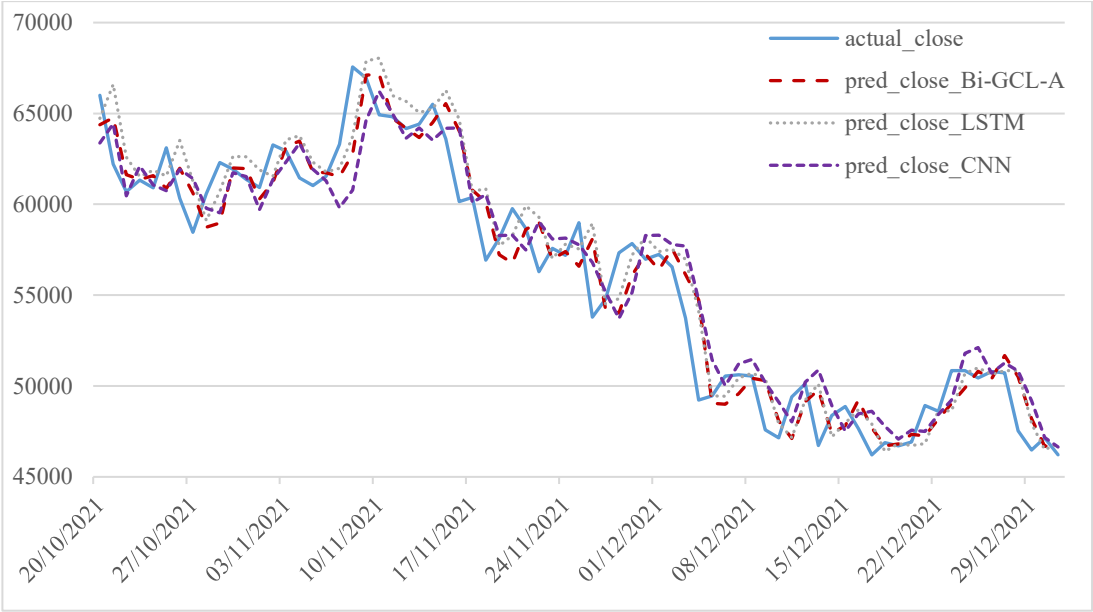


Figure 5.6 Comparison Of Bitcoin Price Prediction Using English News Sentiment

Table 5.4
Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin English News)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	-0.20%	0.037	0.049	0.919	0.917
LSTM	-0.96%	0.054	0.041	0.901	0.900
CNN	-0.63%	0.052	0.039	0.909	0.907

Based on the observation in Figure 5.6, the Bi-GCL-A model's predicted prices closely follow the actual prices, denoting a good accuracy. The predicted price aligns very closely with the actual price trends, indicating that the model effectively captures the trends and movements in Bitcoin's price. Moreover, the LSTM model also shows a well-aligned trend between the predicted prices and the actual prices with good performance. Although there are some discrepancies during harsh price changes, the model still manages to capture the general trend well. The third model, CNN shows a more deviated trend in the predicted prices from the actual prices than the Bi-GCL-A and LSTM models. Even though the model captures the overall trend, there are more instances in which the predicted prices do not align very closely with the actual prices, indicating less accuracy in capturing sudden price changes.

In terms of price volatility, the Bi-GCL-A model handles price volatility effectively, as indicated by the close alignment of the predicted and actual prices even during periods of sharp fluctuations, while the LSTM model handles volatility fairly

well, though not as precisely as the Bi-GCL-A model. It captures major trends but struggles more with the sudden price changes. On the other hand, the CNN model shows the most difficulty in handling volatility among the three models. It tends to have more errors during sharp increases or decreases in price, leading to larger discrepancies.

Throughout the experiments, all three models capture the overall downward trend in Bitcoin prices over the given period. However, the precision of capturing minor trends and fluctuations varies such that the Bi-GCL-A model is best at capturing minor trends and fluctuations, whereas the LSTM model is good at capturing overall trends with moderate accuracy in fluctuations. In contrast, the CNN model performs adequately for overall trends but less accurate for minor trends and fluctuations. Therefore, the Bi-GCL-A model demonstrates the best overall performance, with the closest alignment to actual prices, indicating superior accuracy and robustness in handling both trends and volatility. The LSTM model performs well in apprehending the general trend but has some limitations in accurately predicting sharp price changes. As for the CNN model, it offers a reasonable approximation of the trend but shows larger deviations from actual prices, particularly during volatile periods, making it the least accurate among the three models.

Using the three models, experiments on Ethereum were conducted and the trends of predicted and actual prices are shown in Figure 5.7. The average error difference percentage with the models' performance results are shown in Table 5.5.

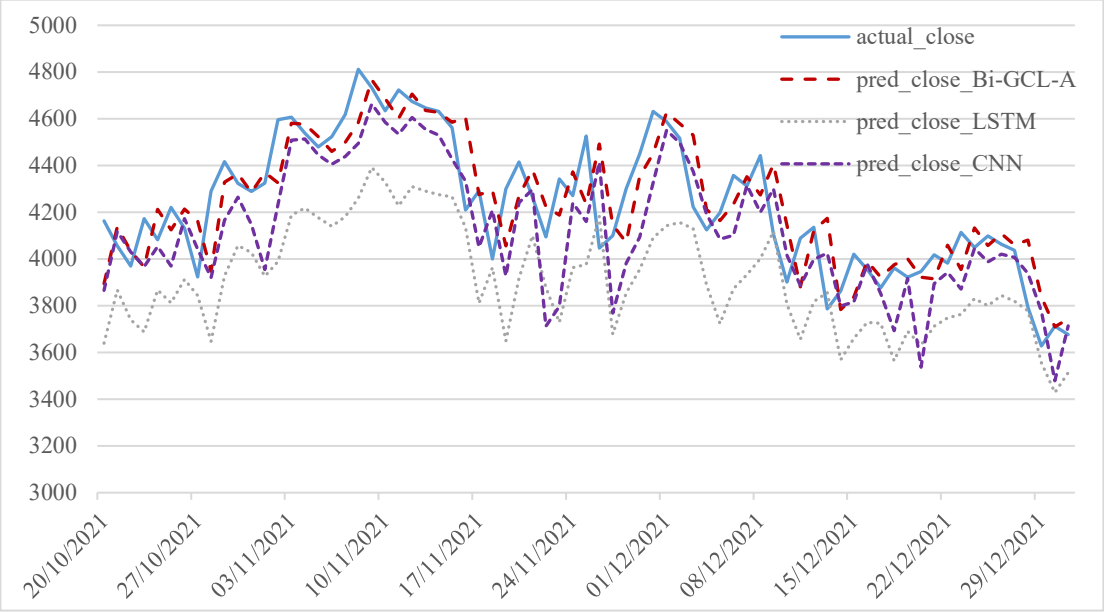


Figure 5.7 Comparison Of Ethereum Price Prediction Using English News Sentiment

Table 5.5
Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum English News)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	-0.12%	0.041	0.031	0.911	0.910
LSTM	8.20%	0.073	0.064	0.889	0.888
CNN	2.72%	0.049	0.040	0.906	0.904

According to the graphs in Figure 5.7, the predicted prices using Bi-GCL-A model demonstrates the best accuracy in following the actual price trends as it captures the peaks and slumps more precisely with minimal deviation. The LSTM model provides smoother predictions but tends to underpredict during severe price movements, thus, showing a delay in response to immediate changes. Whilst the CNN predicted prices follow the general trend, it shows more volatility and digression from the actual prices, predominantly during momentous changes.

In the aspect of volatility, Bi-GCL-A model offers balance between following the trends specifically and maintaining the level of smoothness in the predictions. LSTM model in contrast appears to delay in capturing sharp price movements and less volatile compared to the CNN model that exhibits higher volatility in its predictions. This implies that the Bi-GCL-A model seems to perform best in terms of accuracy and trend-following in predicting Ethereum prices. Nonetheless, the LSTM model provides smoother predictions but may not always capture sudden changes effectively. The CNN model also shows the predicted prices of Ethereum following the trends but shows more volatility and larger deviations at certain points.

From Table 5.5, Ethereum supports the trendlines shown in Figure 5.7. The Bi-GCL-A model demonstrates the lowest average error difference, indicating higher accuracy, along with superior performance metrics, with the lowest RMSE, lowest MAE, and the highest r^2 and adjusted r^2. In contrast, the LSTM model shows the highest average error difference, along with the highest RMSE and MAE, suggesting lower predictive accuracy. The CNN model performs moderately. Overall, Bi-GCL-A outperforms the other models in both error minimization and prediction consistency. Therefore, it can be deduced that the Bi-GCL-A model is reliable in predicting both Bitcoin and Ethereum prices using English news sentiment as the added feature to the historical prices.

Figure 5.8 and Figure 5.9 displays the price trends between the predicted and actual Bitcoin and Ethereum prices accordingly. Table 5.6 and Table 5.7 display the average error difference between the actual and predicted Bitcoin and Ethereum prices respectively.

All three models capture the overall downward trend from late October to late December. However, the degree to which each model captures the smaller fluctuations varies. The predicted prices for Bi-GCL-A model shows to match the trend of the actual prices very closely with very minor deviations throughout the period. It shows the best accuracy in the prediction in comparison to LSTM and CNN models, where the general trend is captured effectively with lesser fluctuations. LSTM predicted prices display a better fit to the actual Bitcoin price trend over the given period with fewer significant deviations than the CNN model, especially during high volatility season. The CNN model shows that the predicted prices generally follow the actual price trend of Bitcoin, but at some range of time, particularly from early to mid-November and late December, there are several conspicuous divergences. This leads to a larger error during periods of high volatility, especially around late October, and early November, where there are some laggings behind the actual price movements.

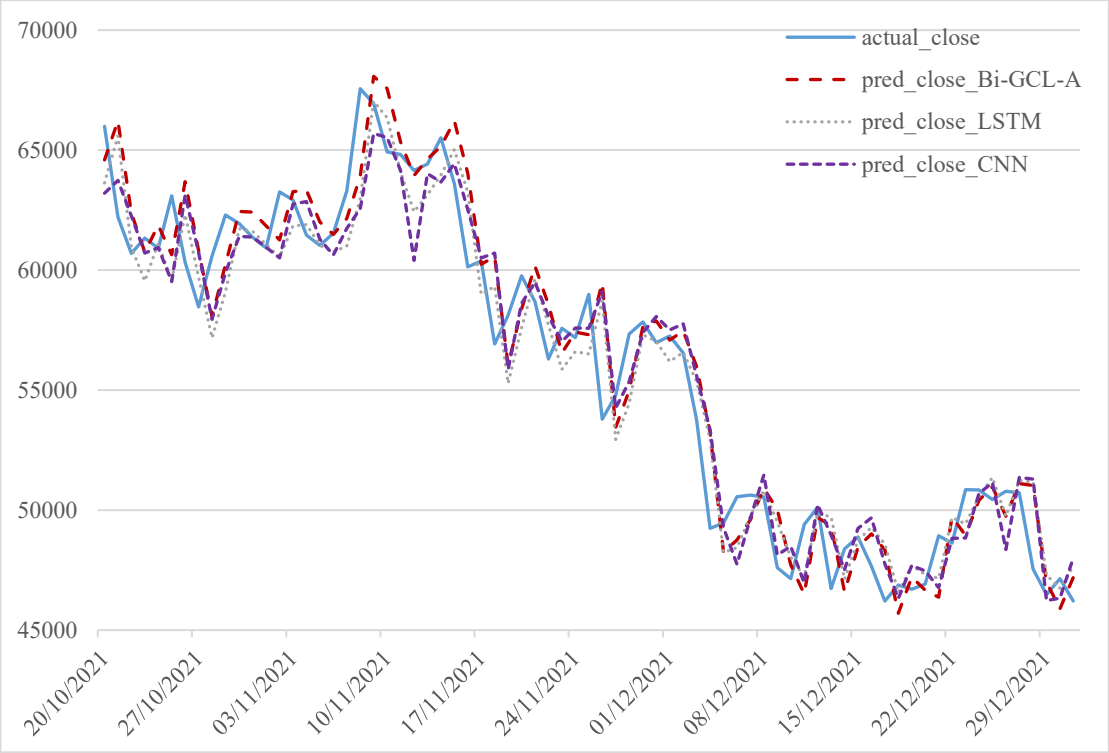


Figure 5.8 Comparison Of Bitcoin Price Prediction Using Malay News Sentiment

Table 5.6
Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin Malay News)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	-0.45%	0.049	0.038	0.917	0.916
LSTM	0.30%	0.040	0.050	0.914	0.913
CNN	0.06%	0.051	0.039	0.912	0.910

Hence, the LSTM and Bi-GCL-A models show a better alignment with the actual price's peaks and troughs, suggesting they are more effective at capturing short-term volatility compared to CNN. This makes the Bi-GCL-A a reliable model for predicting Bitcoin prices using Malay news sentiment as the added feature, and since LSTM also offers a good balance of accuracy and robustness, it can be the next suitable alternative.

Based on the models' performance, the Bi-GCL-A model achieves the lowest average error difference and stands out with strong performance metrics, including the lowest MAE, and the highest r^2 and adjusted r^2, indicating the highest accuracy and reliability. The LSTM model shows a moderate average error difference and RMSE. However, a relatively higher MAE, with slightly lower r^2 and adjusted r^2 are recorded. The CNN model has the smallest average error difference and comparable RMSE and MAE, but the r^2 and adjusted r^2 are lower than Bi-GCL-A and LSTM. Overall, Bi-GCL-A demonstrates the best balance between error reduction and predictive accuracy. Further exploration was done for Ethereum using Malay news sentiment feature and the trends using the three models are shown in Figure 5.9 with the error difference in Table 5.7.

Based on Figure 5.9, Bi-GCL-A model displays close trends between the predicted price and the actual price of Ethereum, without large lags in capturing the market movements. In addition, the predicted prices are balanced, showing neither excessive volatility nor overly smooth trends, showing the reliability of the model. The predicted prices produced by the LSTM model appears to follow the actual prices more closely than the CNN model, with less large deviations during peaks and troughs. LSTM model has also shown to capture the trends more effectively, consistent, and less volatile compared to the CNN model. CNN model on the contrary displays a moderate correlation between the predicted and actual prices but has some points where the prediction deviates significantly. Furthermore, the volatility of the predicted prices can

also be observed, where there are missing sharp rises and drops at certain times. This analysis suggests that the Bi-GCL-A model is the most effective for predicting closing prices in this context, followed by the LSTM and then the CNN model.

From Table 5.7, among the models, Bi-GCL-A shows a balanced performance, achieving a small average error difference, and the lowest MAE, with strong r^2 and adjusted r^2, reflecting its accuracy. The LSTM model, while having the lowest RMSE, demonstrates a higher negative average error difference and a higher MAE, indicating less precision in capturing trends despite a high r^2. The CNN model exhibits the highest error difference and relatively higher RMSE and MAE with slightly lower r^2 and adjusted r^2. Overall, Bi-GCL-A stands out as the most effective model in minimizing errors and maintaining consistency for both Bitcoin and Ethereum Malay news.

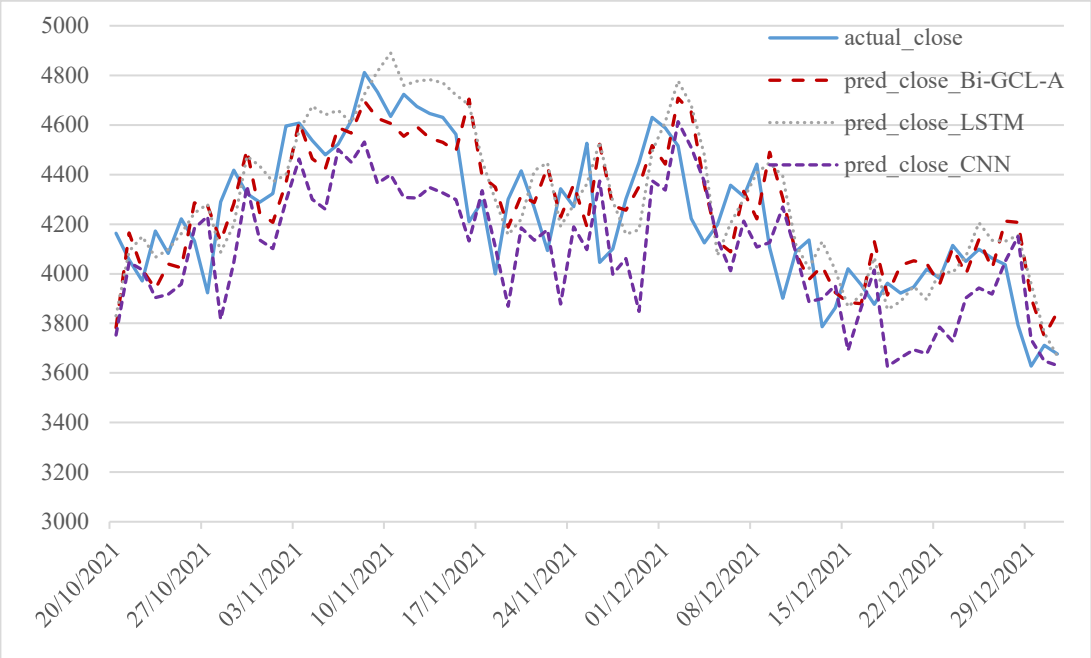


Figure 5.9 Comparison Of Ethereum Price Prediction Using Malay News Sentiment

Table 5.7
Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum Malay News)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	-0.54%	0.042	0.032	0.910	0.909
LSTM	-1.40%	0.037	0.046	0.909	0.908
CNN	3.54%	0.043	0.050	0.905	0.904

Experiments on using English tweets sentiment in addition to the price features were conducted using the same three models: 1) Bi-GCL-A model, 2) LSTM model, and 3) CNN model using similar experimental set-up with the English and Malay news price prediction experiments. The proximity of trends between the predicted prices and the actual prices are shown in Figure 5.10 and Figure 5.11 for Bitcoin and Ethereum, and the average error difference with each model's performances are displayed in Table 5.8 and Table 5.9 accordingly.

