UNIVERSITI TEKNOLOGI MARA

HYBRIDISING DESCRIPTIVE-SEMANTIC TECHNIQUES FOR SENTIMENT ANALYSIS OF MALAYSIA'S ISLANDS VIA YOUTUBE REVIEW

NURAIN BATRISYIA BINTI BIDIN

MSc

February 2026

UNIVERSITI TEKNOLOGI MARA

HYBRIDISING DESCRIPTIVE-SEMANTIC TECHNIQUES FOR SENTIMENT ANALYSIS OF MALAYSIA'S ISLANDS VIA YOUTUBE REVIEW

NURAIN BATRISYIA BINTI BIDIN

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

Faculty of Computer and Mathematical Sciences

February 2026

CONFIRMATION BY PANEL OF EXAMINERS

I certify that a Panel of Examiners met in 4 December 2025 to conduct the final examination of Nurain Batrisyia Binti Bidin on her Master of Science thesis entitled "Hybridising Descriptive-Semantic Techniques for Sentiment Analysis of Malaysia's Islands via YouTube Review" in accordance with Universiti Teknologi MARA Act 1976 (Akta 173). The Panel of Examiners recommends that the student be awarded the relevant degree. The Panel of Examiners was as follows:

Zamali Tarmudi, PhD
Associate Professor
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
(Chairperson)

Sharifalillah Binti Nordin, PhD
Senior Lecturer
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
(Internal Examiner)

Azlan Mohd Zain, PhD
Professor
Faculty of Computing
Universiti Teknologi Malaysia
(External Examiner)

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

Date: 26 February 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	Nurain Batrisyia Binti Bidin
Student ID. No.	2023432318
Programme	Master of Science (Computer Science) - CDCS750
Faculty	Computer and Mathematical Sciences
Thesis Title	Hybridising Descriptive-Semantic Techniques for Sentiment Analysis of Malaysia's Islands via YouTube Review

Signature of Student . . . . . .

Date February 2026

ABSTRACT

Malaysia's island tourism continues to grow, as seen by the extensive reviews of destinations such as Langkawi, Perhentian, and Pangkor on platforms like YouTube. YouTube was selected as the primary data source for this study due to its extensive repository of user-generated video reviews and comments, offering valuable insights into tourist perspectives and satisfaction for sentiment analysis (SA). However, most existing SA research in the tourism sector implements lexicon-based or machine learning (ML) techniques, which frequently overlook the variety of contextual meanings encountered in text reviews, such as sarcasm, local idioms, or mixed-language expressions. This research addressed three main challenges in tourism-related SA. These include the linguistic complexity of online reviews, limited capacity to detect cultural and contextual subtleties, and the lack of labelled data for supervised learning. In order to bridge this gap, this research developed a hybrid SA approach that integrates Descriptive-Semantic Analysis (DSA) techniques with both lexicon-based and ML approaches involving the Valence Aware Dictionary and Sentiment Reasoner and Support Vector Machine. For DSA implementation, significant words were identified using the Term Frequency-Inverse Document Frequency, while dominating topics and thematic insights were extracted using Latent Dirichlet Allocation. The island reviews were extracted using the YouTube Data API, followed by pre-processing steps such as translation, cleaning and normalisation. Findings indicate that the hybrid SA approach achieved accuracy rates of 98.15% for Langkawi, 98.5% for Perhentian, and *91.19%* for Pangkor. Moreover, the model validation using the Area Under the Curve metric showed strong performance. Perhentian had a validation accuracy of 100% and a testing accuracy of 98.84%. Langkawi followed with 99.36%) validation and 98.72%) testing, while Pangkor recorded *95.11%* validation and 97.22%> testing accuracy. Overall, this approach successfully bridged the gap between descriptive and semantic insights, overcoming the limitations of standalone S A techniques, such as struggles to accurately interpret user content and relying on basic techniques that overlook deeper meanings. Future research anticipates broadening the data sources by incorporating platforms such as TripAdvisor, Google Reviews, Facebook, and X (formerly Twitter) to enhance accessibility and diversify sentiment representation. Additionally, subsequent studies could explore deep learning models such as Bidirectional Encoder Representations from Transformers and Long Short-Term Memory to further improve sentiment classification by capturing deeper contextual relationships.

ACKNOWLEDGEMENT

First and foremost, I am grateful to Allah SWT for His blessings, assistance, and endurance throughout the entire process of completing my thesis. A debt of gratitude reaches out to my supervisor, Prof. Madya Ts. Dr. Khyrina Airin Fariza Binti Abu Samah, for her unwavering encouragement, support, and leadership during my research. Her insightful criticism, competent guidance, and patience have proven extremely valuable to the development and accomplishment of this thesis.

My sincere gratitude is also extended to Dr. Raihah Binti Aminuddin, my co-supervisor, for her committed assistance and precise supervision in clarifying unclear concepts, which indirectly facilitated the progression of my research. Her contributions have greatly improved this thesis's overall quality. On the other hand, a special thanks goes out to all of my review panels for the Research Proposal Defence, Mock Viva, and Viva Voce. This thesis became better and was transformed into a higher quality, thanks to the insightful suggestions and feedback.

I would also like to express my gratitude to all of the experts and presenters from all of the webinars I attended over this master journey. Besides, I am also grateful to the Institute of Graduate Studies, Universiti Teknologi MARA, for the Conference Support Fund and Journal, which have been crucial in supporting the continuous development and dissemination of my research. Lastly, I am incredibly thankful to have had such a supportive group of people around me, whose encouragement and faith in my ability to succeed encouraged me to have the determination and persistence to successfully complete this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Symbols

a	Alpha
P	Beta
0	Theta
q>	Phi
X	Total number of word occurrences
|D|	Total number of documents in the corpus
G	Element for documents
log	Logarithm

LIST OF ABBREVIATIONS

Abbreviations

ABSA	Aspect-Based Sentiment Analysis
AUC	Area Under the Curve
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
Bi-LSTM	Bidirectional Long Short-Term Memory
BoW	Bag-of-Words
DL	Deep Learning
DSA	Descriptive-Semantic Analysis
DT	Decision Tree
DV	Data Visualisation
EDA	Exploratory Data Analysis
GDP	Gross Domestic Product
HTML	Hypertext Markup Language
KNN	K-Nearest Neighbours
LDA	Latent Dirichlet Allocation
LR	Logistic Regression
LSA	Latent Semantic Analysis
LSTM	Long Short-Term Memory
ML	Machine Learning

NB	Naïve Bayes
NER	Named-Entity Recognition
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
POS	Part-of-speech
RNN	Recurrent Neural Network
RO	Research Objectives
ROC	Receiver Operating Characteristic
RP	Research Problems
RQ	Research Questions
SA	Sentiment Analysis
SS	Statistical Summaries
SVM	Support Vector Machine
TA	Trend Analysis
TF-IDF	Term Frequency-Inverse Document Frequency
TS	Time Series
t-SNE	t-Distributed Stochastic Neighbour Embedding
URL	Uniform Resource Locators
VADER	Valence Aware Dictionary and Sentiment Reasoner

CHAPTER 1

INTRODUCTION

This chapter examines the research background, problem statements, objectives, scopes, and significance of the research entitled "Hybridising Descriptive-Semantic Techniques for Sentiment Analysis of Malaysia's Islands via YouTube Review".

1.1 Research Background

Tourism is important for Malaysia's economy as it contributes extensively to Gross Domestic Product (GDP) and generates substantial income from both domestic and international tourists. Prior to the COVID-19 pandemic, Malaysia recorded 26.1 million foreign tourists in 2019, generating RM86.1 billion in revenue (Mahavera, 2023). However, the pandemic severely disrupted the sector, reducing tourism's contribution to GDP from 6.7% in 2019 to 2% in 2020 (Statista, 2023) due to global travel restrictions and domestic Movement Control Orders. Despite encountering various challenges, Malaysia's tourism industry has exhibited a remarkable recovery, particularly in domestic tourism. This resurgence has been supported by government initiatives, health and safety protocols, and targeted marketing strategies (Chan, 2021). Recent reports from Salim and Tan (2023) indicate that the sector has exceeded its pre-pandemic performance levels in 2024, in line with national campaigns such as Visit Malaysia Year 2026.

Island tourism constitutes a significant component of Malaysia's tourism sector, offering breathtaking scenery, diverse marine life, adventure activities, cultural heritage, and a variety of recreational opportunities (Ostheimer, 2019). Moreover, the travel decisions of tourists are increasingly influenced by information sourced from online platforms, particularly social media. User-generated content, such as reviews, images, and videos, offers valuable experiential insights that significantly impact tourists' perceptions and their decision-making processes (Morision et al., 2023). Although online content offers abundant information, it often lacks accuracy and may be biased due to commercial promotions or personal exaggeration, leading to the creation or dissemination of false information (Sadiq & Saji, 2022). Consequently, most

of the information on social media has become outdated due to changing travel restrictions or is misused as misinformation, especially during the pandemic (Balis, 2021). Therefore, there is an increasing necessity for systematic methodologies that can efficiently analyse and interpret substantial volumes of online tourism-related content.

In this context, social media platforms such as YouTube represent a significant repository of tourism-related data, where users disseminate their experiences through video reviews in conjunction with textual commentary. YouTube serves a dual purpose, functioning both as a communication platform and as a digital gateway that enables prospective tourists to virtually explore destinations and formulate expectations prior to their trips (Sahu, 2020). Previous studies underscore the significance of analysing YouTube reviews as a means of extracting sentiment-rich expressions that reflect users' satisfaction, preferences, and perceptions regarding various travel destinations (Alhujaili & Yafooz, 2021; Inam et al., 2020). By leveraging this data, stakeholders in the tourism sector can obtain real-time insights into user experiences, thereby facilitating the optimisation of promotional strategies.

On top of that, sentiment analysis (SA) is a subfield of natural language processing (NLP) and machine learning (ML) widely used to identify opinions and emotional polarity within textual data (Dutta, 2021). In tourism, SA has been utilised to analyse user-generated content for the purpose of understanding trends in tourist sentiment, assessing the image of destinations, and aiding in strategic decision-making. A study from Kavitha et al. (2020) showed strong links between sentiment trends and actual tourism occurrences, highlighting the necessity for developing systematic and dependable SA frameworks. Thus, utilising SA on YouTube reviews of Malaysia's islands provides important insights into tourists' perceptions and expectations.

However, conducting SA in the context of Malaysia's islands presents challenges such as data accessibility, linguistic diversity, and cultural nuances. Social media platforms enable users to express opinions in multiple languages, but due to resource constraints, many SA models are built for a single language, which can lead to overlooking critical information (Abdullah & Rush, 2021). Additionally, the way sentiments are expressed can differ greatly across cultural contexts, which may restrict the effectiveness of models that rely on data specific to a certain domain or language (Tan, Lee, & Lim, 2023). These challenges underscore the necessity for additional analytical techniques that can improve sentiment understanding beyond basic polarity classification.

To address these challenges, this study implements a hybrid SA approach that integrates descriptive-semantic analysis (DSA) techniques to analyse YouTube reviews pertaining to island tourism in Malaysia. The hybrid SA framework combines a lexicon-based SA tool, namely the Valence Aware Dictionary and Sentiment Reasoner (VADER), with a ML classifier known as Support Vector Machine (SVM), thereby enhancing the accuracy of sentiment classification. For feature extraction, significant textual features are identified through descriptive analysis using Term Frequency-Inverse Document Frequency (TF-IDF), while semantic analysis using Latent Dirichlet Allocation (LDA) uncovers underlying thematic structures within the reviews. The performance of the model is then evaluated using performance metrics and Area Under the Curve (AUC) for validation to compare model capabilities across categories (Fazrin et al., 2022). As noted by Mandal and Bhardwaj (2023), a hybrid approach effectively manages high-dimensional data, reduces overfitting, and enhances generalisation to new data. Hence, the integration of DSA methodologies with a hybrid SA approach enhances the analytical depth of this research.

1.2 Problem Statements

SA is widely used to evaluate user-generated content such as opinions, reviews, and emotional expressions, including video-sharing websites as articulated on social media platforms, such as YouTube (Akhtar, 2019). Analysing sentiments from YouTube reviews offers valuable feedback on user perceptions and preferences, but it presents several challenges. One major issue arises from the enormous number of reviews written in informal language, such as casual phrases, slang reflecting cultural trends, and jargon, which complicate accurate sentiment classification (Liaqat & Rafique, 2024; Janechek, 2023). The presence of informal language forms contributes to linguistic ambiguity, as the meanings of words and phrases are often heavily influenced by context. This reliance on contextual interpretation can lead to misinterpretations and misclassifications, especially when dealing with unstructured or user-generated content (Abeysiriwardana & Sumanathilaka, 2024).

In the context of Malaysia, the interpretation of sentiment is significantly complicated by cultural and multilingual diversity. YouTube reviews frequently incorporate code-mixing between English and Malay, along with the use of local idioms and expressions that are culturally specific. These variations produce diverse

representations of sentiment, posing challenges for SA models that are tailored to specific languages or cultural contexts (Chen, Lin, Gong, & Tan, 2023; Joshi et al., 2024). In Malaysia's multifaceted communication environment, the inability of models to adapt culturally presents challenges for accurately interpreting sentiment, thereby compromising the reliability and effectiveness of SA results.

Another significant challenge in current SA research is the lack of labelled data, particularly in low-resource languages such as Malay. Labelled datasets are essential in training supervised ML models; nonetheless, the limited availability of annotated multilingual tourism data constrains the models' capacity to learn nuanced sentiment patterns and generalise efficiently (Motlagh et al., 2023). This issue results in biased models, reduced sentiment classification accuracy, and difficulties in evaluating model performance (Kong et al., 2023). Consequently, the performance of conventional ML-based SA models is limited, particularly when it comes to understanding sentiments conveyed informally and in various languages.

Recent studies highlight the potential of DSA techniques to address the challenges of ambiguous and unstructured language. Shah and Bhuvana (2023) demonstrated that NLP can effectively evaluate students' short responses, enhancing semantic similarity detection and reducing ambiguity while improving clarity. Similarly, Amalia et al. (2023) addressed the challenge of processing large COVID-19 news data by using descriptive analysis to categorise sentiments and reduce noise, thus improving sentiment labelling quality. However, the integration of DSA techniques within SA frameworks for the evaluation of YouTube reviews in the context of Malaysia's island tourism remains limited. Hence, to address these challenges, this study combines DSA techniques with a hybrid SA approach that incorporates both lexicon-based and ML techniques to enhance classification accuracy, cultural adaptability, and contextual relevance in sentiment detection. A summary of the research problem (RP) statement is provided in Table 1.1.

Table 1.1

Summary of Research Problem Statement

RP	Research Problem (RP)
RP 1	The presence of informal language forms contributes to linguistic ambiguity, as the meanings of words and phrases are often heavily influenced by context. This reliance on

RP	Research Problem (RP)
	contextual interpretation can lead to misinterpretations and misclassifications, especially when dealing with unstructured or user-generated content.
RP 2	The interpretation of sentiment is significantly complicated by cultural and multilingual diversity. YouTube reviews frequently incorporate code-mixing between English and Malay, along with the use of local idioms and expressions that are culturally specific.
RP3	The lack of labelled data, particularly in low-resource languages such as Malay. Labelled datasets are essential in training supervised ML models; nonetheless, the limited availability of annotated multilingual tourism data constrains the models' capacity to learn nuanced sentiment patterns and generalise efficiently.

1.3 Research Questions

This research aims to enhance the accuracy of sentiment classification for YouTube reviews by integrating DSA with a hybrid SA approach. In this study, SA outcomes were categorised into predefined sentiment classes following text summarisation and analysis. Drawing from the previously identified RP, this research formulated the following research questions (RQ):

i. What are the existing descriptive and semantic techniques applied in SA within the context of Malaysia's island tourism?
ii. How can a hybrid approach that integrates DSA improve the effectiveness of SA applied to YouTube reviews on Malaysia's island tourism?
iii. How can the accuracy of the hybrid SA approach integrating DSA be measured?

1.4 Research Objectives

The section outlines the research objectives (RO), which were formulated based on the corresponding RPs and RQs. This research aims to develop DSA techniques and a hybrid SA approach for YouTube reviews to enhance the accuracy of sentiment classification. The following three RO have been established:

i. To analyse the existing descriptive and semantic techniques used in analysing the SA related to Malaysia's island tourism.

ii. To construct a hybrid approach integrating DSA in improving the SA of YouTube reviews about Malaysia's island tourism,

iii. To evaluate the accuracy of the constructed hybrid SA approach using performance metrics and validate using the AUC metric.

1.5 Research Motivation

Malaysia's island destinations depend significantly on their online presence and digital recommendations, making content created by users a vital influence on public perception. YouTube, in particular, has emerged as a key platform where users share genuine experiences, emotions, and expectations, providing valuable real-time insights that are beneficial for the development of tourism. Leveraging these insights offers a significant opportunity not just for tourism stakeholders but also for furthering research in SA, especially within the multilingual and culturally diverse context of Malaysia. In this context, the multilingual expressions and informal communication styles found in YouTube reviews provide an intriguing pathway to examine how users express their experiences naturally.

Within this context, the integration of multilingual expression and informal communication styles observed in YouTube reviews provides a significant opportunity to examine how individuals authentically articulate their experiences. Studies such as Abdullah and Rush (2021) emphasise the challenges of interpreting sentiments in texts using multiple languages, making the development of effective SA models crucial for capturing public perceptions in Malaysia's diverse linguistic landscape. Furthermore, tourism research further underscores the necessity of obtaining nuanced, aspect-level insights concerning factors such as accommodations, local experiences, and service interactions. Findings from Othman et al. (2024) identify several elements that impact tourist satisfaction in Malaysia, highlighting the importance of understanding sentiment across specific aspects of tourism. Thus, the implementation of DSA techniques offers a valuable approach to gaining insights by uncovering themes and contextual patterns that may be overlooked by conventional sentiment classification techniques.

In addition, classifying text data into positive and negative categories is a fundamental task in SA. A hybrid approach combines lexicon-based and ML models to enhance classification accuracy and feature selection by leveraging the strengths of both techniques while addressing overfitting and underfitting through model diversity.

Previous research by Mandal & Bhardwaj (2023) and Putra et al. (2021) demonstrates that hybrid approaches provide greater robustness, especially when handling complex sentiment expressions like sarcasm and culturally embedded nuances. These developments encourage further exploration of hybrid methodologies aimed at improving both the accuracy and interpretability of sentiment detection.

Figure 1.1 summarises the research motivation process. According to the standard methodology, users start by searching for information about Malaysia's islands on the YouTube platform. The next step involves extracting data from YouTube and analysing it using SA. This process consists of several stages, including data pre-processing, feature extraction, sentiment prediction, and model training. During the data pre-processing stage, translation, cleaning, and normalisation are performed to ensure that the textual content is consistent. Following this, DSA is used for feature extraction, while a hybrid SA approach is applied for sentiment prediction and model training. After these steps, results from the SA are generated based on the outcomes obtained from the DSA implementation and the hybrid SA approach. Finally, analyse the accuracy of the SA classification through the model evaluation and validation.

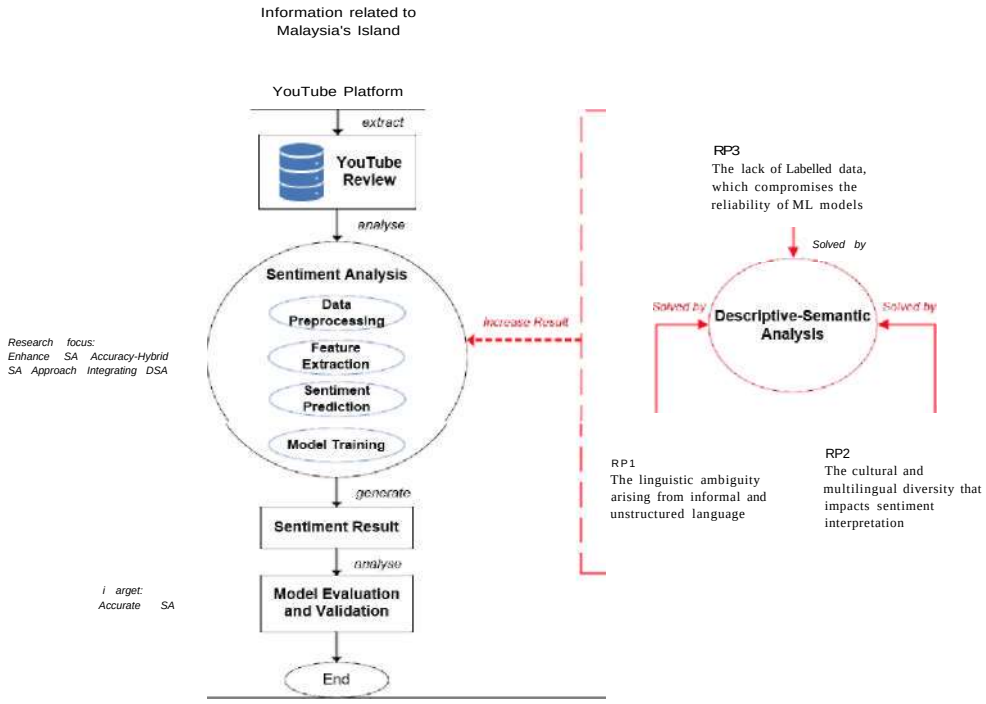


Figure 1.1 Overview of Research Motivation

Furthermore, as stated in the problem statement, this research addresses three key challenges such as 1) the linguistic ambiguity arising from informal and unstructured language; 2) the cultural and multilingual diversity that impacts sentiment interpretation; and 3) the lack of labelled data, which compromises the reliability of ML models. To overcome these challenges, this research integrates DSA techniques with a hybrid SA approach. As a result, it helps to manage the complexities of natural language like linguistic and cultural variations, as well as enhances the overall performance of sentiment classification despite limited labelled data. To the best of the researcher's knowledge, no existing study has combined DSA techniques with a hybrid S A approach specifically for Malaysia's island tourism using YouTube reviews. This integration contributes to the field by enhancing contextual understanding and classification accuracy within multilingual and unstructured review data.

1.6 Research Scopes

i. *Scope of Analysis and Approach:* This research focuses on DSA techniques integrated within a hybrid SA approach, which are applied to YouTube reviews concerning Malaysia's island tourism. SA is NLP method used to determine the emotional tone or polarity of textual content (Agrawal, 2020). Given the length of the textual data, aspect-level SA is preferred in this study, as it enables the division of text into specific aspects for more contextual analysis. In addition to identifying overall sentiment polarity, the research applies DSA techniques to uncover underlying themes and patterns, thereby enhancing the understanding of tourists' perceptions through both quantitative and qualitative insights,

ii. *Scope of Study Area and Data Source:* The empirical analysis is limited to YouTube reviews related to Langkawi, Perhentian and Pangkor islands, which are three well-known Malaysia's islands selected for their natural landscapes and tourism appeal (Raccuia, 2024). These islands were chosen to provide a range of tourism characteristics while keeping contextual relevance. The dataset includes reviews collected over a consistent timeframe, from 1 January 2023 to 31 December 2023,

ensuring both temporal consistency and comparability among the destinations.

iii. *Scope of Language and Data Pre-processing:* This research acknowledges Malaysia's multilingual context, where YouTube reviews frequently involve a combination of Malay and English. To ensure consistency during analysis, non-English reviews were translated into English using the Google Translate library during the data pre-processing stage. Although this approach enables standardisation, potential translation inaccuracies were taken into account when interpreting sentiment outcomes. Consequently, the focus of this research is limited to translated English text representations of the original multilingual reviews.

iv. *Scope of Analytical Techniques and Evaluation:* Within the DSA framework, TF-IDF is utilised to extract key descriptive features, while LDA identifies the underlying semantic topics. For sentiment classification, a hybrid approach combines the lexicon-based tool VADER with the SVM classifier. Model performance is evaluated using standard classification metrics, including AUC, to determine the effectiveness of the hybrid framework in analysing sentiment patterns across the selected island destinations.

1.7 Research Significances

This research integrates DSA techniques with a hybrid SA approach to analyse YouTube reviews related to Malaysia's island tourism. The DSA implementation using TF-IDF and LDA significantly improves the interpretability of sentiment trends, providing greater context for sentiment classification in multilingual tourism reviews. Additionally, the hybrid SA approach addresses limitations of standalone models such as context insensitivity and domain dependence, enabling a more nuanced analysis of complex sentiments. As a result, this research has made the following significances:

i. *Research:* This research is significant as it deepens the understanding of tourist experiences on Malaysia's island tourism. Hybrid SA captures overall emotional responses, while DSA delves deeper into the specific themes underlying those sentiments. This approach provides valuable

insights into tourist preferences and criticisms, highlighting key areas such as service quality and environmental issues,

ii. *Community:* DSA enables island authorities and content creators to better understand audience perceptions of their content and to respond more effectively. Positive reviews help strengthen stakeholder engagement, whereas negative reviews provide opportunities to address and resolve public concerns. Addressing the issues from negative feedback benefits both the community and users, while sharing positive experiences strengthens community ties.

iii. *Government:* DSA can be used in qualitative research on the environment of Malaysia's islands. It helps identify threats like pollution and habitat loss and encourages community involvement in conservation. Subsequently, DSA supports sustainable tourism, highlighting the importance of cultural heritage in these initiatives.

iv. *Environment:* DSA effectively monitors public reactions to environmental issues in Malaysia, such as pollution and climate change. Researchers can use data from platforms like YouTube to pinpoint urgent concerns and emphasise topics like recycling and air quality.

1.8 Thesis Outline

Chapter 1 introduces the research by outlining its background, problem statement, questions, objectives, motivation, scope, and significance. Chapter 2 presents a literature review that discusses Malaysia's island tourism, SA, descriptive analysis, and semantic analysis, along with insights gained to identify research gaps. Chapter 3 explains the six-phase research process design, with each phase including specific deliverables and corresponding results. Furthermore, Chapter 4 presents the results and discussion of the model, while Chapter 5 concludes the study, explains the research limitations and provides recommendations for improvement in further studies.

1.9 Summary

This chapter has presented the research background, problem statements, research questions, objectives, motivation, scope, significance, and an overview of the

thesis structure. To conclude, the key research problems in analysing sentiments from YouTube reviews include linguistic ambiguity caused by informal language, cultural and multilingual diversity, and the lack of labelled data. Accordingly, the objectives focus on analysing DSA techniques, constructing a hybrid approach, and rigorously evaluating its performance and validating it using the AUC metric. As a result, the research implements DSA techniques integrated with a hybrid SA approach to enhance text analysis. Subsequently, an extensive review of relevant research is presented in Chapter 2 to provide a solid theoretical basis for this research. In addition, prior research on SA in tourism contexts was reviewed to identify research gaps and justify the implementation of a hybrid SA approach.

CHAPTER 2

LITERATURE REVIEW

This chapter explores the research regarding sentiment analysis (SA) within the tourism sector from YouTube reviews. It highlights key methodologies, including lexicon-based approaches, machine learning (ML), deep learning (DL), aspect-based sentiment analysis (ABSA), and a hybrid approach. Furthermore, the chapter addresses descriptive-semantic analysis (DSA) as a complementary technique. It concludes by identifying research gaps and insights gained from the literature review.

2.1 Introduction

The analysis of sentiment within YouTube tourism reviews has become a notable area of research, particularly in understanding user perceptions and destination feedback, due to the platform's accessibility and the extensive volume of user-generated content available. These reviews provide authentic insights into tourist experiences, as they are unprompted and voluntarily posted, highlighting both positive attributes and critical concerns associated with various destinations. Moreover, in contrast to conventional survey data, YouTube reviews are spontaneous, allowing tourists to express their genuine emotions and experiences. This natural feedback provides valuable insights into emotional tone, service quality, and overall satisfaction, making it useful for SA in tourism research.

Given the importance of SA in understanding tourist behaviour, this chapter outlines the foundational topics of this research. Figure 2.1 provides an overview of the literature review for this research, illustrating the relationship between SA and DSA techniques. This chapter also consists of seven sections, including the introduction, Malaysia's island tourism, SA, descriptive analysis, semantic analysis, insights gained from the literature review, and a summary. Overall, these sections collectively outline the theoretical and methodological background necessary for supporting the research objectives.

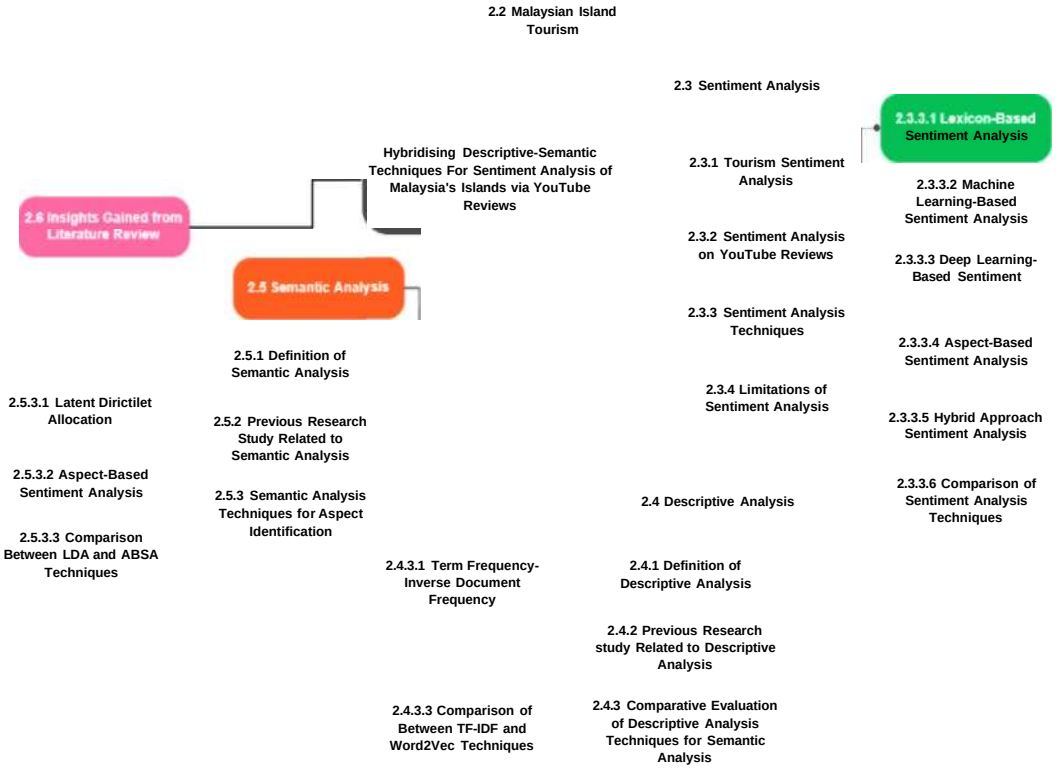


Figure 2.1 Overview of Literature Review

Chapter 2 examines the significant role of SA in the context of tourism and YouTube reviews, with particular emphasis on the implementation of DSA techniques and a hybrid approach within SA. Moreover, it examines YouTube as a valuable data source for user reviews and perceptions, aiming to provide deeper insights through diverse SA approaches. Despite SA's common application across platforms, research on DSA techniques is limited. Thus, this chapter seeks to explore SA and the sequential integration of DSA techniques and a hybrid approach comprehensively.

2.2 Malaysia's Island Tourism

Island tourism is essential for many coastal and island nations, contributing to local development through job creation, infrastructure improvements, and foreign exchange earnings. These destinations attract tourists with their natural beauty and cultural experiences. However, rapid growth poses challenges, including dependence on global markets, environmental threats, and socio-cultural disruptions that may commodify local traditions (Segarra et al., 2024). Thus, sustainable tourism practices focused on planning, environmental responsibility, and community empowerment are vital to ensure long-term viability.

On the other hand, this research focuses on Langkawi, Perhentian, and Pangkor, selected from Malaysia's most notable islands (Ostheimer, 2019). Each island provides distinct tourism experiences and has been selected for this study due to its significant popularity and the abundant availability of online data related to tourism. Langkawi is recognised for its developed infrastructure, luxury resorts, and attractions such as the Sky Bridge and duty-free shopping, making it a premier destination for both international and domestic tourists. The Perhentian Islands are known for their clear waters and vibrant coral reefs, attracting backpackers and eco-tourists for marine activities. Pangkor Island, with its fishing village atmosphere and cultural landmarks, provides a tranquil retreat and insights into local heritage. Collectively, these islands illustrate the multifaceted nature of Malaysia's island tourism.

Furthermore, recent studies show different developmental pathways and challenges across these islands. For instance, Ridzuan et al. (2024) highlight Langkawi's potential to harness both blue and green economy sectors without compromising biodiversity, stressing the need for cohesive policy and public engagement. Yusoh et al. (2023) reveal a mismatch between Pangkor's branding as a scuba-diving site and local stakeholders' views, emphasising the importance of aligning development with authentic experiences. Meanwhile, Tharim et al. (2024) note Perhentian's rapid tourism growth but inadequate public services, with resident satisfaction remaining neutral, underscoring the need for balanced strategies.

Overall, Langkawi, Perhentian, and Pangkor reflect the diversity of Malaysia's island tourism, covering aspects of infrastructure, eco-tourism, and cultural authenticity. Analysing user-generated content provides valuable insights into tourist experiences, complementing conventional research methods. Platforms such as TripAdvisor, Facebook, and Google Reviews vary in accessibility and engagement. In contrast, YouTube merges verbal and visual content with interactive comment sections, offering diverse viewpoints and making it ideal for SA in tourism studies (Uddin et al., 2023; Anastasiou et al., 2023).

2.3 Sentiment Analysis

SA is a subfield of Natural Language Processing (NLP) and ML that identifies emotions and opinions in text (Dutta, 2021). Its main task, sentiment classification, categorises opinions into positive, negative, or neutral classes. SA relies on human

affective states and personality traits, which influence expression in text (A. P. Kumar, Nayak & K, 2019). Furthermore, the SA process requires several stages to extract sentiments from a given text. However, challenges include handling negation, sarcasm, and implicit or explicit features (Birjali, Kasri, & Hssane, 2021).

As illustrated in Figure 2.2, the SA process involves data collection, pre-processing, feature extraction, and classification, followed by outputs such as polarity scores. Following that, SA can be performed at three levels, including the document level, sentence level and aspect level, depending on the required granularity of analysis (Huang et al., 2022). Document-level SA differentiates documents as separate objects with a single sentiment polarity. Sentence-level SA classifies text as either objective or subjective. In contrast, aspect-level SA identifies relationships between aspect terms, categories, and opinions (Zhao et al., 2020).

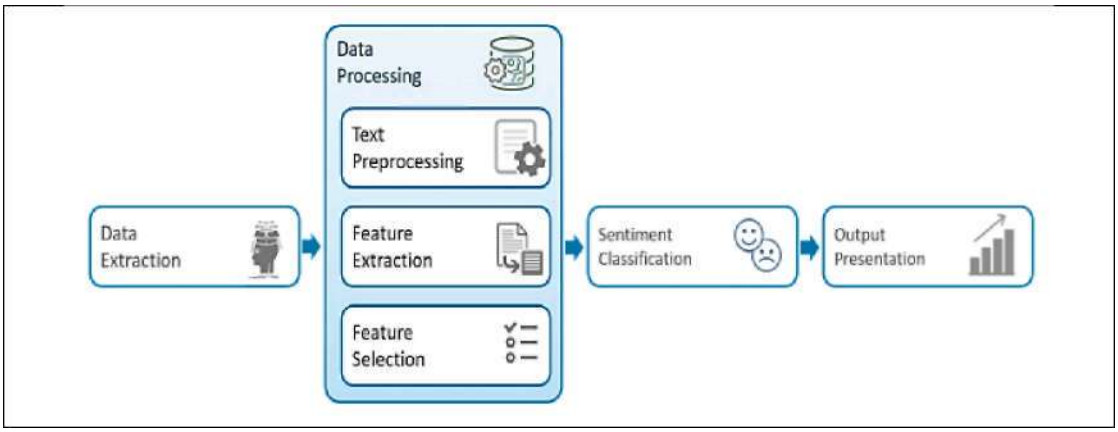


Figure 2.2 Process of Sentiment Analysis
(Source: Birjali et al., 2021)

Besides that, lexicon-based, ML-based, DL-based and hybrid approaches are common techniques used in SA. For instance, SA implements a hybrid of lexicon-based and ML-based techniques to support data-driven decision-making by analysing public opinion, market trends, and user satisfaction. However, challenges such as contextual ambiguity, where the meaning of a word depends on context and can lead to misclassification (Kleppen, 2023). SA also includes tasks such as aspect-based SA (ABSA), fine-grained SA, intent detection, and emotion detection. ABSA captures sentiment about specific components; fine-grained SA distinguishes between levels such as very positive to very negative; intent detection analyses user purpose; and emotion detection identifies underlying feelings. These techniques are especially useful

in tourism to improve marketing and understand tourist feedback (H. Singh & Srivastava, 2023; S. Pradhan, 2022). In recent years, SA has become more prevalent in the tourism sector, as online reviews provide valuable insights into the experiences of tourists. Thus, by evaluating tourist satisfaction and service quality, tourism SA generates well-informed decisions to fulfil tourists' expectations.

2.3.1 Tourism Sentiment Analysis

SA facilitates tourism stakeholders in assessing tourist satisfaction and identifying service gaps by analysing user-generated content like reviews and social media posts, revealing insights into tourist sentiments and perceptions. Stakeholders in the tourism industry, including destination marketers, hotel managers, and tour operators, must capitalise on this information to enhance their services, tailor marketing campaigns, and significantly elevate the tourist experience. Additionally, various methodologies have emerged to improve the accuracy and effectiveness of SA, establishing it as a key resource in the evolving tourism and travel industry. The studies discussed in Table 2.1 highlight the importance of advanced analytical techniques in enhancing tourism recommendations, understanding the impact of social media, and improving SA accuracy.

Despite the increasing interest in SA within the field of tourism research, there are no previous research studies that specifically examined Malaysia's island tourism using SA techniques. For instance, Moud, Nejad and Sadri (2021) introduce an intelligent system for recommending travel destinations that uses SA and semantic clustering to determine user preferences from unstructured review data. Their approach utilises SentiWordNet sentiment scoring, graph-based clustering, WordNet-based semantic similarity, and advanced NLP algorithms. However, their research does not focus on Malaysia or its island destinations, as it primarily addresses global datasets featuring general tourist attractions. In a different approach, Morision et al. (2023) investigated the impact of social media reviews on domestic tourism decisions in Malaysia regarding purchases using a descriptive quantitative approach and structured survey data. Morision et al. (2023) also relied exclusively on structured survey data and did not utilise computational SA techniques for analysing user-generated content. Hence, these two studies neither focus on Malaysia's islands as a specific tourism segment nor investigate rich and unstructured multimedia platforms such as YouTube.

On the other hand, three previous studies present different but complementary methods for SA in the tourism sector. Ali, Omar and Soulaimane (2022) identify sentiment distributions and topic concerns from TripAdvisor reviews using a lexicon-based approach combined with topic modelling. Their method lacks scalability and predictive ability, but it is straightforward, interpretable, and useful for an exploratory study. In contrast, Sano et al. (2023) implemented a structured pipeline that uses Term frequency-inverse document frequency (TF-IDF), Naive Bayes (NB) model, and typical NLP pre-processing to combine primary interview data via online reviews. Their approach provides optimal interpretability and performance and performs well for destination-level comparison and binary sentiment classification. For the hybrid SA approach, Mishra et al. (2021) integrate Long Short-Term Memory (LSTM) for DL-based classification of tweets pertaining to tourism during the COVID-19 epidemic, Valence Aware Dictionary and Sentiment Reasoner (VADER) for lexicon-based sentiment scoring, and Latent Dirichlet Allocation (LDA) for topic modelling. This approach gives deep sentiment insights and performs outstandingly when handling noisy and real-time social media data but requires a large amount of labelled data and high computational power.

Furthermore, Puh & Babac (2023) used both ML and DL models to classify TripAdvisor reviews in English. Their Bi-LSTM model achieved 89% accuracy in sentiment classification and 72% in rating prediction. Another case by Zaman et al. (2024) analysed sentiment using conventional ML models to classify tweets in Bahasa Malaysia, with their best-performing model, which is Support Vector Machine (SVM), achieving 92.4% sentiment classification accuracy. To sum it up, DL models offer theoretical benefits but require extensive computational resources and large datasets. In contrast, ML models can provide accurate predictions for smaller and domain-specific datasets with quicker training times. Moreover, most tourism SA frameworks focus on structured review platforms, limiting their effectiveness with unstructured content from sources like YouTube. Therefore, a more efficient approach that balances resource use, language adaptability, and practical application can be achieved by integrating domain-specific NLP techniques into ML.

Overall, most SA research in tourism has predominantly focused on structured platforms like TripAdvisor, websites and surveys, which limits its applicability in analysing unstructured, user-generated content on platforms like YouTube. Consequently, there is a notable lack of studies regarding the application of SA for

YouTube tourist reviews, particularly in the context of Malaysia's island destinations. This research aims to fill that gap by combining NLP with ML to extract DSA insights from unstructured video reviews. It offers a scalable, multilingual, and context-aware framework to understand public perceptions of tourism in Malaysia's islands. On top of that, this research is noteworthy for its unique geographical focus and methodological innovation, representing the first known application of a hybrid SA framework to YouTube reviews in Malaysia's island tourism context. It aims to uncover the nuanced semantic themes and emotional tones associated with tourists' perceptions of island destinations in Malaysia.

Table 2.1

Previous Research on Tourism Sentiment Analysis

Author (Year)	Aim	Technique									
		Semantic Clustering	Correlation Analysis	VADER	TF-IDF	LDA	LSTM	CNN	SVM	Naive Bayes	Random Forest
1. (Moud et al., 2021)	To develop a context-aware tourism recommendation system.	X									
2. (Mishra et al., 2021)	To analyse the impact of the pandemic, which focuses on the hospitality and healthcare sub-domains.			X		X	X				
3. (Ali et al., 2022)	To analyse tourism reviews by combining topic modelling and SA to extract valuable insights from user reviews.			X	X	X					
4. (Morision et al., 2023)	To explore three main important facets of social media reviews, like quality, quantity and credibility of reviews.		X								
5. (Puh & Babac, 2023)	To explore and compare various ML and DL models for predicting sentiment and ratings from tourist reviews.				X		X	X	X	X	

Author (Year)	Aim	Technique									
		Semantic Clustering	Correlation Analysis	VADER	TF-IDF	LDA	LSTM	CNN	SVM	Naive Bayes	Random Forest
6. (SanoetaL 2023)	To implement a visualised comparative review analysis model and improve tourists' experiences.				X					X	
7. (Zaman et al.. 2024)	To identify positive and negative comments regarding Malaysian Places of Interest using ML methods for SA.								X	X	X
Total											

Tourism SA implements a broad range of techniques, including DL and ML models, as can be shown from the analysis of seven chosen research studies in the table. TF-IDF and LDA were the most commonly used approaches, featuring 4 out of 7, which revealed their effectiveness in feature extraction and topic modelling for unstructured review data. Additionally, 2 to 3 studies implement VADER, SVM, and NB, along with pointing out their efficacy for classification and sentiment scoring tasks, while DL models like LSTM are underutilised due to the higher data and computational demands. In conclusion, an increasing tendency toward hybrid and comparative techniques is generally seen in the table, with some research integrating various techniques to enhance topic comprehension and sentiment classification. In tourism research, this implies a growing emphasis on developing strong and flexible frameworks that achieve a balance between interpretability, computational efficiency, and practical applicability. Moreover, future studies should utilise SA to assess the quality of services and the satisfaction of tourists (Hashim et al., 2020).

2.3.2 Sentiment Analysis on YouTube Reviews

YouTube is a global platform for communication and expression, allowing users to review videos and share their opinions through voting, ratings, favourites, sharing, and reviews. This aids in analysing trends and identifying positive, negative, and neutral sentiments (R. Singh & Tiwari, 2021). Subsequently, YouTube analyses prominent search terms and review polarity using SA to determine the emotional tone of online reviews. Following that, YouTube generates significant unstructured metadata through its videos and is growing in popularity as users engage by liking, disliking, sharing, and commenting, making SA essential for understanding user opinions (R. Pradhan, 2021). As a result, YouTube offers insightful analysis and useful content that can efficiently interpret enormous amounts of reviews, as researchers can use YouTube data and apply visualisation to optimise decision-making (Sahu, 2020).

Figure 2.3 represents a YouTube video with user reviews, which can be collected and analysed in SA. Positive reviews on the video, like *"perfect background music during the swimming scene with turtles"*, show that the music was liked by viewers and that they believed the scene was emotionally appealing. Extracting and evaluating relevant reviews is essential in the feature extraction process, which enables SA to effectively identify key themes and patterns. Overall, YouTube is an essential

platform for SA because of its large user base and diverse data. Users communicate their opinions through reviews, and this platform delivers crucial insights that transcend cultural and linguistic challenges. This makes SA techniques indispensable for understanding diverse perspectives. However, there remains a research gap due to the absence of domain-specific models for YouTube tourism data.

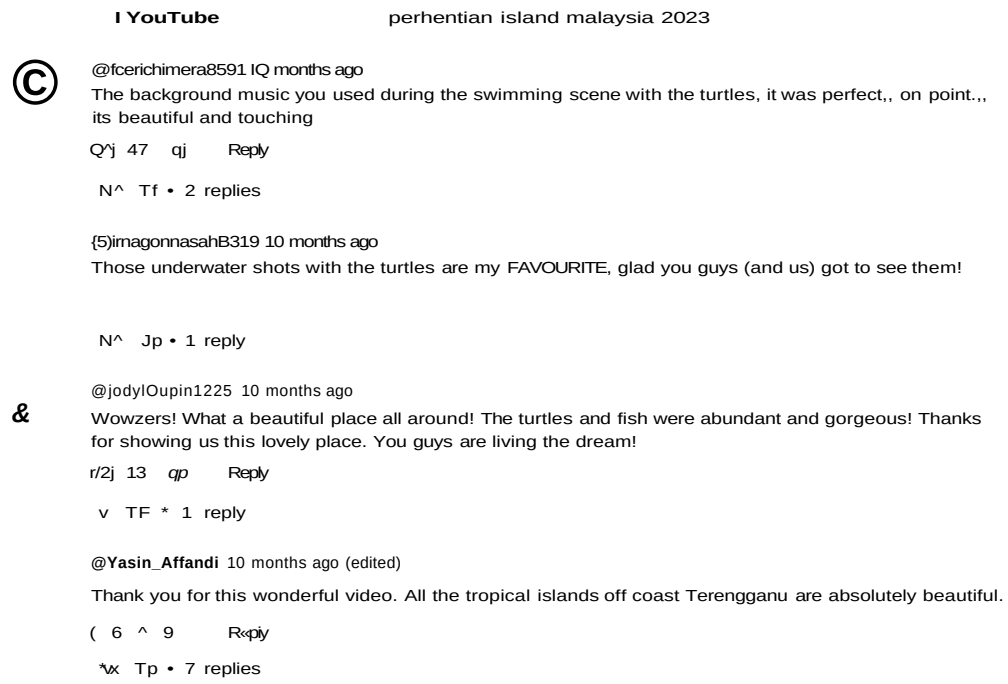


Figure 2.3 Example of YouTube Reviews
(Source: YouTube)

Despite its potential, YouTube is frequently underutilised in tourism SA due to the unstructured nature of its user-generated content, which often features informal language, emojis, and slang. This complexity makes conventional analytical techniques less effective, compounded by the absence of labelled datasets focusing on tourism. Unlike structured review platforms such as TripAdvisor, YouTube's dynamic content makes sentiment extraction challenging. Despite these obstacles, this research chooses YouTube as the primary data source for its authentic and real-time expressions of tourist experiences, allowing for a deeper understanding of public sentiment and perceptions of Malaysia's island tourism through advanced SA techniques.

2.3.3 Sentiment Analysis Techniques

SA techniques are classified into lexicon-based, ML, hybrid approaches and DL-based models, as illustrated in Figure 2.4. Lexicon-based methods assess sentiment scores for tokens by using domain-specific sentiment lexicons, which are usually categorised into dictionary-based and corpus-based approaches. Following that, dictionary-based approaches rely on semantic relationships, while corpus-based methods identify word polarity through statistical patterns (Puri et al., 2023). Although these techniques depend on established rules and do not require labelled data, they often encounter challenges in accurately interpreting context-dependent expressions or identifying nuances such as sarcasm.