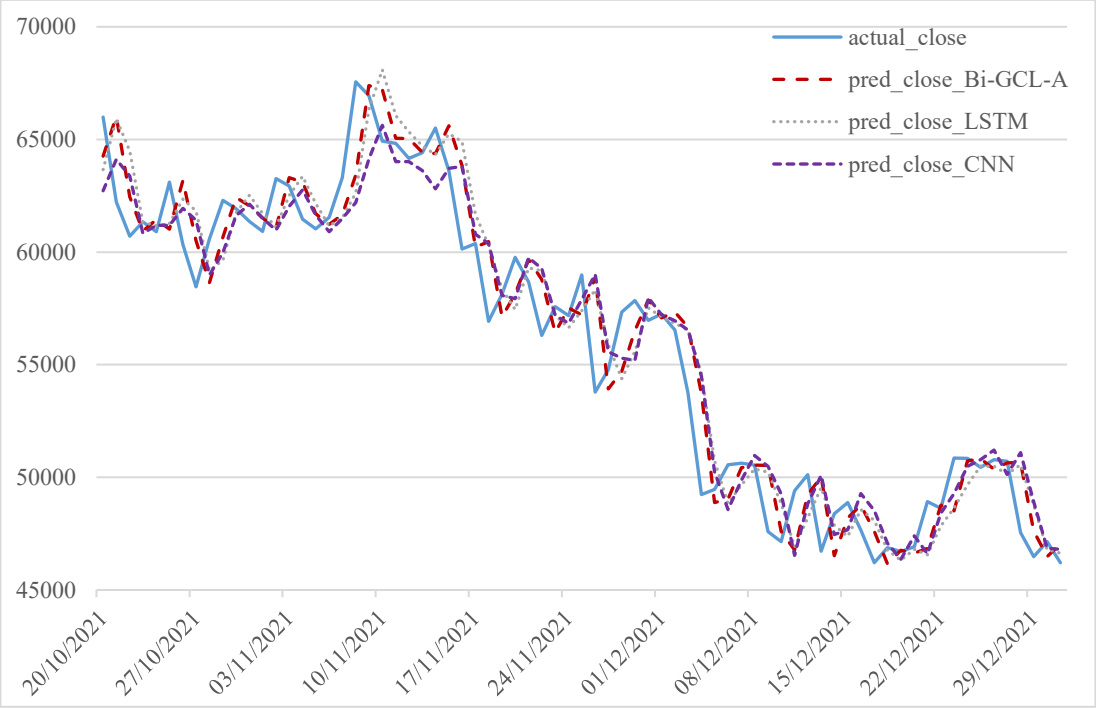


Figure 5.10 Comparison Of Bitcoin Price Prediction Using English Tweets Sentiment

Table 5.8
Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin English Tweets)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	-0.39%	0.038	0.037	0.919	0.918
LSTM	-0.47%	0.055	0.043	0.897	0.895
CNN	-0.27%	0.049	0.050	0.915	0.914

Based on the observations in Figure 5.10, the Bi-GCL-A model demonstrates the closest fit of the predicted prices to the actual closing prices among the three models by capturing both the short-term fluctuations and the overall trend

effectively. With the minimal deviations, the model may have a superior capacity for handling the intricacies and volatility of Bitcoin prices, even during sharp fluctuations. As for the LSTM model, the model's prediction trails the actual price trend closely and have smaller deviations in the mid-November period compared to the CNN model. The model also able to handle the volatility and trends of Bitcoin prices well. The CNN model also shows a good fit in the Bitcoin price prediction overall trend; however, it tends to slightly overestimate during the initial periods but then adjusts well to the actual values at later periods. Notable deviations are observed during mid-November and early December, where the actual prices drop more sharply than predicted. Although the CNN model is shown to be effective, it has the largest deviations among the three, particularly during high volatile periods.

From the overall trend, all three models effectively capture the general downward trend of Bitcoin prices from late October to late December 2021, with the Bi-GCL-A model shows the closest alignment, followed by the LSTM model and the CNN model. This could be due to the Bi-GCL-A model's ability to leverage both convolutional and recurrent neural network architectures, allowing it to better capture both local and long-term dependencies in the data. The LSTM model also performs well, indicating its strength in handling sequential data. The CNN model, while useful, might benefit from additional layers or hybrid approaches to improve its accuracy during volatile periods.

From the observations in Table 5.8, the Bi-GCL-A model performs exceptionally well with the smallest RMSE and MAE, alongside the highest r^2 and adjusted r^2, indicating superior predictive capability. The LSTM model has the highest RMSE which reflects less accuracy compared to the other models. The CNN model achieves the smallest error difference but slightly underperforms Bi-GCL-A in RMSE and MAE, with r^2 and adjusted r^2 values. Overall, Bi-GCL-A outpaces the other models, demonstrating the best balance of error reduction and prediction reliability. Correspondingly, similar experiments are performed for Ethereum, and the trend analysis can be depicted from Figure 5.11.

Based on Figure 5.11, the Bi-GCL-A model provides the closest predictions to the actual price among the three models with minor deviations between the actual and predicted prices, denoting a better fit and more accurate predictions. The LSTM model shows fewer large deviations with close smooth alignment between the predicted and

actual price trends. As for the CNN model, there are visible differences between the predicted and actual price trends, particularly during sharp price changes. However, the model still able to capture the general trend of the actual price.

The Bi-GCL-A model outperforms both the CNN and LSTM models in predicting the closing price of Ethereum. It shows the closest alignment with the actual price, minimal deviations, and better handling of price volatility, making it most robust and accurate prediction model among the three models. The LSTM model also performs well, providing smoother and more accurate predictions than the CNN model, but not as precisely as the Bi-GCL-A model. On the other hand, the CNN model may struggle with volatile price movements as it displays bigger deviations and lower accuracy based on the trends.

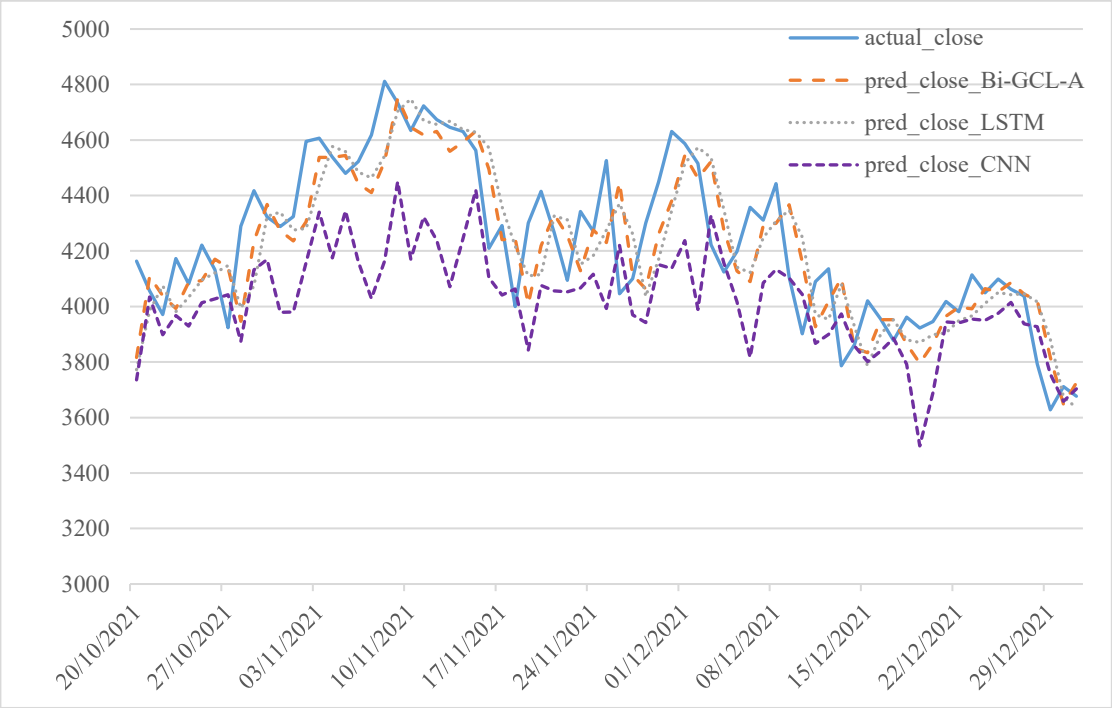


Figure 5.11 Comparison Of Ethereum Price Prediction Using English Tweets Sentiment

Table 5.9
Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum English Tweets)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	0.77%	0.041	0.031	0.912	0.910
LSTM	0.73%	0.045	0.035	0.910	0.909
CNN	5.17%	0.049	0.058	0.899	0.898

The Bi-GCL-A model achieves the best overall performance, with the smallest RMSE and the lowest MAE, along with the highest r^2 and adjusted r^2. The LSTM model performs slightly worse, with an error difference of 0.73%, higher RMSE, and MAE, but maintains strong r^2 and adjusted r^2 values. The CNN model has the highest error difference, along with the highest RMSE and MAE, and lower r^2 and adjusted r^2 values. Hence, Bi-GCL-A demonstrates the most accurate and reliable predictions.

According to Table 5.8 and Table 5.9, it confirms that the Bi-GCL-A model performs the best in comparison to the LSTM and CNN models. However, for Bitcoin, the CNN model appears to have higher adjusted r^2 than the LSTM model which contradicts the visualization of the price trend in Figure 5.9, but LSTM still achieved lower MAE result than the CNN model, indicating its effectiveness is more than the CNN model. Nonetheless, the results shown for Ethereum in Table 5.9 supports the visualized results in Figure 5.11 for all the models' performances.

The price prediction experiment using Malay tweets sentiment used the same experimental set-up as the English tweets, English news and Malay news experiments. The actual price is represented by solid blue line throughout the charts in Figure 5.12 and Figure 5.13, while the predicted price for Bi-GCL-A model is represented with orange dashed line. The predicted prices for LSTM and CNN models are represented with purple dashed line and green dashed line respectively. Hence, the trend analysis for each model on Bitcoin and Ethereum using Malay tweets sentiment feature in addition to the price features is shown in Figure 5.12 and Figure 5.13 correspondingly. The performance results for Bitcoin and Ethereum price prediction are also displayed in Table 5.10 and Table 5.11 respectively.

Based on Figure 5.12, all three models can predict Bitcoin prices to some extent. The Bi-GCL-A model offers the highest accuracy and reliability, followed by the LSTM model, with the CNN model showing the least accurate predictions. The Bi-GCL-A model is shown to be the most reliable model among the three models since it combines the strengths of graph-based learning and attention mechanisms, allowing it to better understand and predict the complex relationships in the data. Meanwhile, the LSTM model demonstrates a closer fit of the predicted prices to the actual prices with less severe deviations on the trend as it captures the general movement of the market accurately. This is due to its efficiency at handling the sequential data and predicting

the price trends. CNN model seems to lack in handling high volatility situations causing a larger deviation in the trendline between the predicted and actual Bitcoin prices. Therefore, the Bi-GCL-A model has shown to be the best performed model in predicting the Bitcoin prices in comparison to the LSTM and CNN models.

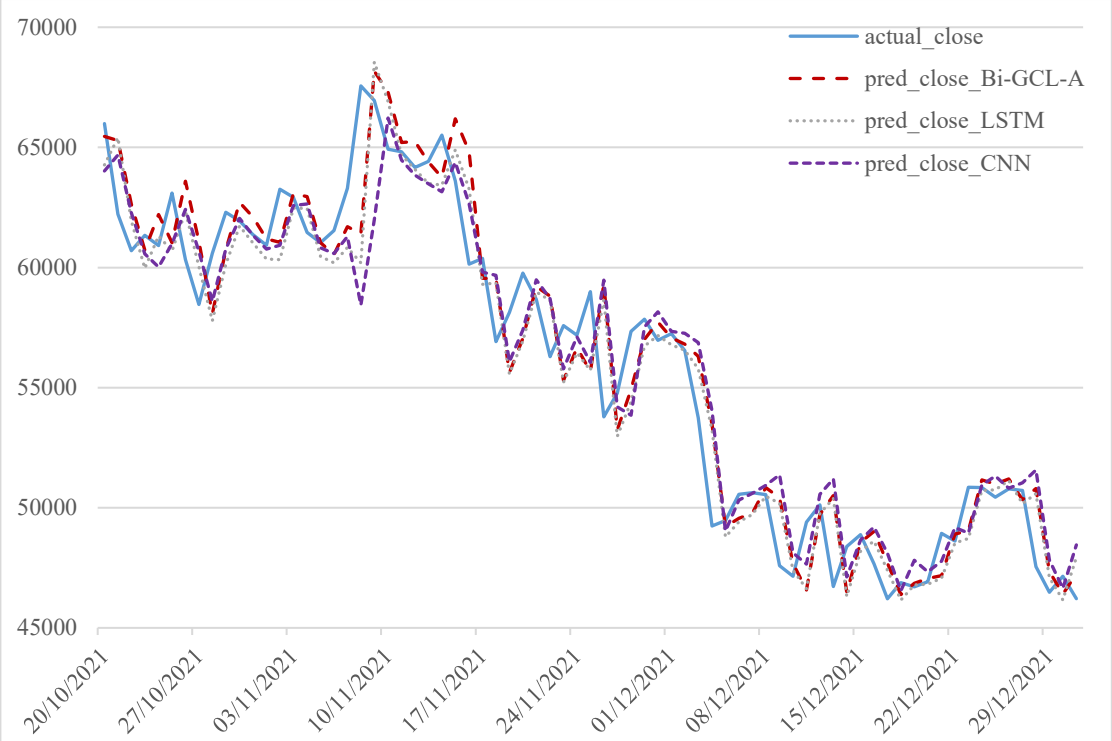


Figure 5.12 Comparison Of Bitcoin Price Prediction Using Malay Tweets Sentiment

Table 5.10
Error Differences Between Predicted And Actual Prices And Performance Result (Bitcoin Malay Tweets)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	-0.29%	0.040	0.038	0.918	0.917
LSTM	0.42%	0.049	0.050	0.914	0.913
CNN	-0.17%	0.051	0.039	0.912	0.911

Bi-GCL-A model performs better than the other models, achieving the lowest RMSE, MAE, and the greatest r^2 and adjusted r^2, while CNN has a marginally higher RMSE and a lower r^2 than Bi-GCL-A and LSTM. However, it has a little better average error difference. For other metrics, LSTM has intermediate values and the biggest average error difference. Therefore, the Bi-GCL-A model performs better than the others. Another similar experiment was conducted for Ethereum and the relationship between the predicted and actual price trends are illustrated in Figure 5.13 using the same representation of trendlines as Bitcoin experiment in Figure 5.12.