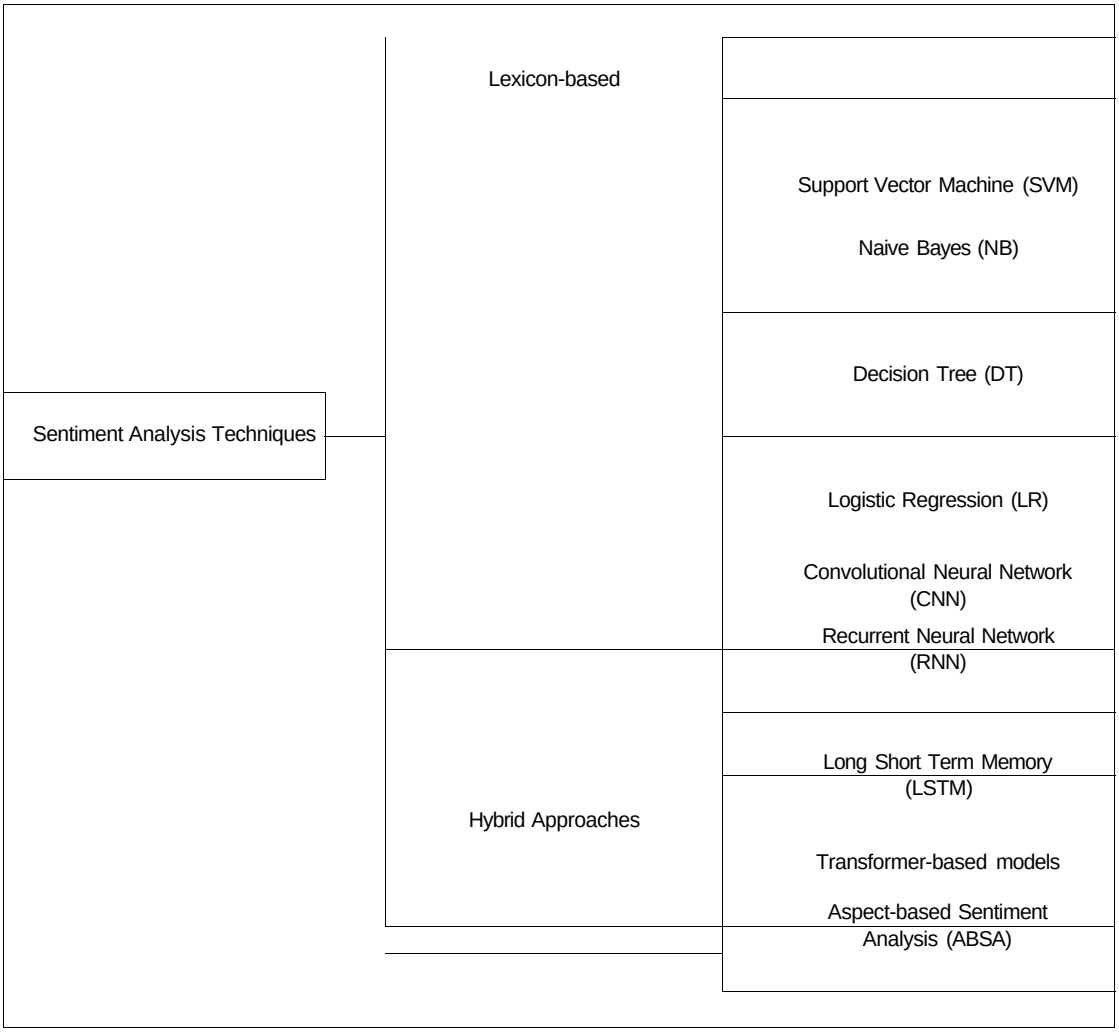


Figure 2.4 Classification of Sentiment Analysis
(Source: Mao, Liu, & Zhang, 2024)

Furthermore, in SA, ML models use computational models to evaluate and categorise textual input, classifying sentiments as neutral, negative, or positive. These models automatically extract sentiments from text using supervised learning, where models are trained on labelled datasets and NLP techniques (Jim et al., 2024). For instance, ML models such as SVM, NB, and Logistic Regression (LR) utilise supervised learning by partitioning the dataset into training and testing sets, which enables the models to learn from labelled examples and generalise patterns to previously unseen data. The supervised technique is typically used for specific category classification tasks, while unsupervised and semi-supervised techniques are applied to unlabelled datasets that include labelled examples. Although ML techniques can detect domain-specific patterns, the best results often require extensive training datasets, and a model trained in one domain may not perform as well in another (Birjali et al., 2021).

Besides that, DL techniques like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), LSTM and Bidirectional Long Short-Term Memory (Bi-LSTM) focus on capturing sequential and contextual relationships in textual data. CNN was initially created for image recognition and has been adapted for text classification, while RNN can predict upcoming words but struggles with vanishing gradients. Next, LSTM units solve this with gating mechanisms that retain long-term dependencies, while Bidirectional LSTM (Bi-LSTM) enhances understanding by processing sequences in both directions, consisting of two parallel layers propagating in opposite directions (Chakravarty & Rengarajan, 2023). Additionally, BERT (Bidirectional Encoder Representations from Transformers) utilises a multi-layer bidirectional Transformer encoder to effectively capture both the syntax and semantics of text. By applying token, positional, and segment embeddings, BERT represents sentences and utilises self-attention mechanisms for enhanced contextual understanding (Mao et al., 2024).

Furthermore, to effectively capitalise on the strengths of various techniques, a hybrid approach that integrates lexicon-based techniques with ML or DL techniques is needed. This integration enhances sentiment classification by enriching feature representation and improving contextual analysis. Additionally, emerging methodologies such as aspect-based sentiment analysis (ABSA), transfer learning, and transformer-based models are gaining prominence for their ability to capture nuanced sentiment elements and transfer knowledge across domains with minimal retraining. Overall, SA techniques encompass a diverse array of techniques, each presenting its

own advantages and limitations. Lexicon-based techniques are known for their simplicity and interpretability, while ML and DL techniques offer scalability and a deeper contextual understanding. Thus, by incorporating a hybrid approach, the performance of SA can be significantly improved, particularly in handling complex and unstructured data environments.

On the other hand, a comparative case study by Ben et al. (2023) presents an evaluation focused on the performance of ML and DL techniques in SA. This study reveals that ML models often require manual feature engineering techniques, such as TF-IDF, to effectively represent textual data. In contrast, DL models autonomously learn semantic and syntactic features from raw data during the training process, minimising the need for manual intervention. ML models like SVM, RF, and NB are generally simpler, more interpretable, and less demanding in terms of computational resources, making them suitable for smaller datasets and rapid experimentation. Conversely, while DL models necessitate greater computational power, they excel at handling complex, large-scale datasets due to their ability to model contextual and sequential information effectively. Finally, ML models may struggle with sequential data, whereas DL models can effectively capture long-term dependencies and contextual nuances. To summarise this, Table 2.2 illustrates the differences between the two techniques, ML and DL.

Table 2.2

Differences between Machine Learning and Deep Learning Techniques

Techniques	Machine Learning	Deep Learning
Feature Extraction	Require manual feature extraction.	Learn features from raw data during training.
Model Complexity	Simpler and require less computational power.	More complex and require significant computational resources.
Sequential Data Handling	Models may struggle with sequential data.	Handle sequential data effectively.
Algorithm	SVM, RF, and NB.	CNN, RNN, LSTM and Bi-LSTM.

(Source: Ben et al., 2023)

Therefore, although DL techniques offer superior performance and flexibility for complex SA tasks, this research focuses on DSA techniques, which necessitate manual feature extraction, as they rely on structured input data for sentiment prediction

using the ML technique. Furthermore, when domain knowledge is effectively applied, manual feature extraction can improve classification performance, leading to highly relevant features. Subsequently, feature extraction assists in identifying key terms and patterns in text, which are crucial for effective sentiment classification. For the next subsections, SA techniques such as lexicon-based SA, ML-based SA, hybrid SA approach, DL-based SA, and ABSA were explained further.

2.3.3.1 Lexicon-Based Sentiment Analysis

Another name for lexicon-based SA is rule-based, which consists of classifying text as positive, negative, or neutral depending on the presence of particular words or phrases using predetermined rules or lexicons (Bonthu, 2024). To find the overall sentiment, the sentiment scores are combined once every text word corresponds with the lexicon (Tho et al., 2021). Following the flowchart shown in Figure 2.5, raw reviews from online platforms are collected before pre-processing the data to create structured input. Then, the cleaned text is input into the lexicon-based sentiment scoring phase, where the VADER library assigns polarity values to effectively label the dataset. These labels generated by the lexicon then serve as training targets for the SVM to learns to distinguish between sentiment classes based on features of the text. The next subsection discusses the lexicon-based SA, including dictionary and corpus-based approaches.

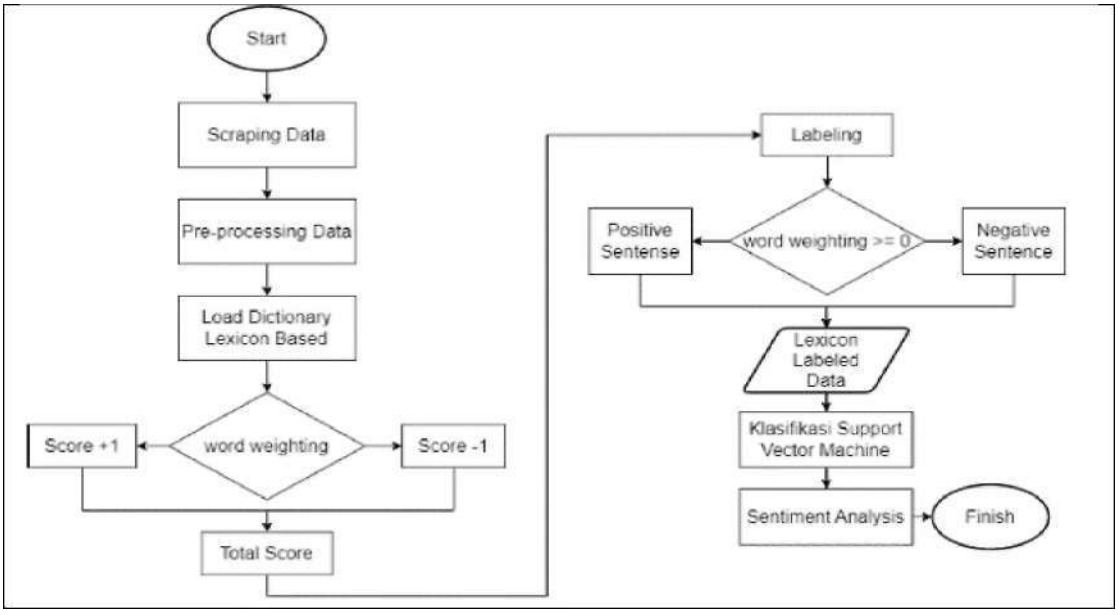


Figure 2.5 Sentiment Analysis Flowchart Using Lexicon-Based
(Source: Kartini et al., 2023)

a) Dictionary-Based

In contrast, the lexicon for dictionary-based information gathers a limited number of expressive phrases with predefined locations. In order to lessen the few-word deficiency problems, the corpus-based strategy uses statistical methods to extract recurring trends or seed sentiment words. Next, the frequency of recurrence of a word in annotated corpora can determine its polarity, with positive polarities implying positive evaluations and negative polarities reflecting negative reviews. Thus, sentiment polarities are produced by the semantic methodology based on conceptual connections between the sentiment polarities (Chauhan et al., 2023).

b) Corpus-Based

Another type of lexicon is corpus-based, which leverages syntactic patterns to extract more opinion phrases from an extensive corpus and detect word polarity. However, it faces a significant drawback that depends on the polarity of the training corpus. Nevertheless, SA discovered that it is useful due to its simplicity of application (Jardim et al., 2021). Moreover, corpus-based techniques have the capacity to evolve as more data becomes available, enabling enhanced accuracy in sentiment detection across various contexts.

c) Lexicon-Based Techniques

On the other hand, unsupervised lexicon-based models are used in S A to identify a text's emotional tone without requiring labelled training data. It also utilised pre-established sentiment lexicons, which comprise terms associated with sentiment scores that are either positive or negative. To ascertain whether a text conveys a positive, negative, or neutral sentiment, this technique examines the text, finds sentiment words, and then adds up their scores (Srivastava, Bharti, & Verma, 2022).

Moreover, text polarity and sentiment in textual data can be evaluated using lexicon-based SA techniques such as Natural Language Toolkit (NLTK), VADER and TextBlob, as illustrated in Table 2.3. For a range of applications, these techniques provide comprehensive SA classification. In other words, NLTK is a robust Python library for NLP which offers a range of tools for tasks such as tokenisation, part-of-speech tagging, named entity recognition, and parsing. NLTK also supports integration

with ML libraries and provides access to valuable linguistic resources like WordNet. Its modular structure allows users to customise and enhance NLP pipelines, making it a versatile toolkit suitable for various language processing tasks (Batta, 2024).

Table 2.3

Comparison of Lexicon-Based Techniques

Techniques	NLTK	VADER	TextBlob
Application	Consists of a collection of lexical resources to provide word processing libraries.	Aggregate sentiment lexicon to list lexical functions with semantic alignment.	Break down the input into sentences for review and provide a simple identification that reflects text sentiment.
Sentiment Scores	Polarity scores range from 0.0 to 1.0 for positive and negative sentiments.	Metric values range from -4 (most negative) to +4 (most positive), with 0 for neutral.	Polarity score ranging from -1 to +1 and subjectivity score from 0 to 1.
Additional Features	Used to extract opinions and assist in inferring polarity details.	Provide reports for the depth of polarity.	Perform a sentence-level analysis.

(Source: R. Sharma et. al, 2022)

Moreover, VADER uses a vocabulary and rule-based methodology to examine sentiments in text, effectively conveying the emotional tone of speech. VADER also relies on a predetermined lexicon of terms that are weighted by emotional intensity and align with particular sentiment values to determine the sentiment of a text (Nurcahyawati & Mustaffa, 2023). Lastly, TextBlob provides a polarity score to indicate the strength of the sentiment, along with a subjectivity score to reflect whether the content is opinion-based or factual. While it is user-friendly, TextBlob often falls short compared to tools like VADER and SentiWordNet due to its limited coverage and its tendency to categorise texts as neutral (Sham & Mohamed, 2022).

2.3.3.2 Machine Learning-Based Sentiment Analysis

ML models are crucial in SA to evaluate the effectiveness of models in classifying sentiments from textual data. ML models are also essential tools for SA,

enabling effective classification and providing valuable insights from textual data (Kumar et al., 2019). Moreover, the two main techniques applied in the SA process are lexicon-based and ML for the SA classification task. In ML, problems are solved by automated techniques based on knowledge and data from the past. This encompasses supervised learning, which utilises labelled data to develop models and predict target classes, and unsupervised learning, which analyses unlabelled input data. Furthermore, using a data-driven methodology, as shown in Figure 2.6, the text data is transformed into feature vectors for feature extraction by removing sentences that provide objective communication and facts and analysing each for subjectivity. However, feature selection can result in data loss due to the tendency to have a lot of significant features in text classification; thus, high-dimensional input spaces are avoided by taking into account only a few significant features.

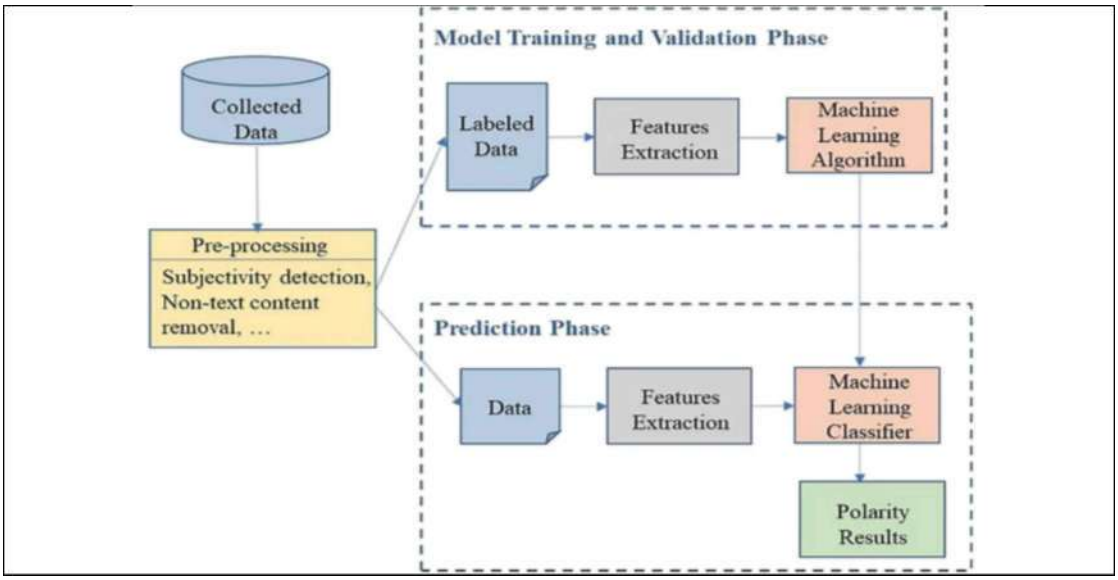


Figure 2.6 Supervised Machine Learning Sentiment Analysis
(Source: Jardim, Mora, & Santana, 2021)

Several common ML models were trained on a pre-processed dataset, including NB, LR and SVM. In addition, NB is a straightforward ML technique based on probability theory that relies on the assumption that each feature is independent and equally important, making it more versatile and user-friendly than LR, which uses a binary logistic curve. Meanwhile, LR computes the probability of the target variable using the natural logarithm (Cam et al., 2024). Moreover, SVM is essential for conventional text classification and regression issues and operates well with various complex datasets. It is also beneficial in ML applications because of its versatility in

handling linear and non-linear separations using kernel functions (Alsemaree et al., 2024).

a) Support Vector Machine

The SVM model is a reliable and efficient supervised learning technique used for text classification. It effectively handles non-linear classification through the use of kernel functions, which facilitate the linear separation of data points. Common kernel types include radial basis function, polynomial, and linear kernels. This allows the model to identify the optimal hyperplane that separates reviews into positive and negative categories, as shown in Figure 2.7. To find the best values for the hyperparameters, a grid search method is applied for hyperparameter optimisation (Octava et al., 2023).

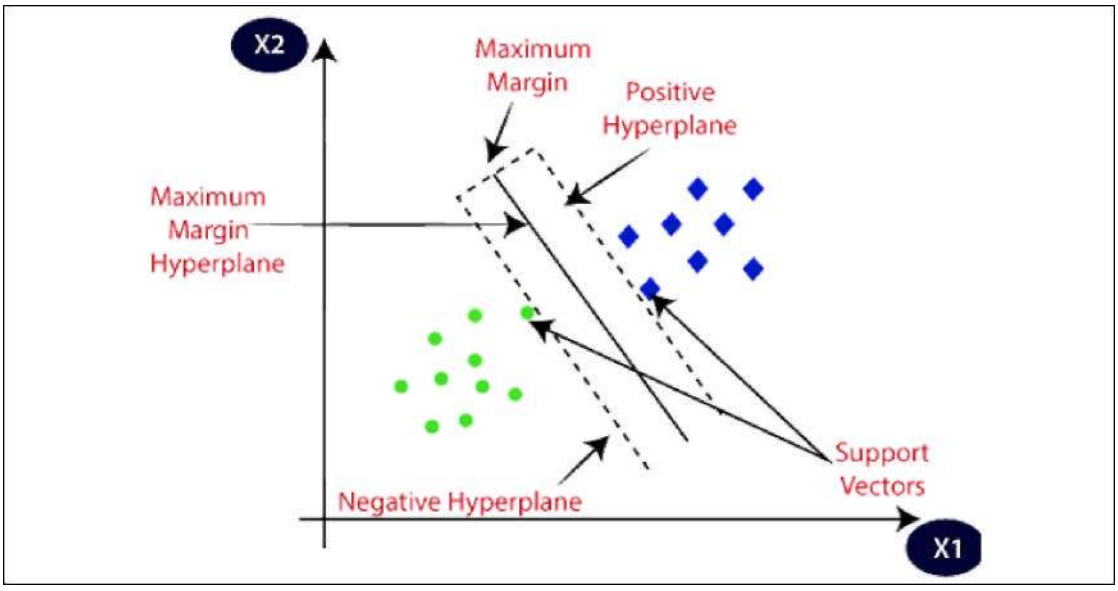


Figure 2.7 Support Vector Machine Model
(Source: Saini, 2024)

Furthermore, the SVM model is designed to find the optimal hyperplane that separates data points belonging to different classes. The points closest to this hyperplane are known as support vectors. During the training process, the model adjusts the position of the hyperplane to maximise the margin while managing the acceptable level of classification errors. This involves modifying parameters to strike a balance between minimising misclassifications and maintaining a wide margin. Additionally, if the data

is not linearly separable, a kernel trick is used to transform the data into a higher-dimensional space, making it easier to separate.

b) Naive Bayes

NB model is a probabilistic ML model that assumes independence between features and is commonly used for classification tasks and SA due to its effectiveness in large feature spaces. Next, the underlying principle involves computing the probability of a specific class given a set of features and selecting the class with the highest probability as the predicted class. NB models are also computationally efficient and require minimal training data, which have been successfully applied in various domains, including text classification, spam filtering, SA and medical diagnosis (Sano et al., 2023).

c) Logistic Regression

Meanwhile, LR is an exponential model that identifies the relationship between a categorical dependent variable, sentiment polarity, and one or more independent variables, such as features extracted using Bag-of-words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF). It also constructs a dictionary of words in the dataset and their frequencies in the negative, neutral and positive sentiment classes. The final probability is calculated using the softmax function, ranging from 0 to 1, and the output is the calibrated probability of the text that belongs to each sentiment class (Sham & Mohamed, 2022).

d) Machine Learning-Based Techniques

For ML-based techniques, Table 2.4 compares three ML techniques used in SA: LR, NB, and SVM. According to the table, SVM uses a hyperplane to separate data into two groups, NB utilises probabilistic methods based on Bayes' theorem, and LR applies a sigmoid function to classify data into binary categories. Each model can be applied to various S A tasks due to its unique approaches to data separation, probability estimation, and iterative optimisation.

Table 2.4

Comparison of Machine Learning-Based Techniques

Model	Support Vector Machine	Naive Bayes	Logistic Regression
Description	A classification technique that finds a hyperplane to separate data into two categories with maximum margin.	A probabilistic model that uses Bayes' theorem and makes strong assumptions about the independence of features.	A statistical method using a sigmoid flattening function to describe model predictions for binary classification.
Key Features	Used to find out the boundary of each plane and predict which side belongs to the margin.	Calculates probabilities to predict the category.	It is used for classification problems to minimise errors until convergence, or a set number of iterations is reached.

(Source: Jardim, Mora, & Santana, 2021)

2.3.3.3 Deep Learning-Based Sentiment Analysis

In addition, DL in NLP incorporates RNN and LSTM networks. RNNs are capable of processing data of arbitrary lengths while maintaining the sequence order; however, they often struggle with capturing long-distance semantic relationships and face issues like the vanishing gradient problem. On the other hand, LSTMs address these challenges by managing the cell state through gates that facilitate the addition or removal of information, allowing them to distinguish between relevant and irrelevant data. This process involves a sigmoid layer and a pointwise multiplication operation. Conversely, Bi-LSTM utilises two LSTM models, which enhance the network's information capacity at each step, thus providing a more nuanced understanding of context. Moreover, CNNs are frequently utilised for tasks such as text classification and SA. These networks use various kernels with weights optimised during training, analysing words and their contexts to generate similar output values. The convolutional layers are often combined with pooling layers, which help eliminate less relevant information. (Puh & Babac, 2023).

2.3.3.4 Aspect-Based Sentiment Analysis

ABSA is an important fine-grained SA problem that focuses on identifying sentiment elements at various levels. These levels include aspect terms, aspect

categories, opinion terms, and sentiment polarity. ABSA tasks can be categorised into two types: single tasks and compound tasks. Single tasks are designed to extract opinion terms from a given text, while compound tasks aim to extract and predict multiple sentiment elements within a sentence (Zhang et al., 2023). An overview of the ABSA tasks is illustrated in Figure 2.8. This figure highlights a single ABSA process that involves aspect term extraction, opinion term extraction, aspect category detection, and aspect sentiment classification. In contrast, compound ABSA tasks extract and predict several sentiment elements simultaneously within a sentence.

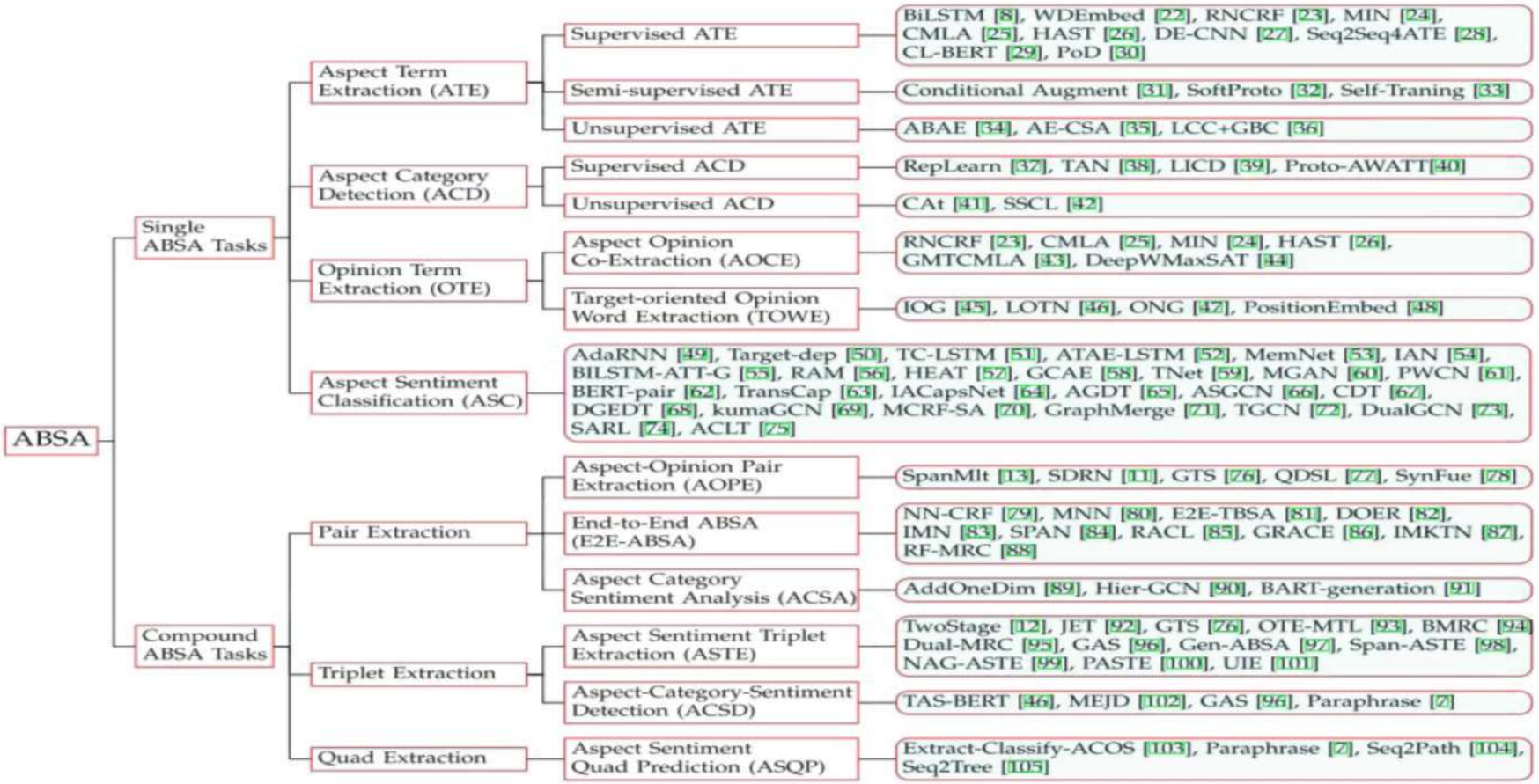


Figure 2.8 Taxonomy of Aspect-Based Sentiment Analysis Tasks
(Source: Zhang et al., 2023)

2.3.3.5 Hybrid Sentiment Analysis Approach

On the other hand, the hybrid SA approach predicts sentiment based on labelled data, allowing it to handle unstructured data well by integrating lexicon and ML techniques. These techniques improve feature extraction and lead to more accurate sentiment classification by addressing issues like sarcasm and negations (Kaur & Sharma, 2023). This is because sentiment polarity scores derived from lexicon-based techniques are crucial target variables for ML. Effectively combining them as a hybrid approach can enhance SA accuracy. Thus, an integration between lexicon and ML is implemented in a hybrid, as seen in Figure 2.9, where every lexicon undergoes hybridisation among models. The probability for each sentiment polarity class is displayed as the result of ML models, as the text's polarity corresponds to its highest probability.

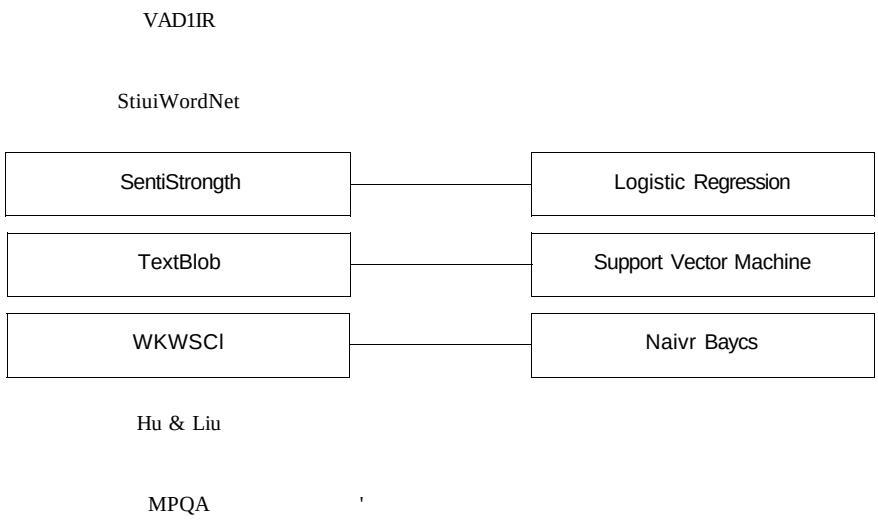


Figure 2.9 Combination of Lexicon and Machine Learning
(Source: Sham & Mohamed, 2022)

Moreover, a hybrid approach can improve ambiguity handling and guarantee stability in challenging applications such as polarity recognition. As a result, these techniques extract novel aspects from natural language and integrate them with designed ones (Polignano et al., 2022). Furthermore, Kaur and Sharma (2023) conducted research utilising a hybrid feature extraction technique that combines aspects and reviews with a DL model like LSTM to classify sentiment. This hybrid approach

enhances SA's reliability and accuracy, particularly when evaluating user reviews. TF-IDF, n-gram, and emoticon-specific features that can be utilised by hybrid approaches to capture linguistic subtleties. Meanwhile, the LSTM model enables the model to retain context and learn from complex data sequences.

2.3.3.6 Comparison of Sentiment Analysis Techniques

A comparative review of relevant research is essential in comprehending the development and variety of SA techniques in recent years. Table 2.5 illustrates the comparison of the SA model implementation for previous studies from 2019 to 2024. There are 47 relevant studies with various techniques to determine the sentiment expressed in text. The common techniques include lexicon-based techniques like VADER, AFINN, and TextBlob. Following that, ML-based techniques include NB, SVM, and LR, whereas DL-based techniques include LSTM, CNN, and BERT. Moreover, some of the advanced NLP techniques include Part-of-Speech (POS) Tagging, Named Entity Recognition (NER) and Latent Dirichlet Allocation (LDA). These techniques can be used individually or in combination to enhance the accuracy of sentiment classification. Each technique has its advantages and disadvantages, and the optimal technique depends on the specifics of the research.

In addition, some research paper evaluates lexicon-based implementations of sentiment extraction, adaptation, and interpretation that implement either a custom-designed sentiment lexicon or SentiWordNet as the fundamental lexicon. Despite having comparable methodology, these studies differ in their approaches to address language-specific difficulties, adapt to different domains, and increase the polarity score. For instance, Bakar, Idris, & Shuib (2019) propose an easy approach involving POS tagging for sentiment score retrieval and translating Malay text reviews into English in accordance with the English-based terminology. Although it is straightforward and computationally lightweight, it lacks language resilience and domain alignment. Another case study by Aye and Aung (2020) proposes a hybrid model that incorporates lexicon-derived polarity values as features in ML framework. Several supervised learning models are applied to classify the sentiment scores, which are calculated using syntactic characteristics and n-gram tokenisation. Despite the fact of increasing classification accuracy, lexicon-based models continue to be limited by semantic ambiguity during translation.

Furthermore, Tho et al. (2021) use SentiWordNet and VADER for sentiment scoring, addressing challenges from code-mixed and complex language. Their findings show moderate accuracy compared to the BERT model. In contrast, Pratama, Utami, & Sunyoto (2019) enhance the lexicon by integrating user star ratings and domain-specific data through POS tagging. This improvement highlights that domain sensitivity can overcome the limitations of conventional sentiment lexicons, resulting in better sentiment representation and classification accuracy. To sum it up, lexicon-based SA has difficulties when applied to datasets that are domain-specific or linguistically varied, despite being comprehensible. Its dependence on pre-defined polarity scores and multipurpose lexicons can result in insufficient semantic alignment and a lack of ability to comprehend contextual subtleties. Thus, lexicon-based models should be enhanced by incorporating domain-specific lexicons or supplementary information into sentiment scores.

An increasing amount of research indicates that DL architectures are outperforming conventional ML models in addressing the classification of text difficulties. P. K. Singh, Sharma, & Paul (2020) classified political tweets using LSTM, RNN, SVM and NB, revealing that the sequence modelling of LSTM performed better than conventional models in detecting sentiment polarity. By using a hybrid approach, Puh & Babac (2023) analysed reviews from tourists and found that Bi-LSTM significantly outperformed other models like CNN, SVM and NB in both sentiment and multi-class rating prediction tasks. However, with 97.6% accuracy using SVM in Bengali sports news classification compared to NB, LR, DT, K-Nearest Neighbours (KNN) and RF, Barua et al. (2021) showed the effectiveness of feature engineering in language-specific datasets. In the study by Mondal et al. (2023), LSTM models performed better in fake news identification than ML, while Datta et al. (2024) achieved 99.8% accuracy by combining TF-IDF-based models with BERT and Bi-LSTM. These findings reveal that most of the previous studies that implement DL models have outstanding performance for SA compared to ML.

Overall, the comparison of previous studies demonstrates that lexicon-based, ML, and DL techniques in SA constitute a complementary and emerging connection. Although lexicon-based approaches are straightforward and interpretable, their efficacy is constrained by their inability to handle linguistic variation and domain-specific lexicons. Some previous studies have improved accuracy by adding lexicon scores as extra features in ML models to address the lack of labelled instances. Likewise, DL

models effectively capture contextual nuances but depend on large amounts of labelled data, increasing the risk of overfitting and complicating domain adaptation due to differences in linguistic expressions. (Islam et al., 2024). On top of that, a hybrid approach that uses lexicon-based insights to facilitate ML and DL models, as well as refining comprehension through contextual learning, could be the most effective for SA frameworks.

Table 2.5

Techniques Comparison Used for Sentiment Analysis Implementation

Author (Year)	Techniques								Results
1. (Azizan et al.. 2019)	X								o Lexicon-Accuracy: 52%
2. (BakaretaL 2019)	X								o Lexicon - Precision: 83.55%; Recall: 84.61%
3. (C. RawatetaL 2019)					X	X		X	o SVM-Accuracy: 66.03% o LR - Accuracy: 65.3 6% o RF - Accuracy: 63.54%
4. (Chang, Wu, & Hwang, 2019)	X								o Lexicon - Average percentage: 55% positive correlation and 20% negative correlation
5. (Chathuranga, Lorensuhewa, & Kalyani, 2019)	X		X		X	X	X		o Lexicon-Accuracy: 87.79% o POS-Accuracy: 91% o NB-Accuracy: 69.23% o SVM - Accuracy: 65.20% o DT-Accuracy: 68.13%
6. (Imane, Kareem, &Faical, 2019)	X				X	X	X	X	o Lexicon-Precision: 82.1%; Recall: 28.5%; Fl-score: 42.3% o SVM - Precision: 76.9%; Recall: 82%; Fl-score: 79.3% o NB - Precision: 87.6%; Recall: 86.2%; Fl-score: 86.9%

Author (Year)	Techniques	Results
7. (Pratamaetal.. 2019)	X X	o LR - Precision: 76.4%; Recall: 81.8%; Fl-score: 79% o RF - Precision: 82.1%; Recall: 83.6%; Fl-score: 82.8% o Lexicon-Accuracy: 88.91% o POS - Accuracy: 77.21%
8. (Schoene & Mel, 2019)	X X	o Lexicon - Accuracy: 78.53% o LDA - Detect topics encapsulated in each tweet and display coherent topics
9. (Aye & Aung, 2020)	X X	Lexicon - Accuracy: 92% with imbalanced class; Weighted average for Precision: 94%; Recall: 92%; F-measure: 93% POS - Determine lexical tag for a particular occurrence of a word in a sentence
10. (Manalu, Tulus, & Efendi, 2020)	X X	o LSTM - Accuracy: 84.68% with an error 0.3667 o RNN - Accuracy: 86.70% with an error of 0.3362
11. (P. K.Singh et al., 2020)	X X X X X	o LSTM - Accuracy: 93.7% o RNN - Accuracy: 81.35% o SVM - Accuracy: 82.56% o NB - Accuracy: 72.75% o LR - Accuracy: 79.54%
12. (Rouhani&	X X X X X	o Lexicon - Accuracy: 51% positive, 26% neutral, 23%

Author (Year)	Techniques	Results
Abedin, 2020)		negative o SVM-Accuracy: 82% o NB - Accuracy: 75% o DT-Accuracy: 46% o KNN - Accuracy: 43 %
13. (Bagherzadeh et al.,2021)	X X	o Lexicon-Accuracy: 87.92% o POS - Accuracy: 88.03%
14. (Baldha et aL 2021)	X X X X	o VADER - Accuracy: 44.37% positive, 36.07% neutral, 19.56% negative o LSTM - Accuracy: 96% o SVM-Accuracy: 87% o NB - Accuracy: 76%
15. (Baruaetal.. 2021)	X X X X X X	o SVM - Accuracy: 97.60% o NB - Accuracy: 97.14% o LR - Accuracy: 97.43% o DT - Accuracy: 96.54% o KNN - Accuracy: 66.53% o RF - Accuracy: 97.40%
16. (Jubair et al.. 2021)	X	o VADER - Accuracy: 34.9% positive, 31.8% neutral, 33.3% negative

Author (Year)	Techniques											Results
17. (Kurniawanetal., 2021)	X						X					o Lexicon - Accuracy: 70%; Recall: 62.5%; F-measure: 76% o LDA - Accuracy: 98% positive; Accuracy: 2% negative
18. (Lim, Li, & Song, 2021)							X	X		X		o LDA-Accuracy: 85% o LSTM - Precision: 85.07%; Recall: 83.5%; Fl-score: 84.28% o RNN - Precision: 79.25%; Recall: 78.01%; Fl-score: 78.63%
19. (Nugroho, 2021)		X										o VADER - Predicted Accuracy: 55% democratic party; Predicted Accuracy: 45% republican party
20. (Thoetal.,2021)	X	X							X			o Lexicon - Precision 60%, recall 77%, accuracy 61%, F-score 67% o VADER - Precision 63%, recall 88%, accuracy 67%, F-score 74% o BERT - Precision 90%, recall 76%, accuracy 83%, F-score 83%
21. (A. Sharma& Ghose, 2021)	X	X		X								o Lexicon - Categorise using number of factors o VADER - Range normalised valence between -1 and 1 o POS - Determine grammatical level by lexicon
22. (Jardimetal.,	X		X									o Lexicon - Accuracy: 87% Portuguese, 92% English,

Author (Year)	Techniques												Results
2021)												X	72% Spanish, 79% French o TextBlob - Determine polarity and subjectivity if a text
23. (Shi & Chen, 2021)					X								o LDA - Accuracy: 80% workplace bullying, 70% workplace support, 66% compensation, 63% career advancement
24. (Trivedi & Singh, 2021)	X												o Lexicon - Accuracy: 24%, 26% and 25% positive; Accuracy: 13%, 12% and 13% negative
25. (Finniganetal., 2022)	X			X									o Lexicon - Accuracy: 63% o POS - Accuracy: 90%
26. (KarametaL 2022)			X						X		X	X	o TextBlob - Used for spelling correction o SVM-Accuracy: 99% o LR - Accuracy: 97% o DT-Accuracy: 97%
27. (Lau et al., 2022)		X	X										o VADER - 37.30% positive, 10.90% neutral, 51.80% negative o TextBlob - subjectivity score for 0.00122 positive, 0.00198 neutral, 0.00093 negative
28. (Park & Lee, 2022)	X			X	X								o Lexicon - Sentiment polarity Amazon reviews have higher standard deviation due to more extreme opinions o POS - Amazon data has higher proportion of adverbs

Author (Year)	Techniques										Results
											while Google News data has more comparative adjectives and adverbs o NER - Accuracy: 23.43%
29. (RamdhanietaL 2022)	X							X	X	X	o Lexicon -Accuracy: 29% positive, 7.1% neutral, 63.9% negative o SVM - Accuracy: 56% o NB - Accuracy: 68% o LR - Accuracy: 67%
30. (Song, Zheng, & Shan, 2022)	X			X							o Lexicon - Probability of textual sentiment score in [0.9,1] is more than 50% and [0,0.1] is less than 30% o LDA - Reviews: Topic 0 positive, topic 1 negative, topic 2 negative and topic 3 negative
31. (Waspodo etal.. 2022)	X							X			o Lexicon - Accuracy: 35% positive, 33% neutral, 32% negative o SVM-Accuracy: 98%
32. (Barrera, 2023)		X		X						X	o AFINN - Identify positive, neutral and negative sentiments o LDA - Identify 8 main topics from online reviews; 75% sentences assigned probability greater than 0.7 o LR - Customer attention, food quality and tasting menu

Author (Year)	Techniques	Results
33. (Cao, 2023)		significantly influenced restaurant ratings o LSTM - Use word embedding techniques to extract semantic information from text o CNN - Hyperparameter tuning of learning rate, batch size and hidden layer size
34. (Gkillas, Simos, & Makris, 2023)		o VADER - Recall: 90%; Fl-score: 60% o TextBlob - Recall 80%; Fl-score: 55% o BERT - Recall 95%; F1 -score: 65%
35. (Hossain et al.. 2023)		o VADER - Mean polarity score: 87% positive, 26% neutral, 2% negative, 87% compound o AFINN - Accuracy: 75.24% positive, 8.89% neutral, 15.87% negative o LR-Accuracy: 88.56% o DT-Accuracy: 76.58% o RF - Accuracy: 82.51%
36. (Kanavos et al.. 2023)		o AFINN - Accuracy: 27.035% positive, 45.93% neutral, 27.035% negative o TextBlob - Accuracy: 33.96% positive, 47.09% neutral, 18.95% negative o NB - Accuracy: 72.8% with 1-gram encoding

Author (Year)	Techniques												Results
												X	o LR - Accuracy: 73.3 % with 1 -gram encoding
37. (Laifa& Mohdeb, 2023)					X					X			o BERT - Accuracy: 67%; Precision: 66%; Recall: 67%; Fl-score: 66% o DT - Accuracy: 60%; Precision: 59%; Recall 60%; Fl-score 59%
38. (Li, 2023)							X	X		X			o SVM - Precision: 83.4%; Recall: 80.6%; Fl-score: 80.5% o NB - Precision: 86.3%; Recall: 83.4%; Fl-score: 85.6% o DT - Precision: 82.5%; Recall: 79.7%; Fl-score: 79.6%
39. (Mahajani, Srivastava, & Smeaton, 2023)	X												o Lexicon-Accuracy: 80%
40. (Manik et aL 2023)	X				X								o Lexicon - Accuracy: 77.3%; Fl-score: 73.9% o BERT - Accuracy: 41.4%; Fl-score: 34.5% without fine tuning
41. (Mondal et aL 2023)				X			X		X	X	X	X	o LSTM-Accuracy: 98.46% train, 91.84% test, 92.38% validation o SVM-Accuracy: 91% o LR-Accuracy: 90% o DT - Accuracy: 79.90%

Author (Year)	Techniques	Results
42. (Puh&Babac, 2023)	X X X X	o KNN - Accuracy: 86% o RF-Accuracy: 90% o LSTM - Accuracy: 66% rating; 87% sentiment o Bi-LSTM - Accuracy: 72% rating; 89% sentiment o CNN - Accuracy: 62% rating; 85% sentiment o SVM - Accuracy: 55% rating; 80% sentiment o NB - Accuracy: 46% rating; 73% sentiment
43. (Silviana, Nababan, & Zarlis, 2023)	X X	o Lexicon - Accuracy: 86.51% with an error of 0.3 3 5 o LSTM - Accuracy: 95.68% with an error of 0.111
44. (Tetteh& Thushara, 2023)	X X X	o VADER - Accuracy: 69% o TextBlob - Accuracy: 69%; Precision: 62%; Recall: 95%; Fl-score: 75% o NB - Accuracy: 73%; Precision: 66%; Recall: 94%; Fl-score: 78%
45. (Alqaryouti et al.. 2024)	X	o SVM - Accuracy: 88.41%; Precision: 84.81%, Recall: 72.19%; F-Measure: 77.99%; Correlation 0.77
46. (DattaetaL 2024)	X X X X X X X	o LSTM-Accuracy: 99.05% o BERT - Accuracy: 99.80% o SVM - Accuracy: 99.60%

Author (Year)	Techniques															Results
																o NB - Accuracy: 95.33% o LR - Accuracy: 98.83% o DT - Accuracy: 99.55% o RF - Accuracy: 99.20%
47. (Kakdeetal., 2024)									X			X	X	X	X	o BERT-Accuracy: 89% o SVM-Accuracy: 88% o NB - Accuracy: 79.40% o LR - Accuracy: 88.40% o RF - Accuracy: 88.90%
Total	23	9	3	6	7	1	6	9	6	2	3	16	13	12	9	

A systematic technique comparison for SA can be seen in the table, which evolves from lexicon-dependent rules to predictions using ML and contextual modelling produced by DL. The effectiveness and applicability of various approaches differ based on domain complexity and computational limitations. Almost half of the studies utilised lexicon-based techniques, valued for their interpretability and low resource requirements. ML models, particularly SVM, generally achieve accuracy rates between 80% and 90% when paired with effective feature engineering. These models perform optimally with domain-specific datasets but necessitate manual pre-processing. In contrast, DL models outperform other techniques in terms of accuracy, occasionally exceeding 99.8%, and while DL is more accurate, several high-performing SVM, NB, and RF models demonstrate that ML models with well-designed features can perform like DL in limited contexts. Also, hybridisation with ML or DL is necessary for lexicon-based techniques to remain competitive in contemporary SA tasks.

In summary, the analysis of previous studies shows that lexicon-based approaches are simple and interpretable but struggle with semantic ambiguity and domain adaptation. In contrast, ML and DL models typically perform better in sentiment classification, particularly when paired with effective feature engineering or large datasets. However, these models can encounter challenges such as overfitting, domain dependence, and considerable computational demands. Therefore, informed by these findings, this research implements a hybrid SA approach that integrates a lexicon-based technique with the ML model by using VADER and SVM for sentiment prediction and classification. This research also utilises DSA techniques for feature extraction to improve the model's ability to capture both surface-level patterns and deeper contextual insights. By harnessing the strengths of lexicon, ML, and feature extraction techniques, this research aims to develop a SA model that is more accurate, interpretable, and adaptable to various domains.

2.3.4 Limitations of Sentiment Analysis

Despite advancements in SA, persistent challenges continue to limit its effectiveness in the tourism sector. Table 2.6 provides a summary of these challenges and their impact on the performance and accuracy of sentiment classification models in tourism research. SA encounters multiple challenges, including the requirement for beneficial findings, diverse data, and context understanding. Sentiment classification in

the tourism sector faces limitations because most datasets are in English, creating language barriers and the lack of comprehensive linguistic resources for languages other than English (Hashim et al., 2020). This previous research also specifies that service types in tourist reviews can lead to less precise analysis, and the evolving nature of the language used in social media poses a challenge for maintaining up-to-date sentiment lexicons and models. Another case study by Ali, Omar, and Soulaimane (2022) highlights the limitations of SA in the tourism sector, such as the inability to identify irony, sarcasm, and connotations, as well as destination-specific data. Since this research also focuses solely on English language reviews, it limits insights from reviews in other languages and suggests that future research should explore advanced ML or DL models and consider multilingual datasets to enhance the robustness and applicability of SA in tourism.

Table 2.6

Summary of Sentiment Analysis Challenges with Multilingual YouTube

Challenge	Description	Impact on SA
Language and Multilingual Barriers	SA relies heavily on English texts and lacks resources for other languages.	Limits the scope of analysis and excludes non-English reviews, reducing inclusivity and cross-cultural insights.
Evolving Language and Contextual Complexity	Social media language evolves rapidly and includes sarcasm, irony, and context-specific expressions.	Makes it difficult to maintain effective sentiment lexicons and models, resulting in misinterpretation of sentiments.
Dependence on Feature Extraction Techniques	ML-based SA requires converting text into numerical features.	Model performance is highly dependent on feature quality, while poor extraction lowers accuracy and generalisability.

Another significant drawback of SA is its dependence on the efficiency of feature extraction techniques, particularly when using ML. The process of converting unstructured text into relevant numerical representations that ML models can comprehend is critically dependent on the quality of feature extraction. As highlighted in the study by Semary et al. (2024), the classification accuracy and overall model performance are significantly impacted by the selected feature extraction techniques.

However, the generalisability and robustness of SA models are limited since no particular method consistently produces the best results across a variety of datasets and application domains. Therefore, incorporating an integration of descriptive and semantic as a DSA technique in feature extraction can lead to improved sentiment classification accuracy by enabling a deeper understanding of the text's content and contextual nuances.