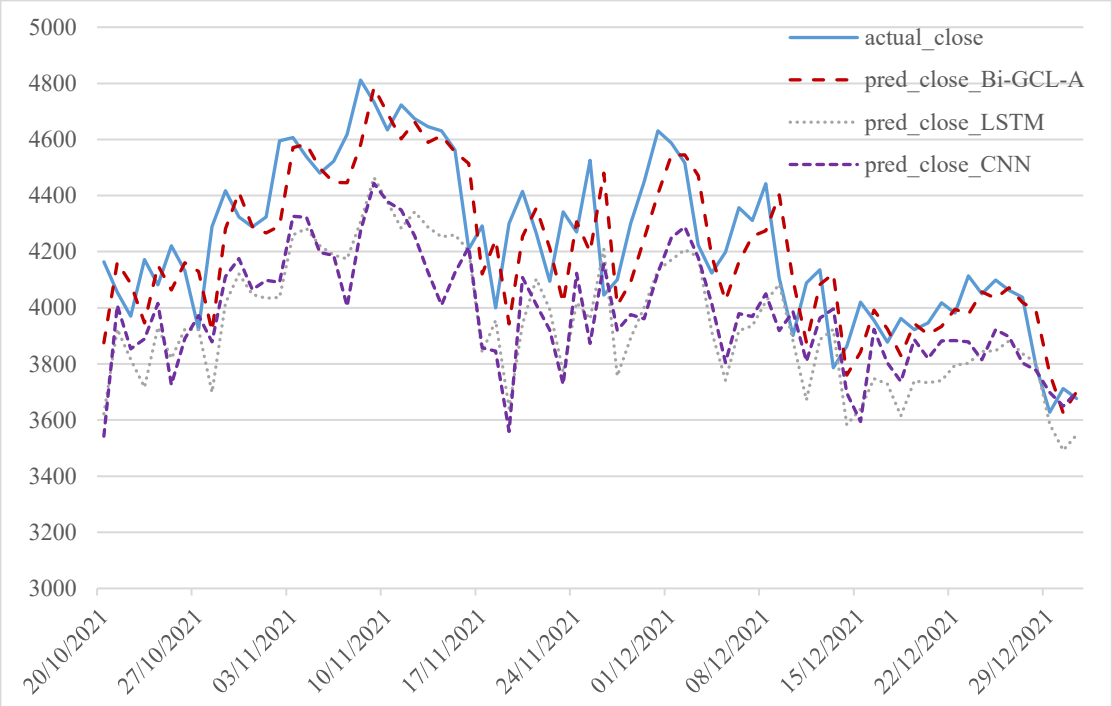


Figure 5.13 Comparison Of Ethereum Price Prediction Using Malay Tweets Sentiment

Table 5.11
Error Differences Between Predicted And Actual Prices And Performance Result (Ethereum Malay Tweets)

Model	Avg. Error Difference	RMSE	MAE	r^2	Adjusted r^2
Bi-GCL-A	0.85%	0.040	0.031	0.909	0.907
LSTM	7.19%	0.069	0.079	0.896	0.894
CNN	6.43%	0.051	0.060	0.906	0.904

Figure 5.13 depicted that the Bi-GCL-A model shows to track the actual prices more closely than the other models. It captures the ups and downs with better accuracy and less error, showing a tighter correlation between the predicted prices and the actual prices. On the contrary, the LSTM and CNN models show a significant deviation of the predicted prices against the actual prices. The LSTM model provides less volatile prediction than the CNN model, while the CNN model does capture some trend, regardless of the significant deviations in the trendline. Therefore, Bi-GCL-A performs the best in capturing the actual price trends, followed by CNN, and then LSTM. The Bi-GCL-A model's predictions are closest to the actual prices, indicating a better understanding of the underlying trends. LSTM model shows the most error difference

in its predictions, often missing sharp spikes. CNN model also lags but to a lesser extent, whereas Bi-GCL-A model exhibits the least error, closely following the actual price movements.

To analyse the performance results extensively, Table 5.11 demonstrates that Bi-GCL-A model achieved the best overall performance with the lowest RMSE, MAE, and the highest r^2 and adjusted r^2. LSTM on the other hand has the highest average error difference and the worst results across all other metrics, including RMSE and MAE. CNN performs better than LSTM but is outperformed by Bi-GCL-A. Overall, Bi-GCL-A is the most accurate model for Ethereum price prediction in this experiment. Moreover, the Bi-GCL-A model shows the lowest RMSE and MAE, indicating the highest accuracy in predicting Bitcoin and Ethereum prices. The high r^2 and adjusted r^2 values further confirm its superior performance. This aligns with the visualized trends in Figure 5.12 and Figure 5.13.

However, for Ethereum evaluation, the LSTM and CNN models results in Table 5.11 contradicts the charts displayed in Figure 5.13. Although LSTM model's predicted price trend is closer to the actual price compared to the CNN model, the adjusted r^2 of CNN model demonstrates 0.01 higher than the LSTM model. This indicates that CNN model predicts Ethereum prices with a slightly higher accuracy than the LSTM model. Hence, the Bi-GCL-A model consistently outperforms the LSTM and CNN models in predicting both Bitcoin and Ethereum prices based on sentiment analysis from Malay tweets. It has the lowest RMSE and MAE values and the highest r^2 and adjusted r^2 values, indicating superior accuracy and reliability. The LSTM model performs the worst, especially for Ethereum predictions, while the CNN model shows intermediate performance.

Besides the visualization of the trendlines and the four evaluation metrices shown for English and Malay news, and for English and Malay tweets, the average error differences for the overall experimentation are also calculated as shown in Table 5.12. Based on the average error difference for one-year duration shown in Table 5.12, it can be observed that Bi-GCL-A model shows the closest predicted price to the actual price for English News (Bitcoin and Ethereum), Ethereum Malay News, and Ethereum Malay Tweets. However, LSTM model has shown to be slightly better in predicting the price than the Bi-GCL-A model for Ethereum English Tweets with a difference of 0.07%

between the two models, while a large percentage error difference was produced by CNN.

Table 5.12
Average Error Differences

SENTIMENT CORPORA	AVERAGE ERROR DIFFERENCE		
	Bi-GCL-A	LSTM	CNN
BTC English news	**-0.20%**	-0.96%	-0.63%
ETH English news	**-0.12%**	8.20%	2.72%
BTC Malay news	-0.45%	0.30%	**0.06%**
ETH Malay news	**-0.54%**	-1.40%	3.54%
BTC English tweets	-0.39%	-0.47%	**-0.27%**
ETH English tweets	0.77%	**0.73%**	5.17%
BTC Malay tweets	-0.29%	0.42%	**-0.17%**
ETH Malay tweets	**0.85%**	7.19%	6.43%

In addition, CNN model shows the least error percentage between the predicted price and the actual price for Bitcoin Malay Tweets compared to the Bi-GCL-A model and LSTM models. Another observation can be deduced when using CNN to predict the price using Bitcoin Malay News and Bitcoin English Tweets. The Bi-GCL-A model has a big difference between the predicted and actual price of 0.45%, while CNN shows the least error difference of only 0.06% for Bitcoin Malay News. But, for Bitcoin English Tweets and Bitcoin Malay Tweets, the difference of error between own model and CNN is only 0.12%.

From the overall results, the Bi-GCL-A model still manage to predict well using the additional sentiment features from the sources. Besides the analysis on the price prediction trends, the relationship and causality between the price and sentiment is also explored in the next section.

5.4 Summary

The chapter compares the performance of models trained on the curated corpora by conducting two types of experiments, where one enriched with sentiment features and one without. The study addresses the research question by investigating how incorporating sentiment features derived from textual data influences the predictive accuracy of deep learning models. From the analysis, it can be inferred that Bitcoin and Ethereum markets are influenced by a multitude of factors including but not limited to sentiment. Market psychology, regulatory news, macroeconomic indicators, and technological advancements all contribute to price movements. The non-linear

relationship suggests that changes in sentiment interact with these factors in complex ways, leading to the observed scatter and high volatility. This implies that while sentiment analysis can provide insights into market movements, it should be combined with other indicators to make more accurate predictions. The comparison of sentiment scores auto labelling and manual labelling highlights the significance of detailed analysis in understanding market sentiment. Moreover, the comparison between the model's performance using price features only and using the combination of price and sentiment features shows that using the added sentiment features to the price features is better. The price-sentiment features combination shows an average improvement in the adjusted r^2 of 0.130 and 0.127 for Bitcoin and Ethereum respectively. The sentiment from news and tweets has a stronger influence on Bitcoin prices compared to Ethereum. This is because Bitcoin receives more media coverage and social media discussion, making investor sentiment more visible and impactful. Additionally, Bitcoin's high retail participation and speculative trading mean that shifts in public sentiment translate more quickly into price movements than for Ethereum.

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Introduction

This chapter presents the conclusion of the research study and suggests potential directions for future work. The primary contribution of the research was to improve cryptocurrency price prediction model by integrating sentiment features derived from English and Malay corpora. In addition. an implementation of a deep learning model, XLNet-GRU+ was done for sentiment classification and regression to validate the sentiment data.

In the following sections, the chapter provides the highlight the major contributions of the research, a summary of the thesis, and discuss the limitations encountered during the study. Furthermore, potential opportunities for future research are outlined. The future work section explores on additional areas for further development. The chapter concludes with an overall reflection on the impact of the research and its relevance to the cryptocurrency domain.

6.2 Contributions

The study yielded multiple novelties to enhance the cryptocurrency price prediction performance by combining sentiment analysis and utilising deep learning models. Several contributions were made, which include the development of new methodologies, datasets, and predictive models that leverage sentiment data for more accurate market forecasting. The contributions of this research are described in the following subsections.

6.2.1 Corpora Creation for English and Malay

This study has curated a total of eight datasets: 1) Bitcoin English news, 2) Ethereum English news, 3) Bitcoin Malay news, 4) Ethereum Malay news, 5) Bitcoin

English tweets, 6) Ethereum English tweets, 7) Bitcoin Malay tweets, and 8) Ethereum Malay tweets. The raw data were extracted from public online news and tweets on Bitcoin and Ethereum in English and Malay language.

The Malay language remains significantly underrepresented in computational linguistics and natural language processing (NLP) research, particularly in sentiment analysis and cryptocurrency-related studies. While English dominates the field, Malay language resources, annotated corpora, and pre-trained language models are scarce, leading to challenges in text processing and sentiment analysis. The lack of comprehensive datasets limits the development of accurate and robust sentiment classification models for Malay, which is critical in domains such as finance and cryptocurrency trading. This study has contributed to address this gap by creating sentiment datasets in Malay. The findings highlight the necessity of developing more Malay language resources to ensure inclusivity in the evolving field of computational finance.

The creation of high-quality corpora for both English and Malay represent a significant contribution to this study. Developing well-curated and linguistically balanced datasets in these two languages enhances the model's ability to capture nuanced semantic, syntactic, and domain-specific patterns inherent in financial discourse. Such corpora support the extraction of more informative features, improving model robustness, cross-linguistic consistency, and generalizability. Furthermore, the establishment of bilingual corpora provides a valuable linguistic resource for future studies in multilingual financial text analysis, enabling more rigorous benchmarking, cross-lingual transfer learning, and the development of language-agnostic predictive models.

6.2.2 Data Annotation and Codebook

A major component of this research involved constructing a high-quality annotated dataset for sentiment analysis in cryptocurrency-related texts. The annotation process was meticulously carried out using a structured rubric and codebook, ensuring consistency and accuracy across human annotators. Currently, there is still no available comprehensive literature on creating a codebook for annotators, particularly on cryptocurrency domain. The codebook played a crucial role in reducing subjectivity by

standardizing sentiment classification guidelines. Without a well-defined codebook, annotation can become subjective and inconsistent, leading to unreliable labels, reduced data quality, and weaker model performance. This systematic approach to annotation ensures that future studies can replicate the methodology and extend the dataset with similar levels of accuracy, further enriching the field of Malay sentiment analysis. Inter-annotator agreement was measured using statistical methods, Krippendorf's alpha to validate the reliability of the dataset.

The data were cleaned and underwent annotation process which involves designing the rubrics to create a codebook for annotators to label each instance with a sentiment score ranging from -1 to 1. TF-IDF and word frequency ranking were applied in the codebook development. The method is used to identify the most significant terms in the text that could influence sentiment categorization. The codebook further guidelines to address the issue of mixed sentiments, where texts might express both positive and negative feelings. In such cases, annotators were instructed to classify the dominant sentiment, ensuring a more refined and accurate approach to sentiment analysis. Another important aspect of the codebook was its treatment of emoticons, which is commonly used in social media posts. These elements were accounted for as significant indicators of sentiment.

The manual annotation process was carefully outlined in the codebook, with a clear step-by-step procedure for annotators to follow. To ensure that the annotations met the codebook's standards, each annotated text was reviewed, and any inconsistencies were resolved through discussions among the research team. The approach ensured that the labelled dataset maintained high quality and accuracy throughout the research. With the guidance from the codebook, three annotators were employed to perform the annotations for a total of four training rounds with a Krippendorf's alpha agreement percentage of more than 60% to determine the final sentiment score for each instance. The annotation was validated by a remisier from Kenanga Bank, who is an expert in trading including his involvement in cryptocurrency.

Achieving a high level of inter-annotator agreement presented a considerable challenge in this study, particularly in reaching Krippendorff's alpha above 60%, which is generally considered the minimum threshold for acceptable reliability. The inherent ambiguity and volatility of cryptocurrency-related textual data often led to different interpretations among annotators, making consistent annotations difficult. Despite the

training efforts, certain data segments remained problematic and consistently generated disagreement. As a result, insignificant ambiguous data were systematically excluded after more than four rounds of training and review to preserve the overall quality and reliability of the annotated dataset. The rigorous process highlights both the complexity of annotating financial discourse and the methodological steps necessary to ensure data integrity.

To validate the effectiveness of the annotation process, several regressors that were priorly implemented in other related works were tested using the created datasets, such as: Linear Regression, Support Vector Machine, Decision Tree, and Random Forest. Based on the achieved results, it can be concluded that the curated corpora are reliable to be used by other related research studies.

In conclusion, the development of the codebook was a critical contribution to this research. It provided a structured, standardized approach to sentiment annotation, ensuring that the sentiment data used in the study was accurate and reliable. By incorporating TF-IDF and word frequency ranking to determine the sentiment categories, the codebook ensured that the annotations were grounded in the most meaningful and relevant linguistic features. This codebook not only facilitated the sentiment analysis in this study but also serves as a replicable framework for future sentiment-driven research in cryptocurrency and other domains.

The study also explored the effectiveness of automated data annotation and compared it with manual annotation. While automated annotation significantly reduces time and effort, it is prone to inaccuracies, particularly in underrepresented languages like Malay, where pre-trained models lack comprehensive training on domain-specific vocabularies. Comparisons revealed that manual annotation, supported by a well-defined codebook, yielded higher quality sentiment labels. However, leveraging pre-trained models and fine-tuning them with human-annotated corpora can improve automated annotation, making it a viable option for large-scale applications in sentiment analysis.

6.2.3 Validation of Sentiment Features using XLNet-GRU+

To enhance sentiment analysis and cryptocurrency price prediction, this study implemented a deep learning model integrating multi-head attention mechanisms. The

model, based on XLNet-GRU, demonstrated superior performance in capturing complex sentiment nuances and dependencies within textual data. Multi-head attention allowed the model to focus on key aspects of sentiment in cryptocurrency-related news and tweets, improving sentiment regression accuracy. The experimental results confirmed that incorporating attention mechanisms significantly enhances model performance by dynamically weighting important words and phrases, leading to a more nuanced understanding of sentiment dynamics in financial contexts.

The development of XLNet-GRU+ sentiment model was implemented with the incorporation of the XLNet language model and the GRU deep learning model, with additional self-attention mechanism. There were very limited existing related studies explored on the self-attention mechanism in sentiment analysis. Hence, this study contributes by exploring the performance gain achieved by the XLNet-GRU+ sentiment model using various number of heads for the self-attention mechanism. Then, the visualization of the context learnt by the self-attention mechanism was also highlighted. Moreover, the dataset from other research studies were also used to be tested with the developed sentiment model to investigate its effectiveness. The study's curated corpora were also experimented with other sentiment models as baselines such as BERT and CNN-LSTM models, thus produced comparable results. This shows the efficiency of the XLNet-GRU+ sentiment model in predicting the sentiment in English and Malay language, particularly cryptocurrency-related domain.

6.2.4 Price Prediction model using Sentiment Feature

The final phase of this research involved integrating sentiment features into a cryptocurrency price prediction model. Traditional financial models primarily rely on historical price data, but this research demonstrated that sentiment indicators derived from news articles and social media (Twitter) can enhance predictive accuracy. The Bi-GCL-A model, incorporating sentiment regression outputs, improved cryptocurrency price forecasting compared to models that relied solely on cryptocurrency price data. The findings suggest that market sentiment, particularly from Twitter and news sources, influences cryptocurrency price movements, making sentiment analysis a valuable addition to forecasting methodologies. This work establishes a foundation for future

research in sentiment-driven financial modelling, emphasizing the growing importance of NLP techniques in economic and investment decision-making.

The Bi-GCL-A price prediction model comprises of hybrid architecture where GRU, CNN, and LSTM were implemented with integrated Bahdanau's self-attention mechanism to produce the study's price prediction model. The price of Bitcoin and Ethereum were predicted at intraday level. The experiments on using price features only and using the added sentiment features were conducted. Then the performance of the model was compared with the two most popularly used in previous research studies, which are LSTM and CNN models. Based on the overall results attained, the Bi-GCL-A outperformed LSTM and CNN models in predicting the next-day Bitcoin and Ethereum prices. The performance results of RMSE, MAE, R^2, and adjusted R^2 were also compared with others' works, and found that the Bi-GCL-A has the potential to be used in predicting cryptocurrency prices well.

Moreover, from the experiments on using price features only and using the combination of price and sentiment features to perform cryptocurrency price prediction, it is found that the additional sentiment features enhance the price prediction model's performance. The performance of using price features only achieved an adjusted r^2 of 0.787 (Bitcoin) and 0.769 (Ethereum), while the average adjusted r^2 achieved across the four sentiment datasets (English news, English tweets, Malay news, and Malay tweets) each for Bitcoin and Ethereum are 0.917 and 0.909 respectively. The moderate improvement of about 0.13 can be seen. The improvement values are considered significant for an adjusted r^2 based on the benchmark of Cohen's F-square calculation (Cohen, 1988; Selya et al., 2012). The increase of 0.13 in the adjusted r^2 result shows a medium but meaningful improvement. Considering the high volatility and multifactor nature of cryptocurrency price movements, even modest gains in predictive accuracy are challenging to achieve. A study by Gurgul et al. (2025) achieved improvements within 0.04 to 0.05 in the r^2 between using price features only and using sentiment features from Twitter. Hence, the sentiment integration contributes significant enhancement to the model's performance.

6.3 Summary of Thesis

This thesis explores the solutions for the problem involving cryptocurrency price prediction using deep learning models. A structured approach to incorporate sentiment analysis into cryptocurrency price prediction is implemented to solve the problem. The process begins with corpora creation, where both sentiment and price data are gathered. This step is crucial for building a dataset that includes both textual sentiment from online news and tweets, and also Bitcoin and Ethereum cryptocurrency price data. The sentiment data address the challenge of limited Malay resources in cryptocurrency-related studies. By ensuring the inclusion of Malay text alongside English, the study acknowledges the importance of mother tongue language sentiment analysis, given that users often express themselves more naturally in their native language.

Following corpora creation, the annotation process is performed. This step ensures that the sentiment data is labelled accurately, providing high-quality features for downstream analysis. There are two types of annotation, which are: manual and automated annotation. Manual annotation was carried out by human annotators who read and labelled each news post and tweet as either positive, negative, or neutral based on predefined sentiment categories. The manual annotation method ensured a high level of accuracy but was time-consuming and resource intensive. In contrast, automated annotation was performed using sentiment analysis model of XLNet-GRU+, which employed machine learning techniques to automatically categorize the sentiment. The comparison between manual and automated annotation allowed for the validation of sentiment features and the effectiveness in predicting cryptocurrency prices. To perform sentiment classification, the texts were labelled as positive, negative, or neutral based on its content. The classification task allowed the research to analyse the general sentiment in the market and its potential effect on price movements. For sentiment regression, the sentiment features were assigned continuous sentiment scores ranging from -1 (very negative) to +1 (very positive), instead of discrete categories. This regression task provided a more nuanced understanding of how the intensity of sentiment could correlate with cryptocurrency price fluctuations, enabling more precise prediction models.

For the feature extraction process, two types of data were considered: price and sentiment data. Price features include historical price data such as Open, High, Low, Close (OHLC), and trading volume, which are commonly used in financial market analysis. These features provided valuable context for understanding price movements based on market trends. Sentiment features, on the other hand, were extracted from the news and tweets related to cryptocurrency. The sentiment scores were taken for sentiment analysis.

The cryptocurrency price prediction is conducted using two approaches which are with price features and combination of price and sentiment features. By integrating both sentiment and price data features into the prediction models, the research provides deeper insights into market behaviour and improving the model's overall performance. The comparison is critical in evaluating the impact of sentiment features on price prediction accuracy. By testing the model with and without sentiment data, the study aims to determine whether sentiment features contribute to improving predictive performance. This aligns with the research objective of developing a deep learning-based cryptocurrency price prediction system and comparing its performance when sentiment features are included versus excluded.

6.4 Limitations and Future Work

The curated corpora are more specific to Bitcoin and Ethereum cases and only caters the language of English and Malay individually. Thus, this can be improved for future work by creating more corpora for other types of cryptocurrency coins (i.e., Litecoin, Ripple, Dogecoin, etc.) to gain more coverage in the cryptocurrency domain. Moreover, additional number of data can be considered for future enhancements.

The XLNet-GRU+ sentiment model did produce comparable results for each corpus; however, the accuracy can be enhanced in a more optimised approach to gain a better performance value. An integration of some optimization techniques into the model may improve the quality of the sentiment analysis as there is still very limited exploration on optimization technique for cryptocurrency-related research studies.

The features used for cryptocurrency price prediction in this study are reasonable but there are still other factors that can be included in the prediction such as the user affinity of tweets or other technical indicators to expand the research on the

price prediction.

6.5 Conclusion

This research aimed to develop a dynamic data-driven cryptocurrency price prediction model by integrating sentiment features extracted from English and Malay text sources. The results of this study demonstrate the significant impact that sentiment analysis can have on cryptocurrency price forecasting, particularly when sentiment features are combined with historical price data in deep learning models. However, there are several aspects that can be improved and expanded in future research to further enhance the prediction models.

- **Corpora Creation and Linguistic Inclusivity**

The research successfully created comprehensive sentiment datasets for cryptocurrency data in both English and Malay, addressing the gap in the available resources for sentiment analysis in non-English languages. The corpora creation is particularly important for promoting inclusivity in sentiment-driven research in financial technology.

Future work could expand the corpus to include other languages, such as: Chinese, Spanish, or Arabic, which could provide a more global perspective on cryptocurrency sentiment and further increase the applicability of sentiment-driven models in diverse markets. Additionally, increasing the volume and diversity of texts within the corpora, by including more cryptocurrency types and sources, such as: blogs, forums, and Reddit, would help improve model robustness.

- **Data Annotation and Codebook Creation**

A detailed codebook was developed to standardize the sentiment annotation process, utilizing TF-IDF and word frequency ranking to define sentiment categories. This ensured a consistent and high-quality dataset for training the sentiment analysis models.

One area for improvement is the scalability of the annotation process. While manual annotation ensured high-quality labels, it is a time-consuming process that could be optimized using semi-automated or crowdsourced approaches. Additionally, future

work could refine the codebook further by integrating more nuanced sentiment categories, such as sarcasm or mixed sentiments, which could improve the granularity of sentiment analysis and its relationship to price prediction.

- **Sentiment Features Validation with XLNet-GRU+**

The XLNet-GRU+ model, which incorporated a multi-head attention mechanism, was validated and demonstrated a significant improvement over traditional sentiment analysis models. The XLNet-GRU+ model showed strong performance in predicting sentiment features, highlighting the importance of integrating sentiment into cryptocurrency price prediction models.

Although XLNet-GRU+ performed well, future work could explore the use of other transformer-based models like GPT or T5 for sentiment analysis tasks. Additionally, the model could be enhanced by experimenting with different pre-training techniques, such as domain-specific fine-tuning, to better handle industry-specific terms and jargon more effectively.

- **Integration of Sentiment Features for Price Prediction**

The integration of sentiment features into price prediction models resulted in improved accuracy, with the Bi-GCL-A model outperforming traditional models that relied solely on price data. This showed that sentiment, especially from social media and news sources, is a crucial factor influencing cryptocurrency price movements.

One possibility for improvement is the real-time integration of sentiment features. While this research used historical data for training, real-time sentiment analysis could provide more timely insights for traders and investors, particularly in volatile market conditions. Future work could focus on streaming sentiment analysis and integrating it with live cryptocurrency price feeds to create an even more responsive prediction model. Furthermore, the incorporation of technical indicators and macroeconomic factors could further enhance the model by adding more dimensions to the prediction process.

In conclusion, while this research made significant contributions to integrating sentiment analysis into cryptocurrency price prediction, there are several areas for improvement. Expanding the corpora, refining the annotation process, and exploring new models and methodologies can further enhance the accuracy and applicability of

the sentiment-driven cryptocurrency price forecasting models. By addressing these areas, future research can continue to develop more sophisticated tools for traders and investors, ultimately improving decision-making in the cryptocurrency market.

REFERENCES

Abou Tanos, B., & Badr, G. (2024). Price Delay and Market Efficiency of Cryptocurrencies: The Impact of Liquidity and Volatility during the COVID-19 Pandemic. *Journal of Risk and Financial Management*, *17*(5), 1-14.

Acosta, R. P., & Arana, A. Z. (2023). Cardano Cryptocurrency Price from Twitter. A Prediction Algorithm from Machine Learning. *International Journal of Electronics and Communication Engineering*, *10*(12), 33–44.

Agarwal, B., Harjule, P., Chouhan, L., Saraswat, U., Airan, H., & Agarwal, P. (2021). Prediction of Dogecoin Price using Deep Learning and Social Media Trends. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, *8*(29), 1-12.

Aggarwal, A., Gupta, I., Garg, N., & Goel, A. (2019). Deep Learning Approach to Determine the Impact of Socio Economic Factors on Bitcoin Price Prediction. *2019 Twelfth International Conference on Contemporary Computing (IC3)*, 1–5.

Akba, F., Medeni, I. T., Guzel, M. S., & Askerzade, I. (2021). Manipulator Detection in Cryptocurrency Markets Based on Forecasting Anomalies. *IEEE Access*, *9*, 108819–108831.

Akyildirim, E., Goncu, A., & Sensoy, A. (2020). Prediction of Cryptocurrency Returns using Machine Learning. *Annals of Operations Research*, *297*(1), 3–36.

Al Mercado. (2023). *BTC Tweets Sentiment* [Computer software]. https://data.world/mercal/btc-tweets-sentiment

Alabdulkarim, N. A., Haq, M. A., & Gyani, J. (2024). Exploring Sentiment Analysis on Social Media Texts. *Engineering, Technology & Applied Science Research*, *14*(3), 14442–14450.

Alghamdi, S., Alqethami, S., Alsubait, T., & Alhakami, H. (2022). Cryptocurrency Price Prediction using Forecasting and Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, *13*(10), 1-10.

Ali, H., Hashmi, E., Yayilgan Yildirim, S., & Shaikh, S. (2024). Analyzing Amazon Products Sentiment: A Comparative Study of Machine and Deep Learning, and Transformer-Based Techniques. *Electronics*, *13*(7), 1-21.

Ali, M. (2024, September 12). *Introduction to Activation Functions in Neural Networks*. Datacamp. https://www.datacamp.com/tutorial/introduction-to-activation-

functions-in-neural-networks

Ali, P. J. M., Faraj, R. H., Ali, P. J. M., & Faraj, R. H. (2014). Data Normalization and Standardization: A Technical Report. *Machine Learning Technical Reports*, *1*(1), 1–6.

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., Abou Elwafa, A., & Kurdi, H. (2021). Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain. *Applied Sciences*, *11*(2), 1-18.

Anastasiou, D., Ballis, A., & Drakos, K. (2021). Cryptocurrencies' Price Crash Risk and Crisis Sentiment. *Finance Research Letters*, *42*, 1-5.

Ante, L. (2023). How Elon Musk's Twitter Activity Moves Cryptocurrency Markets. *Technological Forecasting and Social Change*, *186*, 1-14.

Antoine, J.-Y., Villaneau, J., & Lefeuvre, A. (2014). Weighted Krippendorff's alpha is a More Reliable Metrics for Multi-coders Ordinal Annotations: Experimental studies on emotion, opinion and coreference annotation. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 550–559.

Arslan, S. (2025). Bitcoin Price Prediction Using Sentiment Analysis and Empirical Mode Decomposition. *Computational Economics*, *65*(4), 2227–2248.

Artstein, R., & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, *34*(4), 555–596.

Astuti, I. D., Rajab, S., & Setiyouji, D. (2022). Cryptocurrency Blockchain Technology in the Digital Revolution Era. *Aptisi Transactions on Technopreneurship (ATT)*, *4*(1), 9–16.

Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of Word Embedding Models on Text Analytics in Deep Learning Environment: A Review. *Artificial Intelligence Review*, *56*(9), 10345–10425.

Attila, S. D. (2017). Impact of Social Media on Cryptocurrency Trading with Deep Learning. *Scientific Students' Conference*, 1-47.

Awaysheh, A., & Klassen, R. D. (2010). The Impact of Supply Chain Structure on The Use of Supplier Socially Responsible Practices. *International Journal of Operations & Production Management*, *30*(12), 1246–1268.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordnet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of*

the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2200-2204.

Badev, A., & Chen, M. (2014). Bitcoin: Technical Background and Data Analysis. *Federal Reserve Board*, *2014–104*, 39.

Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural Machine Translation by Jointly Learning to Align and Translate* (No. arXiv:1409.0473). arXiv.

Bakas, D., Magkonis, G., & Oh, E. Y. (2022). What Drives Volatility in Bitcoin Market? *Finance Research Letters*, *50*, 1-8.

Balfagih, A. M., & Keselj, V. (2019). Evaluating Sentiment Classifiers for Bitcoin Tweets in Price Prediction Task. *2019 IEEE International Conference on Big Data (Big Data)*, 5499–5506.

Bashir, I. (2017). *Mastering Blockchain*. Packt Publishing Ltd.

Beheshti, N. (2022, February 5). *Cross Validation and Grid Search*. Towards Data Science. https://towardsdatascience.com/cross-validation-and-grid-search-efa64b127c1b/

Bibi, S. (2023). Money in The Time of Crypto. *Research in International Business and Finance*, *65*, 1-15.

Biedscheid, Z. S. (2023). Bitcoin Price Dynamics: A Multiple Regression Analysis of Market Variables. *Eastern Washington University Symposium*, 1-35.

Bitstamp. (2025). *Bitstamp Exchange Data*. CryptoDataDownload. https://www.cryptodatadownload.com/data/bitstamp/

Borges, T. A., & Neves, R. F. (2020). Ensemble of Machine Learning Algorithms for Cryptocurrency Investment with Different Data Resampling Methods. *Applied Soft Computing*, *90*, 1-20.

Bouteska, A., Abedin, M. Z., Hajek, P., & Yuan, K. (2024). Cryptocurrency Price Forecasting – A Comparative Analysis of Ensemble Learning and Deep Learning Methods. *International Review of Financial Analysis*, *92*, 1-12.

Brunato, D., Dell'Orletta, F., & Venturi, G. (2022). Linguistically-Based Comparison of Different Approaches to Building Corpora for Text Simplification: A Case Study on Italian. *Frontiers in Psychology*, *13*, 1-19.

Budaghyan, L. (2015). Construction and analysis of cryptographic functions. Springer.

Çelik, I. (2020). Can Bitcoin be a Stable Investment? *Financial Studies*, *24*(2 (88)), 19–36.

Cerda, G. N. C. (2021). *Bitcoin Price Prediction Through Stimulus Analysis: On The*

Footprints of Twitter's Crypto-influencers [Master's Thesis, Pontificia Universidad Católica de Chile]. https://repositorio.uc.cl/xmlui/bitstream/handle/11534/60881/TESIS_GCheuque_Firma%20Final.pdf?sequence=1

Chahooki, M. A. Z., Jahanbin, K., & Sutikno, T. (2023). Cryptocurrencies Investment Framework using Sentiment Analysis of Twitter Influencers. *Indonesian Journal of Electrical Engineering and Computer Science*, *30*(2), 1068-1079.

Chauhan, J. K., Ahmed, T., & Sinha, A. (2025). A Novel Deep Learning Model for Stock Market Prediction using a Sentiment Analysis System from Authoritative Financial Website's Data. *Connection Science*, *37*(1), 1-23.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Chen, C. Y.-H., Després, R., Guo, L., & Renault, T. (2019). What Makes Cryptocurrencies Special? Investor Sentiment and Return Predictability During The Bubble. *Comparative Political Economy: Monetary Policy eJournal*, 1–36.

Chen, C.-C., Huang, H.-H., & Chen, H.-H. (2018). NTUSD-Fin: A Market Sentiment Dictionary for Financial Social Media Data Applications. *Proceedings of the 1st Financial Narrative Processing Workshop (FNP 2018)*, 37–43.

Chen, J., Cheng, L., Yang, X., Liang, J., Quan, B., & Li, S. (2019). Joint Learning with both Classification and Regression Models for Age Prediction. *Journal of Physics: Conference Series*, *1168*, 1-12.

Chen, X., Ke, L., Lu, Z., Su, H., & Wang, H. (2020). A Novel Hybrid Model for Cantonese Rumor Detection on Twitter. *Applied Sciences*, *10*(20), 1-12.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Coefficient of Determination R-Squared is more Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation. *PeerJ Computer Science*, *7*, 1-24.