Moreover, in order to further elucidate the significance of these limitations in relation to the current study, Table 2.7 offers a structured mapping of the identified research problems (RP) and their correlation with the existing literature review (LR). This analysis elucidates specific challenges encountered in SA, particularly concerning user-generated content related to tourism. It illustrates how previous studies have either adequately addressed or neglected critical issues such as the linguistic ambiguity arising from informal language (RP 1), the cultural and multilingual diversity that impacts sentiment interpretation (RP 2) and the lack of labelled data, which restricts the models' ability (RP 3). Thus, by aligning the problem statements with literature-identified gaps, this analysis underscores the necessity for more robust, context-aware, and multilingual SA methodologies tailored to Malaysia's tourism sector. Additionally, this alignment reinforces the justification for the DSA framework within this research.

Table 2.7

Problem Statement and Literature Review Relationship

RP vs LR	Research Problem (RP)	Literature Review (LR)
RP 1	Informal language creates linguistic ambiguity as word meanings often rely on context, which can result in misinterpretations and misclassifications.	LR examines the limitations of lexicon-based and ML-based SA in handling informal language. Limited research on SA using YouTube often neglects this complexity.
RP 2	Sentiment interpretation is significantly complicated due to cultural and multilingual diversity. YouTube reviews often feature code-mixing between English and Malay, using culturally specific idioms and expressions.	LR points out that most SA approaches mainly use English data, which inadequately addresses Malay's specific meanings, cultural phrases, and regional dialects,
RP 3	Limited labelled data, especially in low-resource languages like Malay, restricts	LR highlights that ML and DL models require extensive labelled datasets, yet

RPvsLR	Research Problem (RP)	Literature Review (LR)
	the models' ability to learn nuanced sentiment patterns and generalise effectively.	resources in the Malay language are limited, especially for informal tourism reviews.

Overall, the research highlights key issues with SA, particularly regarding the complexity of user-generated data, which often includes slang and context-dependent language. Most SA frameworks focus on English datasets and overlook the multilingual and culturally diverse nature of reviews in Malaysia's tourism, leading to lower accuracy with non-English data. Moreover, the lack of labelled datasets in low-resource languages like Malay hampers the effectiveness of supervised learning models and impacts the performance of SA models. These challenges relate to the problems outlined in Chapter 1 and form the basis for the constructed hybrid approach, which integrates DSA with lexicon and ML techniques. The following section will discuss how DSA techniques can enhance the SA approach.

2.4 Descriptive Analysis

Descriptive analysis is an essential technique in text analytics, particularly valuable during preliminary text mining tasks, as it enables the identification of recurring themes, patterns, and insights within textual data. It entails summarising and interpreting raw data to derive meaningful information. Furthermore, descriptive techniques are especially beneficial in SA, as they deepen the comprehension of text features and structures, including word frequency and co-occurrence patterns. This section defines descriptive analysis and critically examines previous research studies implementing these techniques, thereby framing their relevance to sentiment classification and contextual understanding in this study.

2.4.1 Definition of Descriptive Analysis

Descriptive analysis refers to the statistical summarisation and interpretation of data characteristics, particularly aimed at understanding underlying patterns without inferring relationships or making predictive generalisations. Textual analysis includes metrics such as word frequency, distribution, and co-occurrence. Moreover, finding

trends, patterns, and correlations in past data can be accomplished through descriptive analysis, which can manage and classify data to make it easier to quickly identify outliers or abnormalities (Hall, 2023). Furthermore, descriptive analysis is closely associated with identifying specific textual attributes, such as keyword frequencies, which serve as foundational features in subsequent sentiment classification tasks.

In other words, descriptive analysis examines and summarises data using descriptive statistics and visualisations to understand its main characteristics. Following that, descriptive attributes are specific data characteristics, like keywords, to provide information about the analysed items (Mustika et al., 2020). In contrast, descriptive statistics summarise the main features of a dataset to obtain frequency distributions by revealing how many reviews express positive or negative sentiments in the context of SA (Ramadhan & Gunawan, 2023). Therefore, analysing and comprehending data distributions and patterns enhances the extraction of sentiment-bearing features, such as thematic clusters, which are critical in sentiment polarity detection. Additionally, enhancing SA for Malaysia's islands by implementing descriptive methodologies enables the extraction of data-driven insights, which leads to improved tourism experiences and enhanced decision-making capabilities. The next section discusses a comparative analysis of descriptive techniques from previous research to select which technique to use in descriptive analysis.

2.4.2 Previous Research Study Related to Descriptive Analysis

Previous research studies have widely applied descriptive analysis to uncover meaningful patterns, themes, and trends within textual data, particularly in the context of unstructured user-generated content relevant to SA. Table 2.8 demonstrates the comparative analysis of descriptive techniques such as TF-IDF, exploratory data analysis (EDA), statistical summaries (SS), data visualisation (DV), trend analysis (TA) and time series (TS) in extracting key insights from user-generated content like social media and online websites. Some of the previous research studies explored integrating descriptive analysis with ML models to improve sentiment classification accuracy and contextual understanding. Hence, this section reviews key research contributions in descriptive analysis by emphasising their methodologies and findings for each related research study.

Descriptive analysis comprises a variety of techniques to achieve a deeper comprehension of textual data by identifying patterns, distributions, and insights. TF-IDF is one technique that determines the frequency distribution of words in text documents by calculating the weight of each word based on its term frequency across a collection of documents (Liu, Chen, & Liu, 2022). In contrast, EDA assists in understanding the structure and content of a dataset by revealing patterns and identifying relevant aspects for analysis (Nanayakkara, 2024). Complementing EDA, SS offer a quantitative overview of the dataset's key characteristics, highlighting the range, distribution, and central tendencies, often supported through visual summaries (Puri et al., 2023).

In addition, DV is a technique of visually presenting information to understand trends, outliers, and patterns in data by identifying relationships and insights that may not be apparent from the raw data (Alabi & Bukola, 2023). Furthermore, TA can identify recurring patterns from past data, while TS examines the temporal dynamics of variables over specific time intervals to provide a comprehensive understanding of market behaviour and enable a better understanding of public perception and data predictions (Shamisavi & Jahanshahi, 2022). Thus, descriptive techniques play a pivotal role in extracting valuable insights across various studies, forming a foundational basis for comprehending extensive textual datasets.

Moreover, despite being adapted to different datasets and analytical objectives, there are various studies using descriptive analysis for feature extraction. For instance, C. Rawat et al. (2019) transform unstructured text into numerical features using TF-IDF to identify lexical patterns and subtle clues related to abusive or abnormal user behaviour. Following that, EDA, SS and TA are applied to differentiate behavioural patterns between blocked and non-blocked users. Following the removal of dominating search terms, Lau et al. (2022) utilise TF-IDF to analyse Twitter data and discover contextually meaningful terms. In order to extract features like polarity, subjectivity, and sentiment distribution, they implement EDA and broaden feature extraction using sentiment scoring methods along with TS analysis to provide temporal sentiment patterns for public opinion. These two studies illustrate how descriptive techniques can identify interpretable features from complex textual data, thereby improving model transparency and supporting subsequent classification tasks.

On the other hand, in order to overcome issues like text sparsity and language-specific characteristics, Wang et al. (2020) integrate TF-IDF with topic modelling to

enhance topic interpretation by extracting unique phrases from Chinese Weibo posts. Herqutanto, Zatari, & Sutoyo (2023) implement TF-IDF for extractive summarisation, evaluating sentences based on the cumulative TF-IDF values. However, they recognise constraints when summarising semantically varied information. In addition, Pattnaik, Li, & Nurse (2023) implement TF-IDF to convert cleaned tweets into feature vectors for conventional ML models. Although DL models like BERT eventually provide higher Fl-scores by capturing richer contextual semantics, the TF-IDF-based models perform well in tweet classification. Overall, these studies point out TF-IDF's significance as a descriptive technique for topic modelling, summarisation, and classification, as well as its limitations in conveying complexities across various contexts.

Table 2.8

Previous Research Study Related to Descriptive Techniques

	Author (Year)	Technique						Results
1.	(C. RawatetaL 2019)	X	X	X		X		o TF-IDF - 1500 top features with 1-2-word n-gram o EDA - Non-blocked users make higher number of edits on daily basis compared to blocked users o SS - Comparison blocked and non-blocked users' behaviour o TA - 50% of blocks made by 5 admins and 30% of total blocks made by 1 admin in 2018 which lead to the rise in blocked users in late 2018
2.	(WangetaL 2020)	X				X		o TF-IDF - 38 key topics for positive sentiment, 19 for neutral sentiment and 19 for negative sentiment o TA - Analyse frequency of 11 topics from 1 January 2020 to 18 February 2020
3.	(Baldha et aL 2021)	X	X				X	o TF-IDF - Creates BoW and extracts feature vectors with unigram technique o EDA - Most popular vaccines are extracted and visualisation is created using word cloud o TS - Maximum number of tweets were done from January 2021 to March 2021
4.	(Baruaetal., 2021)	X						o TF-IDF - Extract feature space of different n-gram; all models showed better performance in identifying "Cricket" category than other classes in all feature spaces
5.	(Dornick et al., 2021)	X						o TF-IDF - Top 25 words in abstracts such as patient, infect, case, health, clinic, test, risk and many more
6.	(Harywanto et al.,2021)	X				X		o TF-IDF - TF-IDF score-based method scored higher in BLEU score due to its shorter summary length o TA - Reveals hotel sentiment trends can change over time and eliminate 176 reviews that not have same class
7.	(Widiantoro, Wibowo, &		X		X			o EDA - Word cloud for negative sentiment words such as "gagal", "lama", "nunggu" while positive sentiment words like "bintang" and "kasih"

Author (Year)	Technique						Results
	[in H	Q M	w oo	> Q	< H	oo H	
Harnadi, 2021)							o DV - Among 15 words consists of a total of 257 sentences of word "lama", 184 sentences of word "tolong" and 143 sentences of word "mohon" are the top 3 of most often occurred words
8. (Karametal., 2022)	X	X					o TF-IDF - Each word creates a dimension and identical word lies in the same dimension o EDA - Thailand was the top country, followed by India, United States, United Kingdom whereas Sri Lanka was the least one to tweet about coronavirus among 20 countries
9. (Lauetal., 2022)	X	X		X		X	o TF-IDF - "POTUS" and "vaccine" appear 67000 times in total with 0.59 and 0.73 indicating these terms have relatively low uniqueness o EDA - Top 10 most frequent words among 141k tweets, including the last president "Trump" with 25054 times and "border" appearing 15463 times o DV - Dashboard show compound score, subjectivity score, top 10 for most occurring words and relevant query on data frame in Emoji scatterplot o TS - From 20 January 2021 until 18 January 2022, sentiment generally shares the opposite direction with confirmed cases since Biden's presidency
10. (Motz et al., 2022)	X			X			o TF-IDF - Choose top 5000 most common words then each tweet in dataset turned into sparse vector o DV - SVM has the largest runtime while Naive Bayes has smallest runtime within 87.8 seconds spent analysing
11. (Herqutanto et al., 2023)	X	X					o TF-IDF - Determine value of word frequency in form of an index and words that are represented by numbers along with the frequency value o EDA - Positive sentiment is dominant in property, tour, travel and precious metal categories, while negative sentiment is prevalent in animal care, automotive, body care, carpentry, computers and laptops, food and drinks, office and stationery, other products, party supplies, and craft categories

Author (Year)	Technique				Results
12. (Mondal et aL 2023)	X	X			o TF-IDF - Captured the top 50 topics where each of the topics consisting a combination of similar types of words o EDA - Word cloud analysis reveals significant overlap between phrases in fake and real news which primarily related to US political events
13. (Pattnaik et aL 2023)	X			XX	o TF-IDF - TF-IDF matrix for input into four n-gram based models. o DV - Non-expert Twitter users prioritize VPN use (21%), device account security (13%), laptop security (10%), Wi-Fi and password security (9.2%), and home security (8.6%) o TA - Percentage of 'WifiPass' topic decreased from 10.4% in 2019 to 9.5% in 2021 and replacing with new topics like 'HomePrivSec', TnternetSec' and 'HelpRelated'
14. (PurietaL 2023)		X	X		o EDA - Among 100 persons, 80% of people's voting decisions are influenced by political campaigns o SS - Standard deviation is 3.0582, coefficient of variation is 0.4846, mean is 6.3105 and skewness is 0.3042 respectively
15. (T,Pradeep,& R, 2023)	X	X			o TF-IDF - Combination of LR with TF-IDF achieved highest accuracy with 89.86% for Python Translator o EDA - Similarity index between each translator results in 98.58% of matching
Total	13	9			

As shown in the table of previous studies related to descriptive analysis, TF-IDF is the most widely applied descriptive technique, appearing in 13 out of 15 studies, as it is particularly effective at extracting key features and identifying important words. Following TF-IDF is EDA, utilised in 9 studies, which is instrumental for visual analyses such as word clouds and sentiment frequency distribution. Next, TS and SS are each featured in 4 studies, providing insights into how sentiment evolves over time and facilitating behavioural comparisons. In contrast, DV and TA techniques are less frequently used, appearing in only 2 studies, but they offer valuable visual outputs such as scatter plots and line graphs. Overall, while TF-IDF and EDA dominate, other techniques enhance the understanding of temporal dynamics and user behaviour. In order to gain a clearer understanding of the practical advantages and drawbacks of descriptive techniques, the next section provides a comparative assessment of descriptive techniques related to semantic analysis.

2.4.3 Comparative Evaluation of Descriptive Techniques for Semantic Analysis

This research evaluates descriptive techniques based on their ability to summarise text features and their compatibility with semantic analysis techniques. Given that the next step will involve semantic topic modelling and sentiment interpretation, it is essential to choose a representation that effectively preserves both structure and meaning. Therefore, there are two different feature extraction techniques are compared, which are TF-IDF and Word2Vec, with assessments focused on semantic richness, interpretability, computational feasibility, and suitability for multilingual user-generated data.

Additionally, TF-IDF is identified as the primary method due to its widespread application and effectiveness in text-based SA. In contrast, Word2Vec provides embeddings that are more semantically rich and context-aware. The aim of this comparison is to determine whether the added complexity of these advanced techniques is justified, particularly in a low-resource context such as YouTube tourism reviews, ultimately guiding the selection of the most meaningful technique for future analyses.

2.4.3.1 Term Frequency-Inverse Document Frequency

A method combining data extraction and text mining techniques, known as TF-IDF, is utilised to identify word associations in documents. It is used for extracting key terms, computing similarity degrees, and evaluating search rankings. (S. W. Kim & Gil, 2019). The TF-IDF technique assesses the relevance of a word within a particular text or category of texts. It uses the Inverse Document Frequency (IDF) to indicate how well a word can differentiate between classes of text, while Term Frequency (TF) measures how often words appear within the text (Liang & Niu, 2022). Subsequently, terms in a corpus can be assigned TF-IDF scores based on their significance related to a particular document. (Thakkar & Chaudhari, 2020).

Furthermore, TF-IDF is a quantitative technique that converts textual data into a vector space model, making it applicable for text analysis and NLP. By emphasising both common and uncommon terms, TF-IDF optimises data retrieval systems and facilitates the classification of documents based on their content. The advantages of using TF-IDF for feature extraction include enhancing classification results and improving the performance of ML models through better representation of text data. Additionally, combining sentiment dictionaries with TF-IDF can improve the accuracy, precision, recall, and F1-score of sentiment classification. (Liu et al., 2022).

For instance, Arora et al. (2024) applied VADER to assign sentiment labels, and the text was converted into TF-IDF vectors to be trained by four models. As a result, the accuracy of the NB model is 76.6%, LR is 80.3%, XGBoost is 91.0%, and RF is 92.3%. Overall, the techniques used are VADER for lexicon-based labelling, combined with TF-IDF features, data balancing, and ensemble learning, resulting in an impressive accuracy of up to 92.3%. Therefore, it is recommended by the study to explore sentiment granularity, integrate rating data, and investigate neural network or combined techniques for future research.

In summary, TF-IDF is a highly effective method for feature extraction, converting text into numerical formats by balancing word frequency with term uniqueness. Its ability to emphasise important terms makes it especially valuable for applications like SA, classification, and information retrieval. Studies have shown that combining TF-IDF with sentiment lexicons or ML models can significantly boost performance metrics such as accuracy, precision, recall, and F1-score. The flexibility

of this technique underscores its significance inNLP and facilitates future developments through a hybrid approach and more advanced SA.

2.4.3.2 Word2Vec

Word2Vec is a neural network-based technique that maps words into continuous vector spaces, positioning similar words close together. It features two main architectures, which are Continuous Bag of Words, which predicts a target word from its context, and Skip-gram, which predicts context words based on a target word (Amalia et al., 2023). This approach creates compact embeddings that capture semantic relationships between words, making it more effective for understanding contextual and linguistic nuances than the TF-IDF technique. However, TF-IDF distinguishes itself from Word2Vec by focusing on term specificity rather than capturing the semantics or context of words. Hence, through emphasis on discriminative keywords allows TF-IDF to excel in high-dimensional text classification tasks.

Consequently, experimental results from Nurhaliza et al. (2024) reveal that TF-IDF outperforms Word2Vec when used with the SVM model. Specifically, with an 80:20 train-test split and the RBF kernel, the SVM paired with TF-IDF achieved 84% precision, 85% recall, and Fl-score of 83%. In contrast, the combination of SVM and Word2Vec yielded 83% precision, 82% recall, and a lower Fl-score of 76%. This decline in performance is likely due to the dataset's imbalance and the complexities involved in semantic vector representation. Overall, while Word2Vec provides richer semantic insights, TF-IDF proves to be more effective in this classification scenario. Its simplicity and high interpretability make it a compelling option for S A tasks that depend on sparse feature spaces. This highlights the importance of selecting appropriate feature extraction techniques based on the specific algorithm and the characteristics of the dataset.

To conclude, while Word2Vec offers meaningful semantic representations through dense embeddings, its effectiveness may fluctuate depending on the characteristics of the dataset and the classification techniques used. In particular, it may face challenges with imbalanced datasets or conventional models like SVM. Conversely, TF-IDF tends to produce more consistent results in sparse environments. Therefore, selecting an appropriate feature extraction method should take into account

both the analytical objective and the nature of the data, as performance can significantly depend on the alignment with the specific problem context.

2.4.3.3 Comparison Between Term Frequency-Inverse Document Frequency and Word2Vec Techniques

TF-IDF and Word2Vec are two prominent feature extraction techniques in NLP, each with its distinct advantages and limitations. TF-IDF is a statistical approach that transforms text into sparse vectors based on the importance of words across different documents, making it particularly effective for classification tasks, especially with models such as SVM. Its simplicity and robustness render it well-suited for smaller datasets or those with class imbalances. Conversely, Word2Vec uses neural networks to create dense word embeddings that capture semantic relationships, providing a deeper understanding of language. However, this technique may not perform as well with conventional models like SVM on imbalanced datasets, as evidenced by lower F1-scores in comparative studies. On top of that, while Word2Vec offers richer semantic representation, TF-IDF remains a more reliable choice in sparse feature spaces. Ultimately, the decision between the two should be guided by the specific objectives of the analysis, the compatibility with models, and the characteristics of the dataset. In addition, Table 2.9 provides a comparison of TF-IDF and Word2Vec techniques, highlighting their differences, strengths, weaknesses, and suitability for specific applications in text representation.

Table 2.9
Comparison Between Term Frequency-Inverse Document Frequency vs Word2Vec

Criteria	TF-IDF	Word2Vec
Technique Type	Statistical.	Neural network-based.
Representation	A sparse vector that is based on term frequency and inverse document frequency.	Dense vector since it includes word embeddings in continuous space.
Semantic Understanding	Does not capture the semantics or context.	Captures semantic and contextual relationships.
Interpretability	High as it is easy to interpret due to direct term weighting.	Lower due to the embeddings being abstract and harder to interpret.

Criteria	TF-IDF	Word2Vec
Text Classification	Effective for sparse and high-dimensional classification tasks.	Useful for capturing meaning but may underperform in conventional models.
Strengths	Simple, interpretable, and effective with small or unbalanced datasets.	Rich semantic representation and better with DL models.
Weaknesses	Ignores word context and order.	Requires more data and may not align well with ML models.
Combination Potential	Performs well when combined with lexicons or ML models.	Effective when integrated with neural models.

Hence, TF-IDF has been chosen as the primary feature extraction technique for this research due to its effectiveness and compatibility with both lexicon-based and ML methods for sentiment classification. Given the small and domain-specific dataset of YouTube reviews about Malaysia's islands, TF-IDF efficiently identifies key terms without extensive training data. Its ability to convert text into meaningful numerical representations makes it suitable for models like SVM in the hybrid approach of this research. While TF-IDF will guide the descriptive analysis, a semantic technique will also be introduced to enhance contextual understanding. This approach balances keyword precision with semantic richness, making TF-IDF an optimal choice for SA. Although TF-IDF effectively emphasises the significance of terms for sentiment classification, it fails to capture the deeper semantic relationships present in the text. Hence, to improve the analysis, a semantic technique will be introduced to understand tourist sentiments beyond mere keywords in the following section.

2.5 Semantic Analysis

Semantic analysis aims to uncover the meaning embedded within words, phrases, and sentences, facilitating the extraction of significant patterns and thematic structures from textual data. It also addresses the subtleties of natural language, such as synonyms, antonyms, and contextual meanings. By leveraging semantic analysis, both information retrieval and text comprehension can be significantly improved. This section presents a definition of semantic analysis and examines prior research studies focused on semantic techniques.

2.5.1 Definition of Semantic Analysis

Semantic analysis is a subfield of NLP that aims to interpret the underlying meaning and contextual relationships within textual data, improving sentiment classification. In contextual semantic SA, a word or phrase can express a range of sentiments depending on the words that surround it, where the sentiment attributed to a particular word or phrase is intricately connected to its co-occurring keywords within the sentence (Ramanathan, Hajri, & Ruth, 2024). Hence, this technique enhances the accuracy of sentiment classification in text. Moreover, by examining word associations and grammatical structure, the semantic analysis assists in comprehending and interpreting the text, indirectly assisting with tasks such as SA, feature extraction and word sense disambiguation. However, the challenges include managing semantic relationships to accurately collect sentence meanings and handling ambiguous and polysemic natural language (Goyal, 2021).

2.5.2 Previous Research Study Related to Semantic Analysis

Semantic analysis is a key research area in NLP, and many studies have explored its applications in various domains. Furthermore, semantic analysis techniques have been applied in previous research studies such as Latent Dirichlet Allocation (LDA), Named Entity Recognition (NER), Latent Semantic Analysis (LSA), K-means clustering, and cosine similarity. Most previous research studies reveal its effectiveness in extracting meaningful insights from text using different techniques in semantic analysis. Hence, this section reviews key research contributions in semantic analysis by emphasising their methodologies and findings for each related research study.

Semantic analysis is a key aspect of NLP that aims to understand the meanings of words and sentences. For instance, LDA can effectively align with specific aspects of reviews, such as facilities, locations, and food quality, thereby facilitating sentiment mapping for these key areas in ABSA. LDA is a technique that uncovers hidden topics within text data, and when combined with semantic analysis, it reveals the sentiments associated with these topics, providing deeper insights into user opinions (Sudiro et al., 2021). Conversely, NER involves identifying and classifying named entities in a text, extracting detailed information from structured and unstructured data (Jehangir et al., 2023). In contrast, LSA utilises singular value decomposition on the term-document

matrix to reveal latent semantic structures and conceptual relationships between terms and documents. However, LSA may not effectively capture complex linguistic relationships (Kodicherla et al., 2023). Following that, K-means clustering is an unsupervised ML model that groups a dataset into distinct clusters based on feature similarity, while cosine similarity, a metric ranging from -1 to 1, is commonly used in recommendation systems to measure user preferences, especially in content-based systems (Ravisankar, Marimuthu, & Praveen, 2023).

Table 2.10 outlines various techniques used in text analysis, particularly for semantic techniques, to identify topics within a corpus. Recent studies applied LDA to identify topic patterns from textual data, with the technique adaptable to various data sources and research purposes. Focusing on comprehensive thematic application, Karami et al. (2020) used LDA to analyse a significant amount of Twitter abstracts related to health. Their technique limited interpretability by using conventional pre-processing, although it lacked coherence measures and thorough visualisation. In order to enhance the semantic clarity of topics, Dornick et al. (2021) concentrated on COVID-19 research abstracts and applied a linguistically optimised pre-processing pipeline along with the elimination of prominent domain terms. Khadija et al. (2023) applied LDA to analyse social media reviews on the SRIKANDI application by integrating bigram and adjusting hyperparameters based on coherence scores.

On the other hand, Karam et al. (2022) applied a semi-supervised learning technique to produce pseudo-labels for tweets using K-means clustering, which were then evaluated by annotators and classified by ML models. Although efficient, its flexibility to nuanced emotional subtleties could be constrained by its reliance on a set number of clusters. Meanwhile, cosine similarity was implemented in a taxonomy-based framework by Arslan and Cruz (2023) in a business context to determine the relevance of articles correlated with specified company interests. Though it relies on a static taxonomy that can restrict adaptability to new topics, their technique achieved computational efficiency and semantic clarity by vectorising both news articles and taxonomy ideas using spaCy and scikit-learn.

Mondal et al. (2023) used LSA together with TF-IDF to detect fake news in order to uncover hidden themes in the text, and it was included in several ML and DL models. Despite its comprehensiveness, the linear structure of LSA could limit its capacity to detect more complex semantic relations. Furthermore, Yenkikar et al. (2024) utilised a multi-model NER approach specifically designed for biomedical text, along

with unsupervised topic modelling with LDA. Although their strategy used rule-based techniques and spaCy to obtain high SVM model performance, it performed insufficiently in domain-specific recognition tests, which emphasises the necessity for personalised models.

Overall, these researchers can observe the significance of semantic analysis techniques in acquiring significant insights and trends from unstructured textual data across various domains. Both structured and unsupervised topic modelling techniques provide distinct advantages, from facilitating scalable data classification to revealing hidden theme patterns. Additionally, semantic clarity and model performance can be enhanced by the combination of hybrid models, coherence-based tuning and pre-processing improvements. However, domain flexibility, model interpretability, and sensitivity to subtle linguistic patterns tend to be essential for the effectiveness of these techniques.

Table 2.10

Previous Research Study Related to Semantic Techniques

	Author (Year)	Techniques ^ Q *J	^ W Z	< &o U	g s \ ^	ef l 3 U	-1 o U	£> "i .S ^	Results
1.	(Karami et al., 2020)	X							o LDA-Out of 38 topics, 9 did not have significant trends, while 29 topics consisting of 17 topics had extremely significant with probability < 0.001, 4 topics had very significant with probability < 0.01, and 8 topics had significant with probability <0.05 changes
2.	(Baldha et al., 2021)	X							o LDA - Top 7 most discussed topics between September 2020 to April 2021 such as 'vaccination programme', 'post covid', 'Tokyo Olympics', 'side effects', 'health institute', 'face mask', 'health care' and 'book appointment' followed with distribution of sentiment over topics was found nearly equal for all topics except 'vaccination programme'
3.	(Dornick et al., 2021)	X							o LDA - Classify abstracts into 5 and 4 unique topics using two LDA models
4.	(Gronneberg & Nygard, 2022)	X							o LDA - Identify 12 topics from the corpus and degree of overlap between topic 8 with probability 0.09 and topic 9 with probability 0.08
5.	(Karametal.,2022)				X				o K-Means - Modelled topic into 5 categories such as fear, sadness, disgust, anger and joy where fear has the largest proportion while joy was the least emotion in Covid-19 tweets
6.	(Raman et al., 2022)	X							o LDA - Class 2, 3, 5 and 7 consisting violent content with probability of 0.88, 0.77, 0.8 and 0.66, respectively
7.	(Arslan& Cruz, 2023)						X		o Cosine Similarity - Identify number of topics to find similarity of a text with concepts of a taxonomy
8.	(Herqutanto etal.,	X							o LDA - Most optimum number of topics is 36 and evaluate the coherence score where

Author (Year)	Techniques					Results
2023)						before hyperparameter tuning are more positive and negative after tuning
9. (Khadijaetal., 2023)	X					o LDA - Best number of topics is 2 after the experiment of hyperparameter tuning the model
10. (Mondal et al., 2023)			X			o LSA - Both fake and real news topics include references to prominent figures such as Trump, Clinton, Comey, Putin and Obama as well as both categories include political figure
11. (Pattnaiketal.,2023)	X					o LDA - Analyse topics for tweets from 2019-2021 years and study any changes in topics before and during pandemic
12. (Vipinetal., 2023)	X					o LDA - Identify two topics, which are chief minister's health condition and Apollo Hospital authority
13. (Kakde et al., 2024)	X					o LDA - There is some overlapping among 12 topics and 5 major themes among these topics
14. (Mbooi, Rangata, & Sefara, 2024)	X					o LDA - Trained on a range of topics from 1 to 6 which demonstrated the highest coherence score, indicating its optimal choice for detecting highly repetitive words
15. (Yenkikar et al., 2024)	X	X				o LDA - 5 topics are modelled under both techniques including NMF o NER - SVM outperforms other models with a Fl-score of 55.41% meanwhile SpaCy NER performance is poor and ranking fourth with an F1 -score of 44.59%.
Total	12	1				

In conclusion, the table shows a comprehensive review of the semantic techniques, which reveals a noticeable methodological tendency in favour of unsupervised topic modelling, especially LDA, resulting in 12 of the 15 previous studies. This dominance underscores LDA's effectiveness in extracting themes from various domains and its capability to reveal semantic structures and implicit evaluative aspects in unlabelled data. This allows for scalable analysis by recreating random word distributions across documents. Other techniques like NER, LSA, K-means clustering, and cosine similarity were used less frequently, each serving specific purposes such as entity identification and document organisation. Despite differing methodologies, these techniques enhance the semantic representation of text. In other words, LDA's ability to identify word patterns and assign multiple topics makes it particularly suitable for analysing short texts, such as social media reviews. Its scalability and adaptability solidify its status as a preferred technique in semantic analysis. The next section will discuss how LDA and ADSA work together to extract and group relevant aspects from user-generated content.

2.5.3 Semantic Analysis Techniques for Aspect Identification

Following the review of semantic techniques, this section discusses the application of semantic techniques for aspect identification. Semantic analysis plays a fundamental role in ABSA, particularly in identifying and categorising specific aspects or attributes discussed within user-generated content. Moreover, a widely recognised semantic technique is LDA, an unsupervised probabilistic model that reveals latent topics based on patterns of word co-occurrence. These topics often correspond to aspects such as facilities, customer service, and environmental conditions. LDA is particularly adept at identifying implicit aspects in unstructured and multilingual data without requiring labelled inputs, making it especially valuable for tourism-related analyses, such as assessing YouTube reviews. Therefore, by clustering terms around shared themes, LDA enhances semantic clarity and presents a scalable alternative to conventional techniques that rely heavily on annotated datasets. In this research, LDA is applied not only to uncover hidden thematic structures but also to align sentiment expressions with their respective aspects, positioning it as a versatile tool for extracting insights from unstructured tourism reviews.

2.5.3.1 Latent Dirichlet Allocation

LDA is a generative probabilistic model that identifies hidden topics in a document collection by estimating the probability distributions over words and topics. Within a large collection of documents, or corpus, latent topic information can be found using the well-known generative probabilistic model LDA, where each document is considered as a vector of word frequencies by using the BoW technique, which indirectly converts textual information into easily examined numerical data. Next, through the representation of a document as a collection of topics, LDA efficiently decreases the dimensionality of the BoW model as the document is represented vector-wise in a few hundred dimensions, with a few hundred themes that are often used (Krishnan & Kennedyraj, 2023). Thus, training is faster to complete by reducing dimensions and reducing the chance of overfitting.

On top of that, LDA is an effective unsupervised topic modelling technique that reveals hidden themes in large text corpora, especially in unlabelled data from social media. It models each document as a combination of topics, with each topic represented as a distribution of words, allowing for automated theme detection without manual tagging (Su & Thakur, 2025). LDA typically involves pre-processing steps, such as text normalisation, stop word removal, and tokenisation, followed by vectorisation to create a document-term matrix. The model then learns topic distributions across documents and uses optimisation techniques to determine the best number of topics based on coherence and interpretability (Shekhar et al., 2021). Ultimately, LDA uncovers patterns in unstructured text, making it a valuable tool for thematic analysis in contexts with limited labelling.

2.5.3.2 Aspect-Based Sentiment Analysis

Identifying aspects is a crucial part of semantic analysis, as it allows for the extraction of specific features or dimensions from textual data that are relevant to user opinions. ABSA is a sophisticated sentiment mining technique that goes beyond the simple labelling of entire documents or sentences as positive, negative, or neutral. Instead, ABSA identifies specific opinion targets that are known as aspects or features of an entity, and extracts the sentiment expressed towards each of these aspects. This process generally involves three main subtasks, which are aspect term extraction to

identify aspect-related words or phrases, aspect category detection to classify these aspects into predefined categories, and sentiment polarity classification to determine whether the sentiment expressed towards each aspect is positive, negative, or neutral (Chauhan et al., 2023).

On the other hand, ABSA encompasses two main categories, which are single and compound tasks. Single ABSA tasks focus on predicting individual sentiment elements, such as aspect term extraction and aspect sentiment classification. In contrast, compound tasks jointly identify multiple elements and their relationships, including pairing aspect terms with opinions and predicting triplets of aspect, opinion, and polarity, which these compound tasks offer a more integrated view of sentiment (Zhang et al., 2023). However, there are some key challenges in ABSA, including balancing unified models with computational complexity, ensuring robustness against implicit aspects and ironic language, and enhancing interpretability for extracted opinions.

2.5.3.3 Comparison Between Latent Dirichlet Allocation and Aspect-Based Sentiment Analysis Techniques

In the field of tourism analytics, particularly when analysing unstructured and multilingual platforms like YouTube, effectively identifying key aspects is crucial for understanding public sentiment regarding various dimensions of tourist experiences. Therefore, there are two prominent techniques, which are LDA and ABSA, that provide complementary benefits in this context. Basically, LDA reveals underlying thematic patterns in textual data through unsupervised learning, while ABSA offers a more nuanced analysis by extracting specific aspects and assessing the associated sentiment for each one. Following that, this section explores both techniques, as shown in Table 2.11, to illustrate the advantages of their integration for a more comprehensive semantic and SA.

Table 2.11

Comparison Between Latent Dirichlet Allocation and Aspect-Based Sentiment Analysis

Criteria	LDA	ABSA
Main Objective	Uncover latent thematic structures.	Identify aspect terms or categories

Criteria	LDA	ABSA
		and determine sentiment polarity.
Input Requirement	Unlabelled textual data.	Often requires annotated data for supervised learning.
Output	Distribution of topics and words per document.	A combination of aspect and sentiment.
Strength in Unstructured Data	Highly effective on noisy, unlabelled, or multilingual text.	Moderate since it depends on labelled examples or pre-defined rules.
Dimensionality Handling	Reduces dimensionality using probabilistic topic distribution.	Works with raw text or feature-engineered input.
Aspect Granularity	Often detects broader and implicit aspects.	Captures fine-grained and explicit opinion targets.
Scalability	Highly scalable for large corpora.	May require more computational effort for compound tasks.
Sentiment Integration	Requires additional models to infer sentiment for topics.	Directly classifies sentiment for identified aspects.
Key Challenges	Topic coherence, optimal topic number selection.	Handling implicit aspects, sarcasm, and domain adaptation.

To sum it up, the comparison indicates that LDA is particularly effective at uncovering topics within large, unstructured datasets, making it an excellent choice for initial analysis and aspect generation in environments with limited data, such as YouTube tourism reviews. Conversely, ABSA provides precise sentiment mapping for specific aspects, which enhances our understanding of user sentiments. As a result, the integration of LDA and ABSA forms a promising approach as LDA identifies latent topics or potential aspects, which are then refined and sentiment-tagged using ABSA techniques. This combined strategy leverages the strengths of both models, enabling a comprehensive extraction of semantic and sentiment insights from unstructured tourism-related content.

2.6 Insights Gained from Literature Review

The research on SA in tourism and social media, notably YouTube, provides a clear progression of analytical techniques. Lexicon-based techniques were applied in previous studies due to their simplicity and interpretability. Nevertheless, sarcasm, domain-specific jargon, and context-dependent phrases tend to be challenging for these

types of techniques. This has resulted in an increasing tendency toward ML, DL, and hybrid frameworks that combine different techniques to increase sentiment classification accuracy. The rise and growing recognition of the hybrid SA approach combine supervised learning models such as SVM, NB, LR, and DL models like LSTM networks with lexicon-based scoring tools. This approach has proven very beneficial when handling noisy or unstructured data from online sources. Based on the literature review, the research indicates that lexicon-based techniques are straightforward to implement and interpret, but struggle with adaptability in complex linguistic contexts. In contrast, ML models demonstrate high accuracy but depend on high-quality labelled data, while DL models often yield superior performance but require substantial computational resources.

On top of that, this chapter aligned with the research objective (**RO 1**) as the research analyses DSA techniques and emphasises their role in enhancing SA within the tourism sector, particularly through the integration of TF-IDF and LDA as strategic techniques for transforming unstructured YouTube review data into structured and interpretable features. In descriptive analysis, factual information is clearly and concisely summarised and presented (A. S. Rawat, 2021), while semantic analysis revolves around extracting and interpreting the underlying meaning within natural language (Kanade, 2022). In other words, descriptive analysis offers a quantitative summary of the data, while semantic analysis delves into qualitative aspects and uncovers deeper contextual interpretations (McCombes, 2019). This chapter also underscores a significant research gap concerning the underutilisation of YouTube as a source for SA in the tourism sector. Thus, through the application of these DSA techniques, this research identified their strengths and limitations, thereby paving the way for a more effective hybrid SA approach tailored to YouTube review data. Table 2.12 summarises the comparison of these techniques as they play a crucial role in defining the scope of this research.

Table 2.12

Comparison of Descriptive and Semantic Analysis

Criteria	Type of Analysis	
	Descriptive	Semantic
Definition	Summarising and expressing facts in a	Extraction and interpretation of the

Criteria	Type of Analysis	
	Descriptive	Semantic
	concise and lucid presentation (A. S. Rawat, 2021).	underlying meaning within natural language (Kanade, 2022).
Data Type	Emphasises the quantitative attributes of data (McCombes, 2019).	Delves into its qualitative dimensions (McCombes, 2019).
Differences	Collecting numerical data for statistical techniques to identify patterns, trends, and correlations through measurement without altering variables (Luo, 2019).	Collecting non-numerical data that resists statistical analysis but yields qualitative insights by elucidating the relationships between words and concepts (Luo, 2019).

To sum it up, the lack of YouTube as a data source, multilingual data, and a domain-specific framework are among the research gaps. Hence, by combining TF-IDF, LDA, VADER, and SVM, this framework enhances feature extraction, improves the handling of unstructured YouTube reviews, and strengthens overall sentiment classification performance. This complementary approach therefore contributes meaningfully to reducing the limitations of conventional S A techniques and provides a more robust foundation for analysing tourism-related user-generated content. To the best of the researcher's knowledge, no existing study has combined DSA techniques with a hybrid SA approach specifically for Malaysia's island tourism using YouTube reviews.

2.7 Summary

This chapter reviewed SA techniques and their application in tourism research, particularly using YouTube reviews. It examined lexicon-based, ML, DL, and hybrid approaches, as well as DSA. Each technique has advantages and limitations, such as lexicon-based approaches are simple but context-limited, ML balances performance and interpretability, DL captures deeper semantics but requires resources, while hybrid approaches combine strengths for more robust classification. The review also highlighted the role of DSA through TF-IDF and LDA to extract key features and uncover thematic insights. In tourism contexts, this integration improves sentiment interpretation in multilingual and informal reviews. Research gaps were identified, including the limited use of YouTube as a data source, insufficient evaluation of multilingual data, and a lack of domain-specific frameworks. In summary, SA in

tourism can be enhanced by integrating DSA with hybrid approaches to achieve higher accuracy and contextual relevance. These findings form the foundation for the methodology in Chapter 3, which details data collection and pre-processing, hybrid implementation, as well as model evaluation and validation.

CHAPTER 3

RESEARCH METHODOLOGY

Chapter 3 consists of a methodology section that provides a systematic approach to problem-solving, which leads to the analysis of data collection, the selection of techniques and the evaluation of findings. In order to fill in the research gaps, this chapter outlines the research tasks and methodologies that align with the research objectives. In general, this research provides a comprehensive knowledge of user perspectives by integrating descriptive-semantic analysis (DSA) techniques along with lexicon and machine learning (ML)-based such as Valence Aware Dictionary and Sentiment Reasoner (VADER) and Support Vector Machine (SVM) as a hybrid sentiment analysis (SA) approach.

3.1 Research Design Framework

This research integrates hybrid sentiment classification with DSA techniques to extract meaningful insights from YouTube reviews about Malaysia's islands. The research design framework outlines a structured methodology for the model framework, which consists of six phases, as shown in Figure 3.1, each focusing on specific tasks and outcomes. The first phase involves assessing the problem assessment and research study, which includes a literature review to define the research problem, objectives, questions, scope, and contributions. In the second phase, DSA techniques are selected to analyse data patterns and extract topics from text data.

In the third phase, data collection and pre-processing involve gathering relevant YouTube reviews about Malaysia's islands from a specific timeframe. During the data pre-processing step, the data undergoes translation, cleaning, and normalisation. Following this, the pre-processed data provides significant insights into term and topic relationships through the integration of Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Dirichlet Allocation (LDA) as DSA techniques in the fourth phase, supporting subsequent sentiment classification tasks. This phase also includes manual aspect labelling based on the distribution of terms and topics. Labels were assigned through manual qualitative analysis of the predominant terms identified within each

topic generated by LDA. This approach ensured that each label accurately represented the thematic content associated with its respective topic.

In the fifth and sixth phases, VADER is utilised for sentiment polarity detection, and SVM is applied for sentiment classification. Additionally, model performance metrics such as precision, recall, and Fl-score are evaluated. To ensure a balanced data distribution, stratified *k-fo\d* cross-validation is implemented, as maintaining class balance across folds is essential for imbalanced sentiment datasets. Furthermore, model validation can be significantly enhanced through the utilisation of the Area Under the Curve (AUC) metric. This metric provides a thorough evaluation of the model's capability to effectively distinguish among various sentiment classes. Therefore, this hybrid framework offers a well-structured methodology that integrates term and topic classification with lexicon-based and ML techniques to enhance SA of YouTube reviews.

Besides that, this research design framework also encompasses three primary research objectives. First, it aims to analyse DSA techniques used in the SA related to the tourism sector by identifying their strengths and weaknesses. Second, it seeks to construct a hybrid approach by integrating DSA techniques, such as TF-IDF and LDA, with VADER and SVM. Finally, the hybrid framework evaluates the accuracy of the constructed hybrid SA approach using performance metrics and validates it using the AUC metric, ultimately contributing to a more effective SA framework for understanding tourist perceptions of Malaysia's islands.

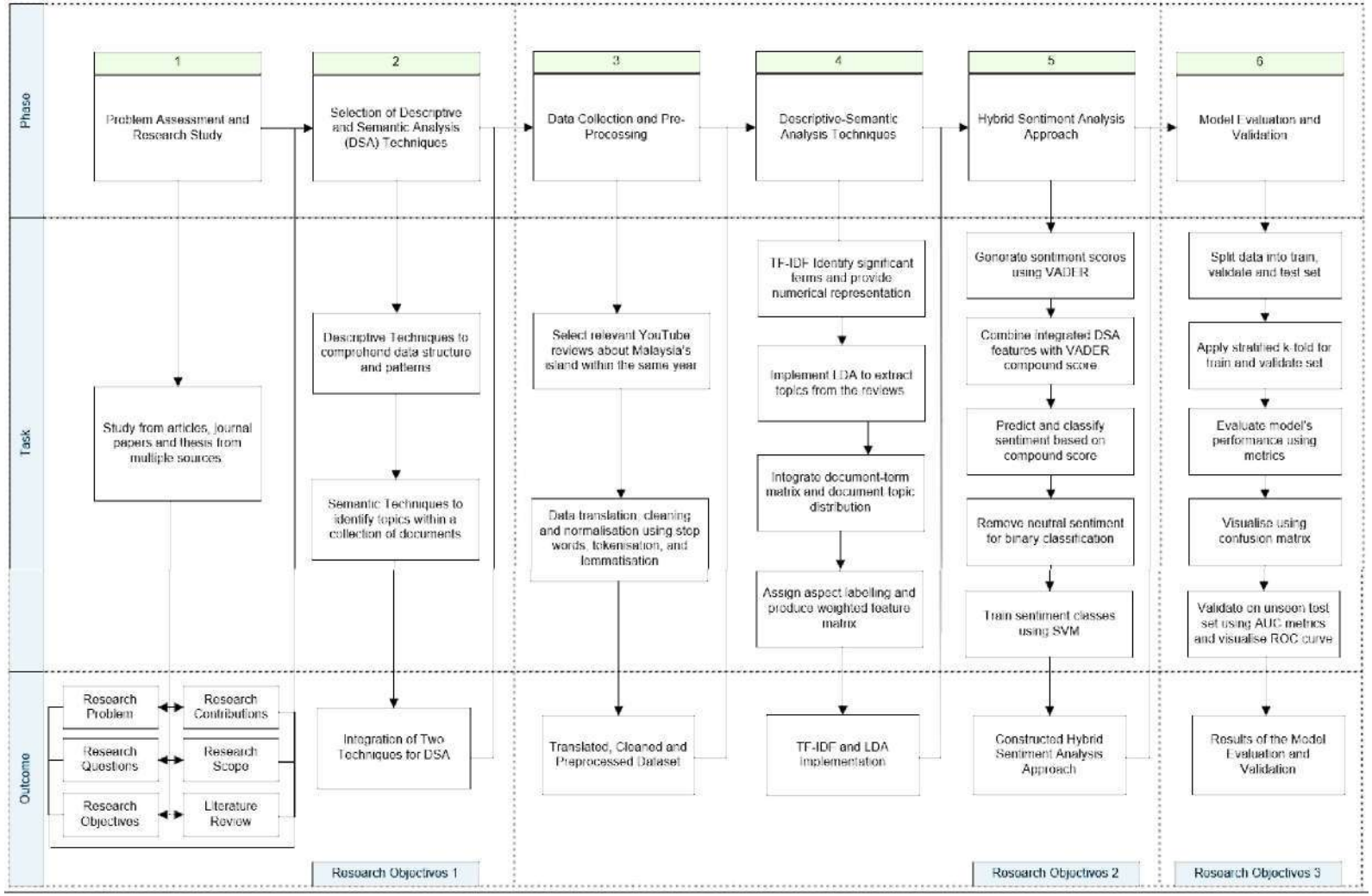


Figure 3.1 Research Design

Overall, the research design consists of six phases, each aligned with the RO presented in the previous chapters. As illustrated in Table 3.1, the first two phases address RO 1, analysing DSA techniques through a literature review and selecting techniques like TF-IDF and LDA for DSA implementation. Following that, Phases 3 to 5 focus on RO 2, where a hybrid approach is constructed by integrating TF-IDF and LDA with VADER and SVM model. Finally, Phase 6 achieves RO 3 by implementing stratified *k-fo\d* validation and a confusion matrix to evaluate performance metrics, while AUC metrics are used to validate unseen data. Collectively, these phases establish a comprehensive strategy for analysing tourist perceptions of Malaysia's islands.

Table 3.1

Alignment of Techniques to Research Objectives

Research Objectives	Methodological Phase	Techniques Used	Justification
RO 1: To analyse the existing descriptive and semantic techniques used in analysing the SA related to Malaysia's island tourism.	1&2	Literature review.	Analyse DSA techniques used in previous research studies.
RO 2: To construct a hybrid approach integrating DSA in improving the SA of YouTube reviews about Malaysia's island tourism.	3,4&5	TF-IDF, LDA, VADER, SVM.	Integrate DSA techniques with a hybrid SA approach using lexicon and ML
RO 3: To evaluate the accuracy of the constructed hybrid SA approach using performance metrics and validate using the AUC metric.		Performance metrics, stratified Mold validation, confusion matrix, AUC metric.	Evaluate using accuracy, precision, recall, Fl -score and validate using unseen data.

3.1.1 Problem Assessment and Research Study

This section provides a comprehensive description of the research, including analysis and discussion of the findings from journal publications, articles and thesis from various sources, as shown in Figure 3.2. It pertains to the topics discussed in this research, including SA, DSA techniques, and other relevant fields. Following that, a

literature review, problem statements, research questions, objectives, scope and significance are among the deliverables acquired during this methodology phase. Additionally, the structured research approach offers several advantages, including the development of a systematic research framework, an in-depth understanding of prior studies, identification of research trends and gaps, and a clear delineation of the research scope and objectives.

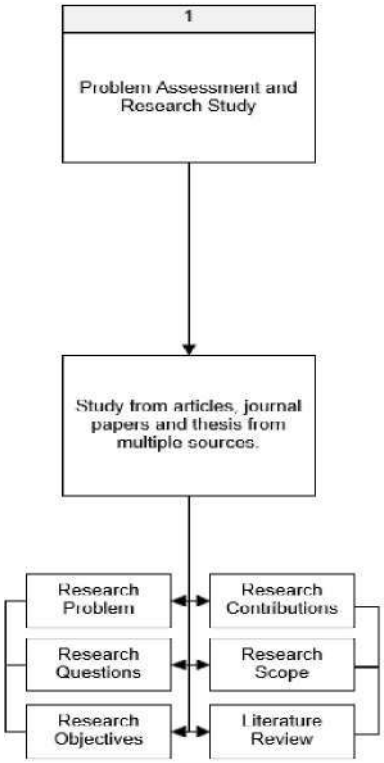


Figure 3.2 Research Design Phase One

This phase also identifies the challenges associated with conducting SA on informal, multilingual, and unlabelled YouTube tourism reviews. A literature review was performed to pinpoint gaps and select appropriate techniques to tackle these challenges. Chapter 1 outlines the problem statements, research questions, objectives, and significance of the research, focusing on using SA to understand tourist perceptions and applying DSA techniques to analyse sentiments in YouTube reviews. The chapter also identifies three main research problems (RPs): 1) linguistic ambiguity due to informal and unstructured language, 2) cultural and multilingual diversity that affects sentiment interpretation, and 3) the lack of labelled data, which undermines the reliability of ML models. In addition, Chapter 2 examines various SA techniques and

provides an overview of the theoretical and practical fundamentals of DSA methodologies within the context of SA.