Chin, C. K., & Omar, N. (2020). Bitcoin Price Prediction Based on Sentiment of News Article and Market Data with LSTM Model. *Asia-Pacific Journal of Information Technology and Multimedia*, *9*(1), 1–16.

Chowdhury, R., Rahman, M. A., Rahman, M. S., & Mahdy, M. R. C. (2020). An Approach to Predict and Forecast The Price of Constituents and Index of Cryptocurrencyusing Machine Learning. *Physica A: Statistical Mechanics and*

Its Applications, *551*, 1-16.

Cochran, W. G. (1977). *Sampling Techniques* (Third). John Wiley & Sons.

Cohen, G. (2020). Forecasting Bitcoin trends using Algorithmic Learning Systems. *Entropy*, *22*(8), 1–11.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed., reprint). Psychology Press.

Cointelegraph. (2021). What Can You Buy with Bitcoin: Places to Spend Your Crypto in 2021. *Cointelegraph*. https://cointelegraph.com/bitcoin-for-beginners/what-can-i-buy-with-bitcoins

Corbet, S., Hou, Y. (Greg), Hu, Y., Larkin, C., Lucey, B., & Oxley, L. (2022). Cryptocurrency Liquidity and Volatility Interrelationships During The COVID-19 Pandemic. *Finance Research Letters*, *45*, 1-9.

Critien, J. V., Gatt, A., & Ellul, J. (2022). Bitcoin price change and trend prediction through twitter sentiment and data volume. *Financial Innovation*, *8*(1), 1-20.

CryptoDataDownload. (2025). *Bitstamp Exchange Data*. Crypto Data Download. https://www.cryptodatadownload.com/data/bitstamp/

Dar, A. B., Bhanja, N., & Tiwari, A. K. (2017). Do Global Financial Crises Validate Assertions of Fractal Market Hypothesis? *International Economics and Economic Policy*, *14*, 153–165.

Daskalakis, N., & Georgitseas, P. (2020). An Introduction to Cryptocurrencies: The Crypto Market Ecosystem. Routledge.

Deutscher, I., Pestello, F. P., & Pestello, H. F. G. (1993). *Sentiments and Acts*. Transaction Publishers.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.

Dolan, R. J. (2002). Emotion, Cognition, and Behavior. *Science*, *298*(5596), 1191–1194.

Dunbar, K., & Owusu-Amoako, J. (2023). Predictability of Crypto Returns: The Impact of Trading Behavior. *Journal of Behavioran and Experimental Finance*, *39*, 1-12.

Durmuş Şenyapar, H. N. (2024). Cryptocurrency on Social Media: Analyzing the Digital Discourse Towards the Coin Market. *İktisadi İdari ve Siyasal*

Araştırmalar Dergisi, *9*(23), 202–223.

Dutta, A., Kumar, S., & Basu, M. (2019). A Gated Recurrent Unit Approach to Bitcoin Price Prediction. *Journal of Risk and Financial Management*, *13*(2), 1-16.

Edgari, E., Thiojaya, J., & Qomariyah, N. N. (2022). The Impact of Twitter Sentiment Analysis on Bitcoin Price during COVID-19 with XGBoost. *2022 5th International Conference on Computing and Informatics (ICCI)*, 337–342.

El Haddaoui, B., Chiheb, R., Faizi, R., & El Afia, A. (2023). The Influence of Social Media on Cryptocurrency Price: A Sentiment Analysis Approach. *International Journal of Computing and Digital Systems*, *13*(1), 1–15.

Erfanian, S., Zhou, Y., Razzaq, A., Abbas, A., Safeer, A. A., & Li, T. (2022). Predicting Bitcoin (BTC) Price in the Context of Economic Theories: A Machine Learning Approach. *Entropy*, *24*(10), 1-29.

Fakharchian, S. (2022). Designing Forecasting Assistant of The Bitcoin Price based on Deep Learning using The Market Sentiment Analysis and Multiple Feature Extraction. *Research Square*, 1-33.

Fama, E. F. (1970). Efficient capital markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417.

Farimani, S. A., Jahan, M. V., Fard, A. M., & Tabbakh, S. R. K. (2022). Investigating The Informativeness of Technical Indicators and News Sentiment in Financial Market Price Prediction. *Knowledge-Based Systems*, *247*, 1-11.

Feizian, F., & Amiri, B. (2023). Cryptocurrency Price Prediction Model Based on Sentiment Analysis and Social Influence. *IEEE Access*, *11*, 142177–142195.

Fiorillo, S. (2018, April 18). How to Use Bitcoin for Purchases. *TheStreet*. https://www.thestreet.com/investing/what-can-you-buy-with-bitcoin-14556706

Fitts, D. A. (2022). Absolute Precision Confidence Intervals for Unstandardized Mean Differences using Sequential Stopping Rules. *Behavior Research Methods*, *55*(4), 1839–1862.

Frankel, R., Jennings, J., & Lee, J. (2022). Disclosure Sentiment: Machine Learning vs. Dictionary Methods. *Management Science*, *68*(7), 5514–5532.

Frohmann, M., Karner, M., Khudoyan, S., Wagner, R., & Schedl, M. (2023). Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting. *Big Data and Cognitive Computing*, *7*(3), 1-20.

Fukushima, K. (1980). Neocognitron: A Self-Organizing Neural Network Model for a

Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, *36*(4), 193–202.

Gadi, M. F. A., & Sicilia, M. Á. (2024). A Sentiment Corpus for The Cryptocurrency Financial Domain: The CryptoLin corpus. *Language Resources and Evaluation*, 1-19.

Galeshchuk, S., Vasylchyshyn, O., & Krysovatyy, A. (2018). Bitcoin response to Twitter sentiments. *ICTERI Workshops*, 160–168.

Girsang, A. S. & Stanley. (2023). Hybrid LSTM and GRU for Cryptocurrency Price Forecasting Based on Social Network Sentiment Analysis Using FinBERT. *IEEE Access*, *11*, 120530–120540.

Giudici, G., Milne, A., & Vinogradov, D. (2020). Cryptocurrencies: Market Analysis and Perspectives. *Journal of Industrial and Business Economics*, *47*(1), 1–18.

Glenski, M., Weninger, T., & Volkova, S. (2019). Improved Forecasting of Cryptocurrency Price using Social Signals. *arXiv:1907.00558 [Cs, q-Fin, Stat]*. http://arxiv.org/abs/1907.00558

Goleman, T. (2018). Cryptocurrency: Mining, Investing and Trading in Blockchain for Beginners. How To Buy Cryptocurrencies (Bitcoin, Ethereum, Ripple, Litecoin Or Dash) and What Wallet To Use. Cryptocurrency Investment Strategies. *Zen Mastery*.

Gomez-Perez, J. M., Denaux, R., & Garcia-Silva, A. (2020). Understanding Word Embeddings and Language Models. In *A Practical Guide to Hybrid Natural Language Processing*. Springer, Cham.

Gong, X.-R., Jin, J.-X., & Zhang, T. (2019). Sentiment Analysis using Autoregressive Language Modeling and Broad Learning System. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 1130–1134.

Gowda, N., & Chakravorty, C. (2021). Comparative study on cryptocurrency transaction and banking transaction. *Global Transitions Proceedings*, *2*(2), 530–534.

Greaves, A., & Au, B. (2015). Using The Bitcoin Transaction Graph to Predict The Price of Bitcoin. *Quoted*, *3*, 1-8.

Gudavalli, H. N., & Kancherla, K. V. R. (2023). *Predicting Cryptocurrency Prices with Machine Learning Algorithms: A Comparative Analysis* [Degree's Thesis]. Blekinge Institute of Technology.

Gunay, S., & Kaşkaloğlu, K. (2019). Seeking A Chaotic Order in The Cryptocurrency

Market. *Mathematical and Computational Applications*, *24*(2), 1-18.

Gupta, H., & Chaudhary, R. (2022). An Empirical Study of Volatility in Cryptocurrency Market. *Journal of Risk and Financial Management*, *15*(11), 1-14.

Gurdgiev, C., & O'Loughlin, D. (2020). Herding and Anchoring in Cryptocurrency Markets: Investor Reaction to Fear and Uncertainty. *Journal of Behavioral and Experimental Finance*, *25*, 1-9.

Gurgul, V., Lessmann, S., & Härdle, W. K. (2025). Deep Learning and NLP in Cryptocurrency Forecasting: Integrating Financial, Blockchain, and Social Media Data. *International Journal of Forecasting*, *41*(4), 1666–1695.

Hadan, H., Zhang-Kennedy, L., Nacke, L., & Mäkelä, V. (2024). Comprehending the Crypto-Curious: How Investors and Inexperienced Potential Investors Perceive and Practice Cryptocurrency Trading. *International Journal of Human–Computer Interaction*, *40*(19), 5675–5696.

Hajric, V., & Greifeld, K. (2021, December 21). Bitcoin Went Mainstream in 2021. It's Just as Volatile as Ever. *Bloomberg*.

Hamza, S. (2020). The Effect of Tweets Made by Cryptocurrency Opinion Leaders on Bitcoin Prices. *Saudi Journal of Economics and Finance*, *4*(12), 569–589.

Han, H., Zhang, J., Yang, J., Shen, Y., & Zhang, Y. (2018). Generate Domain-specific Sentiment Lexicon for Review Sentiment Analysis. *Multimedia Tools and Applications*, *77*(16), 21265–21280.

Haritha, G. B., & Sahana, N. B. (2023). Cryptocurrency Price Prediction using Twitter Sentiment Analysis. *Natural Language Processing, Information Retrieval and AI*, 13–22.

Harris, Z. S. (1954). Distributional Structure. *Routledge Taylor & Francis Group*, *10*(2–3), 146–162.

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing*, *40*(1), 75–87.

Hasan, S. H., Hasan, S. H., Ahmed, M. S., & Hasan, S. H. (2022). A Novel Cryptocurrency Prediction Method Using Optimum CNN. *Computers, Materials & Continua*, *71*(1), 1051–1063.

Hazra, A. (2017). Using the Confidence Interval Confidently. *Journal of Thoracic Disease*, *9*(10), 4124–4129.

Hitam, N. A., Ismail, A. R., Samsudin, R., & Ameerbakhsh, O. (2021). The Influence

of Sentiments in Digital Currency Prediction Using Hybrid Sentiment-based Support Vector Machine with Whale Optimization Algorithm (SVMWOA). *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, 1–7.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780.

Hoffman, M. R., Ibáñez, L.-D., & Simperl, E. (2020). Toward a Formal Scholarly Understanding of Blockchain-Mediated Decentralization: A Systematic Review and a Framework. *Frontiers in Blockchain*, *3*, 1-18.

Hooson, M., & Pratt, K. (2023, August 10). *Top 10 Cryptocurrencies Of August 2023*. Forbes Advisor. https://www.forbes.com/uk/advisor/investing/cryptocurrency/top-10-cryptocurrencies-august-2023/

Hossain, S. S., Arafat, Y., & Hossain, Md. E. (2021). Context-Based News Headlines Analysis: A Comparative Study of Machine Learning and Deep Learning Algorithms. *Vietnam Journal of Computer Science*, *08*(04), 513–527.

Huang, W., Wang, H., & Wang, S. (2024). A Structural VAR and VECM Modeling Method for Open-High-Low-Close Data Contained in Candlestick Chart. *Financial Innovation*, *10*(1), 1-29.

Hutto, C. J., & Gilbert, E. (2015). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM 2014)*, 1-10.

Ibrahim, A. (2021). Forecasting the Early Market Movement in Bitcoin Using Twitter's Sentiment Analysis: An Ensemble-based Prediction Model. *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 1–5.

İçellioğlu, C. Ş., & Öner, S. (2019). An Investigation on The Volatility of Cryptocurrencies by Means of Heterogeneous Panel Data Analysis. *Procedia Computer Science*, *158*, 913–920.

Inamdar, A., Bhagtani, A., Bhatt, S., & Shetty, P. M. (2019). Predicting Cryptocurrency Value using Sentiment Analysis. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 932–934.

Jahjah, F. H., & Rajab, M. (2020). Impact of Twitter Sentiment Related to Bitcoin on Stock Price Returns. *Journal of Engineering*, *26*(6), 60–71.

Jain, A., Tripathi, S., Dwivedi, H. D., & Saxena, P. (2018). Forecasting price of cryptocurrencies using tweets sentiment analysis. *2018 Eleventh International Conference on Contemporary Computing (IC3)*, 1–7.

Jain, P. K., Saravanan, V., & Pamula, R. (2021). A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *20*(5), 1–15.

Jais, I. K. M., Ismail, A. R., & Nisa, S. Q. (2019). Adam optimization algorithm for wide and deep neural network. *Knowledge Engineering and Data Science*, *2*(1), 41-46.

Jaquart, P., Dann, D., & Weinhardt, C. (2021). Short-term Bitcoin market prediction via machine learning. *The Journal of Finance and Data Science*, *7*, 45–66.

Ji, G., Yu, J., Hu, K., Xie, J., & Ji, X. (2022). An Adaptive Feature Selection Schema using Improved Technical Indicators for Predicting Stock Price Movements. *Expert Systems with Applications*, *200*, 1-41.

Ji, L., Gong, P., & Yao, Z. (2019). A Text Sentiment Analysis Model Based on Self-Attention Mechanism. *Proceedings of the 3rd International Conference on High Performance Compilation, Computing and Communications*, 33–37.

Ji, S., Kim, J., & Im, H. (2019). A Comparative Study of Bitcoin Price Prediction using Deep Learning. *Mathematics*, *7*(10), 1-20.

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent Advancements and Challenges of NLP-based Sentiment Analysis: A State-Of-The-Art Review. *Natural Language Processing Journal*, *6*, 1-30.

John, D. L., Binnewies, S., & Stantic, B. (2024). Cryptocurrency Price Prediction Algorithms: A Survey and Future Directions. *Forecasting*, *6*(3), 637–671.

John, V., & Vechtomova, O. (2017a). *SemEval-2017 Task 5 News and Microblogs dataset*. SemEval-2017 Task 5. https://alt.qcri.org/semeval2017/task5/index.php?id=data-and-tools

John, V., & Vechtomova, O. (2017b). UW-FinSent at SemEval-2017 Task 5: Sentiment Analysis on Financial News Headlines using Training Dataset Augmentation. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 872–876.

Johnson, D. (2022). Cryptocurrency and Public Policy: Implications for Democracy and

Governance (1st ed.). Taylor & Francis.

Jung, H. S., Hong Lee, S., Lee, H., & Hyun Kim, J. (2023). Predicting Bitcoin Trends Through Machine Learning Using Sentiment Analysis with Technical Indicators. *Computer Systems Science and Engineering*, *46*(2), 2231–2246.

Jung, H. S., Lee, H., & Kim, J. H. (2025). Detecting Bitcoin Sentiment: Leveraging Language Model Applications in Sentiment Analysis for Bitcoin Price Prediction. *Neural Processing Letters*, *57*(5), 1–25.

Kalodner, H., Carlsten, M., Ellenbogen, P., Bonneau, J., & Narayanan, A. (2015). An Empirical Study of Namecoin and Lessons for Decentralized Namespace Design. *Workshop on the Economics of Information Security (WEIS 2015)*, 1–21.

Kaminski, J. (2016). Nowcasting the Bitcoin Market with Twitter Signals. *arXiv:1406.7577 [Cs]*. http://arxiv.org/abs/1406.7577

Kanat, E. (2023). The Validity of Technical Analysis in The Cryptocurrency Market: Evidence from Machine Learning Methods. *Pressacademia*, 12(3), 102-109.

Kang, H.-J., Lee, S.-G., & Park, S.-Y. (2021). Information Efficiency in The Cryptocurrency Market: The Efficient-Market Hypothesis. *Journal of Computer Information Systems*, 1–10.

Karalevicius, V., Degrande, N., & De Weerdt, J. (2018). Using Sentiment Analysis to Predict Interday Bitcoin Price Movements. *The Journal of Risk Finance*, *19*(1), 56–75.

Karch, J. (2019). *Improving on Adjusted R-Squared* [Preprint]. PsyArXiv.

Karp, A., & Van Vuuren, G. (2019). Investment Implications of The Fractal Market Hypothesis. *Annals of Financial Economics*, *14*(1), 1-27.

Kaur, G., & Sharma, A. (2023). A Deep Learning-Based Model using Hybrid Feature Extraction Approach for Consumer Sentiment Analysis. *Journal of Big Data*, *10*(1), 1-23.

Kehinde, T. O., Adedokun, O. J., Joseph, A., Kabirat, K. M., Akano, H. A., & Olanrewaju, O. A. (2025). Helformer: An Attention-Based Deep Learning Model for Cryptocurrency Price Forecasting. *Journal of Big Data*, *12*(1), 1–39.

Khoo, C. S., & Johnkhan, S. B. (2018). Lexicon-based Sentiment Analysis: Comparative Evaluation of Six Sentiment Lexicons. *Journal of Information Science*, *44*(4), 491–511.

Khursheed, A., Naeem, M., Ahmed, S., & Mustafa, F. (2020). Adaptive Market Hypothesis: An Empirical Analysis of Time–Varying Market Efficiency of Cryptocurrencies. *Cogent Economics & Finance*, *8*(1), 1-15.

Kilimci, Z. H. (2020). Sentiment Analysis Based Direction Prediction in Bitcoin using Deep Learning Algorithms and Word Embedding Models. *International Journal of Intelligent Systems and Applications in Engineering*, *8*(2), 60–65.

Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments And Replies. *PLOS ONE*, *11*(8), 1-17.

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980

Klie, J.-C., Castilho, R. E. D., & Gurevych, I. (2024). Analyzing Dataset Annotation Quality Management in The Wild. *Computational Linguistics*, *50*(3), 817–866.

Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, *65*, 1-22.

Krippendorff, K. (2004a). Content analysis: An introduction to its methodology (2nd ed). Sage.

Krippendorff, K. (2004b). Measuring the Reliability of Qualitative Text Analysis Data. *Quality & Quantity*, *38*(6), 787–800. https://doi.org/10.1007/s11135-004-8107-7

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (Fourth). SAGE Publications.

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, *3*(1), 1-7.

Kruger, S., Maturana, G., & Nickerson, J. (2022). How Has COVID-19 Impacted Research Production in Economics and Finance? *The Review of Financial Studies*, *26*, 3348-3381.

Kshetri, N. (2018). Cryptocurrencies: Transparency Versus Privacy [Cybertrust]. *Computer*, *51*(11), 99–111.

Kudo, T., & Richardson, J. (2018). SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language*

Processing: System Demonstrations, 66–71.

Kukacka, J., & Kristoufek, L. (2023). Fundamental and Speculative Components of The Cryptocurrency Pricing Dynamics. *Financial Innovation*, *9*(1), 61.

Kumar, D., & Rath, S. K. (2020). Predicting the Trends of Price for Ethereum Using Deep Learning Techniques. In S. S. Dash, C. Lakshmi, S. Das, & B. K. Panigrahi (Eds.), *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, *1056*, 103–114.

Kuznetsov, O., Kostenko, O., Klymenko, K., Hbur, Z., & Kovalskyi, R. (2025). Machine Learning Analytics for Blockchain-Based Financial Markets: A Confidence-Threshold Framework for Cryptocurrency Price Direction Prediction. *Applied Sciences*, *15*(20), 1–31.

Lamon, C., Nielsen, E., & Redondo, E. (2017). Cryptocurrency price prediction using news and social media sentiment. *SMU Data Science Review*, 1–22.

Laurence, T. (2019). Introduction to Blockchain technology. Van Haren.

Leirvik, T. (2022). Cryptocurrency Returns and The Volatility of Liquidity. *Finance Research Letters*, *44*, 1-8.

Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using Gradient Boosting Tree model. *Frontiers in Physics*, *7*(98), 1–8.

Li, X., Ding, L., Du, Y., Fan, Y., & Shen, F. (2022). Position-Enhanced Multi-Head Self-Attention Based Bidirectional Gated Recurrent Unit for Aspect-Level Sentiment Classification. *Frontiers in Psychology*, *12*, 1-11.

Li, Y., Zhang, W., Xiong, X., & Wang, P. (2020). Does Size Matter in The Cryptocurrency Market? *Applied Economics Letters*, *27*(14), 1141–1149.

Liang, J., Li, L., Chen, W., & Zeng, D. (2019). Towards an Understanding of Cryptocurrency: A Comparative Analysis of Cryptocurrency, Foreign Exchange, and Stock. *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 137–139.

Liu, Y., Wei, S., Huang, H., Lai, Q., Li, M., & Guan, L. (2023). Naming Entity Recognition of Citrus Pests and Diseases based on The BERT-BiLSTM-CRF Model. *Expert Systems with Applications*, *234*, 1-10.

Livieris, I. E., Pintelas, E., Stavroyiannis, S., & Pintelas, P. (2020). Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series. *Algorithms*, *13*(5), 1-21.

Lo, A. W. (2004). The Adaptive Markets Hypothesis. *The Journal of Portfolio Management*, *30*(5), 15–29.

Loginova, E., Tsang, W. K., Van Heijningen, G., Kerkhove, L.-P., & Benoit, D. F. (2021). Forecasting Directional Bitcoin Price Returns Using Aspect-Based Sentiment Analysis on Online Text Data. *Machine Learning*, *113*, 4761-4784.

Lohr, S. L. (2009). *Sampling: Design and Analysis* (2nd ed.). Cengage Learning.

Loria, S. (2020, April 26). *Textblob Documentation*. https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf

Loughran, T., & Mcdonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks.

Luo, J. (2020). Bitcoin price prediction in the time of COVID-19. *2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID)*, 243–247.

Madan, I., Saluja, S., & Zhao, A. (2015). Automated Bitcoin Trading via Machine Learning Algorithms. *Semantic Scholar*.

Mai, F., Bai, Q., & Shan, J. (2015). The Impacts of Social Media on Bitcoin Performance. *International Conference on Information Systems (ICIS)*, 1–16.

Mai, F., Shan, Z., Bai, Q., Wang, X. (Shane), & Chiang, R. H. L. (2018). How Does Social Media Impact Bitcoin Value? A Test of The Silent Majority Hypothesis. *Journal of Management Information Systems*, *35*(1), 19–52.

Majlinda, A. (2024). The Importance of Manual Inter-Annotation in the Process of Aspect-Term Extraction in Aspect-Based Sentiment Analysis. *South East European Journal of Sustainable Development*, *8*(1), 19-25.

Malik, N., Singh, S., Biswas, P., & Sharan, A. (2024). Fundamentals of Vector-Based Text Representation and Word Embeddings. In *Text Mining Approaches for Biomedical Data*. Springer.

Mallqui, D. C. A., & Fernandes, R. A. S. (2019). Predicting The Direction, Maximum, Minimum and Closing Prices of Daily Bitcoin Exchange Rate using Machine Learning Techniques. *Applied Soft Computing*, *75*, 596–606.

Manaf, S. A., & Nordin, M. J. (2009). Review on Statistical Approaches for Automatic Image Annotation. *2009 International Conference on Electrical Engineering and Informatics*, 56–61.

Maqsood, U., Khuhawar, F. Y., Talpur, S., Jaskani, F. H., & Memon, A. A. (2022). Twitter Mining based Forecasting of Cryptocurrency using Sentimental

Analysis of Tweets. *2022 Global Conference on Wireless and Optical Technologies (GCWOT)*, 1–6.

Matkovskyy, R. (2019). Centralized and Decentralized Bitcoin markets: Euro vs USD vs GBP. *The Quarterly Review of Economics and Finance*, *71*, 270–279.

Matta, M., Lunesu, I., & Marchesi, M. (2015). The Predictor Impact of Web Search Media on Bitcoin Trading Volumes: *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 620–626.

McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin using Machine Learning. *2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, 339–343.

Meng, F., & Wang, C.-A. (2024). Sentiment analysis with adaptive multi-head attention in Transformer. *Proceedings of the 4th International Conference on Signal Processing and Machine Learning*, 133–137.

Miaschi, A., & Dell'Orletta, F. (2020). Contextual and Non-Contextual Word Embeddings: An in-depth Linguistic Investigation. *Proceedings of the 5th Workshop on Representation Learning for NLP*, 110–119.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (No. arXiv:1301.3781). arXiv.

Mikulová, M., Straka, M., Štěpánek, J., Štěpánková, B., & Hajič, J. (2023). *Quality and Efficiency of Manual Annotation: Pre-annotation Bias* (No. arXiv:2306.09307). arXiv. http://arxiv.org/abs/2306.09307

Miraz, M. H., Hasan, M. T., Rekabder, M. S., & Akhter, R. (2022). Trust, Transaction Transparency, Volatility, Facilitating Condition, Performance Expectancy Towards Cryptocurrency Adoption Through Intention to Use. *Journal of Management Information and Decision Sciences*, *25*(1), 1-20.

Mnif, E., Salhi, B., Mouakha, K., & Jarboui, A. (2022). Investor behavior and cryptocurrency market bubbles during the COVID-19 pandemic. *Review of Behavioral Finance*, *14*(4), 491–507.

Mochtak, M., Rupnik, P., & Ljubešič, N. (2022). *The ParlaSent-BCS Dataset of Sentiment-annotated Parliamentary Debates from Bosnia-Herzegovina, Croatia, and Serbia* (No. arXiv:2206.00929). arXiv.

Mohagheghzadeh, A., Amiri, B., & Makui, A. (2024). A Novel Dynamic Model for Ranking Cryptocurrencies in Different Time Horizons Based on Deep

Learning and Sentiment Analysis. *IEEE Access*, *12*, 83022–83042.

Mohamad Zamani, N. A., Kamaruddin, N., & Yusof, A. M. B. (2025). Dataset on Sentiment-Based Cryptocurrency-Related News and Tweets in English and Malay Language. *Language Resources and Evaluation*, *59*(2), 807-842.

Mohammad, S. (2016). A Practical Guide to Sentiment Annotation: Challenges and Solutions. *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 174–179.

Mohanty, P., Patel, D., Patel, P., & Roy, S. (2018). Predicting Fluctuations in Cryptocurrencies' Price using users' Comments and Real-time Prices. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 477–482.

Mohapatra, S., Ahmed, N., & Alencar, P. (2019). KryptoOracle: A real-time cryptocurrency price prediction platform using Twitter sentiments. *2019 IEEE International Conference on Big Data (Big Data)*, 5544–5551.

Mountstephens, J., Quen, M. T. Z., & Hung, L. P. (2023). Bilingual Sentiment Analysis on Malaysian Social Media using Vader and Normalisation Heuristics. *Journal of Theoretical and Applied Information Technology*, *101*(12), 5037-5050.

Mudassir, M., Bennbaia, S., Unal, D., & Hammoudeh, M. (2020). Time-series Forecasting of Bitcoin Prices using High-Dimensional Features: A Machine Learning Approach. *Neural Computing and Applications*, 1–15.

Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, *150*, 6–9.

Nakamoto, S. (2008). Bitcoin: A Peer-To-Peer Electronic Cash System. *Decentralized Business Review*, 1-9.

Nasekin, S., & Chen, C. Y.-H. (2020). Deep Learning-Based Cryptocurrency Sentiment Construction. *Digital Finance*, *2*(1–2), 39–67.

Niederle, M., & Roth, A. E. (2009). The Effects of A Centralized Clearinghouse on Job Placement, Wages, and Hiring Practices. In: Studies of Labor Market Intermediation, *National Bureau of Economic Research Chapters,* 235-271. National Bureau of Economic Research, Inc.

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). *Activation Functions: Comparison of trends in Practice and Research for Deep Learning* (No. arXiv:1811.03378). arXiv. http://arxiv.org/abs/1811.03378

Oikonomopoulos, S., Tzafilkou, K., Karapiperis, D., & Verykios, V. (2022).

Cryptocurrency Price Prediction using Social Media Sentiment Analysis. *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–8.

Oostendorp, M., Little, T., & Berghoff, R. (2024). First language primacy in multilingualism and emotion research: A view from Africa. *International Journal of Bilingual Education and Bilingualism*, 1–15.

Ortu, M., Uras, N., Conversano, C., Bartolucci, S., & Destefanis, G. (2022). On Technical Trading And Social Media Indicators For Cryptocurrency Price Classification Through Deep Learning. *Expert Systems with Applications*, *198*, 1-45.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Library of Congress.

OSL News. (2025, May 19). *How to Use Volume in Crypto Trading?* https://osl.com/academy/article/how-to-use-volume-in-crypto-trading

Pandey, T. D. (2024). Impact of Musk's remarks on Volatility of Bitcoin and Dogecoin Amid COVID-19 Pandemic. *Journal of Digital Economy*, *3*, 85–102.

Pant, D. R., Neupane, P., Poudel, A., Pokhrel, A. K., & Lama, B. K. (2018). Recurrent Neural Network Based Bitcoin Price Prediction by Twitter Sentiment Analysis. *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, 128–132.

Parihar, S. (2025, May 4). *Top 10 Crypto Holding Countries and Governments* [Online news]. The Crypto Times. https://www.cryptotimes.io/articles/explained/top-10-crypto-holding-countries-and-governments/#google_vignette

Park, J., & Seo, Y.-S. (2023). Twitter Sentiment Analysis-Based Adjustment of Cryptocurrency Action Recommendation Model for Profit Maximization. *IEEE Access*, *11*, 44828–44841.

Passalis, N., Avramelou, L., Seficha, S., Tsantekidis, A., Doropoulos, S., Makris, G., & Tefas, A. (2022). Multisource Financial Sentiment Analysis for Detecting Bitcoin Price Change Indications using Deep Learning. *Neural Computing and Applications*, *34*(22), 19441–19452.

Patel, M. M., Tanwar, S., Gupta, R., & Kumar, N. (2020). A Deep Learning-Based Cryptocurrency Price Prediction Scheme for Financial Institutions. *Journal of Information Security and Applications*, *55*(2), 1-14.

Pathak, S., & Kakkar, A. (2020). Cryptocurrency Price Prediction Based on Historical

Data and Social Media Sentiment Analysis. *Proceedings of 7th Innovations in Computer Science and Engineering (ICICSE)*, *97*, 177–192.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Perera, D., Lim, J., Gunraku, S., & Lim, W. (2024). Cryptocurrency Analysis: Price Prediction of Cryptocurrency Using User Sentiments and Quantitative Data: *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*, 210–217.

Peters, E. E. (1994). Fractal Market Analysis: Applying Chaos Theory to Investment And Economics. John Wiley & Sons.

Phaladisailoed, T., & Numnonda, T. (2018). Machine Learning Models Comparison for Bitcoin Price Prediction. *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 506–511.

Pillai, S., Biyani, D., Motghare, R., & Karia, D. (2021). Price Prediction and Notification System for cryptocurrency Share Market Trading. *2021 International Conference on Communication Information and Computing Technology (ICCICT)*, 1–7.

Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T., & Pintelas, P. (2020). Investigating the problem of cryptocurrency price prediction: A deep learning approach. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations*, 584, 99–110. Springer International Publishing.

Podobas, W. J. (2019). Cryptocurrency Encyclopedia: The comprehensive guide through the 100 most important cryptocurrencies. Wojciech Podobas.

Politou, E., Casino, F., Alepis, E., & Patsakis, C. (2021). Blockchain Mutability: Challenges and Proposed Solutions. *IEEE Transactions on Emerging Topics in Computing*, *9*(4), 1972–1986.

Poongodi, M., Nguyen, T. N., Hamdi, M., & Cengiz, K. (2021). Global Cryptocurrency Trend Prediction using Social Media. *Information Processing & Management*, *58*(6), 1-11.

Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*.

Prajapati, P. (2020). Predictive Analysis of Bitcoin Price Considering Social Sentiments. *arXiv:2001.10343 [Cs]*. http://arxiv.org/abs/2001.10343

Prashanth, G., Revanth, M., Narasimhadhan, A. V., & Naganjaneyulu G, V. S. S. K. R. (2023). Multi Indicator based Hierarchical Strategies for Technical Analysis of Crypto market Paradigm. *International Journal of Electrical and Computer Engineering Systems*, *14*(7), 765–780.

PropTrack. (2025). *API vs Data Extracts*. PropTrack. https://www.proptrack.com.au/api-vs-data-extracts/

Protonotarios, G. (2019). Cryptocurrency Trading Guide: Fundamental & Technical Analysis for Cryptocurrency Thinkers. George Protonotarios.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*, 1-12.

Rahman, S. (2021, May 27). *Elon Musk's Controversial Crypto Tweets – Are they Market Manipulation?* Rahman Ravelli. https://www.rahmanravelli.co.uk/expertise/cryptocurrency/articles/elon-musk-s-controversial-crypto-tweets-are-they-market-manipulation/

Ramkumar, G. (2021). Cryptocurrency Portfolio Construction using Machine Learning Models. In K. Jajuga, H. Locarek-Junge, L. T. Orlowski, & K. Staehr (Eds.), *Contemporary Trends and Challenges in Finance*, 103–122. Springer International Publishing.

Ramos, J. (2003). Using TF-IDF to Determine Word Relevance in Document Queries. *Proceedings of the First Instructional Conference on Machine Learning*, *242*, 133–142.