Building upon the three main RPs identified in Chapter 1, these challenges are addressed by implementing DSA techniques such as TF-IDF and LDA, in combination with a hybrid SA approach. Integrating the latent topics from LDA with insights from TF-IDF enables the identification of key themes and trends within the data. The implementation of the TF-IDF technique clarifies data patterns associated with specific topics, while LDA highlights significant themes concerning Malaysia's islands through the analysis of frequently used terms. Moreover, a hybrid SA approach that integrates lexicon-based insights with ML models can enhance the model's ability to capture contextual sentiment patterns, thereby enhancing classification accuracy.

The first challenge of RP 1 is linguistic ambiguity from informal and unstructured language. This challenge can be addressed using TF-IDF, which identifies significant terms by diminishing the impact of commonly used words and emphasising more informative, less frequent terms. Thakkar and Chaudhari (2020) highlight how TF-IDF effectively pinpoints relevant words and decreases data complexity through term weighting to enhance interpretability. Additionally, selecting terms with high TF-IDF scores aids in improving the interpretability of textual data and provides a clearer representation of document content. Meanwhile, the second challenge of RP 2 pertains to the limited ability of SA techniques to capture cultural and multilingual diversity. For instance, Ali et al. (2022) investigated TripAdvisor reviews through LDA and lexicon-based SA. However, they limited their dataset to English reviews and lacked cross-cultural comparisons.

The third challenge of RP 3 concerns the lack of labelled data necessary for effectively training supervised learning models for reliable sentiment classification. A case study from Octava et al. (2023) used labelled data to enhance SA model training and prediction accuracy by combining statistical features with sentiment classifications. Therefore, this research implements a hybrid SA approach that utilises VADER and SVM to mitigate the limitations posed by the lack of labelled data in SA models, as VADER can generate sentiment labels automatically, which are then used by SVM for classification to enhance its ability to learn from data patterns and improve sentiment accuracy. As a result, these techniques reduce the need for manual labelling, making SA more reliable and efficient, even with limited labelled data.

3.1.2 Selection of Descriptive and Semantic Analysis Techniques

Additional text analysis techniques are essential to obtain both structural and contextual insights from complex textual data, particularly in SA tasks. Therefore, DSA techniques are applied, as both approaches play crucial roles in feature extraction and sentiment classification. Descriptive techniques are used to understand data structures and patterns, while semantic techniques analyse the contextual relationships between terms and phrases to determine topics. Figure 3.3 illustrates this process to provide a more comprehensive overview of the data.

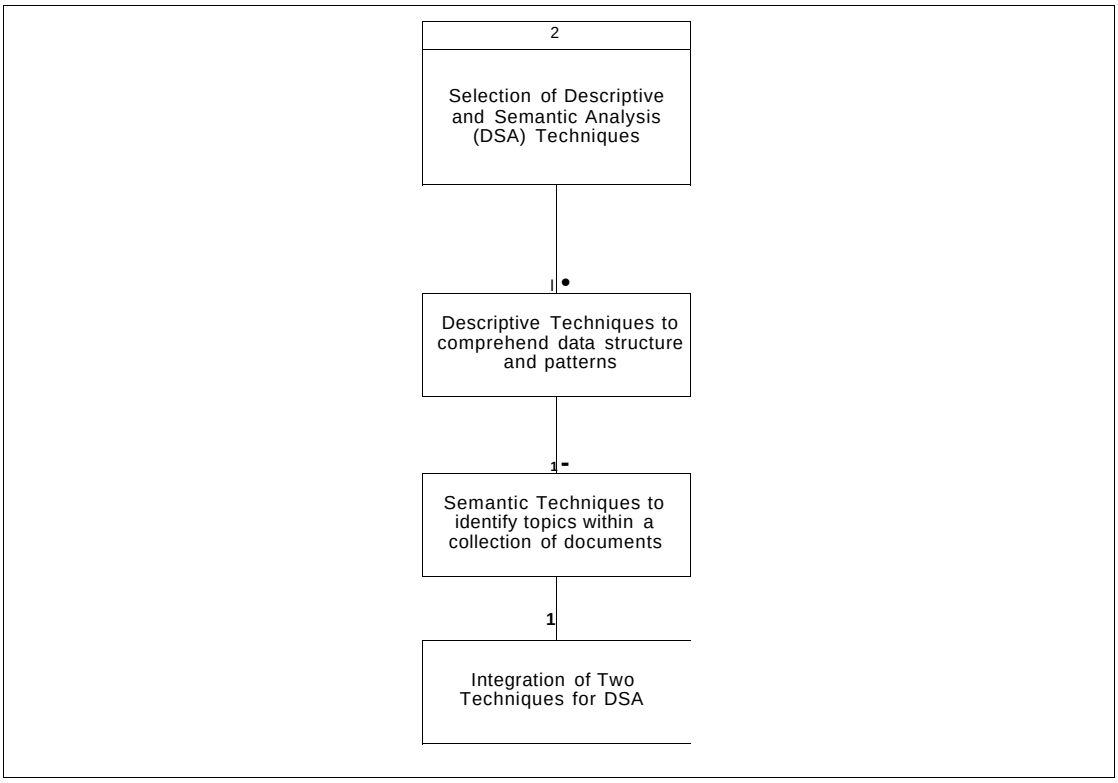


Figure 3.3 Research Design Phase Two

This phase also directly addresses the first RO, which focuses on exploring DSA techniques applied in SA to identify their respective strengths and limitations. For DSA implementation, TF-IDF was chosen for descriptive analysis to identify frequent terms and relevant patterns, while for semantic analysis, LDA was selected to extract contextual meaning from YouTube reviews. The integration of these techniques enhances the efficiency and accuracy of feature extraction and sentiment classification within the SA tasks. The following section will provide more details on the implementation of TF-IDF and LDA.

3.1.2.1 Term Frequency-Inverse Document Frequency for Descriptive Analysis Technique

TF-IDF is a descriptive analysis technique that effectively extracts significant features from textual data, including sentiment distribution and word frequency patterns. This technique focuses on aspects receiving negative and positive reviews by identifying significant words and providing a numerical representation. Moreover, TF-IDF scores can also rank terms based on their relevance within a corpus (S. W. Kim & Gil, 2019), thereby enhancing text mining tasks such as topic detection and SA. Hence, TF-IDF can identify significant terms in documents and provide a descriptive representation of text data for analysis by comparing terms across documents and reducing the dimensionality of text data.

For TF-IDF implementation, the term frequency (TF) computation step counts the occurrences in the data structure of the topics extracted by LDA and the terms specified in a term lexicon. The TF implemented in this research can be defined as in Equation 3.1, where ny denotes the number of repetitions of the word ti, which appears in document dj, whereas the sum over k of n sub, k, as $Y.k^{n}k,j$ is the total number of word occurrences in document dj. Also, k and D represent the number of terms and documents, respectively.

In addition, the document frequency (DF) indicates the total number of times the term appears throughout the collection of documents, whereas the TF represents the number of times a term appears in a single document. For the step during the DF, the total number of documents is divided by the number of documents that contain a certain term within them.

$$\bar{DF}_{U} = \frac{\backslash d;£}{J} \quad \frac{D{:}t;}{|D|} \quad /\text{—} \frac{£d;\backslash}{J} \qquad (3.2)$$

Next, Equation 3.2 demonstrates that |D| represent the total number of documents and dj G D: tj G dj can be determined by the number of documents in which the term tj appears. However, if the terms are frequently seen in a wide range of documents, terms with a high DF score are considered common. Thus, when evaluating

the significance of terms within the collection of documents, the inverted document frequency (IDF) is used, as illustrated in Equation 3.3.

$$IDF_{u} = log \frac{\quad}{\quad} \quad (3.3)$$

$$TFIDF = TF \times IDF \quad (3.4)$$

Thus, by combining Equations 3.1 and 3.3, the TF-IDF score is derived as shown in Equation 3.4. A specific keyword's TF-IDF score increases when it occurs frequently in a document but remains rare in every document featuring it. Consequently, finding terms that appear consistently in documents can be accomplished through this technique (S. W. Kim & Gil, 2019). Furthermore, TF-IDF is used in text analysis to identify the most significant words or terms within a set of documents. It helps determine the relevance of certain topics by summing the TF-IDF scores across documents. Therefore, topic weighting is used to weigh the importance of topics, with higher scores indicating more relevance or frequent discussion. Documents with similar top-weighted terms can be grouped, representing shared topics. In most cases, topics are weighted by adding together all of the terms from TF-IDF scores in all documents that are relevant to that topic. This provides a weight to the topic, revealing its overall significance. For instance, the weight for a topic in a document, *d,* can be determined as Equation 3.5, assuming a collection of related terms as {ti, *ti,* ..., tk}.

$$lf_{=1} TFIDF_{tud} \quad (3.5)$$

In conclusion, TF-IDF is a simple yet powerful technique for evaluating the importance of terms and identifying key topics within a collection of documents, making it particularly valuable for feature extraction in SA. It measures the significance of a term within an individual document and its relevance across the entire set of documents. This technique also encompasses feature extraction, topic modelling, document clustering, and information retrieval. By using TF-IDF, text analysis is enhanced, allowing for better insights into key topics and making it easier to manage large volumes of data.

3.1.2.2 Latent Dirichlet Allocation for Semantic Analysis Technique

Conversely, semantic analysis techniques such as LDA identify topics within text documents, offering insights into underlying themes and contextual meanings. LDA is a probabilistic model representing the relationships between words and topics by capturing semantic structure and enhancing interpretability for quantitative and qualitative analysis (Pattnaik et al., 2023). Besides that, LDA is often a preliminary step for topic context in the SA process. It enriches text data with topic labels and allows a structured approach to understanding user opinions. LDA categorises text into meaningful topics, which can then serve as a foundation for SA to measure user sentiments (Ali et al., 2022).

Moreover, the LDA model generates topics for each term using document-topic and topic-word distributions, representing each document as a distribution over topics identified by sets of words and their probabilities (Nanyonga et al., 2023). In Figure 3.4, the diagram shows that topic and word distributions tend to be features of a document by LDA as it is utilised to determine the topic probability distribution 9 and word probability distribution (p per topic. During the document creation process, word distribution per topic (pk and topic distribution 9 of document samples are obtained from the Dirichlet distribution with hyperparameters P and a, respectively. Next, the latent variable Z_n represents the topic assignment per word and is generated by a multinomial distribution with parameter 9, whilst (pk follows the Dirichlet distribution with parameter P, ultimately yielding the word W_n. Thus, by mapping topics and corresponding documents, LDA can identify topics that most frequently appear throughout the corpus. Next, data collection and pre-processing are the subsequent steps.

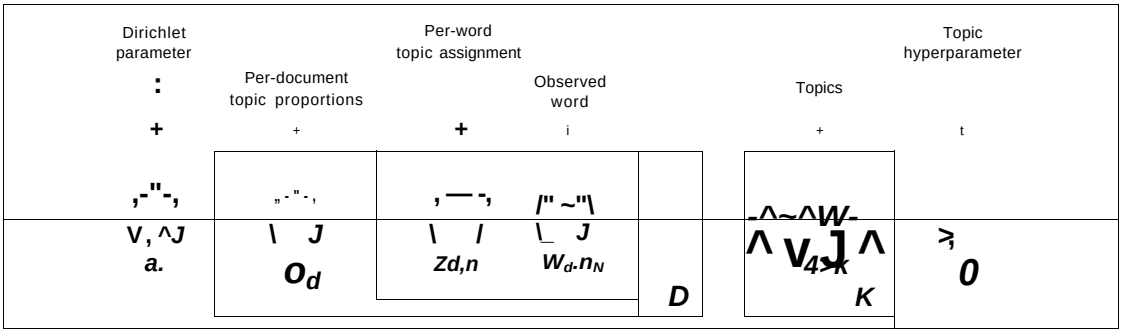


Figure 3.4 Latent Dirichlet Allocation Topic Modelling
(Source: D. Kim et al., 2019)

3.1.3 Data Collection and Pre-processing

For data collection, YouTube reviews serve as a valuable source for S A, offering rich textual content and diverse user perspectives, particularly in the context of tourism. Following that, the data translation, cleaning and normalisation using tokenisation, stop word removal, and lemmatisation are used to prepare the text for analysis, as illustrated in Figure 3.5. These steps facilitate reducing data dimensionality, improving classification accuracy, and enhancing feature extraction for SA.

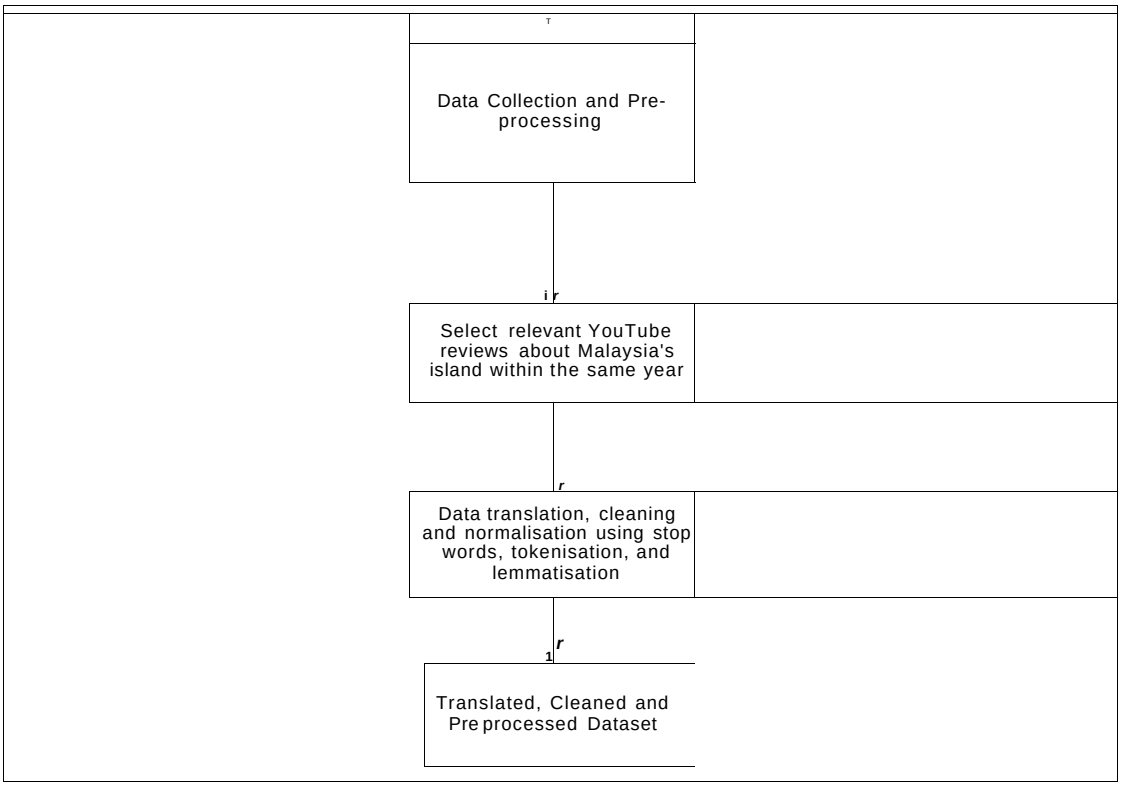


Figure 3.5 Research Design Phase Three

3.1.3.1 Data Collection

YouTube reviews enable the analysis of user engagement, serving as an alternative to surveys for assessing public sentiment (R. Singh & Tiwari, 2021). Hence, the YouTube platform is used to extract pertinent reviews from Malaysia's islands like Langkawi, Perhentian, and Pangkor, within the same year from 1 January 2023 to 31 December 2023. The year 2023 was chosen because it reflects the post-pandemic tourism recovery with eased travel restrictions and more stable tourist trends. Additionally, since this research was conducted in 2024, utilising the complete dataset

from 2023 ensures that all reviews are finalised and publicly accessible, thereby enhancing the integrity of the analysis. Moreover, the YouTube Data API (Application Programming Interface) was utilised to extract user reviews from a YouTube video by obtaining an API key from the Google Cloud Console. After that, the API v3 was activated, and credentials were generated to establish an API connection using Python's 'googleapiclient.discovery' module (Pranayteja, 2023).

The data extraction process is crucial for analysing sentiment to ensure that it is relevant for insightful analysis. By utilising the YouTube Data API, the pseudocode described in Figure 3.6 demonstrates the extraction of user-generated reviews from specific video IDs obtained from YouTube links. This process effectively gathers top-level reviews and their corresponding replies to enable thorough data collection. It begins by initialising a dictionary to categorise reviews by author while tracking the total count of reviews and replies. The method makes API calls to collect top-level reviews for a specified video ID and implements retries to address any issues that arise. Subsequently, replies for all top-level reviews are collected using a nested loop with pagination to ensure that no data is overlooked. This loop continues until all replies are fetched, at which point it advances to the next page of top-level reviews.

```
Algorithm 1 Retrieve All Comments and Replies from a YouTube Video
Require: videoJd
Ensure: commenfs_by_author, totaLcount
 1  Initialise comment&.by-CUtthor as empty dictionary
 2  total-comments <— 0
 3  totaLreplies <— 0
 4  repeat
 5      Retrieve batch of top-level comments for vidm.id
 (i     for each top-level comment do
 7          Store comment, in coiniiicidnJnj.auUior
 8          totaLcomments <— total-comments + 1
 9          repeat
10              Retrieve batch of replies for this comment
11              for each reply do
12                  Store reply in comments-bif-author
19                  totaLreplies <— totaLreplies + 1
14              end for
15          until no more replies
10      end for
17  until no more top-level comments
18  return commtmLsJnj-author, totaLcomments + totaLreplies
```

Figure 3.6 Pseudocode for Data Extraction Using YouTube API

As a result, a total of 43 tourism-related YouTube videos were selected from three islands to create a comprehensive dataset for sentiment analysis. For each video, both top-level reviews and replies were extracted to ensure complete coverage of user interactions. Langkawi contributed the largest number of videos, totalling 17, along with 1,965 reviews. Perhentian followed with 10 videos and 1,760 reviews, while Pangkor had 16 videos and 1,396 reviews. In total, the dataset includes 5,121 reviews and replies, representing a diverse collection of user-generated content that reflects tourist experiences, perceptions, and sentiments across three popular Malaysian island destinations. This extensive dataset provides a solid foundation for conducting DSA feature extraction and hybrid sentiment classification in the subsequent stages of the research.

Table 3.2

Total Extracted YouTube Reviews for All Islands

Type of Island	Number of Videos	Total Reviews + Replies
Langkawi	17	1965
Perhentian	10	1760
Pangkor	16	1396
Overall Total	43	5121

Moreover, the selection of YouTube videos varies among Langkawi, Perhentian, and Pangkor due to differences in the availability of user-generated content suitable for SA. Langkawi, being a prominent tourist destination, boasts a larger number of videos featuring meaningful reviews and clear sentiment. In contrast, Perhentian and Pangkor have fewer videos that meet the necessary criteria, often lacking substantial reviews or audience engagement. Additionally, resource constraints, including time and computational capacity, limited the collection of high-quality videos from Perhentian and Pangkor. Consequently, the final selection reflects variations in content availability while prioritising videos that provide valuable sentiment for analysis. Once the videos were selected and the reviews extracted, the dataset was organised for pre-processing due to the unrefined YouTube reviews containing noise, inconsistencies, and non-standard language, which made thorough pre-processing essential for ensuring data quality and reliability.

3.1.3.2 Data Pre-processing

Data pre-processing is a critical process that encompasses essential steps such as translation, cleaning, and normalisation, ensuring sentiment nuances are preserved during language conversion. These steps are essential for effectively preparing the raw text for thorough analysis and ensuring that it is suitable for the subsequent SA process. After collecting data for research by extracting YouTube reviews of Malaysia's islands, the next step is data pre-processing. This phase is crucial for standardising the text data to prepare it for analysis.

a) Data Translation

Malaysia is a culturally diverse and multilingual nation, featuring Malay as the official language, with English and Chinese widely used. This linguistic richness allows YouTube reviews of the country's beautiful islands to be expressed in multiple languages, capturing a broader range of opinions and emotions for a nuanced understanding of user perceptions. To address this issue, the text is pre-processed to identify and normalise English reviews using the Google Translate library, ensuring consistency throughout the dataset. This process captures a more comprehensive representation of user opinions across different linguistic backgrounds.

During the translation step, as illustrated in Figure 3.7, the reviews published in other languages like Malay, Chinese, Tamil, Russian, Spanish, Japanese and Korean are translated into English using the Google Translate library. The process begins with a thorough verification of reviews that exceed 5,000 characters, ensuring they are effectively reduced and filtered to eliminate any incorrect or empty text entries. To uphold consistency, the reviews are converted to lowercase before translation. In the event that the translation process fails after multiple attempts, the original text will be preserved, and an error notification will be provided.

```
Algorithm 2 Dat;i Translation
Require: text, retries
Ensure: translatedLtext
 1  Initialise Google Translator (ante —i ErigliKh)
 2  for eiich attempt up to retries do
 3      if text is invalid (empty, NaN, non-Htring) then
 •1         return original text
 5      end if
 fi     Preprocess text: lowercase, truncate to 5000 characters
 7      Translate text
 8      if non-ASCII characters detected then
 9          Log mixed-language warning
10      end if
11      return translated text
                                        t> If tmriKkition ffiilw. wait and retry
12: end for
ID: Rcituni original text if fill retries fail
```

Figure 3.7 Pseudocode for Data Translation Pseudocode

Figure 3.8 shows an example of the process of converting multilingual reviews into English for SA on a video about Perhentian Island. Languages such as German, English, and Thai were used to write the initial version reviews, which were then translated into English for cross-linguistic analysis. For instance, a German review like *"Hi, sehr schone footage von den perhantian Ins.."* was translated, which expresses a positive opinion of the Perhentian islands' attractiveness. Meanwhile, "*MUiiiuuhimoMUoanoaoa 'iBuximiT* was translated from Thai, implying satisfaction of having a pleasant vacation without alcohol. In general, the translated reviews provide an insight into the opinions and experiences of the user, as well as pointing out the increasing prominence of Perhentian Island destinations on social media globally. In addition, translations preserve the original text's meaning, including informal expressions, slang and emojis. However, challenges may arise due to the local dialects and informal expressions.

Perhentian data -for video_id: s©sML~tNnEUk

	text	translated text
466	Hi, sehr schone footage von den perhantian Ins...	hi very nice footage from the Perhantian Isla...
591	Thank you Ken for your honest review	thank you ken for your honest review
632	War mir zuerst nicht sicher, ob ich die Perhen...	At first I wasn't sure whether I should includ...
709	i.wuTuuutud'atfimaanEiaasi lEJEiufi^ii	Travel without alcohol. Great!
767	Hallo, Ken. Was fiir eine tolle Insel I Bin ers...	hello, ken. what a great island! I only came a...
976	Sehr schones Video! Danke fur mitnehrnen und un...	very nice video! thanks for taking us along an...

Figure 3.8 Example of Data Translation Output

b) Data Cleaning

After the translation step, the data undergoes a cleaning phase, where irrelevant elements are removed, resulting in a refined and streamlined dataset. Irrelevant elements like usernames, mentions, uniform resource locators (URL), and special characters are removed to improve data accuracy and prevent interference in subsequent analysis stages (Alawi & Bozkurt, 2024). The process of data cleaning in SA involves removing redundancies, managing missing values, reducing unrelated topics and handling inconsistencies. The data cleaning process involves preparing the text to remove irrelevant elements, as illustrated in Figure 3.9.

```
Algorithm 3 Data (/leaning
Require: text
Ensure: cli>ancd.text
1: Replace predefined shorthand words with full forms
2: Remove filler and laughing words (e.g., haha, lol, wow)
.1: Normalise case to Lowercase
<1: Remove HTML tags, entities, and line breaks
5: Remove sequences of dots, special characters, numbers
6: Remove URLs and hashtags
7: Remove extra whitespace
8: Remove any substring starting with "@"
9: return cleancd_text
```

Figure 3.9 Pseudocode for Data Cleaning

The process starts with the input being divided into individual terms, while shorthand conversions are applied using a predefined dictionary to eliminate unnecessary terms. Additionally, terms are replaced whenever a match is found in the dictionary. Once the cleaning is complete, the terms are reassembled into a cohesive string, and the common expressions of laughter are also removed. Moreover, several cleaning steps are used to eliminate noise, where the reviews are converted to lowercase, and leading and trailing spaces are trimmed. Specific Hypertext Markup Language (HTML) elements, dot sequences, special characters, numbers, URL, hashtags, unnecessary whitespace and usernames are also removed.

Figure 3.10 illustrates the example of data cleaning, which involves removing punctuation, special characters, excessive spacing, slang, converting text to lowercase, and eliminating emojis and symbols. For instance, the claim like *"Inpangkor, it's most popular in Teluk Nipah..."* was cleaned through eliminating the comma, apostrophe,

and uppercase letters for easy comprehension, resulting in *"in pangkor its most popular in teluk nipah..."*. Likewise, a review on *"Thankyoufor watching. Don't forget to check ..."* but without the emoji.

	translated_text	cleaned_text
43	we can drive the car to take ferry? or you *ten...*	we can drive the car to take fern/ or you rent...
50	you can drive to either marina or lumut jetty....	you can drive to either marina or lumut jetty ...
51	In Pangkor, it's most popular in Teluk Nipah b	in pangkorits most popular in teluk nipah bee
52	hi antonio, the most happening area in pangkor	hi antonio the most happening area in pangkor
53	thank you nabilah. imsyallah, more travel vide...	thank you nabilah imsyallah more travel videos...
54	great, welcome to pangkor. hope you enjoy your...	great welcome to pangkor hope you enjoy your t...
55	Greetings from Malaysia >* MCMV	greetings from malaysia
5S	Thank you Tor watching --. Don't forget to check...	thank you for watching dont forget to check ou...
57	no brother, someone else's car rental	no brother someone elses car rental
5G	Thank God., more and more people are buying Pa...	thank god more and more people are buying pang...

Figure 3.10 Example of Data Cleaning Output

c) Data Normalisation

Data normalisation is the process of transforming unstructured text into a standardised representation for SA to eliminate irrelevant noise while preserving meaningful linguistic structures. It also involves several data pre-processing processes through natural language processing (NLP) techniques such as tokenisation, stop word removal, and lemmatisation, where it separates text into distinct components, eliminates frequently encountered words, and simplifies words to their basic form (R. Singh & Tiwari, 2021). Additionally, unlike stemming, lemmatisation preserves the meaning of words while reducing them to their dictionary forms (Alawi & Bozkurt, 2024).

Furthermore, the purpose of data pre-processing is to enhance its significance and relevance for analysis through text processing, particularly focusing on slang correction and review filtering. This step is crucial for SA tasks, ensuring that the data is focused on island-related content. As illustrated in Figure 3.11, the process begins with the implementation of a slang correction dictionary, which normalises informal language found in reviews. Each review is broken down into individual words and replaces slang terms with their normalised equivalents. Besides that, predefined keywords such as *"thank"*, *"subscribe"*, and *"hello"* are used by the 'is_relevant' function to identify and filter out reviews that are unrelated to island content.

Algorithm 4 Slang firiti Irrelevant Reviews

Require: text
Ensure: corrected.tcxt
1: Replace slang terms in text using predefined dictionary
2T return corrected .text
Require: text
Ensure: is_relcvant
'h **If** text contains non-island-related keywords (e.g.. "thank". "subscribe", "hello"), return False
4: Else return True

Figure 3.11 Pseudocode for Slang and Irrelevant Reviews

The final step involves a data pre-processing function called 'preprocesstext', which simplifies the text and utilises NLP techniques such as tokenisation, stop word removal, and lemmatisation, as depicted in Figure 3.12. This function ensures that the input is a string, preventing issues related to incorrect data types. When the data is valid, it tokenises the text into words using a pre-trained NLP pipeline. Only significant tokens are retained after removing stop words and punctuation. Additionally, each token is lemmatised before being reassembled into a single string for further analysis. If the provided data is not a string, the function skips processing and issues a warning.

Algorithm 5 Data Pre-processing
Require: text
Ensure: proeessed.tcxt
l: if text is a valid string then
2T Lemmatise text using NLP model
:i: Remove stopwords and punctuation
4: Return defined, lemmatised string
5: else
fi: Return empty string
7: end if

Figure 3.12 Pseudocode for Data Pre-processing

Figure 3.13 illustrates the pre-processed reviews of Pangkor Island. The 'cleanedtext' column shows semi-structured sentences resulting from initial data cleaning steps, while the 'processedtext' column contains modifications made through keyword extraction and text normalisation. This step focuses on condensing the reviews into a format that is both concise and semantically meaningful. For instance, the claim *"we can drive the car to take ferry or you rent... "* was transformed into *"drive car ferry rent car pangkor"* by eliminating unnecessary phrases while maintaining the main

purpose and context. An additional example is *"easy pangkor jetty people offer ferry service"* as a result of simplifying into the important keywords, which can be useful for topic modelling. Hence, this step eliminates noise, reveals important information and prepares data for more complex analytical tasks.

	cleaned text	processed text
37	we can drive trie car to take ferry or you rent...	drive car ferry rent car pangkor
3G	you can drive to either marina or lumut jetty ...	drive marina lumut jetty parte car ferry pangko...
39	in pangkor its most popular in teluk nipah bee...	pangkor popular teluk nipah sea activity stree...
40	hi antonio the most happening area in pangkor ...	hi antonio happening area pangkor teluk nipah ...
41	greetings from malaysia	greeting rnalaysia
42	no brother someone elses car rental	brother else car rental
43	the price at the mosque is indeed cheap but th...	price mosque cheap shop minute pekan pangkor p _
44	its easy at the pangkor jetty many people will...	easy pangkor jetty people offer ferry service ...
45	if at the marina there is a charge nm day with...	marina charge rm day cover

Figure 3.13 Example of Data Pre-processing Output

On the other hand, Table 3.3 provides a comprehensive overview of the dataset attributes utilised in this research. The 'videoid' attribute is generated through the YouTube API to uniquely identify each video, while the 'duration' attribute indicates the length of the video for a given date for temporal context. Following that, the 'text' attribute contains the raw review content extracted from YouTube reviews, forming the foundation for further analysis. To address the multilingual nature of the reviews, the 'translatedtext' attribute holds comments translated into English using the Google Translate library, indirectly ensuring consistency. The 'cleanedtext' attribute contains the reviews after the removal of noise, thereby enhancing data quality. Finally, the 'processedtext' attribute represents the fully refined reviews after utilising tokenisation, stop word removal, and lemmatisation.

Table 3.3

Dataset Attributes Description

Dataset Attributes	Description
videoid	Each of the YouTube IDs is generated using the YouTube API
duration	The duration of a YouTube video
text	The text extracted from YouTube reviews
translatedtext	YouTube reviews after being translated using the Google Translate library
cleanedtext	YouTube reviews after being cleaned
processedtext	YouTube reviews after being pre-processed

Furthermore, the dataset distribution used is summarised in Table 3.4, which outlines the phases of data collection and pre-processing for Langkawi, Perhentian, and Pangkor islands. After completing the data collection and pre-processing, the remaining reviews were retained, which included only significant reviews to serve as the input for SA. The dataset comprises multiple languages from data extraction, emphasising the importance of handling various languages in this research. In addition, the data is separated into three types of datasets, such as training, validation and testing data, following a 70:20:10 ratio.

Table 3.4

Summary of Dataset Distribution

Type of Island	Total Reviews		Language Detected	Total Data (70:20:10)
	Collected	Pre-processed		
Langkawi	1965	1146	English, Malay, Indian, Chinese, Korean, Russian, Arabic.	Training: 543, Validation: 156, Testing: 78
Perhentian	1760	1295	English, Malay, Vietnamese, Korean, Russian, German, Thai.	Training: 599, Validation: 171, Testing: 86
Pangkor	1396	563	English, Malay, Japanese.	Training: 247, Validation: 71, Testing: 36

3.1.4 Descriptive-Semantic Analysis Techniques

DSA techniques substantially enhance SA by combining descriptive statistical methods with semantic contextual analysis to improve feature extraction. This approach effectively addresses the limitations of conventional SA techniques, which typically focus on isolated term-level analysis and provide limited contextual insights. Standalone techniques often overlook relationships between words and struggle to accurately detect sentiment polarity without the use of additional analytical resources. Thus, the implementation of DSA techniques is essential for a more comprehensive and insightful analysis. Figure 3.14 shows the integration of TF-IDF and LDA in identifying significant terms and extracting topics to produce a weighted feature matrix and manual aspect labelling for sentiment classification.

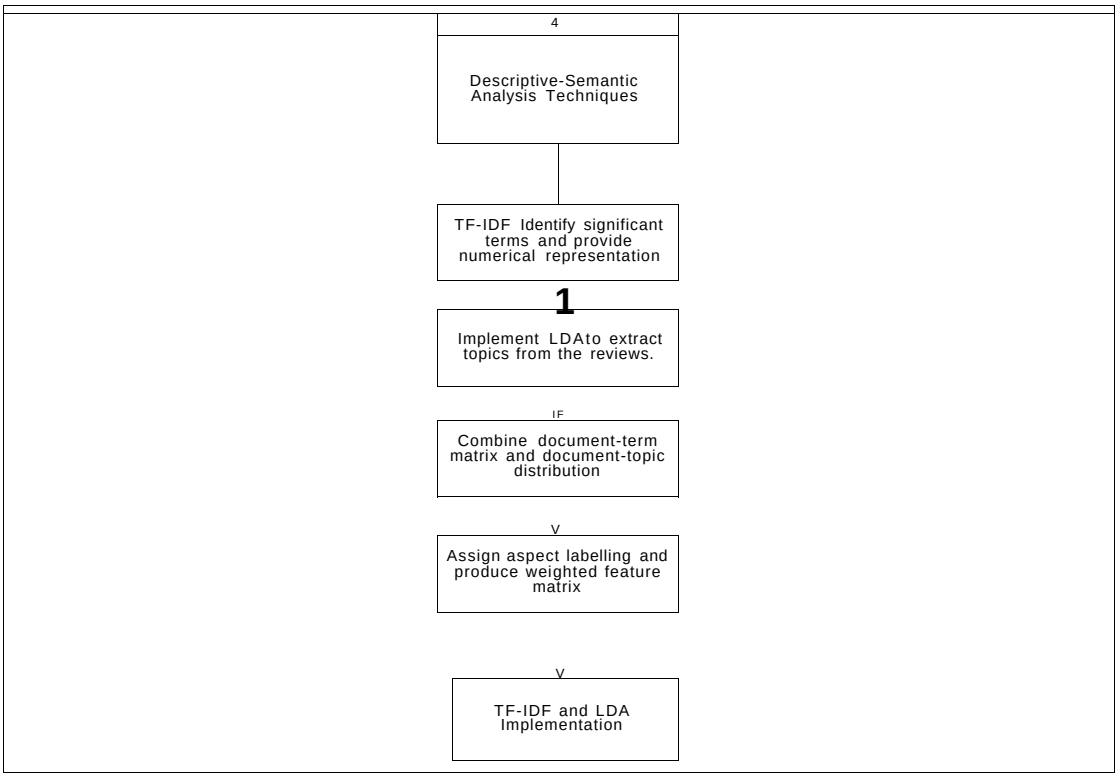


Figure 3.14 Research Design Phase Four

Furthermore, Figure 3.15 depicts the workflow of this research, illustrating the implementation of the conceptual hybrid SA approach and DSA techniques to convert raw textual data into meaningful sentiment insights. The methodology starts with data extraction to generate reviews from YouTube and is followed by pre-processing the raw text. Next, feature extraction begins with TF-IDF captures the importance of terms in a document-term matrix (illustrated by the blue dashed line). TF-IDF first assigns a weight to each term based on the frequency and importance of the pre-processed data. This facilitates emphasising the unique terms and identifying relevant themes or sentiments (Octava et al., 2023). However, TF-IDF does not pinpoint specific topics or aspects, requiring additional analysis to connect terms to themes.

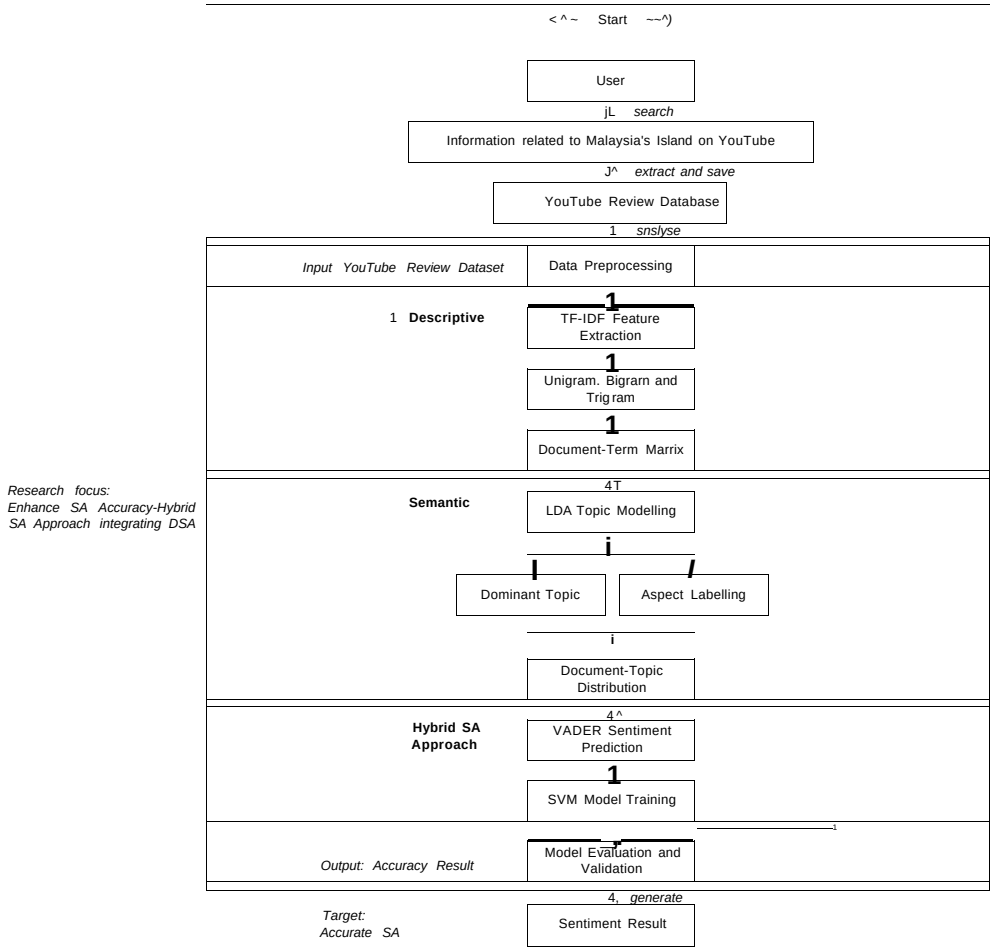


Figure 3.15 Conceptual Hybrid Sentiment Analysis Approach Enhanced Sentiment Analysis Integrating Descriptive-Semantic Analysis Techniques

Following that, LDA is then used to assign accurate probabilities to terms based on their topic relevance to identify the predominant themes or subjects present within the corpus (Kharisudin & Masri'an, 2022). LDA identifies dominant topics and assigns manual aspect labels, resulting in a document-topic distribution (represented by the pink dashed line). The process then progresses to sentiment classification, where VADER is applied to predict sentiment scores and SVM is trained for supervised classification (illustrated by the green dashed line). In addition, a stratified &-fold cross-validation method was used to preserve sentiment distribution. Finally, the performance of the hybrid SA approach is evaluated using performance metrics and validated with the AUC metric to assess its effectiveness in distinguishing sentiment categories.

3.1.4.1 Term Frequency-Inverted Document Frequency

The TF-IDF technique aims to identify the frequency and significance of specific terms within a dataset, such as *"island', "beautifuF, "beach", "place", "good* and *"time",* all of which are directly relevant to the context of islands. This technique identifies the most impactful terms within YouTube reviews, directly contributing to the extraction of keywords relevant to Malaysia's island tourism sentiments. As shown in Figure 3.16, the pseudocode extracts significant terms using the TF-IDF technique. First, it processes the input text and computes the TF-IDF scores using the specified n-gram range, including unigram, bigram, and trigram analyses, which are valuable for multi-word concepts. Then, it aggregates and sorts the TF-IDF scores of all documents in descending order to identify the top ten terms, where elevated scores signify the distinctiveness and relevance of a term to a document. Overall, TF-IDF quantifies the significance of terms or word combinations in relation to their rarity across documents, offering more insightful interpretations than isolated words.

```
Algorithm 1 Get Top TF-IDF Terms
Require: processed-text. iigr;im_range, top_n
Ensure: top_terms
 1: Compute TF-IDF scores for terms in proc.cHmaLierf using spe lined
    ngram-Tunge
 2: Sum scores across fill documents
 ;i: Sort terms by score; in descending order
 4: Return top_n terms
```

Figure 3.16 Pseudocode for Term Frequency-Inverse Document Frequency using n-gram

To improve readability, the sparse TF-IDF matrix is transformed into a document-term matrix data frame, as illustrated in Figure 3.17. In this format, terms and their corresponding TF-IDF values are represented as columns, while documents are displayed as rows. This transformation enhances interpretability and aids in SA tasks. Additionally, it supports visualisation techniques and feature selection processes. Consequently, the frequency distribution of user input plays a critical role in identifying key topics and areas of interest, thereby strengthening topic modelling within the DSA technique.

Algorithm 2 Create TF-IDF Document-Tonn Matrix
Require: proooBsetLtext Ensure: tfidfLdf I: Fit TF-IDF vectorizer to *proceasaLtext* *2*-. Convert TF-IDF output to DataFmme- with terms as columns :i: Output document-term matrix

Figure 3.17 Pseudocode for Term Frequency-Inverse Document Frequency Document-Term Matrix

Table 3.5 displays the results of n-gram for Langkawi, Perhentian and Pangkor islands using TFIDF to determine significant terms and phrases based on frequency and relevance in the dataset. The analysis results uncover distinct patterns in how textual data communicates meaning through the use of unigram, bigram, and trigram. Across all three islands, the unigram consistently exhibits the highest TF-IDF values, indicating that individual terms such as *"langkawi"* (45.87), *"island'* (65.58), and *"good'* (20.63) are the most influential keywords in articulating the core themes of each island review. This observation underscores the importance of location-based identifiers and sentiment-related terms such as *"beautifuP', "good', "nice",* and *"happy"* within tourism-related narratives. On the other hand, bigram has combinations such as *"visit langkawi"* (4.39), *"perhentian island'* (8.76), and *"Pangkor island'* (8.60), which enhance contextual understanding by linking actions or descriptors to specific destinations. In contrast, trigram terms, such as *"wonderfulplace Langkawi"* (2.67) and *"pangkor island beautifuP'* (1.15), provide even deeper semantic insights; however, their TF-IDF scores are generally lower due to increased sparsity and reduced frequency within the dataset.

Table 3.5

Term Frequency-Inverse Document Frequency n-gram Output

Type of Island	Type of n-gram			Terms		
	Unigram	Bigram	Trigram	Unigram	Bigram	Trigram
Langkawi	45.866398	4.390476	2.671302	langkawi	visit langkawi	wonderful place langkawi
	32.028948	4.375779	2.000000	good	beautiful place	good island malaysia
	27.383027	4.093710	2.000000	malaysia	good time	resort world langkawi
	27.326139	3.941297	1.844281	island	roti canai	indian driving license
	26.477238	3.475988	1.686779	not	not know	absolutely beautiful place
	22.330767	3.234939	1.686779	place	island malaysia	month visit langkawi
	22.141731	3.202944	1.618352	beautiful	love langkawi	food people amazing
	21.842859	3.200830	1.605546	time	perhentian island	sabah east malaysia
	20.856866	3.191500	1.443795	nice	good island	greeting sabah east
	17.000894	3.000690	1.389227	visit	month visit	ask price order
Perhentian	65.580994	8.756336	4.184399	island	perhentian island	mabul island sabah
	41.531667	7.885612	3.794828	malaysia	lang tengah	lang tengah island
	39.243321	7.846633	2.720794	beautiful	ringgit malaysia	not wait visit
	32.586593	7.592753	2.539966	good	beautiful island	not want sugar
	27.976314	7.564114	2.211895	beach	island sabah	try redang island
	26.853925	6.253104	2.000000	not	redang island	dusky leaf monkey
	25.248887	4.906050	1.990010	sabah	island malaysia	south east asia
	24.609188	4.844446	1.765977	like	east coast	island sabah malaysia

Type of Island	Type of n-gram			Terms		
	Unigram	Bigram	Trigram	Unigram	Bigram	Trigram
	24.555675	4.755004	1.734432	visit	mabul island	ringgit malaysia coconut
	22.130725	4.520258	1.687124	nice	tioman island	reach kuala lumpur
	20.634942	8.606346	1.649789	good	pangkor island	pangkor island beautiful
	18.773027	3.623087	1.553127	pangkor	rent motorbike	pangkor laut good
	18.269005	3.119363	1.298432	island	long time	go pangkor island
	11.462533	2.909297	1.129699	not	pangkor laut	pangkor laut private
	10.755814	2.323370	1.129699	place	rent car	laut private island
	9.152916	2.239504	1.094916	beautiful	pink van	pangkor island not
	8.797667	1.994363	1.000000	like	deadly fart	change improvement sir
	8.736082	1.910375	1.000000	happy	teluk nipah	kak teah moss
	8.697615	1.869391	1.000000	hornbill	motorbike rental	sea water clear
	8.687759	1.868790	1.000000	want	ringgit malaysia	business saturday sunday

Overall, this analysis highlights a distinct trend, as the progression shifts from unigram to trigram, the specificity of the extracted phrases increases, despite a tendency for both frequency and TF-IDF scores to decline. Unigrams are effective in capturing general sentiment and topical keywords, while bigrams strike a balance by conveying compound meanings without excessive sparsity. In contrast, trigram provides rich contextual information but are observed less frequently, which makes them suitable for capturing unique or highly specific expressions. As a result, using a combination of unigram and bigram can provide the most beneficial approach in SA related to tourism, as it offers both interpretability and meaningful context while maintaining frequency and computational efficiency. Subsequently, after assessing term relevance using TF-IDF, it serves as input for LDA in the next phase.

3.1.4.2 Latent Dirichlet Allocation

A probabilistic model utilises LDA to extract topic structures from a collection of island reviews. This technique identifies the dominant topic within each document, generates a document-topic distribution, and applies aspect labels manually for enhanced semantic clarity. This leads to uncovering hidden patterns and groups related to content, which is valuable for thematic analysis and SA of unstructured datasets. In Figure 3.18, a document-topic distribution matrix has been created by converting TF-IDF-transformed data into a data frame classified by topic numbers. Consequently, this matrix reveals the probability of each topic present in each document and is transformed into a Pandas data frame for trend visualisation and dominant topic classification.

Algorithm 3 Create LDA Document-Topic Distribution

Require: tfidf_data, n_topics
Ensure: document_topic_df
1: Train LDA model with *n_topics* on *tfidf_data*
2: Transform data to obtain document-topic distribution
3: Store results in DataFrame with one column per topic

Figure 3.18 Pseudocode for Latent Dirichlet Allocation Document-Topic Distribution

For LDA implementation, it identifies the most probable topic for each document, as illustrated in Figure 3.19. The process commences with the identification

of the topic that has the highest probability for each document. It is crucial to ascertain the number of documents associated with each dominant topic. Subsequently, the documents are sorted in descending order, which enhances our understanding of the most prevalent topics for further analysis. This approach not only aids in interpreting the results of topic modelling but also facilitates focusing on the most significant topics.

Algorithm 4 Dominant Topic; Calculation using LDA
Rcquire: document_topic_df
Ensure: sorti-H dnTnJnarrt, titpir rninrbi
1: For each document in *document-topic-df.* find the topic with the highest probability
2: Store these topics as *dominant-topics*
3: Count tilt; occurrences of each topic in *dominant.topics*
A: Sort the topic counts in descending order
S: return sort ecLdonrinant .topic .counts

Figure 3.19 Pseudocode for Dominant Topic using Latent Dirichlet Allocation

Following that, LDA describes the extraction of key terms associated with each topic, as illustrated in Figure 3.20. The process begins with LDA extracting and analysing significant topic elements from TF-IDF data. After obtaining terms from the TF-IDF data, LDA creates an undefined dictionary called 'topics' to store the results. Next, it identifies the top ten terms for each topic by iterating through each topic in 'ldamodel' and selecting the terms with the highest probabilities. Consequently, the 'topics' dictionary includes these terms and their corresponding topic labels, which can help users understand the concepts associated with each topic.