Rani, U., Kumar, S., Dahiya, N., Solanki, K., Kuttan, S. R., Shah, S., Shaheen, M., & Ahmad, F. (2024). An Optimized Neural Network with Adahessian for Cryptojacking Attack Prediction for Securing Crypto Exchange Operations of MEC Applications. *Journal of Cloud Computing*, *13*(1), 1-19.

Ranjan, S., Kayal, P., & Saraf, M. (2023). Bitcoin Price Prediction: A Machine Learning Sample Dimension Approach. 61, 1617–1636.

Rateb, M. N., Alansary, Prof. S., Elzouka, Prof. M. K., & Galal, M. (2024). Cryptocurrency Price Forecasting Implementing Sentiment Analysis during the Russian-Ukrainian War. *Research Square*, 1-28.

Raval, S. (2016). Decentralized Applications: Harnessing Bitcoin's Blockchain Technology. O'Reilly Media, Inc.

Ritchie, M. J., Drummond, K. L., Smith, B. N., Sullivan, J. L., & Landes, S. J. (2022). Development of a Qualitative Data Analysis Codebook Informed by the i-PARIHS Framework. *Implementation Science Communications*, *3*(1), 1-16.

Robertson, S. E., & Jones, K. S. (1976). Relevance Weighting of Search Terms. *Journal of the American Society for Information Science*, *27*(3), 129–146.

Robeson, S. M., & Willmott, C. J. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *PLOS ONE*, *18*(2), 1-8.

Rodrigues, F., & Machado, M. (2025). High-Frequency Cryptocurrency Price Forecasting using Machine Learning Models: A Comparative Study. *Information*, *16*, 1-21.

Rognone, L., Hyde, S., & Zhang, S. S. (2020). News Sentiment in The Cryptocurrency Market: An Empirical Comparison with Forex. *International Review of Financial Analysis*, *69*, 1-17.

Sagnika, S., Pattanaik, A., Shankar Prasad Mishra, B., & Meher, S. K. (2020). A Review on Multi-Lingual Sentiment Analysis by Machine Learning Methods. *Journal of Engineering Science and Technology Review*, *13*(2), 154–166.

Salač, A. (2019). *Forecasting of the cryptocurrency market through social media sentiment analysis* [[Student Theses], University of Twente]. https://essay.utwente.nl/78607/

Sasada, T., Liu, Z., Baba, T., Hatano, K., & Kimura, Y. (2020). A Resampling Method for Imbalanced Datasets Considering Noise and Overlap. *Procedia Computer Science*, *176*, 420–429.

Sathiyapriya, K., Vankadara, S., Babu, K. S., & Muralidharan, M. (2023). Performance Comparison of LSTM and XGBOOST for Ether Price Prediction from Spam Filtered Tweets. *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, 650–655.

Sattarov, O., Jeon, H. S., Oh, R., & Lee, J. D. (2020). Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis. *2020 International Conference on Information Science and Communications Technologies (ICISCT)*, 1–4.

Saxena, A., Sukumar, T. R., & Nadu, T. (2018). Predicting Bitcoin Price using LSTM and Compare Its Predictability with ARIMA model. *International Journal of Pure and Applied Mathematics*, *119*(17), 2591–2600.

Schulte, M., & Eggert, M. (2021). Predicting Hourly Bitcoin Prices Based on Long Short- term Memory Neural Networks. *16th International Conference on*

Wirtschaftsinformatik, 1-15.

Sekioka, S., Hatano, R., & Nishiyama, H. (2023). Market Prediction using Machine Learning based on Social Media Specific Features. *Artificial Life and Robotics*, *28*(2), 410–417.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A Practical Guide to Calculating Cohen's F2, a Measure of Local Effect Size, from PROC MIXED. *Frontiers in Psychology*, *3*, 1-6.

Shah, D., & Zhang, K. (2014). Bayesian regression and Bitcoin. IEEE 52nd Annual Allerton Conference on Communication, Control, and Computing, 409–414.

Shah, N. (2021). *Emot Dictionary* [Github]. https://github.com/NeelShah18/emot

Shah, R. (2024, December 19). *Tune Hyperparameters with GridSearchCV*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/tune-hyperparameters-with-gridsearchcv/

Shahzad, M. K., Bukhari, L., Khan, T. M., Islam, S. M. R., Hossain, M., & Kwak, K.-S. (2021). BPTE: Bitcoin Price Prediction and Trend Examination using Twitter Sentiment Analysis. *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, 119–122.

Shakya, D., Deshpande, V., Safari, M. J. S., & Agarwal, M. (2024). Performance Evaluation of Machine Learning Algorithms for The Prediction of Particle Froude Number (F_{rn}) using Hyper-parameter Optimizations Techniques. *Expert Systems with Applications*, *256*, 1-16.

Shintate, T., & Pichl, L. (2019). Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning. *Journal of Risk and Financial Management*, *12*(1), 1-15.

Shrivastava, T., Suleiman, B., Desa, S. A. J., Alibasa, M. J., Tregubov, A., & Zhang, J. (2025). Benchmarking Modeling Architectures for Cryptocurrency Price Prediction using Financial and Social Media Data. *Social Network Analysis and Mining*, *15*(1), 1–16.

Sin, E., & Wang, L. (2017). Bitcoin Price Prediction using Ensembles of Neural Networks. *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 666–671.

Song, Y., & Lu, Y. (2015). Decision Tree Methods: Applications for Classification and Prediction. *Shanghai Archives of Psychiatry, 27*(2), 130-135.

Steinert, L., & Herff, C. (2018). Predicting Altcoin Returns using Social Media. *PLOS*

ONE, *13*(12), 1-12.

Stenqvist, E., & Lönnö, J. (2017). *Predicting Bitcoin Price Fluctuation with Twitter Sentiment Analysis* [Degree Project]. KTH Royal Institute of Technology School of Computer Science and Communication.

Struga, K., & Qirici, O. (2018). Bitcoin Price Prediction with Neural Networks. *International Conference on Recent Trends and Applications in Computer Science and Information Technology*, *2280*, 1-6.

Sun, X., Liu, M., & Sima, Z. (2020). A Novel Cryptocurrency Price Trend Forecasting Model based on LightGBM. *Finance Research Letters*, *32*, 1-6.

Symeonidis, S., Kordonis, J., Effrosynidis, D., & Arampatzis, A. (2017). DUTH at SemEval-2017 Task 5: Sentiment Predictability in Financial Microblogging and News Articles. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 861–865.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, *37*(2), 267–307.

Talaat, A. S. (2023). Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data*, *10*(1), 1-18.

Tan, Z., Li, D., Wang, S., Beigi, A., Jiang, B., Bhattacharjee, A., Karami, M., Li, J., Cheng, L., & Liu, H. (2024). *Large Language Models for Data Annotation: A Survey* (No. arXiv:2402.13446). arXiv. http://arxiv.org/abs/2402.13446

Tandon, C., Revankar, S., Palivela, H., & Parihar, S. S. (2021a). How Can We Predict The Impact of The Social Media Messages on The Value of Cryptocurrency? Insights From Big Data Analytics. *International Journal of Information Management Data Insights*, *1*(2), 1-8.

Taskinsoy, J. (2021). Bitcoin: A New Digital Gold Standard in the 21st Century? *SSRN Electronic Journal*, 1-59.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in The Stock Market. *The Journal of Finance*, *62*(3), 1139–1168.

Thomas. (2025, April 28). Ethereum (ETH): What Is It, How Does It Work and What Is It Used For in 2025? *InvestX*. https://investx.fr/en/learn/crypto/ethereum/

Tilborg, van, H. C. A. (2011). Encyclopedia of Cryptography and Security. In van Tilborg H. C. A. & S. Jajodia (Eds.), *Encyclopedia of Cryptography and Security*, 1194–1195. Springer.

Tin, K. H. (1995). Random Decision Forests. *Proceedings of 3rd International*

Conference on Document Analysis and Recognition, *1*, 278–282.

Tomek, I. (1976). Two Modifications of CNN. *IEEE Transactions on Systems Man and Communications*, *6*, 769–772.

Urquhart, A. (2018). What Causes the Attention of Bitcoin? *Entropy*, *21*(589), 1-12.

Vacca, S., Costerbosa, C. L., Spada, A., Riotta, G., & Uras, N. (2021). Investigation of Coronavirus Impact on Blockchain and Cryptocurrencies Markets. *2021 IEEE/ACM 4th International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*, 56–60.

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price Movement Prediction of Cryptocurrencies using Sentiment Analysis and Machine Learning. *Entropy*, *21*(6), 1-12.

Van Atteveldt, W., Van Der Velden, M. A. C. G., & Boukes, M. (2021). The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms. *Communication Methods and Measures*, *15*(2), 121–140.

Van Horne, J. C., & Parker, G. G. C. (1967). The Random-walk Theory: An Empirical Test. *Financial Analysts Journal*, *23*(6), 87–92.

Vapnik, V., & Lerner, A. Ya. (1963). Pattern Recognition Using Generalized Portrait Method. *Automation and Remote Control*, *24*, 774–780.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need, *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 1-11.

Vo, A.-D., Nguyen, Q.-P., & Ock, C.-Y. (2019). Sentiment Analysis of News for Effective Cryptocurrency Price Prediction. *International Journal of Knowledge Engineering*, *5*(2), 47–52.

Wang, Y., & Chen, R. (2020). Cryptocurrency Price Prediction Based on Multiple Market Sentiment. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 1092–1100.

Wen, N. S., & Ling, L. S. (2023). Evaluation of Cryptocurrency Price Prediction Using LSTM and CNNs Models. *International Journal on Informatics Visualization*, *7*(3–2), 2016–2024.

Wołk, K. (2019). Advanced Social Media Sentiment Analysis for Short-Term Cryptocurrency Price Prediction. *Expert Systems*, *37*(2), 1–16.

Wooley, S., Edmonds, A., Bagavathi, A., & Krishnan, S. (2019). Extracting

Cryptocurrency Price Movements from the Reddit Network Sentiment. *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 500–505.

Wu, C.-H., Lu, C.-C., Ma, Y.-F., & Lu, R.-S. (2018). A New Forecasting Framework for Bitcoin Price with LSTM. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 168–175.

Wu, J. (2017). Introduction to Convolutional Neural Networks. 1-31

Wu, M., Brandhorst, H., Marinescu, M.-C., Lopez, J. M., Hlava, M., & Busch, J. (2023). Automated metadata annotation: What is and is not possible with machine learning. *Data Intelligence*, *5*(1), 122–138. Nanjing University, China

Yahoo! Finance. (2025, May 7). *Crypto* [Crypto market]. Yahoo! Finance. https://finance.yahoo.com/markets/crypto/all/

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding. *Neural Information Processing* Systems, 1-18

Yao, W., Xu, K., & Li, Q. (2019). Exploring the Influence of News Articles on Bitcoin Price with Machine Learning. *2019 IEEE Symposium on Computers and Communications (ISCC)*, 1–6.

Ye, Z., Wu, Y., Chen, H., Pan, Y., & Jiang, Q. (2022). A Stacking Ensemble Deep Learning Model for Bitcoin Price Prediction Using Twitter Comments on Bitcoin. *Mathematics*, *10*(8), 1-21.

Yenidoğan, I., Çayir, A., Kozan, O., Dağ, T., & Arslan, Ç. (2018). Bitcoin Forecasting using ARIMA and PROPHET. *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, 621–624.

Zamani, N. A. M., & Kamaruddin, N. (2023). Crypto-sentiment Detection in Malay Text using Language Models with an Attention Mechanism. *Journal of Information System Engineering and Business Intelligence*, *9*(2), 147–160.

Zamani, N. A. M., Yan, J. L. S., & Yusof, A. M. (2022). Cryptocurrency Price Prediction using Bi-GRU Model with English and Malay News sentiment features. *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 136–141.

Zhang, L., Zhou, Y., Duan, X., & Chen, R. (2018). A Hierarchical Multi-Input and Output Bi-GRU Model for Sentiment Analysis on Customer Reviews. *IOP Conference Series: Materials Science and Engineering*, *322*, 1-6.

Zhou, Z., Song, Z., Xiao, H., & Ren, T. (2023). Multi-source Data Driven Cryptocurrency Price Movement Prediction and Portfolio Optimization. *Expert Systems with Applications*, *219*, 1-30.

Zolkepli, H. (2022). *Malaya Documentation* [Computer software]. https://malaya.readthedocs.io/en/stable/Dataset.html#

Zubair, M., Ali, J., Alhussein, M., Hassan, S., Aurangzeb, K., & Umair, M. (2024). An Improved Machine Learning-Driven Framework for Cryptocurrencies Price Prediction With Sentimental Cautioning. *IEEE Access*, *12*, 51395–51418.

APPENDICES

APPENDIX 1

Demographic of Annotators

Annotator A

NAME:	Dr. Ahmad Muhyiddin
ORGANIZATION:	Universiti Teknologi MARA (UiTM)
OCCUPATION:	Senior Lecturer (PhD)
JOB SPECIALIZATION:	Language studies in English and well-versed in Malay language
WORKING EXPERIENCE:	6 years

Annotator B

NAME:	Muhammad Khidhir
ORGANIZATION:	Centauri Services and Technology Sdn Bhd
OCCUPATION:	P2P Dispute Specialist
JOB SPECIALIZATION:	Handling cryptocurrency transactions among the users
WORKING EXPERIENCE:	3 years (+5 years involvement in cryptocurrency trading)

Annotator C

NAME:	Muhammad Hazwan
ORGANIZATION:	Silverlake Axis
OCCUPATION:	JAVA Developer and a freelance cryptocurrency and Forex trader
JOB SPECIALIZATION:	Developing software and other computing jobs
WORKING EXPERIENCE:	6 years

Validator

NAME:	Kamarulzaman Abdullah Sani
ORGANIZATION:	Kenanga Bank
OCCUPATION:	Remisier
JOB SPECIALIZATION:	Trader and investor
WORKING EXPERIENCE:	55 years

APPENDIX 2

Codebook for Annotators

Annotators must annotate individually by not discussing with the other annotators until the discussion session to finalize the sentiment scoring. There are four sources of data (four different spreadsheet) to be annotate which are English news, Malay news, English tweets, and Malay tweets regarding cryptocurrency matters.

The range of scores will be between -1 to 1, where -1 is considered a very negative sentiment, 0 represents neutral where no sentiment involved in the text, and 1 represents a very positive sentiment. Note that the score must focus on the sentiment toward the issue regarding BTC or ETH (depending on the corpus). The key table shows the guide to annotate the document with sentiment score and the guidance is used for all corpora.

Key Table for Sentiment Score Annotation

Sentiment Score	Meaning
1	This score indicates the maximum positive sentiment for the statement involving top billionaires such as Elon Musk or Bill Gates has bought or invested in Bitcoin or Ethereum, or any positive actions by the president of a country towards the cryptocurrency.
0.9	This score indicates the statement where any top billionaires such as Elon musk or Bill Gates only announce a positive statement regarding Bitcoin or Ethereum and not stating any form of investment actions.
0.8	This score indicates any positive activities such as investments on Bitcoin or Ethereum by other popular figures such as Michael Saylor or other crypto whales, or positive statements by ex-president of a country towards the cryptocurrency.
0.7	This score indicates any positive matters involving banks of a country.
0.6	This score indicates any price of the crypto market going up more than usual such as containing the words “soar”, “high”, or any equivalent meaning for both in English and Malay language.
0.5	This score indicates any statement of showing the price is going up, or any standard positive actions by the public that can potentially cause the price to increase.
0.4	This score indicates more to an individual’s (normal citizens) very positive events or statements towards cryptocurrency. For example, obtaining a good profit after investing into Bitcoin or Ethereum.
0.3	This score indicates a launching of a certain positive events by the cryptocurrency

	community or an individual who has just bought or invest in Bitcoin or Ethereum.
0.2	This score indicates any positive events of matters regarding NFTs or an individual positive opinion towards a cryptocurrency that do not quite affect the cryptocurrency prices.
0.1	This score indicates a slightly positive events on Bitcoin or Ethereum that do not really affect the cryptocurrency price and almost neutral.
0	The sentiment score that shows NEUTRAL statement (without having any sentiment) in the text.
-0.1	This score indicates a slightly negative events on Bitcoin or Ethereum that do not really affect the cryptocurrency price and almost neutral.
-0.2	This score indicates any negative events of matters regarding NFTs or an individual negative opinion towards a cryptocurrency that do not quite affect the cryptocurrency prices.
-0.3	This score indicates a launching of a certain negative events by the cryptocurrency community or an individual who has just sold their cryptocurrency asset.
-0.4	This score indicates more to an individual's (normal citizens) very negative events or statements towards cryptocurrency. For example, obtaining a loss after investing into Bitcoin or Ethereum or perhaps using some swearing words towards the cryptocurrency expressing anger and disappointment.
-0.5	This score indicates any statement of showing the price is going down, or any standard negative actions by the public that can potentially cause the price to drop.
-0.6	This score indicates any price of the crypto market going down more than usual such as containing the words "plunge", "low", or any equivalent meaning for both in English and Malay language.
-0.7	This score indicates any negative matters involving banks of a country such as restricting the cryptocurrency activities.
-0.8	This score indicates any negative activities such as pulling out of investments on Bitcoin or Ethereum by other popular figures such as Michael Saylor or other crypto whales, or negative statements by a country towards the cryptocurrency such as the banning of cryptocurrency tradings.
-0.9	This score indicates the statement where any top billionaires such as Elon musk or Bill Gates only announce a negative statement regarding Bitcoin or Ethereum and not stating any form of trading actions.
-1	This score indicates the maximum negative sentiment for the statement involving top billionaires such as Elon Musk or Bill Gates has sold their Bitcoin or Ethereum asset, or any negative actions by the president of a country towards the cryptocurrency.