Algorithm 5 Extract Top Words for Each Topic from LDA Model
Require: lda.modcl. tfidf.data, n.top.words
Ensure: topics
1: Extract feature: names (terms) from *tfidf.data*
2: Initialise; empty dictionary *topics*
:i: for each topic index and its corresponding component in *lda-rnod<d* do
4: Sort terms by their weight in descending order
5: Select the top *n.top-iiiords* terms
6: Store these terms in *topics* with key as *Topic.i*
7: end for
8: return topics

Figure 3.20 Pseudocode for Topic and Terms Extraction using Latent Dirichlet Allocation

After thoroughly analysing the most significant terms and extracting top terms for each topic, Figure 3.21 presents the leading terms associated with each topic, mapping them to specific predefined aspect labels manually. The process begins by initialising an empty dictionary to store aspect labels. For each topic and its associated top terms, a set of labels is created. The algorithm checks each word against a predefined keyword-to-label mapping, adding any matching labels to the set. If no labels are identified, the topic is marked as *"Unlabeled'*. After processing all topics, the top terms and their assigned labels are presented. The dataset is then organised according to the dominant topics, and aspect labels are aggregated into broader categories known as *"Aggregated Aspects"*. To assess the significance of each topic, the mean dominant topic weight is calculated. Finally, the results are compiled into a thematic structure, delivering both the aspect labels and the thematic organisation. This approach systematically links topics to meaningful aspects and synthesises them into higher-level themes.

```
Algorithm 6 Aspect Labeling and Thematic Structure Extraction
Require: topics, keywonLtoJabel, data
Ensure: aspect ..labels, thematic structure!
 1: Initialise empty dictionary aspect Jabe Is
 2:  for each topic and its associated words in topics do
 3:       Initialise empty set assignedAabels
 4:       for each word in topic's top words do
 5:          if word exists in keywordJ.oAabel then
 6:              Add corresponding label to assignedAabels
 7:          end if
 8:       end for
 9:       Store; concatenated labels in aspect-labels for the current topic
10:       If no label found, assign label "Unlabeled"
11:  end for
12:  Display topics, their top words, and assigned aspect labels
13:  Group dataset data by "Dominant Topic"
14:  Aggregate associated aspect labels into "Aggregated Aspects"
15:  Compute mean "Dominant Topic Weight" for each topic
16:  Store result in thematic-structure
17:  return aspectJabels, thematic-structure
```

Figure 3.21 Pseudocode for Topic and Aspect Labelling using Latent Dirichlet Allocation

Moreover, Figure 3.22 presents the results of LDA, illustrating the dominant topics and their associated labels for reviews of Langkawi, Perhentian, and Pangkor

islands. The distribution and relevance of these topics in the reviews are strengthened by considering the average topic weight and the number of documents associated with each topic. Based on the results, the underlying concepts of Langkawi indicate that aspects labelled like *"Destination, Seasonality, Experience"* represent the destination's popularity as a seasonal getaway for the dominant 'Topic_*V* with an average weight of 0.099 of 126 reviews. Meanwhile, Perhentian reviews point out passions for outdoor activities, representing aspect labels like *"Activity, Wildlife, Destination, Experiences"* for dominant 'Topic_l' with an average weight of 0.154 out of 243 reviews. Lastly, Pangkor reviews mainly focus on the *"Social, Time, Culinary, Quality, Location"* aspect labels, which emphasise cuisine and social interactions for the dominant 'Topic_*V* with an average weight of 0.105 of 68 reviews. In addition, these results were documented for further analysis in Chapter 4.

LANGKAWI

	Dominant Topic	Aggregated As.pec s	Average Weight	Common Themes	Doc ume n t Count
O	Topic_1	Destination. Seasonality. Experience. Destinat	0 098762	[Destination. Seasonality. Experience]	126
1	Topic_10	Action, Emphasis, Activity, Natural Attraction	0 084826	[Action, Emphasis, Activity, Natural Attractio...	92
2	Topic_2	Activity, Practical Advice, Social Interaction	0 085557	[Activity, Practical Advice, Social Interactio..	85
3	Topic_3	Destination, Seasonality, Experience. Social I	0 092250	[Destination Seasonality, Experience, Social	104
4	Topic_4	Emotion, Preference, Seasonality, Destination.	0 092528	[Emotion, Preference. Seasonality. Destination	100
5	Topic_5	Emphasis, Emotion, Affirmation, Duration, Des1	0 090390	[Emphasis, Emotion, Affirmation, Duration, Des	97
6	Topic_6	Knowledge, Activity, Practical Advice, Natura	0111991	[Knowledge, Activity, Practical Advice, Natura	131
7	Topic_7	Emotion Activity, Social Interaction, Negatio	0.140468	[Emotion. Activity. Social Interaction, Negati...	179
8	Topic_8	Destination, Direction. Experience, Negation,	0.112983	[Destination. Direction. Experience, Negation]	136
9	Topic_9	Culinary Experience, Emotion, Preference, Soc	0 090242	[Culinary Experience, Emotion, Preference, Soc	96

PERHENTIAN

	Dominant Topic	Aggregated Aspeci s	Average Weight	Common Themes	Document Count
0	Topic_1	Activity, Wildlife. Destination, Location, Exp	0.154091	[Activity. Wildlife, Destination. Location, Ex...	243
1	Topic_10	Taste, Wildlife, Destination, Location, Experi	0091500	[Taste, Wildlife, Destination, Location, Exper...	110
2	Topic_2	Emotion. Preference. Cultural. Destination, Ex	0 078168	[Emotion. Preference, Cultural, Destination. E...	87
3	Topic_3	Emotion, Activity Negation Affirmation, Acco	0.080074	[Emotion Activity, Negation, Affirmation, Ace.	91
4	Topic_4	Reflection, Activity, Social Interaction, Natu	0.142185	[Reflection, Activity, Social Interaction, Nat...	205
5	Topic.5	Culinary, Preference, Activity, Social Interac	0.100279	[Culinary, Preference, Activity, Social Intera...	131
e	Topic_6	Culinary Experience, Preference, Social Intera	0 068684	[Culinary Experience, Preference, Social Inter. .	70
7	Topic_7	Reflection, Cultural, Affirmation, Cost, Desti	0088119	[Reflection, Cultural, Affirmation, Cost, Dest.	105
8	Topic_8	Preference, Experience, Negation, Preference,	0 087463	[Preference, Experience, Negation]	101
9	Topic_9	Culinary Experience, Currency. Natural Attract	0 109438	[Culinary Experience, Currency. Natural Attrac...	14S

PANGKOR

	Dominant Topic	Aggregated As pec s	Average Weight	Common Themes	Document Count
0	Topic_1	Social, Time, Culinary, Quality. Location, Exp	0 104809	[Social, Time, Culinary, Quality, Location]	68
1	Topic_10	Action, Social, Knowledge, Service, Relationsh	0.095571	[Action, Social, Knowledge, Service, Relations.	50
2	Topic_2	Time, Emotion, Relationship, Wildlife Accommo	0.101237	[Time, Emotion, Relationship. Wildlife, Accomm...	57
3	Topic_3	Social, Emotion, Negation, Quality, Cultural,	0.079759	[Social, Emotion, Negation, Quality, Cultural]	40
4	Topic_4	Social, Emotion. Humor, Preference. Quality, L	0 090058	[Social, Emotion, Humor. Preference, Quality]	48
5	Topic_5	Action, Social. Humor, Narrative, Relationship	0 122178	[Action, Social, Humor, Narrative, Relationship]	77
6	Topic_6	Action, Social. Emotion, Preference, Accommoda	0 082744	[Action. Social, Emotion. Preference. Accornmod.	40
7	Topic_7	Action, Time, Culinary, Negation, Quality, Loc	0 114365	[Action, Time, Culinary, Negation, Quality]	66
8	Topic_8	Action, Service, Time, Transport, Quality, Loc	0 124546	[Action, Service, Time, Transport, Quality]	75
9	Topic_9	Action, Social. Time. Relationship, Accommodat	G.034733	[Action, Social, Time, Relationship, Accommoda...	42

Figure 3.22 Latent Dirichlet Allocation Output

In conclusion, this research implements a hybrid feature integration technique to enhance the performance of sentiment classification by merging TF-IDF with LDA representations. TF-IDF vectorisation evaluates the significance of words within texts, while LDA reveals latent semantic structures by representing each document as a collection of topics. By weighting the TF-IDF vectors according to their corresponding LDA topic probabilities, these techniques effectively identify important terms that correspond to the thematic context of each document. Following the DSA techniques phase, the next section explores the hybrid SA approach as the subsequent phase of the research.

3.1.5 Hybrid Sentiment Analysis Approach

This research evaluates a hybrid SA approach that integrates the VADER lexicon-based technique with ML using SVM model. This hybrid approach outperforms conventional SA techniques. As illustrated in Figure 3.23, the hybrid SA approach consists of VADER and SVM for sentiment scoring and classification. After integrating TF-IDF and LDA features, the compound sentiment score obtained from the VADER lexicon is incorporated as an additional feature in the SVM model to enhance classification performance. To reduce misclassification risks and improve prediction accuracy, neutral sentiment instances are excluded prior to vectorisation and model training, as their inherent ambiguity increases class overlap and noise within the feature space, thereby reducing the discriminative capability of the SVM model between positive and negative sentiments. As a result, this hybrid SA approach effectively combines VADER lexicon-based sentiment scoring with the capabilities of SVM, enhancing model performance, sentiment classification, and prediction efficiency.

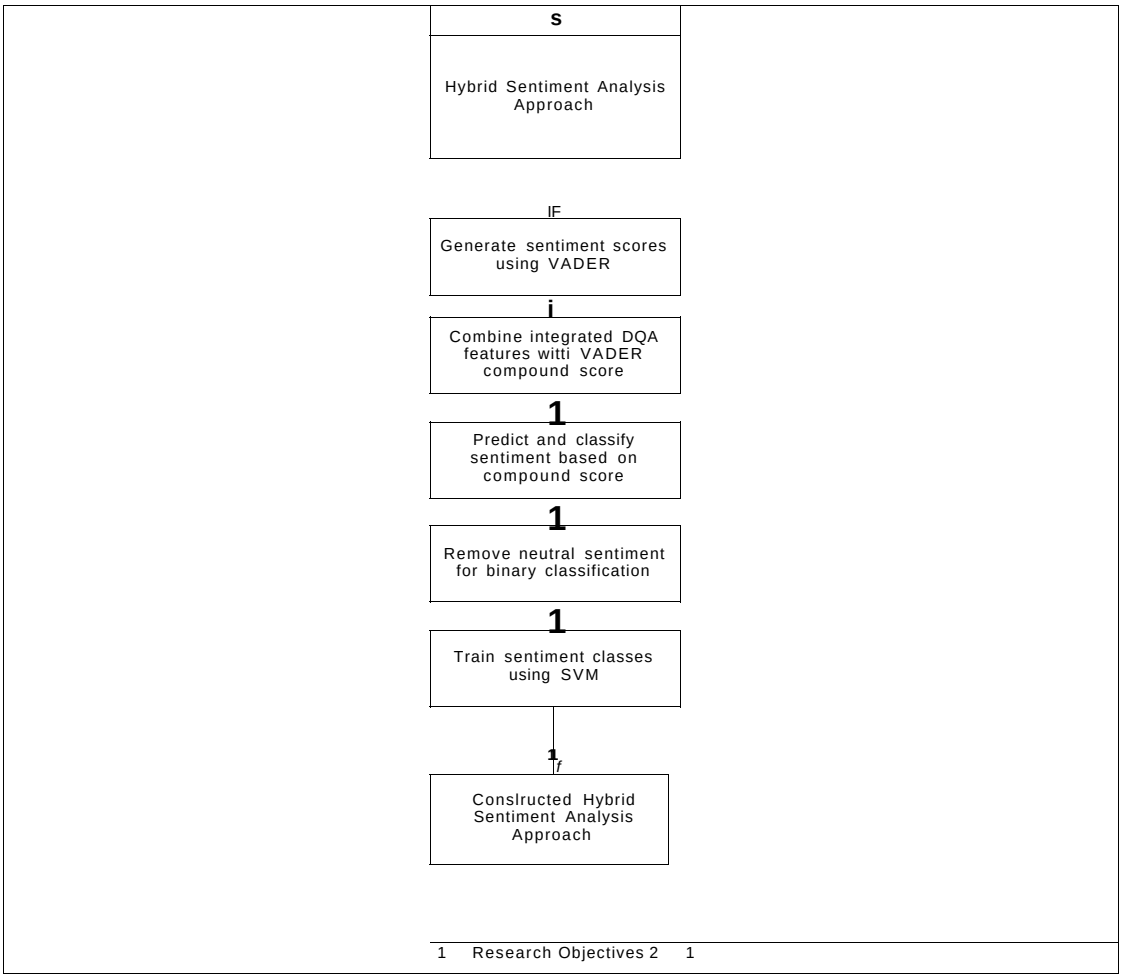


Figure 3.23 Research Design Phase Five

In other words, the process starts with VADER analysing pre-processed text data to generate sentiment scores and classify sentiment, while SVM is applied for sentiment training using binary classification, such as positive and negative. After that, the integration of unsupervised lexicon-based and supervised ML techniques utilises the polarity score of each lexicon's dataset as a target variable for ML models during the training process. Ultimately, the hybridisation across models yields the probability that the text corresponds to each sentiment polarity class, with the highest probability indicating the text's sentiment polarity (Sham & Mohamed, 2022). Consequently, this hybrid SA approach effectively combines a lexicon-based technique and ML models, specifically VADER and SVM, as illustrated in Figure 3.24. This approach balances lexicon-based sentiment scoring with contextual learning to enhance sentiment classification.

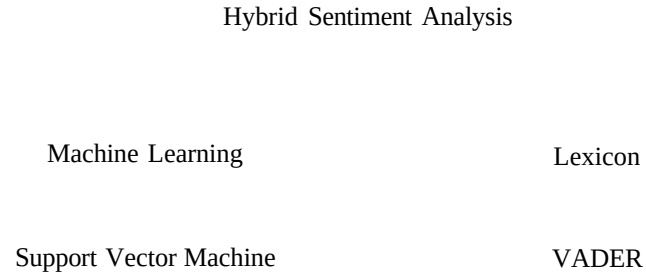


Figure 3.24 Hybrid Sentiment Analysis Classification

3.1.5.1 Valence Aware Dictionary and Sentiment Reasoner

Following the integration of TF-IDF and LDA, VADER was included as an additional component to classify sentiment by directly analysing pre-processed text. The VADER lexicon functions as a sentiment classification tool that accounts for the linguistic and contextual characteristics of input sentences, providing sentiment scores along with detailed analyses of individual words within the text (Sham & Mohamed, 2022). Moreover, VADER analyses text to identify words in its lexicon and calculates the polarity index using the 'polarity scores()' function. The compound score is subsequently calculated by normalising the lexicon ratings within a range of -1 to +1, where scores closer to -1 indicate extreme negative sentiment and scores near +1 indicate extreme positive sentiment. Standard thresholds are then applied to categorise sentences accordingly (Fazrin et al., 2022). Overall, this technique ensures a comprehensive evaluation of sentiment polarity.

As shown in Figure 3.25, the process begins by downloading and loading the VADER lexicon from NLTK. The dataset handles missing values in the *'processedJexf* column, which are replaced with empty strings, ensuring all entries are formatted as strings. After that, the VADER Sentiment Intensity Analyser is initialised, and a function named '*PredictSentimenf* is defined to compute sentiment scores for text inputs. Based on the compound score, sentiments are classified as positive, negative or neutral. The function is then applied to each review in the *'processed texf* column, eliminating neutral reviews and removing duplicates to maintain data quality. This process effectively cleans and categorises the dataset, making it ready for model training.

```
Algorithm 7 VADER Sentiment Analysis and Data Preproccssing
Download and load the vader_lexicon from NLTK.
Import dataset from CSV file into a dataframe.
Handle missing values in processed_text column by replricing with empty
strings and converting all entries to string type;.
Initialise the VADER. Sentiment Intensity Analyser.
function PREDICTSENTIMENT(text)
    Compute sentiment polarity scores using VADER
    if compound score > 0.05 then
        Assign label "Positive".
    else if compound scon; < -0.05 then
        Assign label "Negative".
    else
        Assign label "Neutral".
    end if
end function
Apply the PREDICTSENTIMENT function to each review in processed.text.
Filter out reviews classified as "Neutral".
Remove duplicate reviews based on processed_text.
```

Figure 3.25 Pseudocode for Sentiment Polarity using Valence Aware Dictionary and Sentiment Reasoner

In addition, to ensure optimal model performance and interpretability in binary sentiment classification, it is crucial to maintain two distinct classes, such as positive and negative. Since VADER utilises compound polarity scores, it categorises texts as positive, negative, or neutral. This can lead to confusion in sentiment classification using SVM model in the decision-making process, resulting in decreased certainty due to the presence of neutral sentiments. To avoid misclassifications and enhance prediction accuracy, any data labelled as neutral is removed prior to vectorisation and model training. This approach guarantees that the dataset aligns with the binary classification framework, incorporating only documents that express either positive or negative sentiment for sentiment classification tasks.

3.1.5.2 Support Vector Machine

After implementing VADER, the next phase involves training for the SVM model. The integration of SVM following VADER aims to strengthen the SA framework by leveraging SVM's capacity to detect complex sentiment patterns that may not be captured by lexicon-based techniques alone. SVM is a ML model that can map feature spaces to higher-dimensional spaces through nonlinear transformations for

optimal hyperplanes (Hasanati et al., 2023). According to Kristiyanti et al. (2022), SVM is a classification algorithm that uses extreme points to create a hyperplane, searching for a support vector from two classes. The goal of SVM is to classify data points using a hyperplane or separator function between classes. One of SVM's key advantages is its strong generalisation capability, enabling it to accurately classify unseen or out-of-sample data beyond the training set.

Therefore, SVM is applied in sentiment classification due to its effectiveness in managing high-dimensional text features and clearly separating different classes. However, evaluating various data partitions is essential before sentiment classification in order to identify the most effective split for reliable SA of Malaysia's island tourism reviews. Effective data partitioning is crucial for the training and validation of SVM models, as it directly influences performance and the risks of overfitting or underfitting. Moreover, by examining how these data partitions affect accuracy, precision, recall, and Fl-score across datasets from Langkawi, Perhentian, and Pangkor, this research seeks to determine an optimal configuration that achieves a balance between efficient learning and reliable evaluation, ultimately enhancing model stability and predictive power for sentiment classification.

The selection of data partition is supported by prior studies, including those by Satrya et al. (2022) and Prastiani et al. (2023), who utilised the holdout method to split datasets into training and testing subsets with various ratios such as 90:10, 80:20, and 70:30. Their findings indicated that splits of 70:30 and 80:20 often achieve a strong balance between accuracy and generalisation. Furthermore, Table 3.6 shows the evaluation of three data partition strategies, which are 60:20:20, 70:20:10, and 80:10:10 for Langkawi, Perhentian, and Pangkor, yielding noteworthy insights. In Langkawi, the 60:20:20 partition achieved the highest validation accuracy of 100% and a test accuracy of 99.36%, along with an Fl-score of 98.83%). The 70:20:10 partition also demonstrated strong performance, recording a test accuracy of 98.73%) and an Fl-score of 97.77%, indicating stable performance without overfitting. Although the 80:10:10 configuration produced satisfactory results, it did not significantly exceed the performance of the 70:20:10 partition, and its smaller validation and test sizes may compromise the robustness of the evaluation.

Table 3.6

Comparison of Data Partition Accuracy Results

Type of island	Data Partition	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Precision (%)	Recall (%)	F1 -Score (%)
Langkawi	60:20:20	97.86	100	99.36	98.08	99.62	98.83
	70:20:10	98.15	99.35	98.73	96.43	99.24	97.77
	80:10:10	98.07	98.72	98.72	96.15	99.24	97.62
Perhentian	60:20:20	98.64	100	100	100	100	100
	70:20:10	98.50	100	98.84	96.67	99.31	97.93
	80:10:10	97.95	98.84	97.67	94.12	98.59	96.16
Pangkor	60:20:20	96.66	92.96	91.55	82.35	95.00	86.65
	70:20:10	97.19	95.77	97.22	92.86	98.33	95.31
	80:10:10	97.55	97.14	100	100	100	100

In the analysis of Perhentian Island, all three data partitions demonstrated high scores. The 60:20:20 partition achieved perfect scores across training, validation, and testing, which may indicate overfitting due to the small size of the test set. The 70:20:10 partition also delivered excellent results, achieving a test accuracy of 98.84% and an F1-score of 97.93%, reflecting strong generalisation capabilities. However, the 80:10:10 partition saw a slight decline in the F1-score to 96.16%, likely due to the smaller evaluation dataset. Meanwhile, the model performance for Pangkor Island was generally lower. The 60:20:20 partition recorded the weakest test accuracy and F1-score, suggesting potential instability and insufficient evaluation data, while the 70:20:10 partition showed improved results, achieving a test accuracy of 97.22% and F1-score of 95.31%, indicating better generalisability. Although the 80:10:10 partition attained perfect scores, these results may not be widely applicable due to the limited size of the test set, which can distort performance assessments.

Based on these insights, the 70:20:10 data partition was selected for this research, as it consistently provided a reliable balance between training, validation, and testing across all island datasets. This configuration ensures sufficient data for training while retaining ample samples for comprehensive evaluation. In contrast to the 60:20:20 split, which exhibited tendencies toward overfitting, the 80:10:10 split, which risked underrepresenting the test data, while the 70:20:10 approach demonstrated stable and high metrics in accuracy, precision, recall, and F1-scores, particularly for more variable

datasets such as Pangkor. Consequently, the 70:20:10 partition was deemed the most appropriate and representative framework for evaluating the hybrid SA approach integrating DSA in this research.

Moreover, Figure 3.26 illustrates the process of constructing hybrid features by integrating TF-IDF, LDA, and VADER representations for SA, as well as partitioning the dataset. It begins by eliminating duplicate reviews from the *'processed texf* column and defining the input text alongside the corresponding sentiment labels. The dataset is divided into training (70%), validation (20%), and test (10%) sets using stratified sampling to ensure class balance. Sentiment labels are then encoded numerically, and TF-IDF vectorisation, incorporating both unigram and bigram to produce sparse feature matrices. In contrast, the LDA model with 10 topics is subsequently trained on the training TF-IDF features, while the validation and test sets are transformed into LDA topic distributions. Additionally, VADER compound sentiment scores are computed for all reviews across the training, validation, and test sets.

Algorithm 8 Hybrid Feature Construction ;md Dataset Partitioning
1: Remove: duplicate reviews based on processed_text.
1- Define input *X* — processed_text and target *y* = Sentiment.
;i: Split data into training (70%), validation (20%), and test (10%) sets using stratified sampling, ensuring no overlap between sets.
IT Encode sentiment labels into numerical format using LabelEncoder.
ft: Apply TF-IDF vectorisation with unigram and bigram features to obtain sparse; feature matrices for train, validation, and test sets.
fi: Train LDA topic model with *ri* = 10 topics using the training TF-IDF features.
7: Transform validation and test sets into LDA topic distributions.
8: Compute VADER. compound sentiment scores for each review in train, validation, and test sets.
9: **function** CREATEHYBRID FEATURES (TF-IDF, LDA, VADER)
10: for each document *i* in dataset do
11: Extract TF-IDF vector J}.
12T Extract LDA topic distribution *L.;*.
13: Extract VADER, compound score K-
14: Weight *Ti* by the mean of $L.\text{-}_L$ to capture semantic adjustment.
IS: Concatenate weighted T;, ij, and $V.\text{-}_L$ into hybrid feature vector *H^*.
Ifi: end for
17: return Hybrid feature matrix *H*.
18: end function
19: Construct final feature sets: $X_{ITIIi_{7t}}$. $X_{va}i$, and $X_{LK!ll}$ using CREATEHYBR.IDFEATURES.

Figure 3.26 Pseudocode for Hybrid Feature and Dataset Partitioning

Following that, a function named '*CreateHybridFeatures'* is defined to merge these diverse feature types. For each document, it extracts the TF-IDF vector, LDA

topic distribution, and VADER compound score. The TF-IDF vector is weighted by the average of the LDA distribution to effectively capture semantic context. These weighted vectors, combined with the topic distributions and sentiment scores, are concatenated into a hybrid feature vector. The function also outputs a hybrid feature matrix that merges linguistic, semantic, and sentiment information. Finally, hybrid features are constructed for the training, validation, and test sets, resulting in enriched representations (X_train, X_val, X_test) for subsequent sentiment classification tasks.

Besides that, overfitting in ML model can occur when an algorithm fits too closely to the training data, resulting in poor performance on unseen data. This occurs when the model confuses random noise with meaningful patterns (Zhu, 2024). Hence, to overcome this, an analysis of degrees of freedom utilising the SVM model was conducted in Table 3.7 to examine the impact of various values of the regularisation parameter *C* on the training and validation accuracy for three island datasets. The parameter *C* regulates the balance between minimising training error and reducing testing error, which is crucial for identifying tendencies of overfitting or underfitting in the model.

Table 3.7

Degree of Freedom Analysis of Support Vector Machine Accuracy Results

Type of island	C Values	Training Accuracy (%)	Validation Accuracy (%)
Langkawi	0.001	16.21	83.79
	0.01	96.69	96.13
	0.1	97.61	97.42
	1	98.16	98.15
	10	100	100
	100	100	100
	0.001	83.47	83.47
	0.01	98.00	97.66
	0.1	99.00	98.67
	1	99.00	98.50
	10	100	100
	100	100	100
	0.001	84.21	62.80
	0.01	100	99.21
	0.1	97.98	96.78

Type of island	C Values	Training Accuracy (%)	Validation Accuracy (%)
	1	97.98	97.19
	10	100	99.60
	100	100	100

For Langkawi Island, the model demonstrates significant underfitting at a C = 0.001, with a training accuracy of 16.21% compared to a higher validation accuracy of 83.79%. This considerable discrepancy indicates that the model is overly constrained and unable to capture the complexity of the data. As the C value increases from 0.01 to 1, both training and validation accuracies improve steadily, approaching perfect scores. At C values of 10 and 100, both accuracies reach 100%, suggesting that the model effectively learns the underlying patterns in the data. Importantly, since the training and validation performances remain consistent, there is no strong indication of overfitting.

Similarly, the Perhentian Island dataset demonstrates a balanced performance across various C values. Starting at C = 0.001, both training and validation accuracies are equal at 83.47%, indicating a limited capacity for learning. As C increases to 0.01, 0.1, and 1, both accuracies improve, ultimately reaching 100% at C values of 10 and 100. This concurrent rise in training and validation accuracy suggests effective generalisation without signs of overfitting. In contrast, Pangkor Island exhibits early signs of overfitting at lower C values. At C = 0.001, training accuracy is relatively high at 84.21%, while validation accuracy falls to 62.80%, indicating challenges in generalisation. As C increases, validation accuracy improves, narrowing the gap between the two metrics. By C = 1 and beyond, both accuracies exceed 97%, indicating that overfitting is alleviated and model performance stabilises.

To sum it up, this analysis demonstrates that the selection of the C parameter is essential for managing model complexity and ensuring effective generalisation. For Langkawi and Perhentian, the model achieves strong generalisation even as C values increase, whereas Pangkor requires careful tuning to prevent early overfitting. These findings endorse the use of mid to high C values, such as C = 1, as optimal for SVM modelling, effectively balancing the trade-off between underfitting and overfitting across different island datasets. Following the completion of sentiment classification utilising the SVM model, the subsequent step entailed evaluating and validating its performance. This process is essential to ensure the reliability and generalisability of the results obtained.

On the other hand, stratified cross-validation and learning curve analysis are applied after defining the data partition, as represented in Figure 3.27. The procedure begins by initialising the SVM model with a linear kernel, setting the regularisation parameter C to 1, utilising balanced class weights to address class imbalance, and enabling probability estimates for evaluation. The model is then trained on the hybrid feature sets for the training, accompanied by the corresponding sentiment labels. Subsequently, a stratified *k-fo\d* cross-validation strategy with $k = 9$ folds is implemented to ensure balanced class representation during validation. To evaluate model performance across varying data sizes, a learning curve is constructed by progressively increasing the training set proportion from 10% to 100%. For each subset, both training accuracy and validation accuracy are calculated, and the mean and standard deviation of these accuracies across the folds are determined to provide insights into model stability. Finally, the learning curve is visualised with the training set size on the x-axis and accuracy on the y-axis. The plot features training and validation accuracy curves, confidence intervals, and essential titles and labels for clarity. This visualisation offers a statistical assessment of the hybrid features and the SVM model's generalisation, aiding in the identification of potential overfitting or underfitting issues.

Algorithm 9 Training and Validation of the Hybrid SVM Model

1: Initialise SVM classifier with:
- Linear kernel
- Regularization parameter $C = 1$
- Class weights balanced
- Probability estimates enabled

2: Train SVM model on hybrid feature set X_{train} with labels y_{train}.
3: Define a Stratified k-Fold cross-validation strategy with $k = 9$ folds, ensuring balanced class representation.
4: Generate a learning curve using incremental training set proportions (10% to 100% of training data):
- Compute training accuracy for each subset size.
- Compute validation accuracy for each subset size.

5: For each training set size:
- Calculate mean and standard deviation of training accuracy.
- Calculate mean and standard deviation of validation accuracy.

6: Plot the learning curve:
- X-axis: training set size
- Y-axis: accuracy
- Display training accuracy curve with confidence interval.
- Display validation accuracy curve with confidence interval.
- Add title, labels, legend, and grid.

Figure 3.27 Pseudocode for Support Vector Machine Training and Learning Curve

In summary, the integration of DSA techniques with a hybrid SA approach creates a well-structured pipeline for SA. The provided pseudocodes enhance textual representation through a hybrid feature that incorporate TF-IDF and LDA, generate sentiment-labelled data using VADER, and ensure reliable classification with SVM model supported by cross-validation and learning curve analysis. Together, these processes present a comprehensive approach that improves representation, enhances generalisation, and yields robust results in sentiment classification. Building upon this established pipeline, the following section presents the evaluation and validation of the hybrid approach, underscoring its effectiveness in classifying sentiment.

3.1.6 Model Evaluation and Validation

After training the model, its performance is systematically evaluated using standard performance metrics, followed by validation through the AUC metric.

Evaluation is a crucial process that assesses a model's effectiveness by comparing its predictions with actual outcomes using distinct datasets. In this research, the data is strategically divided into three segments such as training, validation, and testing to provide a structured and unbiased evaluation of the model's performance. Additionally, the validation data plays a vital role in fine-tuning hyperparameters and mitigating overfitting. Subsequently, the testing data is used to assess the model's ability to generalise to unseen data. The performance of the model is then evaluated using performance metrics, including accuracy, precision, recall, and Fl-score, along with being validated using the AUC metric as illustrated in Figure 3.28.

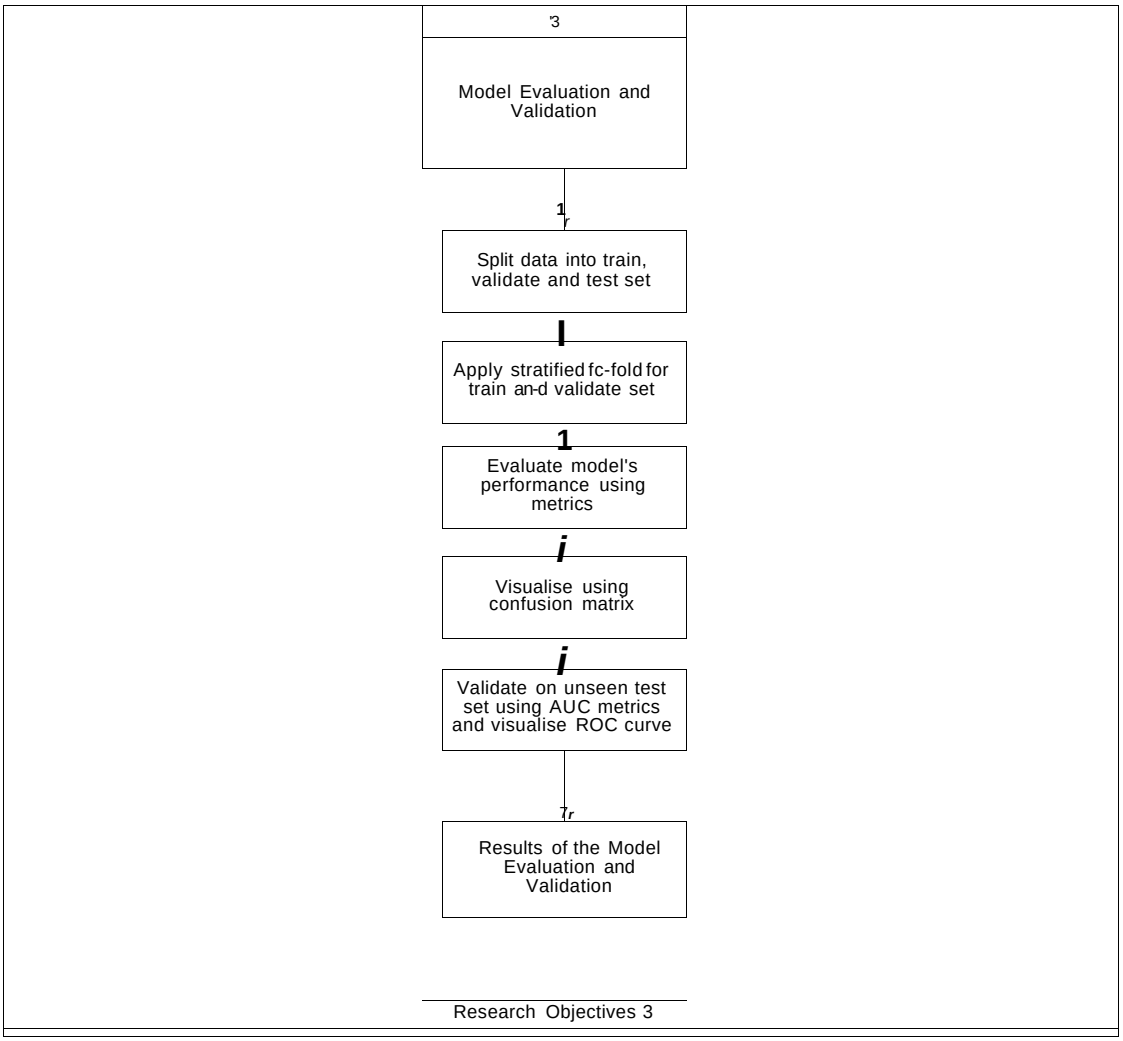


Figure 3.28 Research Design Phase Six

3.1.6.1 Evaluate using Performance Metrics

This section evaluates the effectiveness and reliability of the classification model using stratified &-fold cross-validation and confusion matrix analysis to assess predictive performance. Stratified &-fold cross-validation preserves the original class distribution within each fold, ensuring a balanced evaluation while minimising sampling bias. Meanwhile, the confusion matrix provides a detailed account of correct and incorrect predictions, facilitating the calculation of key metrics such as accuracy, precision, recall, and Fl-score. Together, these methodologies offer a comprehensive and reliable assessment of the model's predictive performance.

a) Stratified &-fold Cross-Validation

Stratified &-fold cross-validation is an essential technique in ML for evaluating model performance, particularly when data is limited for forming separate training and testing data. This technique ensures representativeness and prevents overfitting by randomly dividing the data into *k* folds. Subsequently, *k* training and validation iterations are conducted, where *k* parts are utilised for testing and *k* - 1 parts are used for training. The final results are derived by averaging the outcomes from all tests (Brandao & Calixto, 2019). However, the choice of & substantially influences both the reliability of the evaluation and the associated computational cost. Recent studies in SA have investigated various values of *k* to achieve an optimal balance between bias, variance, and efficiency. Notably, research emphasises that the selection of folds can influence model accuracy and performance across various classes.

Kristiyanti et al. (2022) applied stratified 10-fold cross-validation to analyse a dataset of 1,869 tweets, divided into 10 folds with 1,443 training data and 426 testing data each. This method ensures adequate training data and accurate measurements, resulting in an overall mean accuracy of 91.89%. The test folds achieved accuracies between 91.0% and 94.0%, with a weighted Fl-score of 0.92. By selecting *k*=*Q*, the research study effectively mitigates potential variations among individual folds, as fewer folds can lead to increased evaluation variance. In contrast, Verma and Thakur (2023) found that maintaining all attack classes can be achieved by minimising estimation bias and controlling computational costs by using 9 folds. The results from the 9-fold cross-validation demonstrated that several ML techniques had high accuracy

with low variance, such as Decision Tree, Random Forest and Logistic Regression, which had accuracy values of 94.16%, 95.12%, and 80.50%, and the recall rates for each class remained strong even for the rarest attacks.

Furthermore, the selection of 9-fold cross-validation for this research is based on a thorough evaluation of several critical factors, including the distribution of review data across three distinct islands and the potential for class imbalance. As presented in Table 3.4 in the data collection and pre-processing section, there is considerable variation in the number of pre-processed reviews among the islands, such as Langkawi (1,146), Perhentian (1,295), and Pangkor (563) reviews. This uneven distribution presents challenges in determining the appropriate number of cross-validation folds. For smaller datasets like Pangkor, the implementation of 10-fold cross-validation, as highlighted by Kristiyanti et al. (2022), could lead to a further reduction in the number of samples in each test fold, resulting in less reliable performance estimates for minority sentiment classes.

Conversely, research conducted by Verma and Thakur (2023) illustrates that 9-fold cross-validation effectively balances evaluation stability, class preservation, and performs effectively even in scenarios with minor class distributions, highlighting its robustness in practical applications. Moreover, 9-fold cross-validation slightly increases the sample size per fold, enhancing the stability of the evaluation compared to the 10-fold approach. In this study, the 9-fold approach is preferred because it achieves a better balance between training and validation data, thereby reducing variance in performance estimates while still preserving sufficient data for model learning. Consequently, this research adopts stratified 9-fold cross-validation to ensure consistent evaluation outcomes while maintaining computational efficiency. This method maintains proportional class distribution and minimises the risk of underrepresentation in smaller datasets.

The results in Table 3.8 demonstrate the effectiveness of stratified 9-fold cross-validation for evaluating a classification model. This method divides the dataset into 9 folds, maintaining the original class distribution to ensure balanced representation during training and testing. The model is trained on 8 folds and tested on the remaining one, repeating this process across all folds, which enhances reliability and mitigates bias from uneven class distributions. In Fold 1, the confusion matrix [[10, 0], [1, 50]] indicates accurate classification of all 10 instances of the first class and 50 of the second class, resulting in one misclassification. This leads to a high accuracy of 98.36%, along

with precision, recall, and F1-score values of 95.45%, 99.02%, and 97.12%, respectively. Folds 2 and 9 achieved perfect scores, while Fold 4 recorded an accuracy of 96.67% due to misclassifications. Fold 7 achieved 93.33% accuracy, which was the lowest among folds, illustrating consistent performance across all folds.

Table 3.8

Results of Stratified *k*-fold Cross-Validation

k	Fold	Confusion Matrix	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
9	2	[[10 0] [1 50]]	98.36	95.45	99.02	97.12
	3	[[10 0] [0 51]]	100	100	100	100
	4	[[10 0] [2 48]]	96.67	91.67	98.00	94.43
	5	[[10 0] [1 49]]	98.33	95.45	99.00	97.11
	6	[[10 0] [0 50]]	100	100	100	100
	7	[[10 0] [4 46]]	93.33	85.71	96.00	89.58
	8	[[9 0] [1 50]]	98.33	95.00	99.02	96.87
	9	[[9 0] [0 51]]	100	100	100	100
10	2	[[9 0] [2 44]]	96.36	90.91	97.83	93.89
	3	[[9 0] [0 46]]	100	100	100	100
	4	[[9 0] [1 44]]	98.15	95.00	98.89	96.81
	5	[[9 0] [1 44]]	98.15	95.00	98.89	96.81
	6	[[9 0] [1 44]]	98.15	95.00	98.89	96.81
	7	[[9 0] [2 43]]	96.30	90.91	97.78	93.86
	8	[[9 0] [2 43]]	96.30	90.91	97.78	93.86

k	Fold	Confusion Matrix	Accuracy (%)	Precision (%)	Recall (%)	F1 -Score (%)
	9	[[8 0] [1 45]]	98.15	94.44	98.91	96.51
	~10	[[8 0] [0 46]]	100	100	100	100

To sum it up, the findings show that both 9-fold and 10-fold stratified cross-validation achieve high accuracy, precision, recall, and Fl-score, indicating the model's robustness. However, the 9-fold method slightly outperforms 10-fold in stability and consistency, with higher accuracy performance across all folds except for Fold 4 and 7. This stability in the 9-fold process is due to its larger training set per iteration, which allows the model to generalise better, particularly benefiting moderately sized datasets. Therefore, stratified 9-fold cross-validation consistently achieves high performance, supporting its reliability and generalisability, justifying the selection of the evaluation technique for this research.

Figure 3.29 outlines the implementation of the SVM model with 9-fold cross-validation to evaluate model performance. First, the training feature matrix and encoded labels are provided as input, and a "*StratifiedKFold*" with nine splits is initialised to maintain balanced class distribution across folds. The SVM classifier is then configured with a linear kernel, balanced class weights to handle class imbalance, and probability estimates enabled for probabilistic predictions. During the process, the dataset is split into 9 folds, where 8 folds are used for training and one for validation in each iteration. Accuracy scores are recorded for each fold, followed by calculating the mean accuracy and standard deviation across folds to assess consistency and robustness. The results are then visualised by plotting fold numbers against accuracy scores, where individual fold accuracies are connected by lines, a horizontal dashed line indicates the mean accuracy, and a shaded region highlights ±1 standard deviation. Finally, the algorithm outputs the mean cross-validation accuracy, standard deviation, and a graphical representation, ensuring a comprehensive evaluation of SVM performance.

Algorithm 1 SVM with 0-Fold Cross-Validation

```
Input: Training feature matrix Xtroin, encoded label ... !Jtrain-fin:
Initialise StratifiedKFold with n.spiit >• = 0, shuttle enabled fixed random Bead
& Initialize SVM classifier:
◁: Kernel: Linear
Ge ("lass weight: Balanced
fi: Probability estimates: Enabled
7: Perform 9-Fold Cross-Validation on training set using accuracy as the metric
8: Record:
fl: Accuracy score for each fold
10: Mean accuracy across folds
11: Standard deviation of accuracy
12: Plot Cross-Validation Results:
l.i: X-axis: Fold number
M: Y-axis: Accuracy
15: Plot individual fold accuracies as points connected by lines
lfi: Draw a horizontal dashed line for mean accuracy
17: Shade region representing ±1 standard deviation
18: Add title, labels, legend, and grid
Ifl: Output: Mean CV accuracy, standard <leviation, and visual repres*•ntation of fold accuracies
```

Figure 3.29 Pseudocode for Stratified &-fold Cross-Validation

b) Confusion Matrix

Following that, the performance of the SVM model is further evaluated using a confusion matrix to evaluate the effectiveness of the classification models. It shows a table that visualises a model's performance, as illustrated in Figure 3.30. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are the four categories represented in a confusion matrix to compare actual and predicted values (Jayaswal, 2020). The evaluation metrics are then derived from the confusion matrix, where the accuracy measures the ratio of correctly predicted instances to total instances, while precision measures the ratio of true positive predictions to total positives and F1-score balances precision and recall (Octava et al., 2023). Moreover, the confusion matrix offers a detailed analysis of predictions across each sentiment category. This analysis is particularly significant when integrating VADER with TF-IDF and LDA, as it can result in subtle overlaps between classes. Hence, by visualising the confusion matrix along with precision, recall, and F1-score, the evaluation offers a comprehensive understanding of the model's predictive behaviour, ensuring a reliable assessment of its performance.

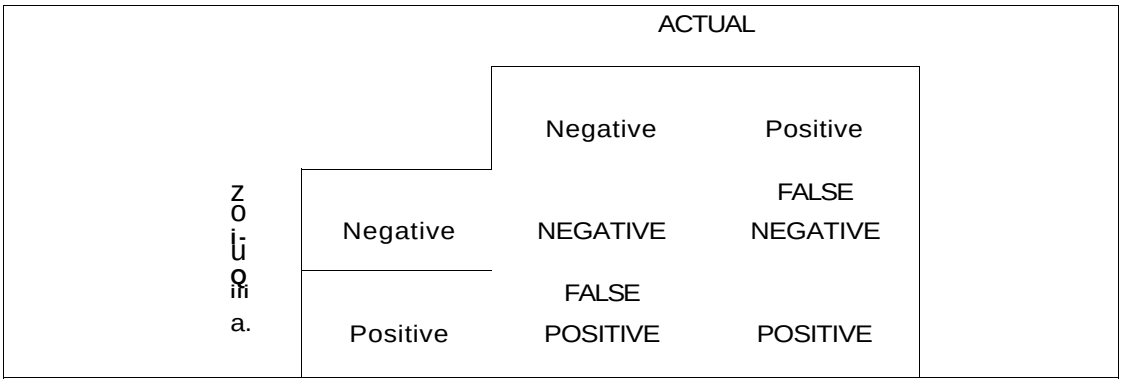


Figure 3.30 Confusion Matrix

(Source: Jayaswal, 2020)

Equation 3.6 shows accuracy for the percentage of all predictions that turned out correct, while Equation 3.7 represents precision that quantifies the accuracy of positive predictions. In contrast, Equation 3.8 is recalled to evaluate a model's capacity by identifying relevant data within a dataset, and it also refers to the percentage of real positives that were accurately classified. Models with high accuracy but low recall avoid getting unnecessarily optimistic performance results due to the Fl-score in Equation 3.9, which corresponds to the harmonic mean of precision and recall and balances precision and recall in datasets with unbalanced values.

$$accuracy \quad \frac{True\ \ Positive\ +\ True\ \ Negative}{True\ \ Positive\ +\ True\ \ Negative\ +\ False\ \ Positive\ +\ False\ \ Negative} \tag{3.6}$$

$$precision\ =\ \frac{True\ \ Positive}{True\ \ Positive\ +\ False\ \ Positive} \tag{3.7}$$

$$recall\ =\ \frac{True\ \ Positive}{True\ \ Positive\ +\ False\ \ Negative} \tag{3.8}$$

$$Fl\ \ score\ \ \frac{2\ \ xprecision\ \ xrecall}{precision\ \ \ xrecall} \tag{3.9}$$

Figure 3.31 evaluates pseudocode that describes a method for visualising a confusion matrix alongside key performance metrics in classification tasks. It starts by taking true labels, predicted labels, and a class label encoder as input. The confusion matrix is computed from the true and predicted labels, followed by the calculation of metrics such as accuracy, macro-averaged precision, recall, and Fl-score. A heatmap is

then created to represent the confusion matrix, with predicted labels on the x-axis and actual labels on the y-axis, and each cell annotated with counts using a consistent colour scheme. Class names from the label encoder are used for clarity, and a summary box displaying the calculated metrics is positioned next to the confusion matrix. The visualisation is completed with a title, axis labels, and layout adjustments. The final output is a well-annotated confusion matrix heatmap that provides a clear evaluation of the classification model's performance.

Algorithm 2 Confusion Matrix and Performance Metrics> Visualisation

1: Input: True test labels J/[B,,/,_K,M;, predicted labels *yi$_{KJ}$. t-predi* ClaSS label encoder
2: Compute confusion matrix using *yu,$_S$i -tint',* ^{d}
Calculate evaluation metrics: and *yiexi_pr,*
4: Accuracy
5: Precision (macro-averaged)
(i: Recall (macro- aver aged)
7: F 1-score (macro-averaged)
8: Create heatmap visualization:
9: X-axis: Predicted labels
10: Y-axis: Actual labels
11: Annotate cells with counts
12: Use consistent color scheme (Bines)
13: Apply class names from label encoder
14: Generate metrics summary text box:
15: Display accuracy, precision, recall. an< 1 F 1-score
Hi: Position text box beside the confusioii matrix
17: Add plot title, axis labels, and layout adjustments
18: Display the plot
19: Output: Confusion matrix heatmap with annotated performance metricK

Figure 3.31 Pseudocode for Confusion Matrix with Performance Metric

In other words, the confusion matrix analysis provides an evaluation of the SVM model's performance on Malaysia's islands testing dataset by comparing actual and predicted classifications. It reveals the model's ability to differentiate between two sentiment classes and identifies both correctly and incorrectly classified instances. Next, key performance metrics, such as accuracy, precision, recall, and F 1-score, are calculated to gauge the model's overall correctness and reliability. The Fl-score effectively balances precision and recall, which is particularly important in situations with class imbalances. By integrating these metrics into the confusion matrix visualisation, including a heatmap and annotated scores, the analysis offers a clear assessment of the SVM model's classification effectiveness.

Furthermore, training supervised learning models for sentiment categorisation often encounters the challenge of limited labelled data (RP 3). Hence, a hybrid SA approach can effectively address this issue by automatically generating sentiment scores using a word dictionary like VADER, which are subsequently utilised to train SVM model. Evaluation metrics such as precision, recall, accuracy, and Fl-score are then applied to assess the model's performance. However, accuracy can be misleading in unbalanced datasets, as it may disproportionately represent the sentiments of the dominant class. Therefore, the Fl-score is essential for evaluating precision and recall across different sentiment classes, as it offers a balanced perspective on model performance. Additionally, the AUC metric in the final phase is used to assess how well the model differentiates between positive and negative sentiments.

3.1.6.2 Validate using Area Under the Curve Metric

In the validation phase, the AUC of the Receiver Operating Characteristic (ROC) curve is used as an additional evaluation metric to complement conventional performance measures to assess the performance of the SVM model. AUC ranges from 0 to 1, with higher values indicating better classification performance (Cam et al., 2024). Contrary to metrics based solely on accuracy, the AUC delivers a more nuanced evaluation, particularly in cases of class imbalance. It assesses the balance between the TP rate and the FP rate across a range of threshold settings. In this research, AUC evaluation is applied to the hybrid SA approach using stratified &-fold cross-validation and an independent unseen dataset to comprehensively assess the model's generalisation performance. This approach ensures a robust evaluation of the model's generalisation performance by capturing its discriminative ability across multiple data splits as well as in previously unseen data.

Figure 3.32 assesses the performance of the SVM model using validation and test sets, incorporating ROC curve visualisation for assessing classification performance. It begins by taking the trained SVM model, along with the validation and test sets, and a label encoder as inputs. The evaluation is divided into two phases, which are the validation and testing phases. In the validation phase, the model predicts class labels and computes decision scores, generating a classification report that includes precision, recall, Fl-score, validation accuracy, and the AUC. Meanwhile, in the testing phase, the same steps are applied to the test set. The ROC curve visualisation is created

by calculating the false positive rate on the x-axis and true positive rate on the y-axis from the test scores, while including the AUC score in the legend alongside a reference line (y = x). Additionally, titles, labels, and grid lines are added to enhance clarity before displaying the plot. The final output consists of validation metrics, test metrics, and the ROC curve plot, providing a comprehensive evaluation of the model's performance.

Algorithm 3 Validation and Test Sot Evaluation with ROC Curve
1: Input: Trained SVM model, validation set $\{X_{va}i, y_{va}i\_sac)_1$ tost sot *(Xtest,yteat-etus),* label encoder
2: Validation Phase:
3: Predict class labels for validation set
4: Compute decision scores for validation set
5: Generate classification report {precision, recall, Fl-score) using label encoder classes
fi: Calculate validation accuracy
7: Calculate validation AUC
8: Testing Phase:
H: Predict class labels for test set
10: Compute decision scores for test set
11: Generate classification report (precision, recall, Fl-score) using label encoder classes
12: Calculate test accuracy
13: Calculate test AUC
14: ROC Curve Visualization:
15: Compute False Positive Rate (FPR) and True Positive Rate (TPR) from test decision scores
Ifi: Plot HOC curve with:
17: X-axis: FPR
18: Y-axis: TPR
19: AUC score in legend
2G: Diagonal reference line *(y = x)*
21: Afld title, axis labels, legend, and grid
22: Display the; plot
23: Output: Validation metrics, test metrics, ROC curve plot

Figure 3.32 Pseudocode for Area Under the Curve Metric

3.2 Summary

In summary, this study conducted SA on the islands of Malaysia using a dataset of YouTube reviews sourced from publicly available content on the platform. The research methodology is structured into six distinct phases, each playing a vital role in the development of the model framework. The first phase consists of an extensive literature review concerning SA and DSA techniques, establishing the theoretical and methodological foundation for the research. The second phase focuses on selecting DSA techniques, particularly the integration of TF-IDF and LDA for feature extraction and topic modelling.

Meanwhile, the third phase entails data collection and pre-processing of YouTube reviews, incorporating lemmatisation, stop word removal, and tokenisation to prepare the data for accurate sentiment trend analysis. Ethical considerations were strictly adhered to during this phase by collecting only publicly accessible data, anonymising user identities, and excluding personally identifiable information to ensure user privacy. There was no direct interaction with users, and the data were analysed solely for academic research purposes. The fourth phase addresses the limitations of conventional SA techniques by applying TF-IDF and LDA to identify significant terms, latent topics, and aspect labels, thereby enhancing contextual understanding.

Following that, the fifth phase constructs a hybrid SA approach by combining the lexicon-based tool VADER with the ML classifier SVM to improve sentiment prediction accuracy. Finally, the sixth phase evaluates the reliability and performance of the model using standard evaluation metrics, including accuracy, precision, recall, and F1-score, complemented by validation on unseen data utilising the AUC metric. Building upon this structured and ethically guided methodology, Chapter 4 presents a detailed explanation of the results and discussions related to the model framework.

CHAPTER 4

RESULT AND DISCUSSION

This chapter presents the findings of the research and analyses the results from the implemented models to uncover patterns, trends, and relationships within the data. It also discusses how these findings relate to the research objectives and explores potential implications and recommendations for future enhancements. Consequently, this chapter plays a crucial role in connecting data analysis with meaningful conclusions that lead to actionable recommendations and insights.

4.1 Research Overview

This chapter presents an overview of the implementation and findings of the model framework applied to YouTube reviews of Malaysia's islands, as illustrated in Figure 4.1. The framework operationalises advanced text analytics to extract meaningful textual and semantic features, followed by a hybrid sentiment analysis (SA) approach that integrates lexicon-based and machine learning (ML) techniques for enhanced sentiment detection. In the descriptive-semantic analysis (DSA) phase, textual patterns and thematic structures within the review datasets are uncovered through n-gram analysis and document-term matrix construction using Term Frequency-Inverted Document Frequency (TF-IDF). Additionally, topic modelling and manual aspect labelling were conducted using Latent Dirichlet Allocation (LDA), to generate document-topic distributions and provide insights into dominant themes within the review datasets. This phase is crucial to establishing a basis for assessing the effectiveness of the hybrid SA approach in capturing the nuanced sentiments expressed by tourists, thereby deepening the comprehension of users' perceptions of Malaysia's island destinations.

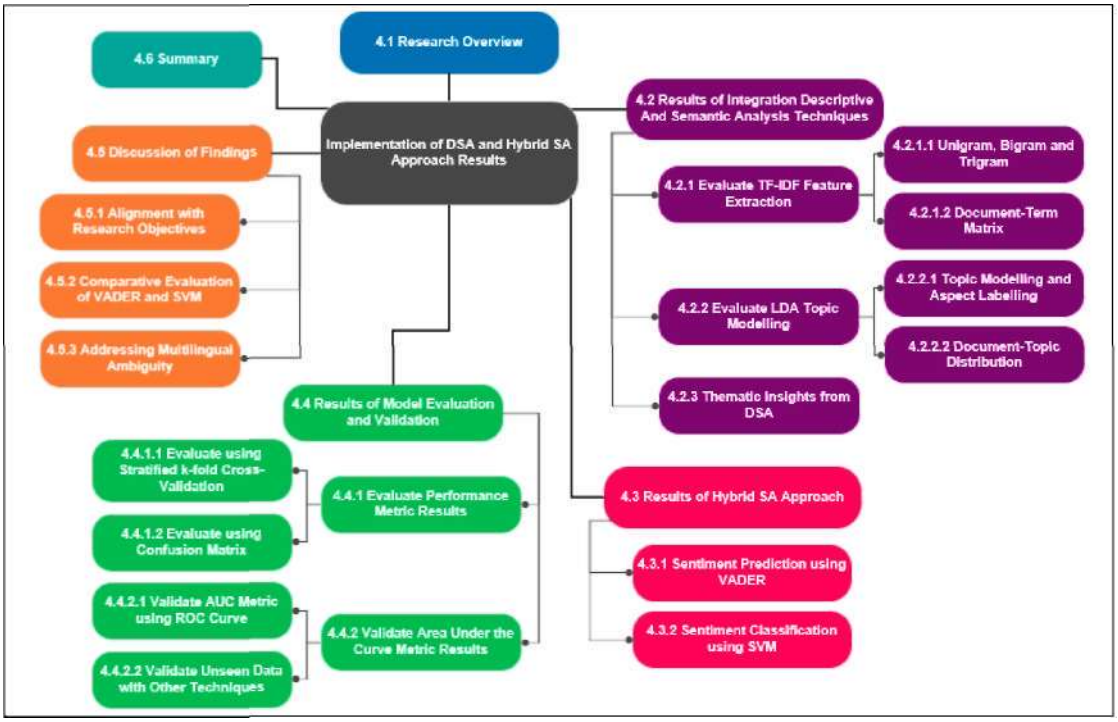


Figure 4.1 Overview of Results and Discussion

Furthermore, for sentiment prediction, the hybrid SA approach begins with the Valence Aware Dictionary and Sentiment Reasoner (VADER) to generate sentiment polarity scores, which are then integrated with textual features into a Support Vector Machine (SVM) model. After that, model evaluation involves the use of performance metrics such as accuracy, precision, recall, and Fl-score. A confusion matrix aids in providing detailed insights into the prediction outcomes for each sentiment class. Furthermore, the Area Under the Curve (AUC) metric is used to validate the model's ability to effectively differentiate between sentiment classes and imbalanced data. Next, the chapter discusses the findings and concludes by emphasising the effectiveness of integrating DSA techniques with a hybrid SA approach.

4.2 Results of Integration Descriptive and Semantic Analysis Techniques

The integration of DSA techniques with a hybrid SA approach was utilised to evaluate tourist sentiments and perceptions related to Malaysia's islands. This section details the primary analytical methods applied to derive insights from the review datasets, emphasising TF-IDF feature extraction, LDA topic modelling, and thematic interpretation through aspect labels. TF-IDF was used to identify significant terms in each review by assessing word importance across the corpus, effectively filtering out

common and non-informative words. Following this, LDA was applied to uncover hidden thematic structures within the reviews. To enhance clarity, aspect labels were manually assigned to categorise these topics into high-level themes, resulting in a more structured presentation of user concerns and preferences. Together, these techniques provided a nuanced understanding of tourists' experiences and sentiments regarding Malaysia's island destinations.

4.2.1 Evaluate Term Frequency-Inverse Document Frequency Feature Extraction

Descriptive analysis utilising the TF-IDF technique effectively identifies the most significant terms in text reviews by evaluating their importance within datasets. When applied to n-gram results, TF-IDF highlights words that are particularly relevant or unique to Langkawi, Perhentian, and Pangkor Islands. Additionally, integrating TF-IDF-weighted n-gram features with SVM is essential for assessing the effectiveness of unigram, bigram, and trigram in sentiment classification.

Following that, to enhance this process, a document-term matrix was developed utilising TF-IDF scores, thereby converting qualitative textual data into a structured numerical format. This matrix captures the term significance, weighted by its frequency across all documents, which facilitates consistent comparison and analysis. Overall, this analysis offers valuable insights into the perceptions, preferences, and experiences of tourists. Filtering out frequently used keywords and focusing on contextually significant terms enhances our understanding of the recurring themes and distinctive characteristics associated with all three of Malaysia's islands.

*4.2.1.1 Unigram, **Bigram and Trigram***

A dataset containing n-words is called an n-gram; n = 1 is a unigram, n = 2 is a bigram, and n = 3 is a trigram. The n-gram findings are then utilised to compute the inverse document frequency for word lists from datasets using TF-IDF (Kaur & Sharma, 2023). The n-gram technique captures the sequence of words by concentrating on multiword tokens within a specified context window. This approach is particularly effective because the meanings of words can change based on their combinations and can be represented in a compact vector format (Semary et al., 2024). In addition, to

identify the most significant features for SA, unigram, bigram, and trigram are assessed using SVM classification. SVM efficiently manages the high-dimensional feature space generated by TF-IDF-weighted n-gram and distinguishes between different sentiment classes. This evaluation identifies the most informative n-gram structure, thereby enhancing model performance by leveraging both individual word contributions and their contextual relationships.

Furthermore, to attain a more nuanced understanding of the linguistic structures within the dataset, analyses of unigram, bigram, and trigram were performed. These n-gram models represent varying levels of contextual information, such as unigram, which focuses on the frequency of individual words, while bigram and trigram examine sequences of two and three words, respectively. This approach not only identifies commonly occurring terms but also reveals meaningful combinations of words that enhance the semantic interpretation of the text. By analysing these n-gram patterns, underlying relationships between words can be uncovered, offering deeper insights into the textual content that may not be readily apparent through single-word analysis alone.

Figure 4.2 illustrates the top 10 terms for Langkawi, Perhentian, and Pangkor Islands based on their TF-IDF scores. Higher scores indicate more significant terms in island-specific reviews and provide insights into tourists' perceptions and discussions of the three destinations. The results show that Langkawi's most prominent term, *"langkawi"*, with a TF-IDF score of 45.87, indirectly reveals the frequent mentions of the island's name in the reviews. Other significant terms, such as *"good'* and *"malaysia"*, with 32.03 and 27.38, resulted in positive sentiment and potential references to Langkawi's significance as a tourist destination in Malaysia. Next, the term *"island"* is highly ranked at 27.33, based on the island's geographical identity. Negative sentiment is observed with the term *"not"*, which serves as a negation operator and can indicate positive or negative sentiment depending on the context, whereas other terms like *"place"*, *"beautiful'*, *"nice"*, and *"visit'* reveal Langkawi's popularity for scenic beauty and beaches.

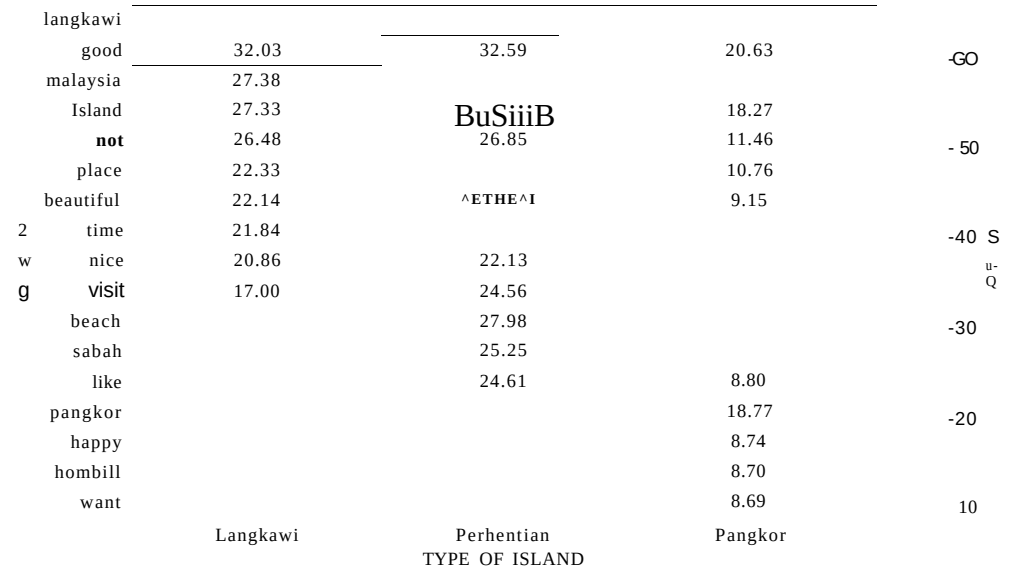


Figure 4.2 Unigram of Three Islands

Meanwhile, Perhentian has the highest TF-IDF score of 65.58 for the term *"island'* due to its popularity as a tourist destination, whereas the prominent term *"malaysia"* has a score of 41.53, which is related to its association with Malaysia's tourism appeal. Next, positive descriptors such as *"good'* and *"beautifuF,* with 32.59 and 39.24, refer to the island's scenic beauty, while the term *"beach"* has 27.98 as a beach destination. Other terms like *"visit"* and *"like"* refer to positive personal preferences, while *"sabah"* suggests comparisons with islands in Sabah. On the other hand, Pangkor has lower TF-IDF scores compared to Langkawi and Perhentian, whilst terms such as *"good', "pangkor",* and *"island",* with scores of 20.63, 18.77 and 18.27, respectively, refer to positive sentiment about the island. Next, terms like *"place", "not",* and *"beautiful"* with 10.76, 11.46 and 9.15 could be mixed reviews about the island destination. Also, the presence of *"hornbills"* at 8.70 is unique to Pangkor, which is well-known for its hombill population and is indirectly an ecological aspect of Pangkor tourism.

Moreover, bigram analysis, as in Figure 4.3, provides a more particular context by focusing on the pair of terms. This leads to more comprehensive insights into term relevance for the Langkawi, Perhentian, and Pangkor islands based on TF-IDF scores. In the case of Langkawi, the most frequently mentioned bigram is *"visit langkawi",* with 4.39 for its popularity as a tourist destination. Other terms include *"beautiful place", "good time", "good island',* and *"roti canai",* with scores of 4.38, 4.09, 3.19 and 3.94, which refer to positive sentiment for the island's scenic beauty, experience

and local delicacies. Next, the presence of *"not know"* with 3.48 can be referred to as uncertainty or mixed experiences among tourists, whereas *"love langkawi"* and *"island malaysia"* illustrate strong positive sentiments from tourists who enjoyed their stay on the island.

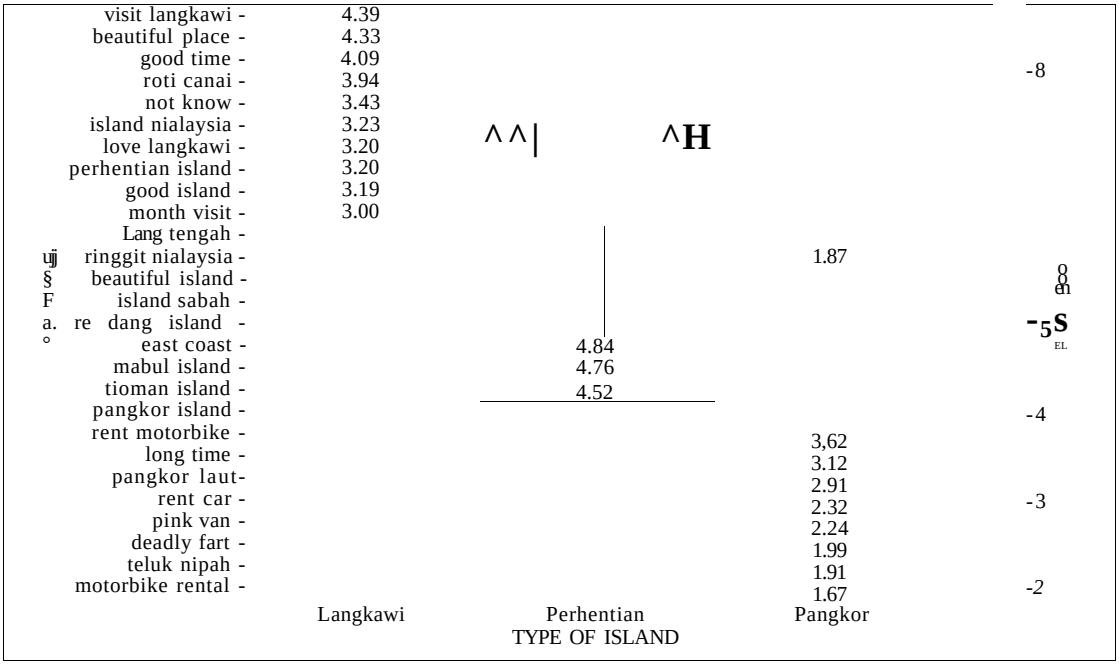


Figure 4.3 Bigram of Three Islands

For Perhentian, the highest-ranking bigram is *"perhentian island'*, with 8.76 as the highest TF-LDF score that refers to the island destination. Following that, the bigram *"beautiful island'* with 7.59 illustrates a strong positive sentiment towards Perhentian's natural beauty. Other prominent terms, including *"long tengah"*, *"redang island'*, *"island sabah"*, *"mabul island'*, and *"tioman island'*, are often mentioned for comparisons with other Malaysia islands. However, compared to Langkawi and Perhentian, Pangkor has fewer highly ranked bigram, leading to less frequent discussions about this island. The most significant bigram for Pangkor Island is *"pangkor island'*, with 8.61 showing frequent mentions in tourist reviews. Meanwhile, the presence *of "rent motorbike"*, *"rent car"*, and *"pink van"* with 3.62, 3.12 and 2.24 is a common topic among tourists about renting vehicles as an important aspect while travelling in Pangkor.

On top of that, bigram provide a deeper connection between terms and topics compared to unigram by pairing terms within specific contexts, such as activities,

sentiments, or attractions. While unigram offer a basic summary of important concepts, they lack specificity regarding how these terms are used. However, although bigram can yield more useful and insightful findings, they may also introduce noise from uncommon or unnecessary pairings that arise due to limited data. As demonstrated in the heatmap results, bigram contribute to a more comprehensive understanding of the significance of terms, especially in descriptive and thematic analyses.

Collecting phrases with three terms provides a more detailed contextual meaning, as the trigram analysis in Figure 4.4 significantly simplifies term relevance. Subsequently, trigram provides more context than unigram and bigram in providing better insights into tourists' experiences on each island. The highest-ranked trigram for Langkawi is *"wonderful place langkawi", "good island malaysia", "absolutely beautiful place",* and *"resort world langkawi"* with 2.67, 2.00, 1.69 and 2.00, which demonstrate positive perception of the island destination. Other term phrases include *"indian driving license", "month visit langkawi", "foodpeople amazing", "sabah east malaysia",* and *"greeting sabah east",* which refer to the context of renting vehicles, travelling duration, hospitality, along with comparisons with other Malaysia islands.

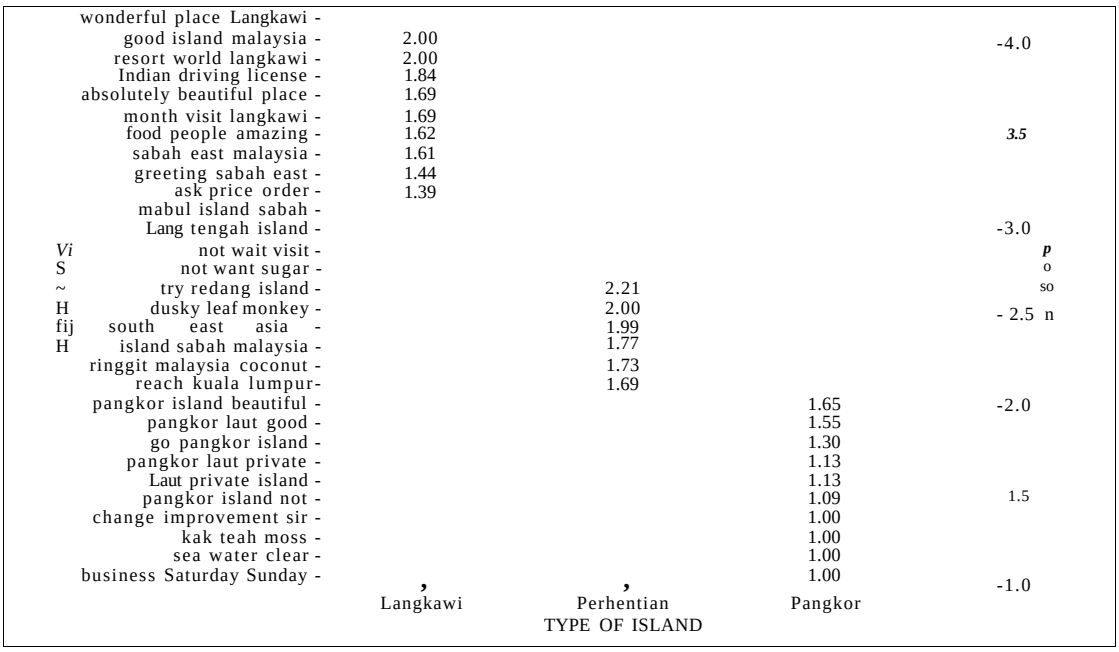


Figure 4.4 Trigram of Three Islands

On the other hand, the trigram for Perhentian and Pangkor focuses on the popularity of each destination. Perhentian is often compared to Mabul Island, Sabah, Lang Tengah Island and Redang Island, with 4.18, 3.79 and 2.21, often considered an

alternative or complementary destination to Perhentian. Other phrases like *"not want visit"* and "not want sugar" could refer to negative experiences or dietary preferences. On the other hand, Pangkor is ranked highest for its beauty as the term *"pangkor island beautifuF* with a score of 1.65, while *"pangkor lout good\ "go pangkor island"* and *"pangkor lout private"* with 1.55, 1.30 and 1.13, indicating the positive reviews about travel recommendations and experiences. Meanwhile, the term phrases *"pangkor island nof* and *"change improvement sir"* refer to mixed reviews or negative experiences. Last but not least, the terms for *"sea water clear"* imply a positive opinion of Pangkor island's sea, while the term *"business Saturday Sunday"* indicates intense business dealings on weekends that can overwhelm tourists, but it also conveys a neutral or mixed sentiment.

Overall, unigram analysis is inadequate for more detailed analysis as it summarises commonly used terms without providing context. For instance, terms like *"beautifur, "visit",* and *"island'* are commonly used, but their actual meanings are still unclear. In order to better comprehend tourists' interactions, bigram analysis integrates terms together to uncover clearer patterns, provide more detail throughout reviews, and be more effective in identifying theme structures. Phrases consisting of *"visit langkawi", "beautiful island',* and *"island malaysia"* offer a clearer idea that refer to a certain place. Meanwhile, trigram analysis serves as the most specific by identifying complete phrases to reveal clear purpose, such as *"wonderful place langkawi", "pangkor island beautifur,* and *"island hopping cost",* yet trigram is exceptionally particular and appear less frequently, resulting in a smaller amount of relevant terms.

Therefore, in addition to the descriptive analysis, a performance comparison was conducted using SVM model with unigram, bigram, and trigram features to assess their effectiveness in sentiment classification. As presented in Table 4.1, the unigram model achieved the highest mean accuracy of 63.85%, accompanied by a standard deviation of 1.136%, which demonstrates that individual word features provided a stable and reliable level of performance. In contrast, the bigram model followed with a mean accuracy of 53.26%, but it displayed slightly higher variability at 1.560%). This suggests that while it incorporated richer contextual information through word pairings, it also introduced increased complexity and sparsity. Lastly, the trigram model, which sought to capture phrase-level semantics, recorded the lowest mean accuracy at 44.67%), with a low standard deviation of 0.309%), resulting in limited generalisability, which is likely due to its sparsity and low term frequency. In summary, unigram and

bigram representations demonstrated commendable performance along with distinct advantages, where unigram provided stable accuracy and bigram offered crucial contextual insights. Consequently, both features were chosen for integration with LDA to effectively enhance topic modelling and classification tasks.

Table 4.1

Comparison of Performance for Unigram, Bigram, and Trigram with Support Vector Machine

Type of n-gram	Mean Accuracy (%)	Standard Deviation Accuracy (%)
Unigram	63.85	1.136
Bigram	53.26	1.560
Trigram	44.67	0.309

In other words, the analysis of unigram, bigram, and trigram models utilising the SVM model reflects a balance between contextual richness and classification performance. While bigram and trigram models offer more specific representations, their effectiveness is often constrained by data sparsity and increased dimensionality. In this case, the unigram model exhibited the highest accuracy and the least variability, while the bigram model provided significant contextual information, despite its slightly lower performance. Consequently, the combination of unigram and bigram strikes a suitable balance between semantic richness and statistical reliability, making them the ideal choice for integration with LDA and classification tasks. Next, the upcoming section will focus on constructing a document-term matrix, which is vital for converting textual data into a numerical format.

4.2.1.2 Document-Term Matrix

The document-term matrix technique transforms text into a numerical format, emphasising terms that are frequent in individual documents but rare across the entire dataset, thereby highlighting important features. This matrix effectively captures the significance of individual words while maintaining computational efficiency. It also serves as the input for training the SVM model, allowing for an evaluation of its predictive performance. Moreover, the inclusion of n-gram features in the document-term matrix improves the representation of textual information by recognising both

single words and contextual clusters of words, which facilitates more effective feature extraction and classification.

In this matrix, each document corresponds to a row, while each n-gram term represents a column, with values indicating the TF-IDF scores that reflect the importance of terms within documents in relation to the entire corpus. The TF component measures the frequency of a term in a specific document, whereas the IDF component downweighs common terms across the corpus, thereby emphasising those that are more distinctive. Higher TF-IDF scores suggest that a term is both frequent in a document and rare in others, making it a strong identifier; conversely, lower scores imply less significance. By analysing scores at different n-gram levels, key themes, contextual patterns, and vocabulary within documents can be identified. This approach offers deeper insights into semantic structures for effective sentiment classification.

For Langkawi Island in Figure 4.5, a document-term matrix was developed to analyse the key themes. Each row corresponds to an individual review, while each column denotes a specific term. The values within the matrix represent the TF-IDF score of each term, reflecting its significance relative to the entire collection of reviews. These scores were visualised using a heatmap. Additionally, assessing the relevance of a TF-IDF score requires context rather than fixed thresholds. A high score is significant when compared to other terms within the same document, while a low score suggests commonality or absence. This relative comparison enables the identification of representative terms in each review.

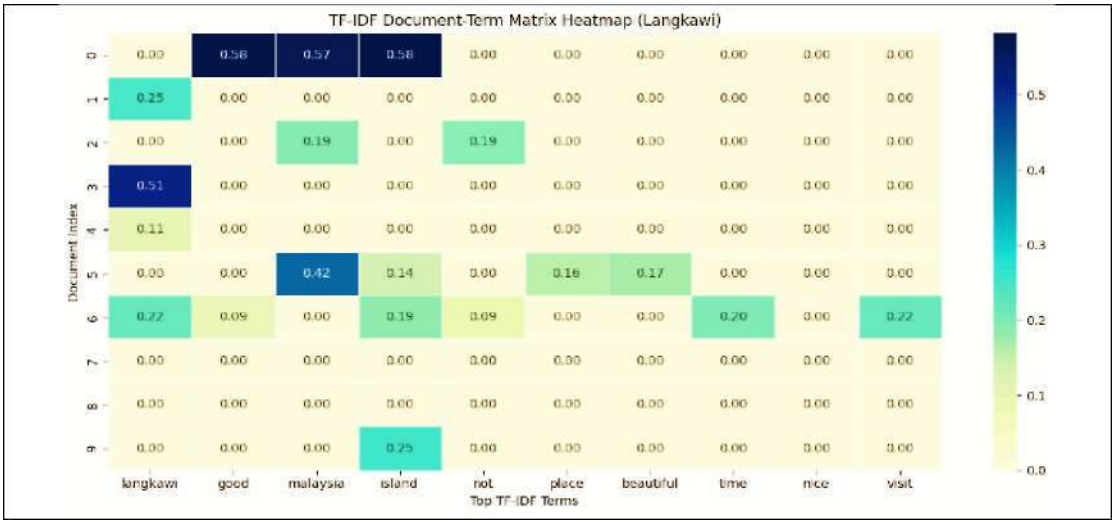


Figure 4.5 Document-Term Matrix for Langkawi

For instance, in Document 0, the terms like *"langkawi", "good",* and *"malaysia"* achieved TF-IDF scores exceeding 0.57, indicating their significance in that particular review. In contrast, Document 5 emphasises significant terms such as *"island", "place",* and *"beautifuF,* with *"malaysia"* standing out as the most significant, achieving a score of 0.42. Overall, each document contains its own significant terms, while some documents scored zero due to the absence of those terms. Subsequently, this document-term matrix provides valuable insights into the language patterns present in tourist reviews of Langkawi. It highlights distinctive terms and serves as a robust foundation for further analyses, such as topic modelling or sentiment classification. By identifying significant terms, this technique uncovers themes and perceptions associated with the travel experience in Langkawi.

Meanwhile, Figure 4.6 presents a document-term matrix for Perhentian Island, indicating a sparse distribution of term importance among tourist reviews. The heatmap shows that most terms, such as *"island', "beautifuF', "beach", "not", "sabah", "like",* and *"visit",* have TF-IDF values close to zero, suggesting these terms are either absent from specific reviews or too common to be significant. There is a limited number of documents that emphasise terms with higher scores. For instance, Document 4 features the term *"good*\ which has a score of 0.48. Similarly, Document 7 highlights the term *"malaysia"* with a score of 0.40, and Document 2 includes the term *"nice"* with a score of 0.31, as these scores indicate the significance of these terms in the reviews. The overall low density of high TF-IDF scores suggests a repetitive vocabulary where tourists often use similar descriptive words. No single term appears consistently across multiple documents with a high TF-IDF value, indicating a lack of dominant keywords.

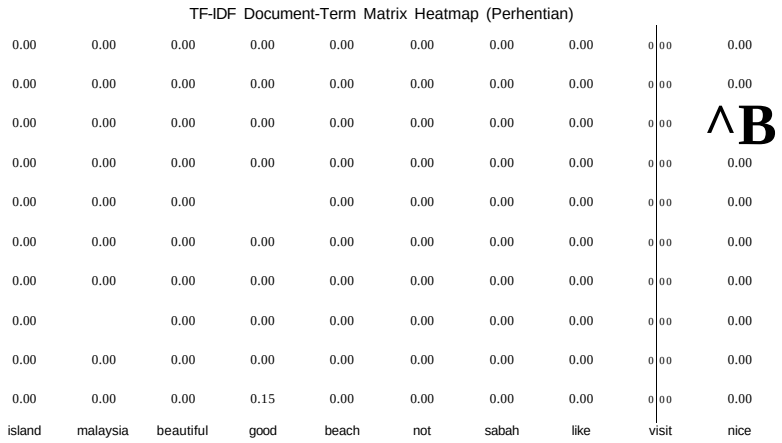


Figure 4.6 Document-Term Matrix for Perhentian

Figure 4.7 illustrates a document-term matrix for Pangkor Island, which shows a sparse distribution of term importance across reviews. Most terms like *"good\ "not", "place", "like", "happy"* and *"hornbiir* have TF-IDF scores close to zero, indicating these terms are either absent or too commonly used to be significant in individual documents. Notably, Document 0 highlights *"beautifuF* with the highest TF-IDF score of 0.32, suggesting it is particularly significant in that review. Document 4 emphasises *"pangkor"* and *"island"* with scores of 0.18 and 0.19, indicating a focused mention of the location. Document 7 scores 0.21 for "want", which may reflect a personal preference expressed in the review. Overall, the matrix reveals a consistent pattern across reviews, with few standout terms. While the language used is often repetitive, some terms still carry distinctive weight. These insights can help identify key expressions and provide a basis for further analysis.

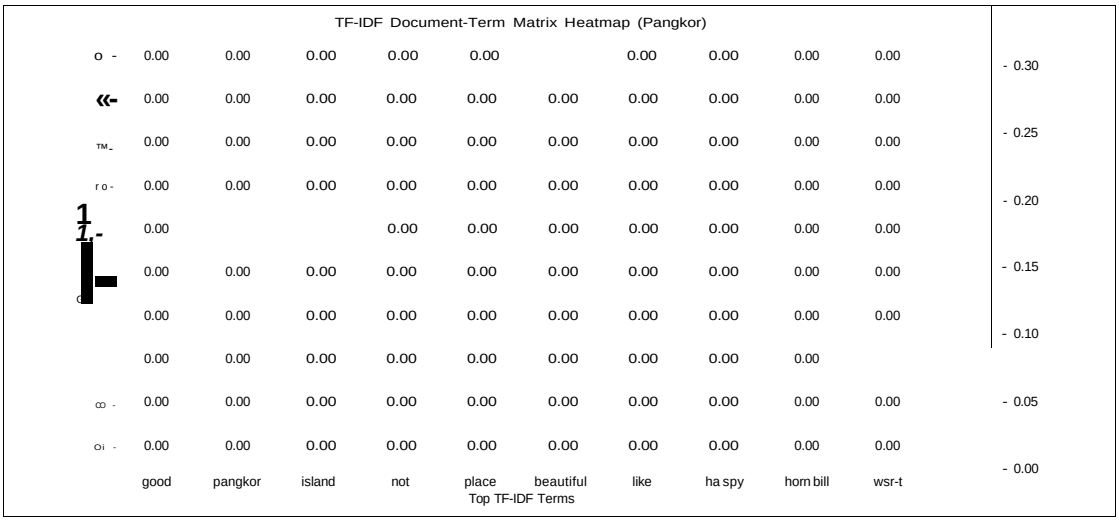


Figure 4.7 Document-Term Matrix for Pangkor

In conclusion, the analysis of the TF-IDF document-term matrix for Langkawi, Perhentian, and Pangkor Islands reveals distinct patterns in tourist reviews. All three islands show sparsity, with only a few terms having high TF-IDF scores, indicating a limitation on the diversity of unique and high-weighted terms within individual documents. Langkawi demonstrates the most varied term relevance, as documents frequently feature terms like *"good", "malaysia",* and *"island'* with scores above 0.57. This reflects its popularity and diverse attractions, prompting more elaborate reviews. In contrast, Perhentian Island has a more concentrated distribution; high TF-IDF scores for *"good'* are found only in a few documents, indicating repetitive phrasing among

reviewers. Meanwhile, Pangkor Island follows a similar trend but with even fewer standout documents, primarily featuring *"beautifur* in Document 0 and *"want"* in Document 7, indicating limited lexical diversity and fewer tourist engagements.

To sum it up, the comparative analysis of n-gram features with SVM demonstrates that unigram and bigram models are the most suitable for further analysis. This finding is further reinforced by the TF-IDF results, where the document-term matrix revealed significant lexical distinctions across the three islands. Specifically, the unigram model achieved the highest accuracy and stability, whereas the bigram model contributed meaningful contextual relationships that enriched the representation of text, although slightly less effective in terms of accuracy. In contrast, trigram features introduced sparsity and increased dimensionality without yielding notable performance improvements. Consequently, unigram and bigram representations were selected for implementation within the TF-IDF framework and integration with LDA in order to ensure both statistical reliability and semantic depth in uncovering latent themes within the reviews.

4.2.2 Evaluate Latent Dirichlet Allocation Topic Modelling

A semantic analysis technique known as LDA is designed to uncover dominant topics within textual reviews by analysing patterns of word co-occurrence. Unlike TF-IDF, which prioritises term frequency, LDA organises related terms into latent topics, facilitating a deeper contextual interpretation of user content. Once these dominant topics are identified, each topic is assigned an aspect label manually that reflects its broader semantic category. Subsequently, LDA generates a document-topic distribution that illustrates the proportion of each topic present in every review, enriching the understanding of tourists' perceptions and experiences across various island destinations.

4.2.2.1 Topic Modelling and Aspect Labelling

This section outlines the process of identifying semantic patterns in user-generated content through topic modelling and aspect labelling. The objective is to uncover themes and categorise reviews for a clearer understanding of user sentiment. By applying the LDA technique, hidden structures can be extracted from the text

without reliance on predefined labels. This approach facilitates the identification of recurring topics, which can then be mapped to broader categories through aspect labelling. As a result, this methodology enhances the ability to monitor tourism trends, prioritise visitor concerns, and inform destination management decisions using data-driven insights.

a) Dominant Topic

LDA is a topic modelling technique utilised to uncover hidden themes within textual review data. It analyses patterns of word co-occurrence, based on the idea that each document contains a combination of topics, with each topic represented by a distribution of words. By analysing the entire dataset, LDA reveals dominant topics that capture the key themes present in user reviews. Additionally, the dominant topic for each document is identified based on the highest probability score, allowing for effective classification of reviews according to their key themes.

Table 4.2 provides a comparative overview of the dominant topic distributions and their weights for Langkawi, Perhentian, and Pangkor. Each island shows a unique pattern from the LDA analysis, where average weight indicates the average probability of a topic across all documents, highlighting the overall significance of each topic to avoid bias toward frequently occurring topics. In Langkawi, Topic_7 stands out as the most prominent, appearing in 179 documents with a total weight of 160.97 and an average topic weight of 0.140. Topic_8 and Topic_6 follow closely, with average weights of 0.113 and 0.112, indicating a moderate influence from several themes. Moreover, Perhentian exhibits a clear dominance of Topic_1, which appears in 243 documents and carries a total weight of 199.55, leading to the highest average weight of 0.154. Topic_4 ranks second, appearing in 205 documents with a weight of 184.13, highlighting a concentration on key topics. In contrast, Pangkor demonstrates a more balanced topic distribution. Topic_8 has the highest average weight at 0.125, while Topic_5 and Topic_7 follow at 0.122 and 0.114, respectively. No single topic significantly overshadows the others, suggesting a more equitable spread of influence. In summary, Langkawi and Perhentian showcase strong dominance in specific topics, whereas Pangkor exhibits a more balanced distribution across multiple topics.

Table 4.2

Comparison of Dominant Topic Weights Across Islands

Type of Islands	Dominant Topic	Dominic Topic Count	Weight	Average Weight
Langkawi	Topic_1	126	113.18	0.099
	Topic_2	85	98.05	0.085
	Topic_3	104	105.72	0.092
	Topic_4	100	106.04	0.093
	Topic_5	97	103.59	0.090
	Topic_6	131	128.34	0.112
	Topic_7	179	160.97	0.140
	Topic_8	136	129.48	0.113
	Topic_9	96	103.42	0.090
	Topic_10	92	97.21	0.085
	Topic_1	243	199.55	0.154
	Topic_2	87	101.23	0.078
	Topic_3	91	103.70	0.080
	Topic_4	205	184.13	0.142
	Topic_5	131	129.86	0.100
	Topic_6	70	88.95	0.069
	Topic_7	105	114.11	0.088
	Topic_8	101	113.26	0.087
	Topic_9	146	141.72	0.109
	Topic_10	116	118.49	0.092
	Topic_1	68	59.01	0.105
	Topic_2	57	57.00	0.101
	Topic_3	40	44.90	0.080
	Topic_4	48	50.70	0.090
	Topic_5	77	68.79	0.122
	Topic_6	40	46.58	0.083
	Topic_7	66	64.39	0.114
	Topic_8	75	70.12	0.125
	Topic_9	42	47.70	0.085
	Topic_10	50	53.81	0.096

On the other hand, the bar chart in Figure 4.8 visually represents the dominant topic weights summarised in the table for three islands. In Langkawi, Topic_7 is the most dominant, with an average normalised weight of 0.14, indicating a strong focus

on a specific theme. This is followed by Topic_8 and Topic_6, reflecting a moderate concentration around a few dominant topics. Perhentian shows a similar pattern, with Topic_1 having the highest weight at 0.15, closely followed by Topic_4 at 0.14. This indicates a strong emphasis on specific aspects of visitor experiences. Conversely, Pangkor exhibits a more balanced distribution. Topic_8 (0.12), Topic_5 (0.12), and Topic_7 (0.11) are the most prominent, but none dominate, showcasing a wider range of tourist interests. Overall, the bar chart highlights how Langkawi and Perhentian focus on specific themes in user reviews, while Pangkor reflects more diverse feedback, offering insights into differing tourist perceptions by destination.

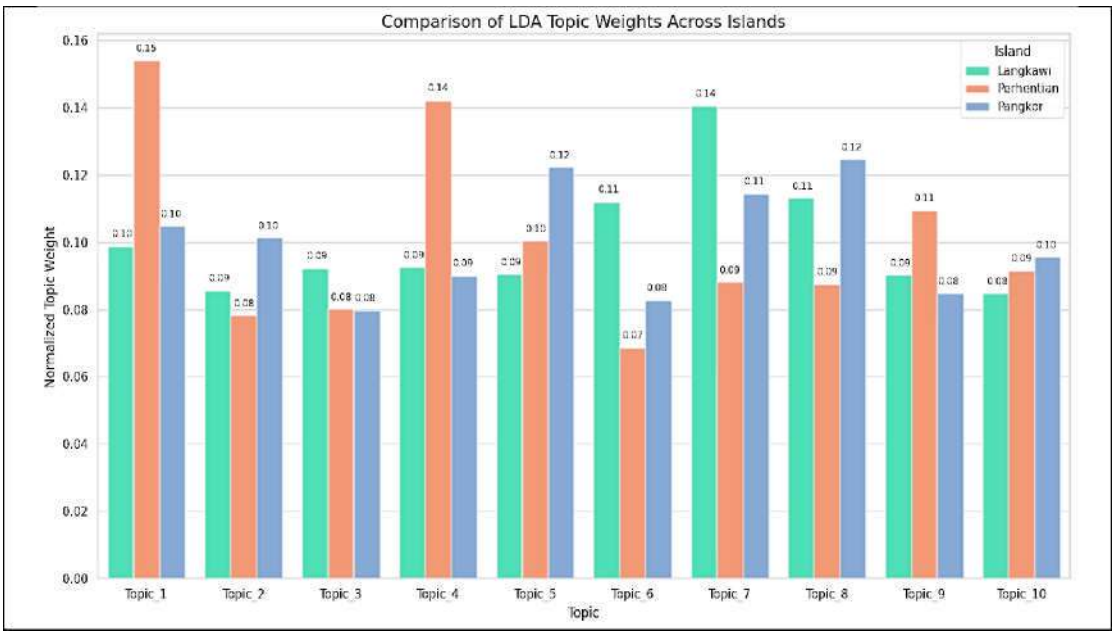


Figure 4.8 Comparison of Latent Dirichlet Allocation Topic Weights Across Islands

Furthermore, to enhance the clarity of the topics, word cloud visualisations are utilised to emphasise the key terms associated with each topic. The font size reflects the importance of each term based on its probability weight. Following the LDA analysis that identified dominant topics and their corresponding weights, Figure 4.9 shows a word cloud that was generated to illustrate the most frequently mentioned words associated with Langkawi. This visualisation enhances the quantitative findings by highlighting the significant terms that define the themes within the reviews.

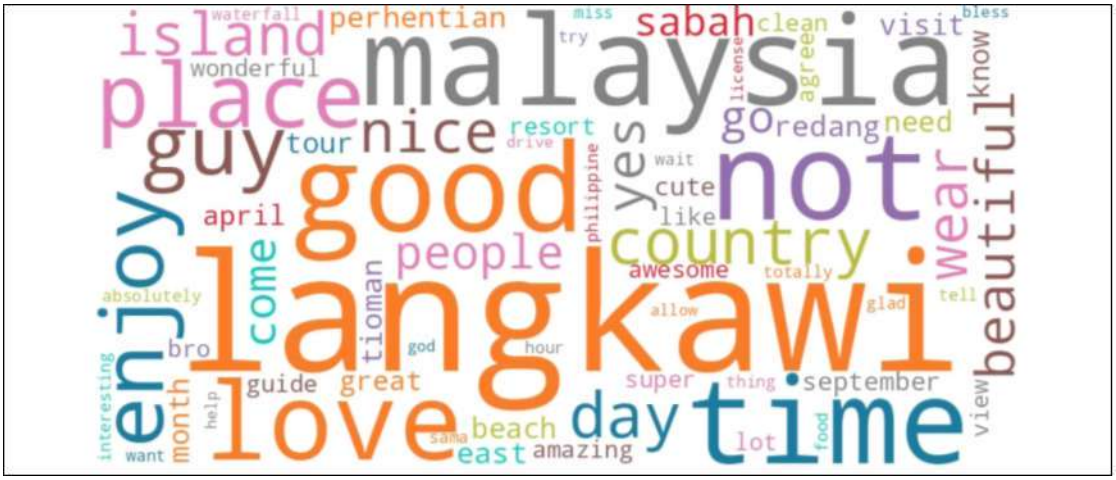


Figure 4.9 Langkawi Dominant Topic

For instance, significant terms in the Langkawi word cloud include *"langkawi"*, *"love"*, *"good\ "time"*, *"place"*, *"enjoy"*, and *"day"*, all of which convey a generally positive sentiment among tourists. Terms like *"love"*, *"good\ "enjoy"*, *"nice"*, *"awesome"*, and *"beautifuF* emphasise the emotional connections and enjoyable experiences visitors have had, while time-related terms suggest discussions about trip durations and the best seasons to visit. Geographical references such as *"malaysia"*, *"island'*, *"tioman"*, *"perhentian"*, *"redang"*, and *"sabah"* imply comparisons with other destinations. The inclusion of the word *"not"* indicates that some reviews contain negative aspects or offer balanced perspectives. Overall, the Langkawi word cloud effectively captures common sentiments and experiences of tourists, visually reinforcing the insights derived from the LDA analysis.

For Perhentian Island, the word cloud in Figure 4.10 reveals significant terms extracted from user-generated content. Prominent terms such as *"malaysia"*, *"island'*, and *"sabah"* suggest that Perhentian is frequently discussed alongside other popular islands. Although Sabah is a distinct geographical region, its inclusion indicates that reviewers often draw comparisons between various islands in their travel narratives. Positive descriptors like *"beautifuF*, *"amazing"*, *"good"*, and *"clear"* reflect the generally favourable perceptions of Perhentian Island, associating it with natural beauty and enjoyable experiences.

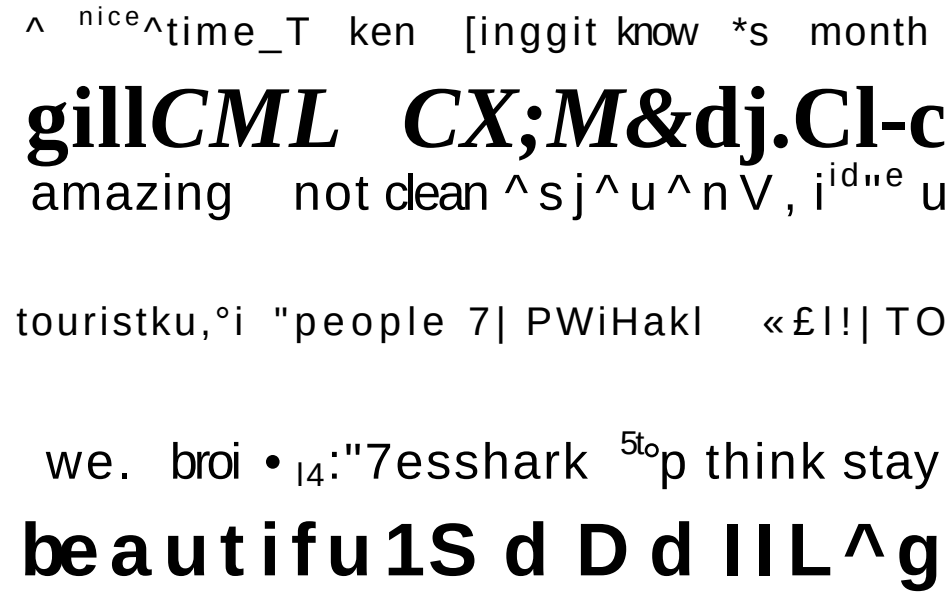


Figure 4.10 Perhentian Dominant Topic

Phrases such as *"delicious", "enjoy",* and *"paradise"* further reinforce its image as a tranquil retreat. Additionally, the word cloud underscores key attractions, with terms like *"fish", "shark", "turtle",* and *"beach"* highlighting the island's appeal for marine exploration and activities such as diving and snorkelling. Travel-related terms like *"visit", "stay",* and *"hoteF* emphasise a focus on tourism infrastructure, while financial terms such as *"expensive"* and *"ringgit"* suggest that reviewers often discuss budgeting and travel expenses. Overall, the word cloud reveals that discussions about Perhentian Island are rich in positive sentiment, with a strong emphasis on natural beauty, marine activities, and the travel experience, positioning it as a standout destination among other notable islands like Langkawi and Sabah.