1. Annotating Sentiment based on the News Headline

For the first task, annotation will be performed for English and Malay news. In these two sources, each row represents a document. The columns are labelled with “Title” and “Sentiment”. Each column represents as follows:

Title: The headline of the news (the text in which annotators refer to in order to score the sentiment).
Sentiment: The scoring of the sentiment of range between -1 to 1 (only 1 decimal place allowed).

Annotators are required to score the sentiment for each document based on the text in the “Title” column and state the sentiment score in the “Sentiment” column. Some examples of scoring the sentiment of the text for each of the document is shown below as a guide to perform the annotation for both English and Malay news.

Example of Annotating Sentiment for English News Headline

Example	Crypto	Sentiment Score
Tesla invests about $1.50 billion in bitcoin - Reuters.com	BTC	1
Elon Musk says bitcoin is slightly better than holding cash - Reuters	BTC	0.9
Michael Saylor has invested over $1 billion of MicroStrategy's funds in Bitcoin. The software CEO-turned Bitcoin whale explains why he is making such a massive bet on the digital asset.	BTC	0.8
Former Canadian PM: Bitcoin and CBDC could replace fiat in the future	BTC	0.7
XRP price soars 55% to 'crucial' level as Bitcoin notches new high at $38.5K	BTC	0.6
Bitcoin, ether hit fresh highs - Reuters India	BTC, ETH	0.5
Gemini is launching a credit card with bitcoin rewards	BTC	0.3
Rare alien CryptoPunk digital collectible sells for 605 ETH	ETH	0.2
Ethereum targets $1K after ETH, altcoins rally versus Bitcoin	ETH, BTC	0
Ethereum, Litecoin, and Ripple s XRP Daily Tech Analysis January 1st, 2021	ETH	0
Litecoin Price Prediction: Following Bitcoin, both for the good and bad	BTC	0

Lost Passwords Lock Millionaires Out of Their Bitcoin Fortunes	BTC	-0.1
Bitcoin sits below all-time high after U.S. ETF debut - Reuters	BTC	-0.2
Invesco reportedly refrains from plans to launch Bitcoin Futures ETF	BTC	-0.4
Ethereum Is Falling Behind Solana, Cardano and Polygon	ETH	-0.5
JPMorgan CEO Says Bitcoin Has No Intrinsic Value After Claiming That Its Price Could Rise 10X	BTC	-0.6
Chinese local government auctions seized bitcoin mining machines	BTC	-0.8
Elon Musk breaking up with Bitcoin? Cryptocurrencies slide after cryptic tweet - CNET	BTC	-0.9
Report: Tesla sells $272 million of its bitcoin holdings	BTC	-1

Example of Annotating Sentiment for Malay News Headline

Example	Crypto	Sentiment Score
Elon Musk Usik Bitcoin Sampai Cecah Lebih $37,000!	BTC	1
Saya Peminat Bitcoin Elon Musk	BTC	0.9
Bekas PM Kanada Ramalkan Bitcoin, CBDC Gantikan Dolar AS	BTC	0.8
Tinjaun Bank of America Ini Berjaya Pengaruhi Pelabur Bitcoin!	BTC	0.7
Litecoin Atasi XRP, #3 Terbesar Selepas Bitcoin Dan Ethereum	BTC, ETH	0.5
Grayscale Lancarkan Valuing Ethereum	ETH	0.3
Terlupa Kunci Wallet Kripto? Ini Cadangan Pengasas Ethereum!	ETH	0.1
Niaga AWANI: Prestasi Bitcoin dan komen Gates, Yellen	BTC	0
Polis serbu premis haram 'melombong' bitcoin di Serdang	BTC	-0.1
Terlupa Kata Laluan Wallet Bitcoin, Lelaki Ini Bakal Kehilangan Lebih $260 Juta!	BTC	-0.1
Pelabur Institusi Bitcoin Dalam Keadaan Berjaga-Jaga!	BTC	-0.2
Bitcoin SV (BSV) Hampir Musnah, Diserang Serangan 51%!	BTC	-0.4
Pelabur Bitcoin Semakin Takut?	BTC	-0.5
Bitcoin: Pelabur Lakukan Panic Selling Selepas Amaran Lembaga Kawalan Kewangan Dikeluarkan?	BTC	-0.6
Microsoft Tidak Berminat Dengan Pelaburan Bitcoin	BTC	-0.7
Tak Macam Negara Lain, India Bakal Sekat Bitcoin, Ethereum!	BTC, ETH	-0.8
Tergempar! Syarikat MicroStrategy Inc Beli Lebih 19,000 Bitcoin Pada Hari Ini!	BTC	-0.9
Nilai Bitcoin | Tesla tidak terima Bitcoin kerana kos alam sekitar	BTC	-1

2. Annotating Sentiment based on the Tweets contents

For the second task, annotation will be performed for English and Malay tweets. In these two sources, each row represents a document. The columns are labelled with "Tweets", and "Sentiment". Each column represents as follows:

Tweets: The tweets contents posted by the users (the text in which annotators refer to in order to score the sentiment). **Sentiment:** The scoring of the sentiment of range between -1 to 1 (only 1 decimal place allowed).

Annotators are required to score the sentiment for each document based on the text in the "Tweets" column and state the sentiment score in the "Sentiment" column. Hence, some examples of scoring the sentiment of the text for each of the document is shown below as a guide to perform the annotation for both English and Malay tweets.

Example of Annotating Sentiment for English Tweets

Example	Sentiment Score
equity monday tesla buys bitcoin nexthink raises and bumble	1
tesla investing in btc the rise of doge are we about to devalue federal reserve notes either this musk be a joke or its elon time coming	0.9
Ethereum Whales are on a Spending Spree as Addresses with at least 10,000 ETH Hit All-TimeHigh	0.8
survey shows of russian investors prefer bitcoin to gold ctel	0.7
BITCOIN IS THE FUTURE !	0.4
@StackerSatoshi What do you expect for ETH?	0
he tried to teach me how to buy ethereum and i got so stressed out and cried a lil bit my brain is too small to understand imaginary internet money	-0.1
selling some more eth	-0.3
I think btc looks shyte	-0.4
Crypto Markets Suffer Heavy Losses, Bitcoin Price Sinks More Than 25% in 24Hours	-0.6
big news as china bans cryptocurrency bitcoin btc ethereum eth and others are no longer allowed in china what does this mean for the hype surrounding crypto well give you our thoughts in this video	-0.7
vitalikbuterin ethereum the reason why elonmusk will not involved himself in eth is	-0.8

because of high gas fee ethereum gas fee is tooo much	
elon musk has confirmed that tesla has not sold any bitcoin despite speculation over the weekend that the company may have done so or will be soon cryptocurrency	-0.9
The ripples have violent effects and the landscape becomes chaos. What kind of peace is that? There are no words to describe the pain of being your secret. To love in whispers, when our love is fire. A storm. A landslide. A tornado. Catastrophe demanding to be recognized. 5/6	N/A

Example of Annotating Sentiment for Malay Tweets

Example	Sentiment Score
Lepas DOGE, Elon Musk pilih Bitcoin! Agak-agak boleh naik lagi tinggi tak? Baca di sini:	1
Elon musk tukar status, letak logo bitcoin..mencanak harge bitcoin naik beribu2...have mercyyyyy	0.9
Wah..wah..Rothschild pun berminat dengan Bitcoin? Baca di sini:	0.8
Gila la BTC dah tembus paras tertinggi sepanjang masa, melepasi $30,000	0.7
BTC naik 1% sejak 24H yang lalu Rp404.895.000 - Rp410.000.000	0.5
tak sia sia peram bitcoin 8 tahun dari harga 20 ringgit, tapau 200 bijik. skarang aku dah pencen. hahahaha	0.4
BTC $BTC TIME IS COMING.. TIK TIK	0.3
harap terus menuju ke wahai ethereum ku sayang	0.2
desentralisasi sudah ada sejak dulu sejak emas ada, sekarang era tersebut hanya dibangkitkan kembali oleh bitcoin dan oleh blockchain.	0.1
Bitcoin tu apa?	0
BTC KO GILA?	-0.2
aku rasa btc ni turun lagi ni	-0.3
Baru nak untung BTC dah LHDN nak kacau	-0.4
Orang banyak dah dump bitcoin sebab ni antara highest peak dia kot early 21.	-0.6
Kejatuhan 2017 menyebabkan BTC seakan kehilangan momentum selama 1 tahun. Baca disini:	-0.7
bill gates tidak berminat melabur bitcoin	-0.8
Pelabur institusi 'berehat' seketika dari membuat sebarang pembelian Bitcoin. Baca di sini:	-0.9
elon jual bitcoin nak beli safemoon	-1

APPENDIX 3

Example usage of the codebook

EXAMPLE STATEMENT: "Bitcoin rallies above $30,000 for first time – Reuters"

POSITIVE SENTIMENT RUBRIC

CATEGORY	STATEMENT	SCORE	CHECK	TOTAL SCORE
Influential person	Crypto influencer who made a statement/action			
	Has more than 1 million followers (Mega-influencer)	0.4		
	Has between 100,000 to 1 million followers (Macro-influencer)	0.3		
	Has between 10,000 to 100,000 followers (Micro-influencer)	0.2		
	Has below than 10,000 followers (Nano-influencer)	0.1		
	Billionaire	0.3		
	Other non-crypto influencer or authority who made a statement/action			
	Leader/government	0.4		
	International celebrity (well-known)	0.3		
	Local celebrity	0.2		
	High post in a company	0.1		
	Action			
	Buy, purchase, invest, hold or HODL	0.2		
	Launching	0.1		
	Planning	0.1		
	Opinion or feelings			
	Extreme positive	0.2		
	Moderate positive	0.1		
				0
Country	Country status based on GDP per capita[18]			
	High-income (developed) country	0.4		
	Newly emerging economies (emerging) country	0.3		
	Low-income (developing) country	0.2		
	Action			
	Buy, purchase, invest, hold or HODL	0.2		
	Launching	0.1		
	Planning	0.1		
	Opinion			
	Extreme positive	0.2		
	Moderate positive	0.1		
				0
Organization	Bank			
	Top world bank (developed country – high market cap)	0.3		
	Local bank (from other countries)	0.2		
	Company			
	International (well-known – Top 100 based on market cap)	0.4		
	Local	0.2		
	Action			
	Buy, purchase, invest, hold or HODL	0.2		
	Launching	0.1		
	Planning	0.1		
	Opinion			
	Extreme positive	0.2		
	Moderate positive	0.1		
				0
Event/News	Positive announcement (gaining profits or anything positive)	0.3		

[18] https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD

	Promote crypto trading/investment	0.2		
	Planning to launch a promotional event	0.1		
	Free giveaways during sign-up in a crypto coin investment	0.1		
	Price is going up			
	A stronger emphasis on price going up (e.g., "to the moon", "rocket", etc)	0.4	/	
	General statement of price going up (with no strong emphasis)	0.3		
	Pumping	0.2		
	Stating price value	0.1	/	
				0.5
			FINAL SCORE (positive)	0.5

NEGATIVE SENTIMENT RUBRIC

CATEGORY	STATEMENT	SCORE	CHECK	TOTAL SCORE
Influential person	Crypto influencer who made a statement/action			
	Has more than 1 million followers (Mega-influencer)	-0.4		
	Has between 100,000 to 1 million followers (Macro-influencer)	-0.3		
	Has between 10,000 to 100,000 followers (Micro-influencer)	-0.2		
	Has below than 10,000 followers (Nano-influencer)	-0.1		
	Billionaire	-0.3		
	Other non-crypto influencer or authority who made a statement/action			
	Leader/government	-0.4		
	International celebrity (well-known)	-0.3		
	Local celebrity	-0.2		
	High post in a company	-0.1		
	Action			
	Sell crypto asset	-0.2		
	Launching	-0.1		
	Planning	-0.1		
	Opinion or feelings			
	Extreme negative	-0.2		
	Moderate negative	-0.1		
				0
Country	Country status based on GDP per capita[19]			
	High-income (developed) country	-0.4		
	Newly emerging economies (emerging) country	-0.3		
	Low-income (developing) country	-0.2		
	Action			
	Against cryptocurrency, boycott, or banning	-0.3		
	Planning	-0.1		
	Opinion			
	Extreme negative	0.2		
	Moderate negative	0.1		
				0
Organization	Bank			
	Top world bank (developed country – high market cap)	-0.3		
	Local bank (from other countries)	-0.2		
	Company			
	International (well-known – Top 100 based on market cap)	-0.4		
	Local	-0.2		
	Action			
	Against cryptocurrency, boycott, or banning	-0.3		
	Sell crypto asset	-0.2		
	Planning	-0.1		
	Opinion			
	Extreme negative	-0.2		

[19] https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD

	Moderate negative	-0.1		
				0
Event/News	Negative announcement (having loss or losing something)	-0.3		
	Against cryptocurrency (boycotting or banning	-0.2		
	Planning to boycott or ban cryptocurrency	-0.1		
	Price is going down			
	A stronger emphasis on price going down (e.g., “slide”, “rock bottom”, etc)	-0.4		
	General statement of price going down (with no strong emphasis)	-0.3		
	Dumping	-0.2		
	Stating price value	-0.1		
				0
			FINAL SCORE (negative)	0

APPENDIX 4

Operational Definition

	Definition
Open Price	Open price for cryptocurrency price data is the price taken at 12.00am daily since cryptocurrency price is operating 24 hours.
High Price	High price for cryptocurrency price data is the highest price reached for the current day between 12.00am until 11.59am.
Low Price	Low price for cryptocurrency price data is the lowest price reached for the current day between 12.00am until 11.59am.
Close Price	Close price for cryptocurrency price data is the price taken at 11.59pm daily since cryptocurrency price is operating 24 hours.
Trading Volume	Trading volume is the total number of trading made for the current day between 12.00am until 11.59am.
Sentiment Score	The weighted sentiment based on post-level of news or tweets in continuous value for each day.
Sentiment Positive-Negative Count	The total number of positive/negative sentiment based on the weight, where sentiment score > 0 is positive and sentiment score < 0 is negative, for each day.

AUTHOR'S PROFILE



Nur Azmina Mohamad Zamani obtained Diploma in Computer Science in 2012 and BSc in Computer Science (Hons.) in 2014 from Universiti Teknologi MARA, Shah Alam. Then she pursued her MSC Advanced Computer Science (2016) from The University of Sheffield, United Kingdom. Currently she is working on English and Malay sentiment-based cryptocurrency dataset, involving deep learning methods for her current PhD study. She is also a computer science lecturer at Universiti Teknologi MARA Perak Branch, Tapah Campus, teaching programming and algorithm subjects, particularly JAVA for Diploma and Degree students.

LIST OF PUBLICATIONS:

Mohamad Zamani, N. A., Kamaruddin, N, and Yusof, A. M. Dataset on sentiment-based cryptocurrency-related news and tweets in English and Malay language. Language Resources and Evaluation, Vol.59, 2025:p. 807-842.

Zamani, N. A. M., and Kamaruddin, N. Crypto-sentiment Detection in Malay Text Using Language Models with an Attention Mechanism. Journal of Information Systems Engineering and Business Intelligence, Vol. 9, No. 2, 2023: p. 147-160.

Zamani, N. A. M., Yan, J. L. S., and Yusof, A. M. Cryptocurrency Price Prediction using Bi-GRU Model with English and Malay News Sentiment Features. In 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2022: p.136-141. IEEE.

Zamani, N. A. M., Liew, J. S. Y., & Yusof, A. M. XLNET-GRU Sentiment Regression Model for Cryptocurrency News in English and Malay. In Proceedings of the 4th Financial Narrative Processing Workshop@ LREC2022, 2022:p. 36-42.