Besides that, the word cloud for Pangkor Island in Figure 4.11 shows a range of terms derived from dominant topics identified through LDA. The term *"pangkor"* serves as the focal point, representing discussions about the island, while other frequently mentioned words, such as *"island', "good\ "malaysia",* and *"time",* contribute to a predominantly positive perception of the Pangkor experience. The frequent occurrence of the word *"good'* indicates that many visitors express positive sentiments about their experiences. Supporting this positive outlook are terms like *"nice", "beautifur, "happy",* and *"stay",* which reflect enjoyment and satisfaction with accommodations.

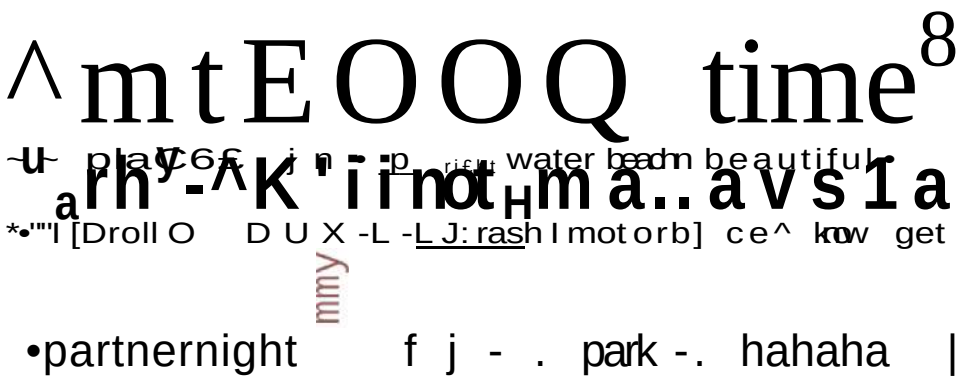


Figure 4.11 Pangkor Dominant Topic

Moreover, terms such as *"partner"*, *"honeymoon"*, and *"couple"* suggest that Pangkor is regarded as a romantic destination appealing to tourists in search of leisure, while terms including *"rental"*, *"motorbike"*, *"car"*, and *"park"* emphasise the importance of transportation logistics for tourists. Additionally, the terms *"room"*, *"place"*, and *"hotel"* indicate ongoing discussions about accommodations available to tourists. Hence, discussions about Pangkor Island are characterised by positive and practical expressions as tourists appreciate the island's natural beauty, tranquil environment, and accessible travel options that contribute to a laid-back atmosphere conducive to enjoyable experiences.

In conclusion, the word cloud visualisations effectively illustrate the key terms found in tourist reviews, providing an accessible overview of recurring themes. These visualisations also enhance the coherence of the topics identified by the LDA model and uncover the defining semantic patterns of each destination. To sum it up, Langkawi Island presents experiences for tourists, while Perhentian Island is primarily associated with natural beauty, marine activities, and travel experience, and Pangkor Island showcases practical tourism concerns. These insights reveal the emotional tone and unique identities of each island, which are crucial for informing tourism development and marketing strategies. The subsequent step is aspect labelling, which categorises each dominant topic into broader themes which ultimately offer a structured understanding of tourist perceptions.

b) Aspect Labelling

Once the dominant topics have been identified, a semantic labelling process is conducted to assign each topic an aspect label manually that captures its broader meaning or thematic focus. In this approach, each significant term within a topic is carefully examined and assigned a specific aspect label, which is then organised under a broader category. For example, terms such as *"happy"*, *"sad*\ and *"enjoy"* are classified under the same category of *"emotion"*. This qualitative interpretation not only bridges the gap between the outputs of LDA with actionable insights but also ensures consistent categorisation of similar terms. Additionally, it facilitates structured comparisons across various islands, enabling more in-depth analysis. By examining the relationship between dominant topics and their associated aspect labels, deeper insights can be gained into the semantic structure and thematic focus of user-generated content.

A bar chart in Figure 4.12 illustrates the average dominant topic weight based on aspect labels for Langkawi Island. By emphasising dominant topics revealed in the reviews, aspect labels offer a contextual analysis of every topic. As a result, the highest average of weight among the dominant topics is Topic_7 with 0.14, which is also the most influential in Langkawi tourism discussions, along with its associated aspects, which are emotion, activity, social interaction, negation, duration, geography, destination and experience. Following that, the comparatively balanced weights of the remaining topics, which range from 0.08 to 0.11 points, indicate that various topics produce significant contributions to the SA.

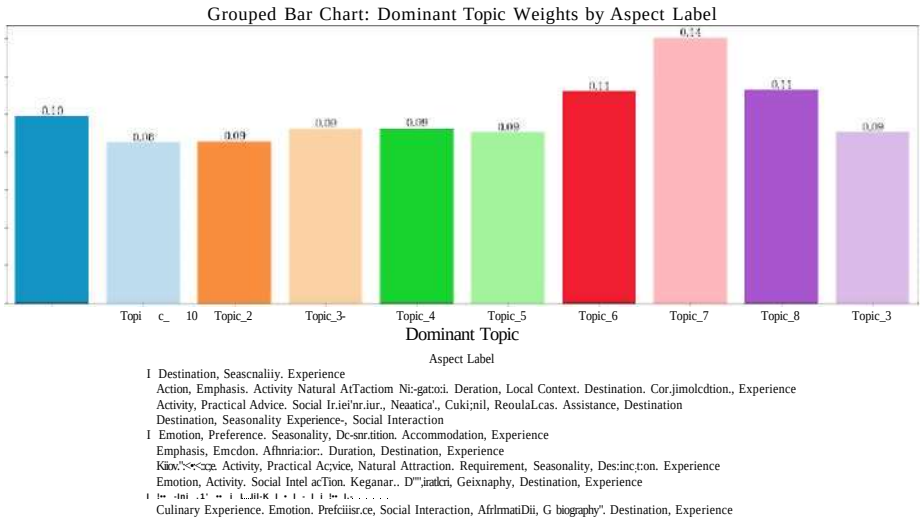


Figure 4.12 Langkawi Aspect Labels

Furthermore, the recurrence of aspect labels among several dominant topics is caused by the probabilities of LDA, which allows terms to be assigned to multiple topics. The overlapping is typical of user-generated material such as YouTube reviews, as discussions frequently focus on several distinct topics within a single review. As an illustration, the aspect label *"Destination, Seasonality, Experience"* can be found in several topics, particularly in Topicl, Topic_3, Topic_4 and Topic_6, implying that discussions regarding travel locations and seasonal experiences happen in various contexts but exhibit necessary similarities. This implies that although certain topics predominate over others, their specific contextual meanings differ based on recurrence terms.

Moreover, the correlation between aspect labels and dominant topics further shows semantic comprehension of these topics. For instance, Topic_7 correlates with aspect labels *"Emotion, Activity, Social Interaction, Negation, Duration, Geography, Destination, Experience"* which refers to the evaluations on topic that focus on opinions, activities and geographical features. In other words, the dominant topic weights emphasise the relevance of distinct topics, and the fact that the same aspect labels exist in several topics reveals that discussions are interrelated, as well as comparable topics appearing in various contexts.

On the other hand, Figure 4.13 represents the dominant topic weights and aspect labels for Perhentian Island. The most significant topics are Topicl and Topic_4 with weights of 0.15 and 0.14, revealing constant discussions on wildlife, activities and opinions on natural resources. Next, the remaining topics have a balanced thematic distribution with weights ranging from 0.07 to 0.11. However, similar aspect labels like *"Culinary Experience, Preference, Social Interaction, Negation, Duration, Destination, Location, Experience"* are found in Topic_6, whereas *"Culinary Experience, Currency, Natural Attraction, Expectation, Location, Experience"* appear in Topic_9. Both of these topics represent discussions about food and experience that occur in two distinct contexts, such as user engagement and cost. Thus, the overlap reflects discussions among users on relevant topics from various perspectives.

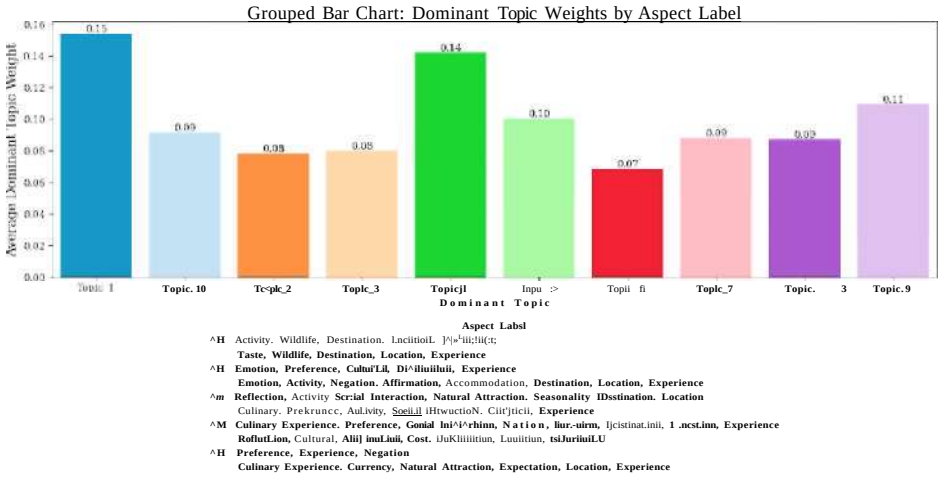


Figure 4.13 Perhentian Aspect Labels

For Pangkor Island, Figure 4.14 illustrates the dominant topic weights and aspect labels as the dominant topics are Topic_5 and Topic_8, which represent frequent discussions with a weight of 0.12. Meanwhile, the remaining topics for weightage results range from 0.08 to 0.11. Additionally, there are important contributions and well-distributed discussion on a variety of aspects in other topics, especially in Topic_l, Topic_2, Topic_7 and Topic_10. At the same time, aspect names are repeated throughout various dominant topics. As an instance, *"Action, Quality, Location"* appears throughout various topics with slightly distinct contextual interpretations within Topic_5, Topic_7, Topic_8 and Topic_10. This occurs as a result of aspect label overlap caused by users discussing opinions regarding accommodation quality within multiple topics.

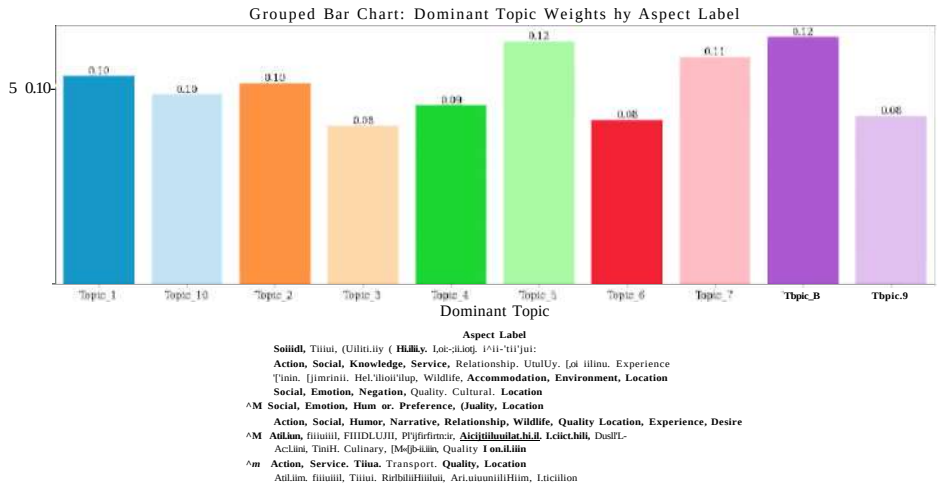


Figure 4.14 Pangkor Aspect Labels

From a semantic context perspective, identifying dominant topics and aspect labels provides deeper insights into tourists' perceptions and expectations. Higher-weighted topics indicate frequent and significant topics, whereas the overlap of aspects becomes apparent by the existence of identical aspect labels across many topics. Analysing the aspect labelling results from Langkawi, Perhentian, and Pangkor Islands reveals distinctive semantic structures. Overall, there are overlaps among the three islands due to the nature of user-generated reviews. Langkawi emphasises diverse travel experiences, Perhentian highlights natural beauty and cost considerations, and Pangkor centres on practicalities and service quality. These distinctions provide valuable insights for targeted tourism development.

4.2.2.2 Document-Topic Distribution

In addition, the document-topic distribution provides a detailed perspective on how each review aligns with the extracted topics. This distribution reflects the proportion of each topic present within individual documents, capturing the intricate nature of user experiences. Instead of assigning a single theme to a review, the document-topic distribution reveals multiple contributing topics, allowing for a nuanced understanding of tourist perspectives. This probabilistic allocation is a key advantage of LDA, aiding in the identification of overlapping themes and diverse viewpoints. As shown in Figure 4.15, distinct patterns emerge across Langkawi, Perhentian, and Pangkor islands. Topicl is the most dominant overall, particularly in Perhentian, which has 243 documents compared to Langkawi's 126 and Pangkor's 68. Topic_4 also concentrates on Perhentian with 205 documents, which indicates a higher density of coherent themes in Perhentian reviews.

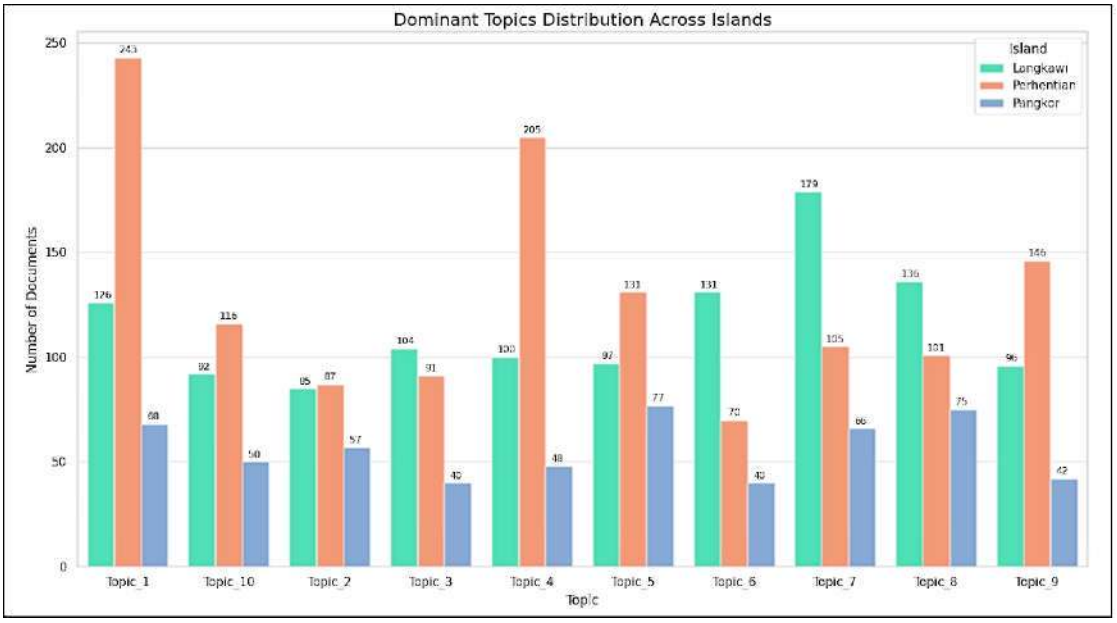


Figure 4.15 Document-Topic Distribution Across Islands

In contrast, Topic_6 and Topic_7 are more evident in Langkawi, with 179 and 136 documents, respectively, while Topic_5 shows a balanced presence across Langkawi and Perhentian, with 131 documents each and some in Pangkor with 77 documents. Additionally, Topic_9 supports Perhentian's dominance with 146 documents, while Langkawi and Pangkor have 96 and 42, respectively. Topic_2, Topic_3, Topic_8, and TopiclO are uniformly distributed and clearly represent general themes that apply across various locations. Overall, the document-topic distribution quantifies each island's thematic focus and provides a foundation for mapping these topics to meaningful aspect labels, enhancing the understanding of tourist experiences and further comparative analyses.

Moreover, another visualisation like t-SNE is used to classify document clusters by their dominant topics within the island dataset. In this visualisation, distinct colours represent different topics ranging from Topicl to TopiclO, with each point corresponding to an individual document. The x-axis (TSNE-1) and y-axis (TSNE-2) illustrate a two-dimensional representation of high-dimensional data, enhancing clearer insights into the relationships among documents based on their dominant topics. These axes indicate the degree of similarity in topic distributions among the documents. Documents with comparable themes are positioned closer together, while those with differing topics are located farther apart, effectively grouping the ten distinct topics.

Thus, this visualisation enhances the understanding of the distribution and proximity of topics within the dataset, highlighting potential areas for further analysis and research.

For instance, Figure 4.16 shows that, for Langkawi, topics for 3, 4, 5, and 9 demonstrate a high concentration of data points in specific areas, indicating significant themes within the documents. In contrast, topics for 1, 2, and 8 exhibit a more dispersed arrangement, suggesting either lower levels of representation or a greater variety of content within those clusters. Additionally, topics for 6, 7, and 10 are situated close to one another, indicating potential similarities in content or overlapping themes. This arrangement provides a clear overview of the relationships among the various topics, facilitating further analysis of the underlying data, including possible thematic overlaps and the identification of unique discussions within each topic area.

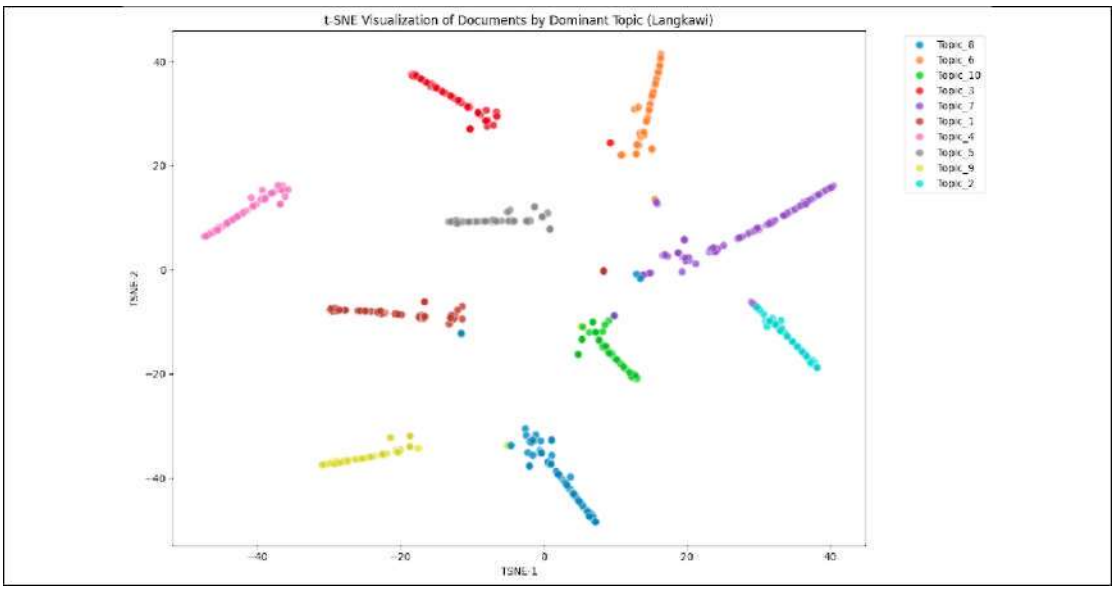


Figure 4.16 Topic Clusters for Langkawi using t-SNE Visualisation

Next, Figure 4.17 presents the distribution of topic clusters for reviews of Perhentian Island. There is a clear separation among the clusters, which emphasises the effectiveness of the topic modelling process. Clusters like Topic_2, Topic_5, Topic_6, and Topic_8 show strong similarity in themes and vocabulary. The balanced distribution across topics reflects a diverse dataset without a single predominant theme. Additionally, clusters such as Topic_1 and Topic_4 are positioned near each other, suggesting possible thematic overlap.

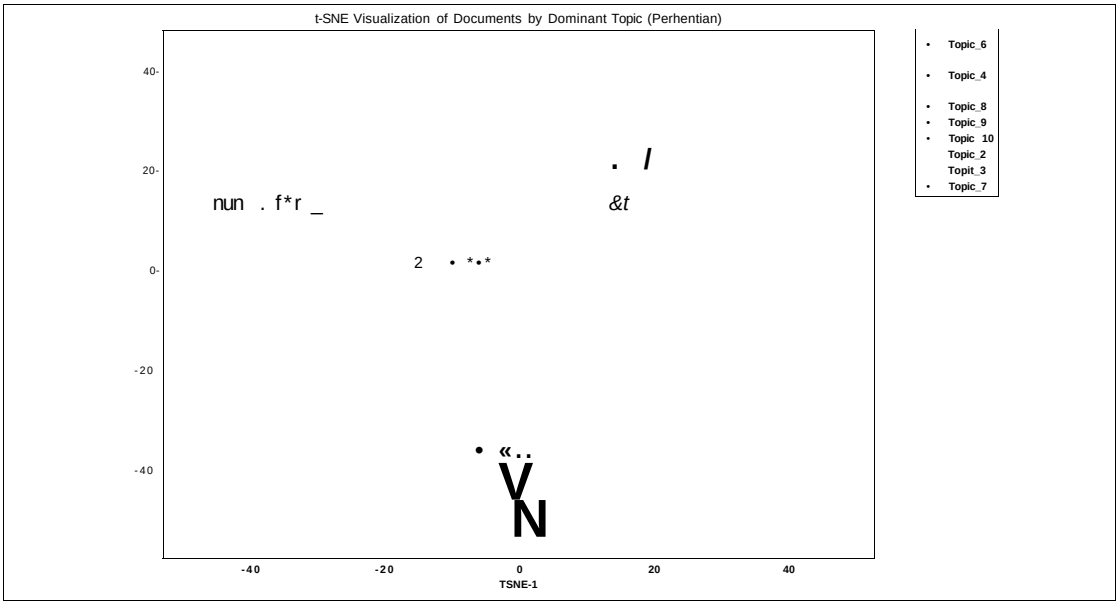


Figure 4.17 Topic Clusters for Perhentian using t-SNE Visualisation

Furthermore, Figure 4.18 shows the distribution of topic clusters for Pangkor Island. The visualisation reveals ten distinct clusters, each tightly grouped with minimal overlap, suggesting that the topic modelling approach effectively assigned meaningful labels based on underlying themes. Clusters like Topicl, Topic_4, Topic_7 and Topic_9 are notably far apart and internally coherent. Compared to Perhentian, Pangkor's clusters are even better separated, indicating a clearer thematic segmentation in the reviews. The absence of overlapping clusters further affirms the quality of the topic modelling.

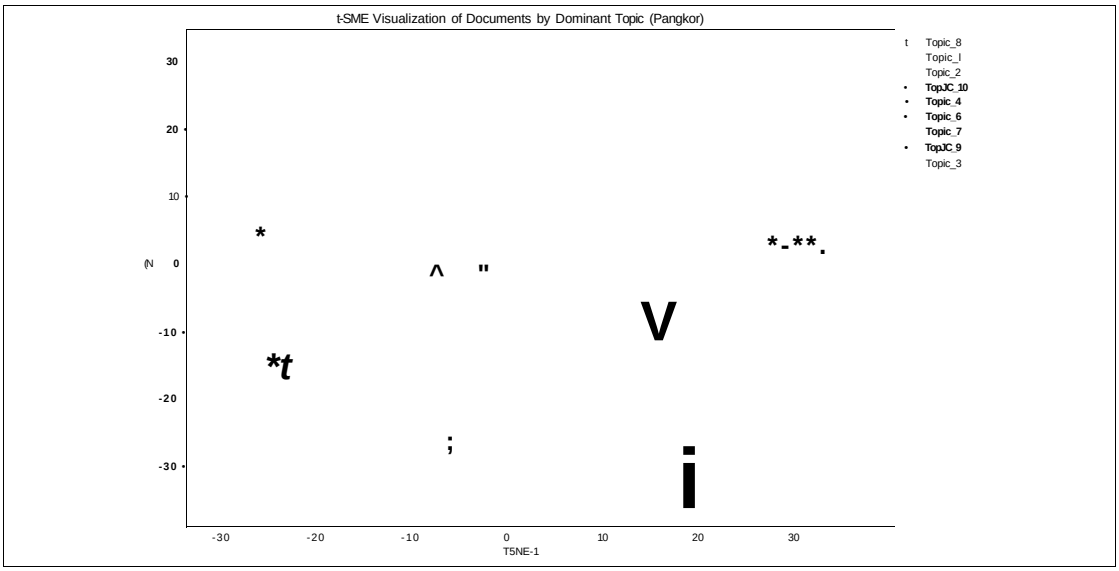


Figure 4.18 Topic Clusters for Pangkor using t-SNE Visualisation

In conclusion, t-SNE visualisation has proven to be an effective tool for qualitatively validating the results of topic modelling derived from user-generated reviews of various island destinations. The distinct and well-defined clusters observed in the island datasets indicate that the topics extracted using LDA are both semantically coherent and distinguishable. This two-dimensional mapping enhances interpretability and underscores the effectiveness of both feature extraction and topic modelling in capturing essential themes in tourist discourse. Furthermore, deeper insights were obtained through aspect labelling utilising a Venn Diagram, which identifies both overlapping and unique aspects across the islands. This technique allows for the comparison of shared themes while also highlighting island-specific concerns or attractions.

4.2.3 Thematic Insights from Descriptive-Semantic Analysis

The integration of DSA techniques through TF-IDF and LDA yields valuable insights into user sentiment. TF-IDF identifies significant keywords based on their frequency of occurrence, emphasising the specific aspects highlighted in user reviews. Conversely, LDA uncovers underlying semantic patterns by clustering related terms into coherent topics, which can then be designated as thematic categories. By analysing the outputs of both techniques, it becomes possible to identify overlapping themes that are frequently mentioned and contextually significant. These intersections illuminate the core concerns and expectations of users. Hence, a Venn diagram illustrates the shared and distinct thematic insights derived from DSA, facilitating a more nuanced understanding of sentiment.

The aspect labelling for island tourism in Malaysia reveals unique and common user experiences across the three islands. Figure 4.19 illustrates a Venn Diagram where each island exhibits unique themes defining sentiment-driven discussions. For instance, Langkawi focuses on practical advice, regulations, geography and communication, while Perhentian focuses on taste, expectation and cost. Following that, the overlap, including social interaction, destination, activity, culinary experience, duration, natural attraction, and seasonality, refers to the key factors between Langkawi and Perhentian in terms of sentiment expression. On the other hand, Pangkor features aspects related to quality, environment, time, transport, and service, which refers to tourists' focus on the overall atmosphere and travel logistics. Subsequently, common aspects across all

three islands include experience, accommodation, preference, emotion, culture and negation, which appear to be common among tourists regardless of the destination. On top of that, negation means the reviews frequently contain contrasts or dissatisfaction in certain aspects, leading to emphasising the need for further analysis into areas requiring improvement.

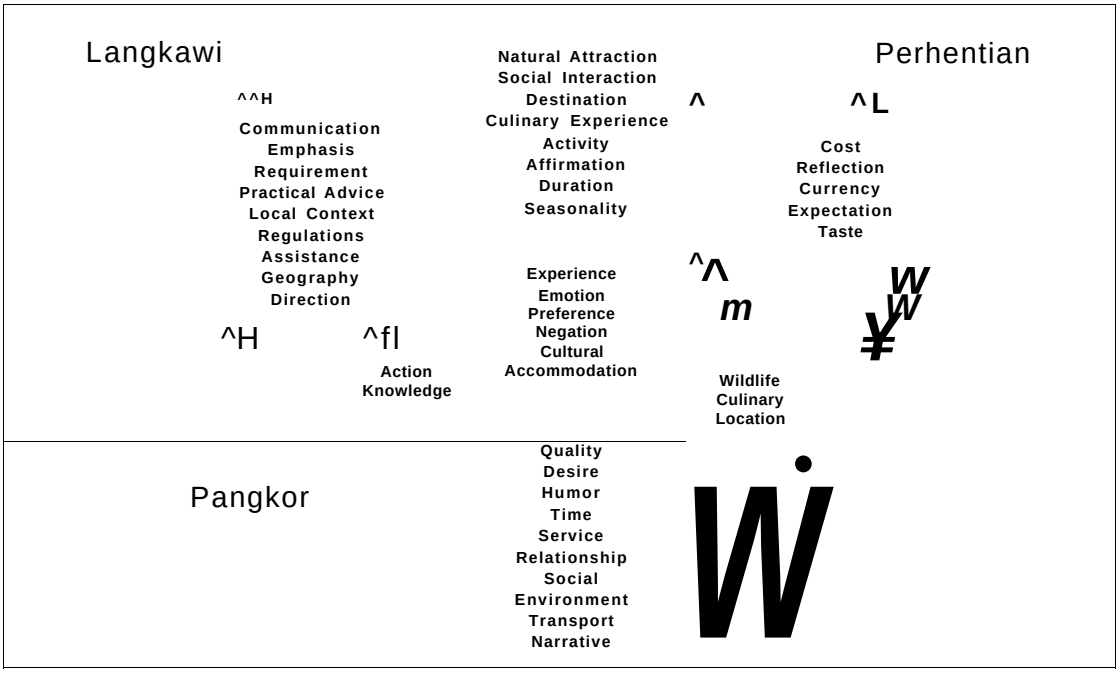


Figure 4.19 Venn Diagram for Aspect Labels of Each Island

Overall, aspect labelling provides deeper insights into tourists' perceptions of each island and the factors influencing overall sentiment. The presence of distinct themes suggests that tourism stakeholders handle marketing strategies and service improvements based on tourists' dominant concerns and interests. For overlapped aspects, the labels indicate fundamental attributes that significantly enhance tourist satisfaction across all three islands. Hence, by combining DSA techniques, this analysis effectively bridges sentiment polarity with contextual understanding and provides a more comprehensive evaluation of Malaysia's island tourism sentiment trends.

4.3 Results of Hybrid Sentiment Analysis Approach

This section describes a hybrid SA approach for classifying sentiments in YouTube reviews of Malaysia's islands. By integrating lexicon-based techniques with ML, this research enhances accuracy and contextual relevance. Initially, VADER, a

lexicon-based tool, reveals that all three islands exhibit predominantly positive sentiments, with Langkawi having the highest at nearly 60%>. Neutral sentiments are moderate, while negative sentiments are minimal, reflecting generally positive tourist experiences. This research extends VADER analysis by integrating SVM together with techniques like TF-IDF and LDA, resulting in high accuracy rates of 98.15%) for Langkawi, 98.5%> for Perhentian, and 97.19%) for Pangkor. Overall, this hybrid SA approach yields robust predictions for tourism reviews, demonstrating the effectiveness of integrating lexicon-based with ML techniques tailored to specific contexts.

4.3.1 Sentiment Prediction using Valence Aware Dictionary and Sentiment Reasoner

This section examines the effectiveness of VADER sentiment prediction in identifying sentiment polarity. VADER is a lexicon-based tool designed to evaluate the emotional tone of text by utilising a comprehensive vocabulary of words and phrases. It operates by applying specific rules and assigning sentiment scores, which together contribute to calculating an overall sentiment score. Additionally, VADER is particularly effective in handling intensifiers and negations, which enhances the accuracy of sentiment measurement. Besides that, VADER is tailored for English text; hence, any reviews in other languages must be translated into English to ensure accurate polarity detection.

The VADER lexicon considers input sentence characteristics and returns sentiment metrics with word-level analyses. Table 4.3 shows the sentiment polarity distribution in percentage for YouTube reviews of Langkawi Island, revealing a predominantly positive sentiment distribution across all topics, as the percentages consistently exceed 50%, with high satisfaction levels in topics like Topic_5 and Topic_9 for 70.10%) and 67.71%). Neutral sentiment peaks at 36.64%) for Topic_6 of mixed opinions, while negative sentiment remains low across most topics, with Topic_2 for 20%) showing the highest dissatisfaction level. Overall, high positive sentiment may be influenced by Langkawi's reputation as a well-developed tourist destination with beautiful landscapes, convenient facilities, and established hospitality services. Meanwhile, negative feedback could be linked to concerns about crowd management or environmental issues, and moderate neutral sentiment may arise from reviews that focus on facts rather than personal opinions.

Table 4.3

Sentiment Polarity Results using Valence Aware Dictionary and Sentiment Reasoner for Langkawi Island

Type of Island	Dominant Topic	Sentiment Polarity (%)		
		Negative	Neutral	Positive
Langkawi	Topic_l	10.317460	30.158730	59.523810
	Topic_2	20.000000	29.411765	50.588235
	Topic_3	11.538462	30.769231	57.692308
	Topic_4	10.000000	25.000000	65.000000
	Topic_5	11.340206	18.556701	70.103093
	Topic_6	6.870229	36.641221	56.488550
	Topic_7	11.731844	32.960894	55.307263
	Topic_8	10.294118	33.088235	56.617647
	Topic_9	8.333333	23.958333	67.708333
	Topic_10	10.317460	30.158730	59.523810

Furthermore, Table 4.4 shows that Perhentian Island has a predominantly positive feedback distribution, with Topic_8 showing the highest positive sentiment at 60.40%. Although the overall sentiment distribution is slightly less favourable than Langkawi's, other topics maintain positive sentiment between 55-60%). Topic_6 shows the highest proportion of neutral sentiment at 41.43%, while negative sentiment reaches 20%). Generally, the island's serene beaches and diving experiences likely contribute to positive sentiment. In contrast, higher neutral sentiment may reflect reviews about logistics. Also, the peak in negative sentiment for Topic_6 may be due to challenges like accessibility issues, high costs or seasonal overcrowding.

Table 4.4

Sentiment Polarity Results using Valence Aware Dictionary and Sentiment Reasoner for Perhentian Island

Type of Island	Dominant Topic	Sentiment Polarity (%)		
		Negative	Neutral	Positive
Perhentian	Topic_l	5.761317	35.802469	58.436214
	Topic_2	10.344828	34.482759	55.172414
	Topic_3	9.890110	32.967033	57.142857
	Topic_4	14.146341	26.829268	59.024390

Type of Island	Dominant Topic	Sentiment Polarity (%)		
		Negative	Neutral	Positive
	Topic_5	13.740458	30.534351	55.725191
	Topic_6	20.000000	41.428571	38.571429
	Topic_7	14.285714	33.333333	52.380952
	Topic_8	13.861386	25.742574	60.396040
	Topic_9	10.273973	32.191781	57.534247
	Topic_10	5.172414	34.482759	60.344828

Next, Table 4.5 reveals that Pangkor Island displays a balanced distribution of positive and neutral sentiments, with a peak of 64% for TopiclO, while neutral sentiment is high in Topic_7 and Topic_8 with 39.39% and 37.33%, respectively. Next, negative sentiment is the lowest among all three islands, as most values are under 14%. Therefore, high neutral sentiment may be due to Pangkor's peaceful nature as a tourist destination, while positive sentiment may be influenced by its tranquillity, scenic beauty, and cultural experiences. Minimal negative sentiment suggests fewer operational challenges or lower visitor expectations compared to more commercialised destinations.

Table 4.5

Sentiment Polarity Results using Valence Aware Dictionary and Sentiment Reasoner for Pangkor Island

Type of Island	Dominant Topic	Sentiment Polarity (%)		
		Negative	Neutral	Positive
Pangkor	Topic_l	5.882353	45.588235	48.529412
	Topic_2	5.263158	47.368421	47.368421
	Topic_3	10.000000	32.500000	57.500000
	Topic_4	14.583333	31.250000	54.166667
	Topic_5	14.285714	28.571429	57.142857
	Topic_6	12.500000	27.500000	60.000000
	Topic_7	10.606061	39.393939	50.000000
	Topic_8	13.333333	37.333333	49.333333
	Topic_9	11.904762	33.333333	54.761905
	Topic_10	2.000000	34.000000	64.000000

On the other hand, by aggregating results across all topics, VADER highlights the overall emotional tone present in the dataset and detects dominant sentiment trends while considering topic-specific nuances, offering a clearer understanding of the sentiment landscape. Hence, Figure 4.20 illustrates the performance of the VADER lexicon across the three Malaysia islands. The polarity results indicate that Langkawi leads with the highest percentage of positive sentiment at 59.86%, followed by Perhentian at 55.47% and Pangkor at 54.28%). Additionally, Pangkor exhibits a slightly higher neutral sentiment percentage of 34.66%) compared to Langkawi and Perhentian. This elevated neutral sentiment can be attributed to factors such as a tendency for positive bias in destination reviews, as the model likely categorises these reviews as neutral when sentiments are more subtle. Lastly, negative sentiment scores are notably low for all three islands in comparison to the other sentiment polarity results. This minimal prevalence of negative sentiment may be driven by the allure of these destinations, as tourists tend to focus more on enjoyable experiences.

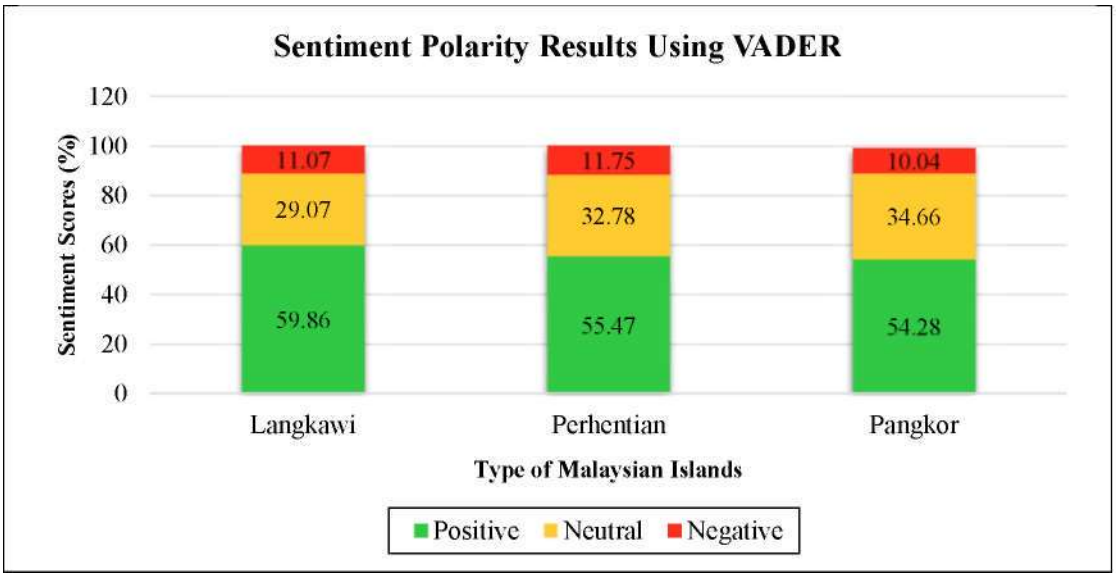


Figure 4.20 Performance of Valence Aware Dictionary and Sentiment Reasoner for Sentiment Polarity

To sum it up, the findings reveal a generally positive sentiment toward all three islands, indicating that tourists have a favourable response to the reviews. The minor differences observed between neutral and negative sentiments suggest only slight variations in tourists' experiences or perspectives. In summary, the VADER lexicon is a valuable tool for predicting sentiment polarity, which is useful in analysing YouTube

reviews as it helps to identify both strengths and areas for improvement. Additionally, comparative SA can enhance tourism management strategies and improve tourist experiences across Malaysia's islands. Furthermore, integrating the VADER lexicon with ML models for sentiment classification can enhance prediction accuracy, providing more detailed and scalable insights for strategic planning.

4.3.2 Sentiment Classification using Support Vector Machine

The significance of SVM in SA lies in its strong classification capabilities and its effective management of data distribution and learning behaviour. In this research, the dataset is systematically divided into training, validation, and testing data using a 70:20:10 ratio. This technique ensures balanced evaluation and helps mitigate overfitting. By structuring the data in this form, the model can learn effectively from the training data, adjust hyperparameters with the validation data, and assess generalisation on the testing data. Additionally, the learning curve of SVM can evaluate the model's performance across different data sizes, providing insights into its generalisation as the amount of training data increases. Therefore, to evaluate the learning behaviour and generalisation capacity of the SVM model, two analyses were conducted to provide both numerical insights and an intuitive understanding of the model's performance trends.

Table 4.6 shows the training and validation accuracy across various training sizes to provide more detailed insights into the model's accuracy at each training increment, allowing for a clearer comparison of performance progression. A 9-fold stratified cross-validation technique was utilised to ensure balanced class distribution, with training and validation accuracy evaluated at different training sizes and results averaged across the folds. This analysis was conducted for each island dataset, such as Langkawi, Perhentian, and Pangkor, to examine the model's consistency across diverse linguistic patterns. The findings offer insights into the model's scalability and serve as a foundation for discussions regarding performance comparisons among the island-specific sentiment datasets.

Table 4.6

Training and Validation Accuracy by Training Size

Type of Island	Training Size	Mean Accuracy (%)	
		Training	Validation
Langkawi	48	100 ± 0.00	97.79 ± 1.92
	102	100 ± 0.00	98.16 ± 1.82
	156	98.79 ± 0.64	98.53 ± 1.83
	210	97.30 ± 0.81	97.97 ± 2.05
	265	97.57 ± 0.54	97.97 ± 2.05
	319	98.12 ± 0.55	97.79 ± 2.08
	373	97.94 ± 0.54	98.34 ± 2.08
	427	98.13 ± 0.33	98.34 ± 2.08
	482	98.32 ± 0.32	98.15 ± 2.00
Perhentian	53	100 ± 0.00	99.17 ± 0.75
	113	100 ± 0.00	99.83 ± 0.47
	172	99.74 ± 0.29	99.00 ± 1.23
	232	99.38 ± 0.21	98.83 ± 0.94
	292	99.16 ± 0.17	98.66 ± 1.32
	352	98.64 ± 0.18	98.16 ± 0.95
	412	98.87 ± 0.16	98.16 ± 0.95
	472	98.87 ± 0.17	98.33 ± 0.85
	532	98.96 ± 0.37	98.50 ± 1.22
Pangkor	21	96.30 ± 1.98	96.77 ± 3.17
	46	98.31 ± 0.90	97.57 ± 2.45
	71	97.65 ± 0.66	96.37 ± 2.95
	95	97.31 ± 0.52	96.77 ± 2.68
	120	97.78 ± 0.56	97.18 ± 2.86
	145	97.09 ± 0.78	97.19 ± 2.26
	169	97.30 ± 0.49	96.78 ± 2.03
	194	97.65 ± 0.69	96.78 ± 2.03
	219	97.56 ± 0.57	97.19 ± 2.26

In Langkawi, the training accuracy starts at 100% for the smallest training sizes, suggesting early signs of overfitting as the model easily memorises the limited dataset. As the training size increases, the training accuracy gradually declines to approximately 97.3% to 98.3%, while validation accuracy remains consistently high, ranging from 97.7% to 98.5%. This trend indicates effective generalisation and a well-balanced

model, with no evident signs of significant overfitting or underfitting. Notably, validation accuracy peaks at training sizes between 373 and 427 data, reaching a maximum of 98.34%. This suggests that any additional data beyond this point may result in diminishing returns in terms of performance improvements.

In contrast, the Perhentian dataset demonstrates the highest overall validation performance. The training accuracy is perfect at the early stages, and even as the training size increases up to 532 samples, it remains impressively high, with more than 98.6% and minimal variance. Validation accuracy consistently exceeds 98.0%>, reaching a peak of 99.83%> with just 113 training samples. This indicates that the Perhentian dataset is highly learnable and consistent, potentially due to greater sentiment clarity or reduced linguistic variability. The small gap between training and validation accuracy across all folds signifies excellent generalisation with minimal overfitting.

Meanwhile, Pangkor exhibits slightly more variability in its results. The training accuracy begins at a lower level with 96.3%> at 21 samples and stabilises between 97.0% and 97.8%) as the training size increases. Validation accuracy mirrors this pattern, ranging from 96.3%> to 97.8%. The comparatively lower training and validation scores relative to the other two islands suggest that the Pangkor dataset may encompass a wider range of expressions, contain more ambiguity, or present noisier sentiment patterns. Nevertheless, the model demonstrates strong generalisation capabilities, as evidenced by the close alignment between training and validation performance, which indicates effective learning from a limited and potentially noisier dataset.

On the other hand, the SVM learning curves chart for each island was created based on the results of training and validation accuracy across different training sizes. The shaded areas around the lines represent standard deviation, illustrating the model's reliability with different training data sizes. Figure 4.21 shows a learning curve for Langkawi Island utilising the SVM model that integrates VADER with TF-IDF and LDA, providing valuable insights into the model's generalisation capabilities and training dynamics. Initially, the training accuracy is exceptionally high; however, when the training data is small, it can be prone to overfitting. Besides, as the training data size increases, a slight decline in training accuracy is observed, accompanied by a steady rise in validation accuracy. This indicates that the model starts to generalise more effectively with larger datasets. Both accuracy curves converge above 400 samples, ultimately stabilising up to 98%>, suggesting that the model achieves a well-balanced

state with minimal overfitting at larger dataset sizes. The relatively narrow gap between training and validation accuracy throughout most of the dataset, along with low variance, demonstrates consistent performance. This trend reinforces the robustness of the SVM model in effectively learning sentiment patterns, particularly when bolstered by rich textual features and a well-structured training process.

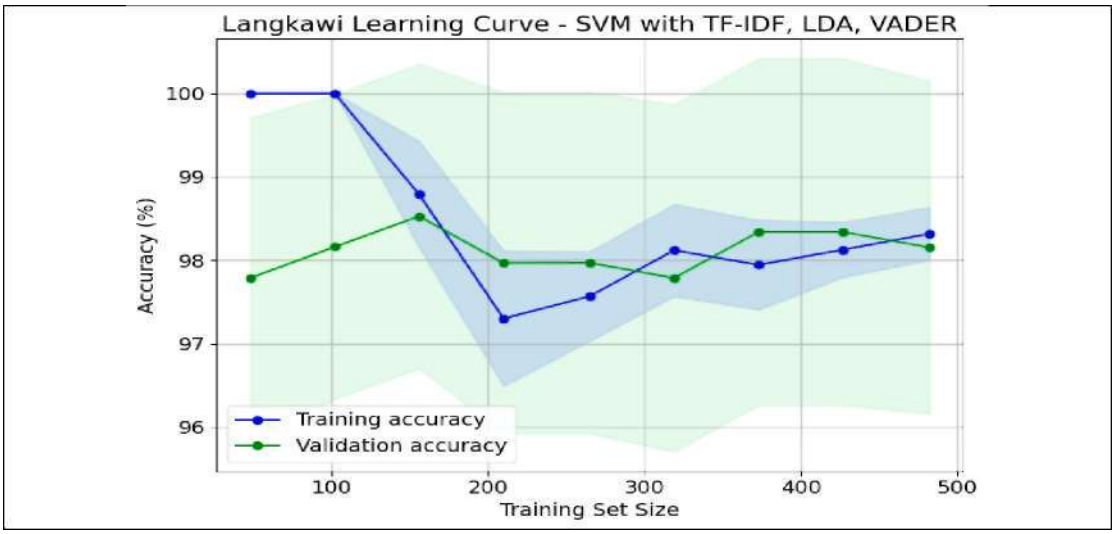


Figure 4.21 Support Vector Machine Learning Curve for Langkawi

Following that, Figure 4.22 illustrates the SVM learning curve for the Perhentian dataset and indicates that training accuracy remains consistently high, approaching 100% with smaller sets, before slightly declining to approximately 99.7% for larger training sets. In contrast, validation accuracy shows more variability, starting around 99% for smaller sets and decreasing to about 98% as the training data size reaches more than 300 samples. This trend suggests a potential overfitting issue, where the model performs exceptionally well on the training data but struggles with generalisation to unseen data. The shaded areas around the accuracy lines reflect confidence intervals, highlighting the need for improvement in generalisation while maintaining strong training performance. These findings emphasise the trade-off between achieving high training accuracy and ensuring effective generalisation.

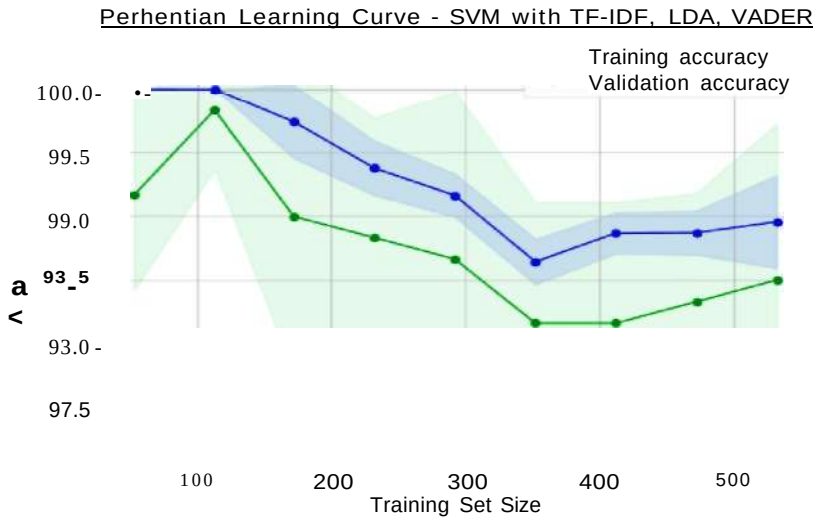


Figure 4.22 Support Vector Machine Learning Curve for Perhentian

Moreover, Figure 4.23 illustrates the SVM learning curve on the Pangkor dataset, which presents two distinct lines representing both training and validation accuracy against varying sizes of the training data. The training accuracy demonstrates a generally high level of performance, consistently hovering around 98%, while displaying some fluctuations with increasing training data. Meanwhile, the validation accuracy shows solid performance, stabilising around 97%. Despite a slight gap between the training and validation accuracy, the results suggest that the SVM model effectively generalises the knowledge acquired from the training data. Overall, the consistent performance in both accuracy metrics highlights the model's robustness in handling the Pangkor data, thus indicating its potential for practical applications.

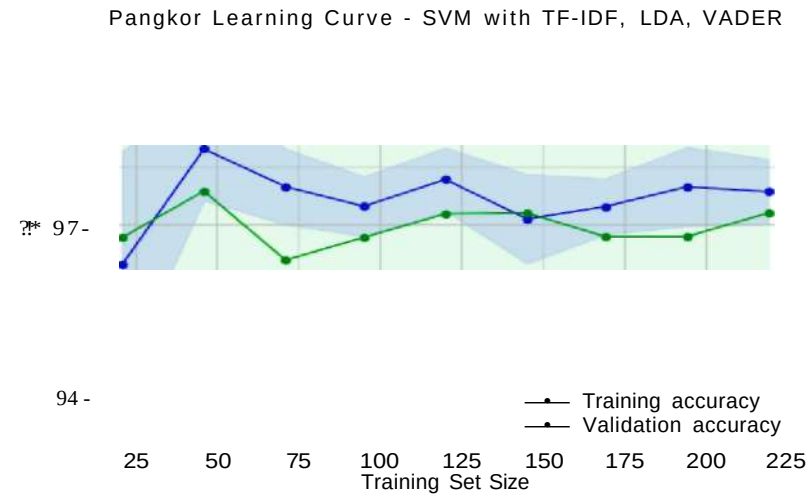


Figure 4.23 Support Vector Machine Learning Curve for Pangkor

In summary, the analysis of the SVM learning curve is essential in assessing the model's generalisation capability and identifying potential issues such as overfitting or underfitting across different training data sizes. By evaluating the trends in both training and validation accuracies, the learning curve provides valuable insights into how effectively the SVM model learns from the available data and how its predictive performance evolves with the incorporation of additional data. Hence, this diagnostic process is essential for understanding whether the model benefits from an increase in data volume and whether it maintains consistent performance across various subsets.

Overall, the analysis demonstrates that the SVM model performs reliably across all three island datasets, with particularly notable results for the Perhentian and Langkawi datasets. The narrow and stable gap between training and validation accuracies for these datasets indicates effective learning and a minimal risk of overfitting. Although the Pangkor dataset presents slightly lower scores, the SVM model still achieves commendable accuracy, highlighting its robustness in managing diverse sentiment structures and review contexts. Notably, the Perhentian dataset stands out as the most conducive to learning, generating generalisable patterns that enhance the model's performance. Building on these findings, the next step entails a more thorough evaluation and validation of the SVM model using comprehensive performance metrics and AUC metrics to further affirm its classification effectiveness and reliability across all datasets.

4.4 Results of Model Evaluation and Validation

In order to ensure the reliability and generalisability of the sentiment classification model, a comprehensive evaluation and validation process was carried out. Stratified *k-fo\d* cross-validation was utilised to preserve the original class distribution, resulting in balanced training and testing datasets. Key performance metrics, such as accuracy, precision, recall, and Fl-score, were calculated for each fold and averaged to provide a robust estimate of performance. Furthermore, a confusion matrix was used to visualise the distribution of correctly and incorrectly classified instances, thereby offering valuable insights into the model's behaviour. The AUC metric was calculated to evaluate the model's capacity to differentiate between sentiment classes across various thresholds. This thorough evaluation approach

effectively mitigates the risk of overfitting and provides a clearer understanding of the model's performance.

4.4.1 Evaluate Performance Metric Results

Performance metrics and the confusion matrix are used to evaluate the performance of DSA techniques integrated with a hybrid SA approach. Moreover, to ensure a balanced and equitable assessment, stratified *k-fo\d* cross-validation is applied. This approach preserves the original class distribution across each fold during both the training and testing phases and enhances the reliability of performance evaluations by reducing variability and mitigating the risk of biased splits. Additionally, a detailed analysis of the model's classification results can be conducted through the confusion matrix, which measures true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The model's effectiveness can be demonstrated by analysing its ability to distinguish between different sentiment classes, identifying potential misclassifications, and highlighting areas for improvement. Consequently, this research delves deeper into the model's capacity to reliably identify user sentiment regarding the Langkawi, Perhentian, and Pangkor islands.

4.4.1.1 Evaluate using Stratified k-fold Cross-Validation

Stratified &-fold cross-validation is a technique for evaluating model performance by ensuring that each fold maintains the same class distribution as the original dataset. This is crucial for maintaining class proportions, ensuring fair evaluation and minimising bias from uneven representation. Moreover, this method allows for training and testing on various data subsets, minimising the risk of overfitting or underfitting. Overall, it strengthens the reliability of model evaluations, especially in scenarios with skewed class distributions.

The accuracy results obtained from stratified k-fold cross-validation for Malaysia's islands in Table 4.7 underscore the stability and effectiveness of the hybrid SA approach. The findings reveal consistently high classification accuracy across all folds, with minimal variance, which reinforces the robustness of the VADER, SVM, TF-IDF, and LDA configurations. For Langkawi, the model demonstrates consistently high accuracy, ranging from *91.19%* at *k=2* to 98.16%) between *k=4* and £=8, with a

minor decline to 98.15% at *k=9.* Although the standard deviation shows a slight increase with higher folds (from ±0.74 to ±2.00), this variation remains within acceptable limits, suggesting the model's stability. In addition, the confusion matrix for *k=9* reveals no false positives, with only nine false negatives, indicating a classification pattern that is both precise and conservative.

Table 4.7

Accuracy Results using Stratified &-fold Cross-Validation for Malaysia's Islands

Type of Island	Mean Accuracy Results for k-fold (%)								Confusion Matrix (k=9)
	2	3	4	5	6	7	8	9	
Langkawi	97.79 ±0.74	98.34 ±0.90	98.16 ±1.10	98.16 ±1.02	98.16 ±1.38	98.16 ±1.36	98.16 ±1.76	98.15 ±2.00	[[0 9] [0 51]]
Perhentian	98.83 ±0.17	98.50 ±0.41	98.83 ±0.87	99.00 ±0.97	98.83 ±0.90	98.83 ±1.25	98.83 ±1.04	98.50 ±1.22	[[11 0] [0 55]]
Pangkor	95.55 ±1.20	97.17 ± 1.49	96.36 ±2.09	96.78 ±2.04	96.78 ±2.65	96.77 ±3.16	96.77 ±3.23	97.19 ±2.26	[[4 0] [0 23]]

On the other hand, Perhentian demonstrates the highest performance among the three islands, achieving a peak accuracy of 99.00% at *k=5* and consistently maintaining values above 98.5% across all folds. Its low standard deviation, varying from ±0.17 to ±1.22, indicates a strong capacity for generalisation, while the confusion matrix at *k=9* shows only one false positive and no false negatives, indicating excellent sensitivity and specificity. In contrast, Pangkor begins with a lower accuracy of 95.55% at *k=2* but steadily improves to 97.19% at *k=9.* Although the standard deviation is higher (reaching ±3.23), this suggests that increasing the number of folds improves the model's reliability. Furthermore, the confusion matrix for Pangkor shows no false negatives. However, it indicates four false positives, suggesting a slight tendency to over-predict positive sentiment while rarely missing actual positive reviews.

On top of that, the choice to implement a 9-fold stratified cross-validation setup is both practical and supported by empirical evidence. This technique ensures that each fold reflects the same class distribution as the entire dataset, which is critical for SA, particularly in cases of class imbalance. Using 9 folds in cross-validation provides more training data in each fold compared to lower *k* values, while also reducing the high variance that can come with larger *k* values. Additionally, the consistent performance

observed across folds 6 to 8 indicates that 9-fold cross-validation strikes an optimal balance for model generalisation and reliability.

Besides, the performance of the stratified *k*-fold cross-validation (*k*=9) is visualised for Langkawi Island, as depicted in Figure 4.24, which illustrates the accuracy of each fold throughout the validation process. The x-axis shows folds numbered 1 to 9, and the y-axis indicates accuracy for each fold. Across the analysed 9 folds, the accuracy consistently approached the 1.0 mark, reflecting strong model performance. The overall mean accuracy indicated by the green dashed line for the *k*-fold process is approximately 98.15%. The accompanying graph also showcases the standard deviation, represented by the ±1 standard deviation range, which demonstrates the variation in accuracy across the different folds. This consistency in results indicates that the model exhibited robust performance, thereby affirming its effectiveness in predicting outcomes for Langkawi Island.

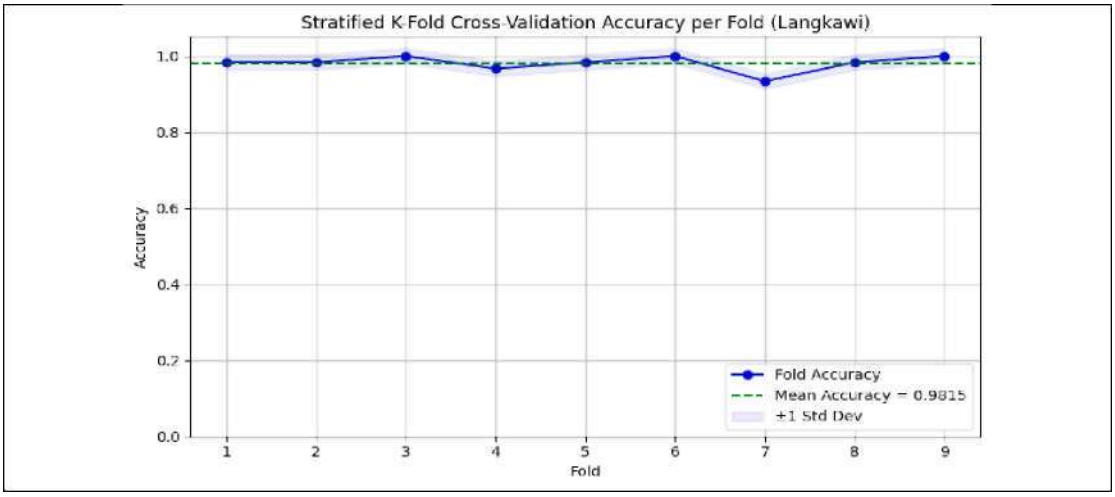


Figure 4.24 Stratified *k*-fold Performance for Langkawi Island

On the other hand, the performance of each fold for Perhentian Island is shown in Figure 4.25. This graph illustrates the accuracy achieved across the 9 folds, where each fold consistently demonstrates high accuracy, with a mean accuracy of 98.50%. The variations in accuracy are minimal, evidenced by the closely clustered points for each fold. The standard deviation, represented by the shaded area around the mean accuracy line, confirms a tight distribution, indicating stability and reliability in the performance across the folds. To sum it up, these results suggest a robust predictive

model resulting from the stratified cross-validation approach applied to the Perhentian dataset.

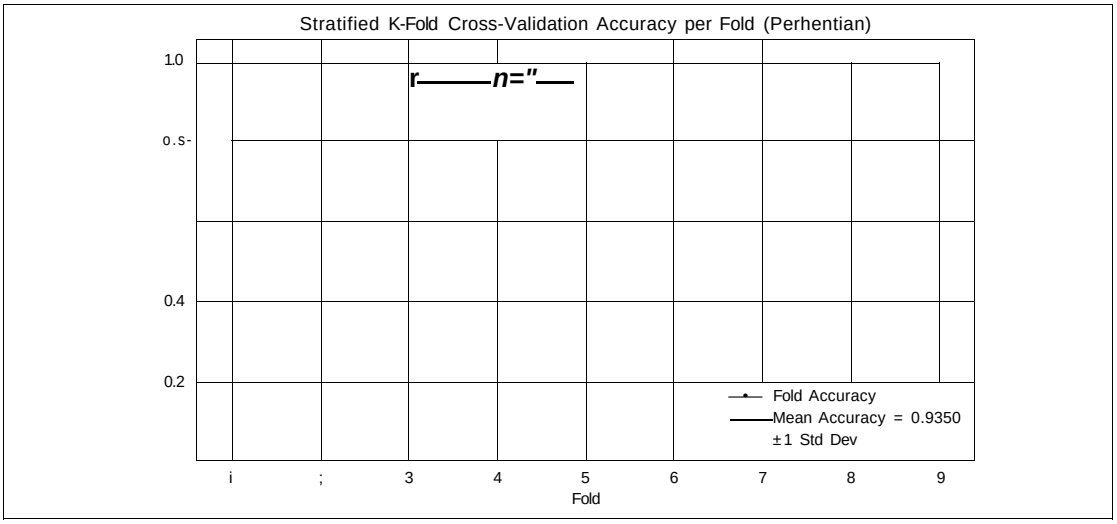


Figure 4.25 Stratified *k*-fold Performance for Perhentian Island

Lastly, Figure 4.26 shows the performance of each fold for Pangkor Island. The accuracy for each fold demonstrates a consistently high level of accuracy throughout all iterations, while the mean accuracy is approximately 97.19%, reflecting a strong overall performance. The shaded area surrounding this line represents ±1 standard deviation, indicating that the accuracy values are not only high but also stable, with minimal variation between the folds. These results confirm the reliability of the model in effectively classifying data for Pangkor Island.

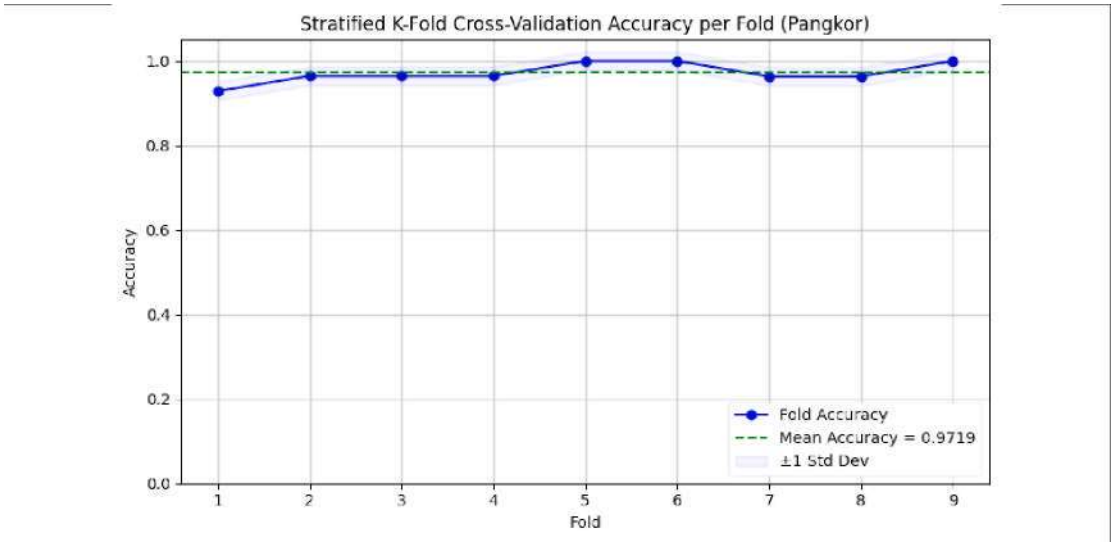


Figure 4.26 Stratified *k*-fold Performance for Pangkor Island

In conclusion, the implementation of stratified &-fold cross-validation provided a reliable and balanced assessment of the hybrid SA approach across all three island datasets. By preserving the class distribution within each fold, the model's performance was evaluated consistently and fairly. The results revealed impressive classification accuracy, with Langkawi achieving 98.15%, Perhentian 98.50%, and Pangkor 97'.19%. These findings suggest that the model is highly effective in capturing and classifying sentiment patterns in user-generated reviews across diverse tourism contexts, further affirming the robustness and generalisability of the model approach.

4.4.1.2 Evaluate using Confusion Matrix

The confusion matrix is another vital tool for evaluating the performance of a sentiment classification model. It provides a comprehensive breakdown of prediction outcomes, enabling the calculation of key performance metrics such as accuracy, precision, recall, and Fl-score. By contrasting the model's predicted sentiment labels with the actual labels, the confusion matrix delineates TP, TN, FP, and FN. This visualisation of the model's classification behaviour not only highlights patterns of misclassification but also evaluates its capability to accurately identify each sentiment class. Consequently, it offers insightful perspectives on the model's overall effectiveness and reliability.

Figure 4.27 illustrates the confusion matrix for the Langkawi testing data demonstrates the overall performance of the classification model. In this analysis, there were a total of 78 instances (13 negative and 64 positive), with only one FP. The model achieved an impressive accuracy of 98.72%, indicating that it effectively classified most predictions correctly. Furthermore, the precision score of 96.43%) signifies that a high percentage of positive predictions were indeed correct. The recall metric, at 99.23%), highlights the model's ability to correctly identify positive samples. Lastly, the Fl-score of *91.16%* reflects a strong balance between precision and recall, confirming the model's robustness in performance.

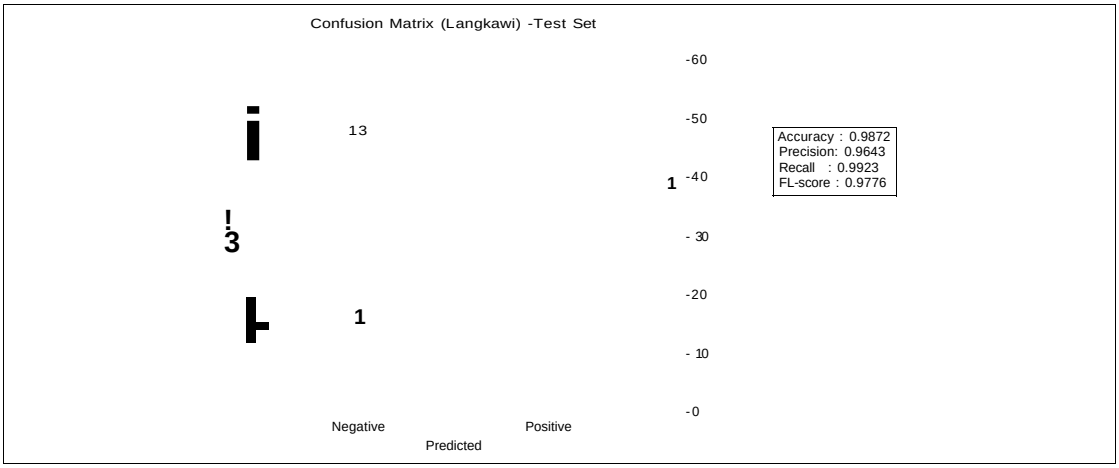


Figure 4.27 Confusion Matrix for Langkawi

In contrast, Figure 4.28 presents the confusion matrix for the Perhentian testing data. It shows that out of 86 total predictions, the model correctly identified 14 negative samples and 71 positive samples, leading to a total of 85 correct predictions. The matrix indicates 1 FN, where a positive sample was misclassified as negative. The model then achieves an overall accuracy of 98.84%, reflecting its high reliability in classification tasks. It has a precision of 96.67%, meaning that when it predicts a positive outcome, it is correct about 96.67% of the time. The recall score stands at 99.31%, indicating that it successfully identifies 99.31% of the actual positive cases. Finally, the Fl-score, which balances precision and recall, is calculated at 97.93%, showcasing the model's effectiveness in achieving consistent results between both metrics. Overall, these figures demonstrate the robustness of the model in classifying the testing dataset accurately.

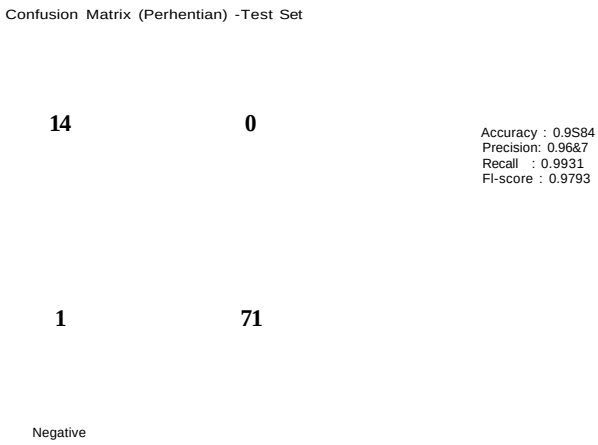


Figure 4.28 Confusion Matrix for Perhentian

Lastly, the confusion matrix analysis for Pangkor Island is illustrated in Figure 4.29. The matrix reveals that out of 36 total predictions, 6 instances were correctly classified as negative, while one instance was misclassified as positive when it should have been negative. Conversely, the model accurately identified 29 positive cases, with no FP. From the metrics provided, the model achieved an overall accuracy of 97.22%, demonstrating high effectiveness across the testing data. Further, the precision was calculated at 92.86%, indicating the proportion of true positive identifications against all positive predictions made. The model's recall stood at 98.33%, reflecting its ability to capture the majority of actual positive cases. Finally, the Fl-score, which balances precision and recall, was reported at 95.31%, showcasing a strong overall performance in binary classification for the Pangkor data.

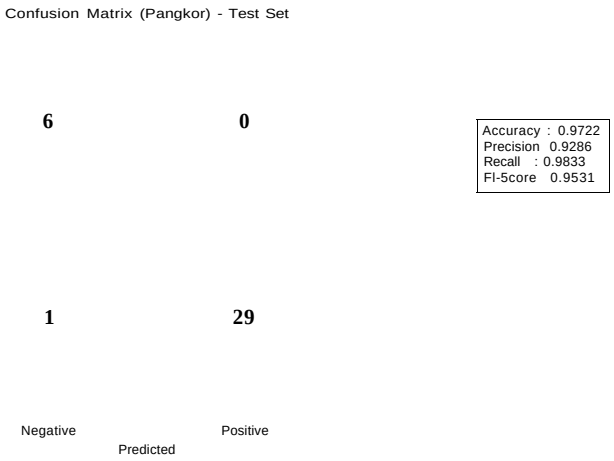


Figure 4.29 Confusion Matrix for Pangkor

On the other hand, Table 4.8 shows the details of performance metrics for each sentiment class on three Malaysia islands using the test dataset. The model achieved an accuracy of 98.72% in Langkawi, demonstrating high reliability in sentiment predictions. It excelled in identifying positive sentiments, with perfect precision, a recall of 98%, and an Fl-score of 99% over 65 test samples. For negative sentiments, performance was slightly lower, with a precision of 93%, recall of 100%), and Fl-score of 96%) based on 13 samples, indicating some misclassification of positive reviews as negative.

Table 4.8

Performance Metric Results for Malaysia's Islands

Type of Island	Sentiment Class	Accuracy (%)	Performance Metric (Test Set)			
			Precision (%)	Recall (%)	Fl-Score (%)	Support
Langkawi	Positive	98.72	100	98	99	65
	Negative		93	100	96	13
Perhentian	Positive	98.84	100	99	99	72
	Negative		93	100	97	14
Pangkor	Positive	97.22	100	97	98	30
	Negative		86	100	92	6

Meanwhile, the model for Perhentian Island reached the highest accuracy at 98.84%o, showcasing superior generalisation capabilities. Positive sentiment detection was flawless, with a perfect score for precision, recall, and Fl-score from 72 samples. Negative sentiment performance also remained strong, achieving a precision of 93%, a recall of 100%>, and an Fl-score of 97% from 14 samples, with some false positives likely due to class imbalance. For Pangkor Island, the model's accuracy of 97.22%> indicates solid performance. It maintained a precision of 100%>, a recall of 97%, and a Fl-score of 98% for positive sentiment across 30 samples. However, for negative sentiment, precision dropped to 86% while maintaining a recall of 100%), leading to an Fl-score of 92% from 6 samples, suggesting more false positives due to the small and imbalanced dataset.

To sum it up, the sentiment classification model demonstrates notable effectiveness across all islands, particularly in detecting positive sentiments, which consistently yielded high precision and recall scores. The slight reduction in precision for negative sentiment, particularly observed in Pangkor, may stem from a limited sample size and class imbalance. These findings validate the robustness of the hybrid S A approach and highlight the importance of balanced data representation to enhance the classification of negative sentiments. Additionally, to evaluate the model's predictive capability on unseen data, the AUC metric was applied.

4.4.2 Validate Area Under the Curve Metric Results

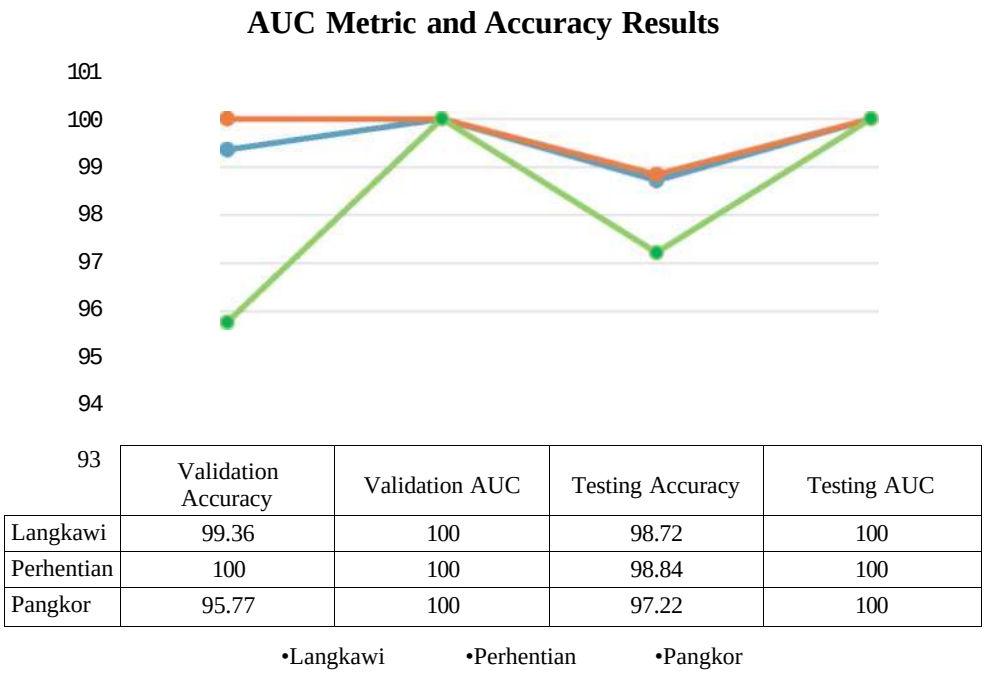
To ensure the robustness and generalisability of the sentiment classification models, various evaluation techniques were used. Stratified *k-fo\d* cross-validation was utilised to segment the dataset into multiple folds while maintaining the original class distribution, thereby minimising bias and variance during both training and testing. This technique facilitated reliable performance estimation across different subsets. Following that, key performance metrics were calculated, including accuracy, precision, recall, and Fl-score, which provided a comprehensive assessment of the model's effectiveness. Additionally, the AUC of the Receiver Operating Characteristic (ROC) curve was used to validate the model's capacity to distinguish between sentiment classes at various threshold levels, with a higher AUC signifying superior classification performance. In summary, the evaluation strategies applied established a comprehensive validation framework, thereby ensuring that the models developed were both statistically robust and effective for the tasks associated with SA.

4.4.2.1 Validate Area Under the Curve Metric using Receiver Operating Characteristic Curve

In order to strengthen the reliability and generalisability of the hybrid S A model, this research decisively enhances its performance evaluation by incorporating the AUC of the ROC curve. While precision, recall, and Fl-score offer valuable insights into classification accuracy at fixed thresholds, AUC effectively measures the model's capability to distinguish between sentiment classes across all threshold levels. This aspect is crucial in real-world scenarios where sentiment class imbalance and ambiguity frequently arise. The AUC evaluation was rigorously conducted on a distinct, unseen dataset comprising YouTube reviews that were excluded from the model's training and validation phases.

The implementation of the AUC metric through the ROC curve serves as an effective approach for evaluating classification performance, particularly in the context of imbalanced datasets. The ROC curve illustrates the relationship between the TP and the FP rates across various threshold settings, while the AUC quantifies the model's capacity to differentiate between classes. A perfect AUC score of 100 signifies flawless classification, whereas a score of 50 indicates performance equivalent to random

guessing. In the evaluations for Langkawi, Perhentian, and Pangkor islands in Figure 4.30, all three achieved a perfect AUC score of 100 in both the validation and testing phases, showcasing the model's exceptional discriminative power in identifying sentiment polarity. However, slight discrepancies emerge when examining accuracy. Perhentian recorded the highest validation and testing accuracy, reaching 100% and 98.84%, respectively, closely followed by Langkawi at 99.36% and 98.72%. In contrast, Pangkor exhibited slightly lower accuracy, with *95.11%* for validation and 97.22%) for testing. Overall, the observed variations indicate that, although the model effectively ranks predictions despite having identical AUC scores, it may still misclassify specific instances based on the specified threshold settings.



	Validation Accuracy	Validation AUC	Testing Accuracy	Testing AUC
Langkawi	99.36	100	98.72	100
Perhentian	100	100	98.84	100
Pangkor	95.77	100	97.22	100

Figure 4.30 Area Under the Curve Metric

Furthermore, the inclusion of the AUC score using ROC evaluation on unseen data further validates the robustness and generalisability of the hybrid SA approach. The ROC curve for the testing data in Figure 4.31 clearly illustrates the model's outstanding performance across all classification thresholds. Represented by an orange line, the curve closely follows the left and top edges of the plot, indicating an exceptional classification ability. This is further substantiated by an impressive AUC score of 1.0, which signifies flawless discrimination between positive and negative

instances. The proximity of the ROC curve to the top-left corner highlights the model's effectiveness in accurately identifying TP while minimising FP. Next, the dashed diagonal line, serving as a benchmark for random chance performance, illustrates that the model's predictions significantly outperform random guessing. Overall, the ROC curve reinforces the model's robustness and reliability in classifying the testing data, showcasing its impeccable performance without any errors.

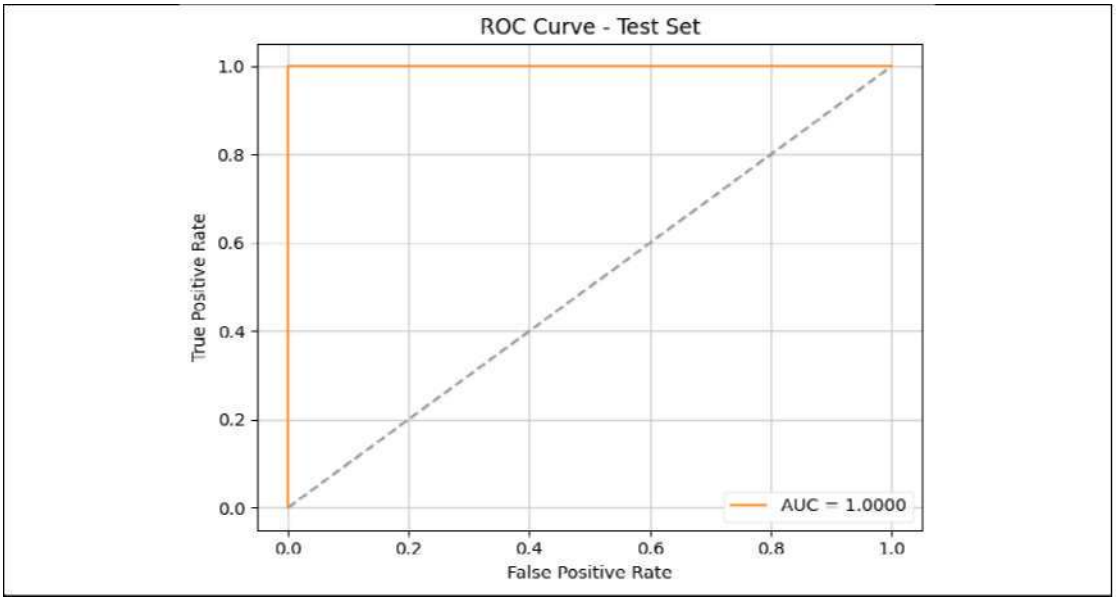


Figure 4.31 Receiver Operating Characteristic Curves for Langkawi, Perhentian, and Pangkor

4.4.2.2 Validate Unseen Data with Other Techniques

Ensuring the robustness and generalisability of the hybrid SA model involved further validation using unseen data and alternative techniques. This step is essential for evaluating the model's ability to maintain high performance beyond the initial training and cross-validation phases. By applying the trained SVM model that integrates VADER sentiment scores along with combined TF-IDF and LDA features to unseen datasets, it allows for assessing the model's real-world applicability. Moreover, comparing this hybrid approach with other SA techniques offers valuable insights into its effectiveness and consistency across diverse contexts and datasets specific to different islands.

The SA results for Malaysia's islands, such as Langkawi, Perhentian, and Pangkor, in Figure 4.32 demonstrate notable differences based on the techniques

utilised to train the SVM model with stratified k-fold cross-validation (*k*=9). Among the five configurations evaluated, the hybrid SA approach that integrates VADER with SVM and implements both TF-IDF and LDA features consistently demonstrated superior performance compared to all other models across the three islands. This methodology achieved accuracy rates of 98.15% for Langkawi, 98.5% for Perhentian, and 97.19% for Pangkor. These findings indicate that the combination of lexical-based sentiment features (VADER) and statistical representations (TF-IDF and LDA) effectively captures a more comprehensive and nuanced understanding of the sentiments expressed in reviews.

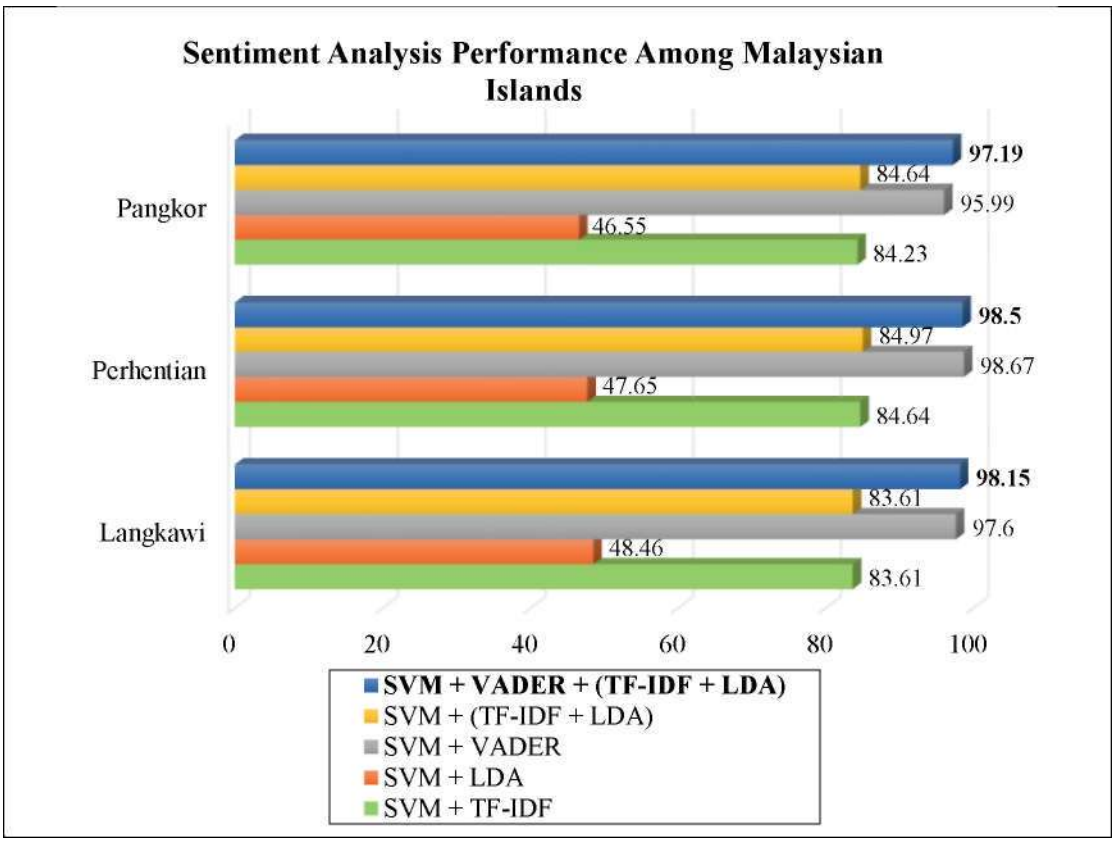


Figure 4.32 Sentiment Analysis Performance of Support Vector Machine Model with Other Techniques

In comparison, the implementation of LDA features alone resulted in the lowest performance metrics for all three islands, registering accuracies of merely 48.46% for Langkawi, 47.65% for Perhentian, and 46.55% for Pangkor. This indicates that LDA, an unsupervised topic modelling technique, is unable to effectively capture sentiment-specific information on its own unless paired with more targeted methods. Conversely,

TF-IDF demonstrated significantly better performance, achieving over 83% accuracy for each island, emphasising its effectiveness as a dependable baseline feature set. Interestingly, while combining TF-IDF and LDA without VADER yielded a slight improvement over using TF-IDF alone, however, this combination did not surpass the performance achieved by the hybrid approach that incorporated VADER.

Throughout all islands, the relative ranking of feature set performance exhibited consistency, which underscores the robustness of the hybrid approach and highlights the importance of integrating both sentiment-aware and context-aware features. The superior performance achieved by the VADER, SVM, TF-IDF and LDA techniques illustrates the benefits of amalgamating diverse feature types, particularly when analysing tourism-related reviews where explicit sentiment expressions and thematic context are paramount. This rigorous analysis supports the assertion that a hybrid approach is essential for attaining high sentiment classification performance in real-world applications, such as evaluating experiences related to island tourism.

4.5 Discussion of Findings

This section presents significant findings regarding the sentiment of YouTube reviews pertaining to Malaysia's island tourism. Through the integration of DSA techniques utilising TF-IDF and n-gram analysis, significant keywords and phrases have been identified, offering insights into tourist preferences and sentiment responses. Meanwhile, LDA has further refined semantic understanding by clustering related terms into coherent topics, which facilitated a structured interpretation of user concerns and revealed thematic distinctions among the islands.

On the other hand, the hybrid SA approach demonstrated robust performance. VADER served as a reliable baseline for sentiment prediction, while the SVM, enhanced by features derived from TF-IDF and LDA, enhanced sentiment classification accuracy by learning contextual cues. This approach exhibited consistent improvements in accuracy, particularly in the context of stratified *k-fo\d* cross-validation and AUC evaluation, indicating its effectiveness and scalability. Hence, these findings enhance the understanding of public sentiment within the tourism sector and underscore the value of integrating DSA techniques with a hybrid SA approach for multilingual SA. The subsequent subsections will explore the implications of these findings in relation

to research objectives, comparative evaluations of models, and the challenges encountered throughout the study.

4.5.1 Alignment with Research Objectives

This research demonstrates a clear alignment with all three predefined research objectives, as evidenced by the analytical results and evaluations presented in Table 4.9. The first objective aimed to analyse descriptive and semantic techniques used in analysing the SA related to Malaysia's island tourism. This was effectively addressed through the comprehensive implementation of TF-IDF and LDA. These techniques facilitated the extraction of significant keywords, identification of dominant topics, and aspect labelling through topic modelling analysis. The findings provided valuable insights into tourists' discourse patterns and highlighted key themes such as natural attractions, accommodation, and emotional experiences across Langkawi, Perhentian, and Pangkor islands.

Table 4.9

Alignment with Research Objectives

Research Objectives	Description	Techniques Applied	Findings and Outcomes
RO 1	Analyse the existing descriptive and semantic techniques used in analysing SA related to Malaysia's island tourism.	TF-IDF n-gram analysis, LDA topic modelling and aspect labelling.	Successfully identified key terms, thematic patterns, and aspect labels for Langkawi, Perhentian, and Pangkor islands.
RO 2	Construct a hybrid approach integrating DSA to improve the SA of YouTube reviews about Malaysia's island tourism.	A hybrid SA approach combining VADER and SVM, integrated by TF-IDF and LDA techniques.	Enhanced sentiment classification performance and captured contextual and multilingual nuances.
RO 3	Evaluate the accuracy of the constructed hybrid SA approach using performance metrics and validate using the AUC metric.	Stratified Mold cross-validation, confusion matrix, and AUC metric.	Achieved high model performance with accuracy of more than 97% and strong AUC and showed robustness and minimal overfitting across three island datasets.

The second objective focused on constructing a hybrid approach that integrates DSA techniques to improve the SA of YouTube reviews about Malaysia's island tourism. This was achieved through the development and application of a hybrid SA approach that combines lexicon-based sentiment scoring via VADER with ML classification using SVM. The integration of TF-IDF and LDA features into the SVM model has resulted in a significant enhancement of its capacity to process nuanced expressions, informal language, and multilingual content that is frequently present in user-generated reviews. Subsequently, this hybrid approach effectively bridged the gap between semantic context and sentiment polarity, resulting in accurate and context-aware sentiment classification.

Lastly, the third objective involved evaluating the accuracy of the constructed hybrid SA approach using performance metrics and validating it using the AUC metric. This was thoroughly evaluated through performance metrics using stratified *k-fo\d* cross-validation and a confusion matrix, along with validation on unseen data using the AUC metric. As a result, the hybrid SA approach achieved high accuracy and robustness across various data sizes and island-specific datasets, with validation scores consistently exceeding 97%. Additionally, AUC-based evaluations confirmed strong class separation capabilities and demonstrated their effectiveness in handling real-world sentiment data with variable structures.

In summary, all three research objectives have been effectively addressed through the integration of DSA techniques within the hybrid SA approach. The findings indicate that the integration of TF-IDF, LDA, VADER, and SVM is effective in capturing both lexical and contextual sentiment dimensions, thereby evaluating and validating the robustness of the model through the performance and AUC metrics. These results align with the overarching goal of developing a reliable and scalable SA model tailored for user-generated tourism content, particularly within the context of Malaysia's multilingual and culturally diverse island reviews. The subsequent section will provide a comparative analysis of the primary components of SA, which are VADER and SVM, emphasising their respective strengths, limitations, and complementary functions in achieving accurate sentiment predictions and classifications.

4.5.2 Comparative Evaluation of Valence Aware Dictionary and Sentiment Reasoner and Support Vector Machine

The comparative analysis of VADER and SVM elucidates the distinct yet complementary strengths of lexicon-based and ML-based SA techniques. VADER effectively captures general sentiment trends in YouTube reviews, particularly in the identification of strong positive sentiments, allowing for straightforward interpretation of sentiment polarity. However, its performance is limited when addressing ambiguous or complex expressions, as well as multilingual elements often encountered in Malaysia's tourism reviews. Conversely, the SVM model exhibits superior adaptability and precision in detecting contextual sentiment patterns while accommodating linguistic diversity, especially when integrated with DSA features.

This contrast is clearly illustrated in Table 4.10, which summarises the accuracy results for each island implementing with or without DSA techniques. In the analysis of all datasets, the hybrid SA approach with DSA techniques demonstrated superior performance. For instance, in Langkawi, the accuracy improved from 97.60% with only SVM and VADER to 98.15% when enriched with DSA features. Likewise, while Perhentian achieved high accuracy with both techniques, the hybrid approach yielded a slight enhancement and greater generalisability. Notably, Pangkor experienced the most substantial performance improvement, increasing from 95.99%) to 97.19%), thereby underscoring the value of incorporating enriched features to effectively manage more ambiguous or less expressive reviews.

Table 4.10

Summary of Accuracy for Malaysia's Islands

Type of Island	Technique	Accuracy (%)
Langkawi	SVM, VADER	97.60
	TF-IDF, LDA, VADER, SVM	98.15
Perhentian	SVM, VADER	98.67
	TF-IDF, LDA, VADER, SVM	98.50
Pangkor	SVM, VADER	95.99
	TF-IDF, LDA, VADER, SVM	97.19

These findings reinforce the assertion that while VADER is effective for rapid and surface-level sentiment assessment, SVM and integrated DSA inputs lead to superior analytical performance. This synergy between lexicon-based and ML methodologies facilitates a more accurate and nuanced interpretation of sentiment. In addition, this is particularly significant within the context of YouTube reviews, where the incorporation of multiple languages and informal expressions can pose challenges to straightforward SA.

To sum it up, the comparative analysis of VADER and SVM reveals the distinct advantages of each approach within the SA framework. VADER functions as a lexicon-based tool for sentiment polarity detection, whereas SVM offers advanced learning-based insights that are particularly effective for nuanced sentiment classification, especially when utilised alongside comprehensive DSA features. The enhanced accuracy exhibited by the hybrid approach across all island datasets underscores the significance of integrating both lexicon-based and ML methodologies to achieve a scalable and robust SA of tourism-related user-generated content.

4.5.3 Addressing Multilingual Ambiguity

Multilingual ambiguity presents a substantial challenge in SA, particularly in user-generated content related to tourism, where individuals frequently switch between English, Malay, and other local languages. Hence, this research examined YouTube reviews for Malaysia's islands, where informal phrasing hindered the efficacy of standard lexicon-based tools like VADER, which struggles with non-English phrases and local slang. To overcome these challenges, the research applied the ML approach using SVM trained on TF-IDF and LDA features. This technique identifies sentiment patterns based on term distribution rather than relying on predefined dictionaries, which allows it to adapt to a diverse vocabulary, including informal expressions. The results showed a significant improvement in sentiment classification accuracy across all datasets, particularly for Pangkor Island, where linguistic diversity was particularly pronounced.

Moreover, by incorporating topic modelling through LDA, the technique captured semantic relationships that transcend linguistic boundaries, facilitating the organisation of multilingual terms into shared thematic clusters. This capability enabled the SVM model to infer sentiment even when textual elements were fragmented or

expressed in multiple languages. Overall, the implementation of this hybrid approach has significantly improved classification accuracy while demonstrating substantial resilience to the limitations commonly associated with standard lexicon-based tools. Consequently, it is particularly well-suited for conducting SA on digital platforms in Southeast Asia.

4.6 Summary

This chapter analysed the implementation of the model framework on YouTube reviews of Malaysia's islands. By implementing DSA techniques such as TF IDF, n-grams, and LDA, the research identified significant terms, themes, and semantic patterns. Visualisations, including t-SNE, word clouds, and Venn diagrams, highlighted both overlapping and unique aspects of tourist experiences in Langkawi, Perhentian, and Pangkor. These results provided meaningful insights into tourist sentiments and thematic structures across the islands.

For sentiment classification, the hybrid approach that combines VADER with SVM outperformed the standard lexicon-based method. Although VADER faced challenges with multilingual and informal expressions, the SVM model enhanced with TF DDF and LDA features achieved high accuracy and strong generalisation. This performance was confirmed through cross-validation accuracy and consistently high AUC values. The chapter also compared VADER and SVM, addressed multilingual ambiguity, and linked outcomes to the research objectives. Overall, the model framework proved effective in enhancing sentiment classification. Chapter 5 will present the summary of findings, limitations, and recommendations for future research.

CHAPTER 5

CONCLUSION AND FUTURE WORK

This chapter presents an overview of the research findings, which are provided in the last chapter of this research, with an emphasis on the implementation of hybrid sentiment analysis (SA) and descriptive-semantic analysis (DSA) methodologies. The limitations that were observed throughout this research approach are also addressed, and particular possibilities for future research are recommended. Moreover, this chapter demonstrates the significance of the hybrid approach in determining the user's opinion of Malaysia's islands, along with providing recommendations for future developments.

5.1 Research Contribution

This research contributes to the field of SA in tourism by systematically integrating Descriptive-Semantic Analysis (DSA) techniques with a hybrid SA approach. The research focused on three major Malaysia islands, Langkawi, Perhentian, and Pangkor, using YouTube reviews as the primary data source, demonstrating how linguistic, thematic, and sentiment-driven insights can be jointly extracted to better understand tourists' perceptions. By integrating Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Dirichlet Allocation (LDA) for feature extraction with a hybrid SA approach that implements Valence Aware Dictionary and Sentiment Reasoner (VADER) and Support Vector Machine (SVM), this research establishes a robust and scalable framework for tourism SA.

On the other hand, Research Objective 1 (RO 1) was achieved by analysing the existing descriptive and semantic techniques, which directly address Research Question 1 **(RQ1)** in identifying the descriptive and semantic techniques applied in SA within the context of Malaysia's island tourism. The implementation of TF-IDF in conjunction with n-gram models (unigram, bigram, and trigram) successfully extracted significant keywords and meaningful word combinations from YouTube reviews, highlighting notable linguistic patterns associated with tourist experiences. Consequently, the combination of unigram and bigram features demonstrated the strongest performance. Moreover, LDA topic modelling identified the dominant topics discussed by tourists,

enabling the assignment of aspect labels that encapsulated essential experiential dimensions such as destination appeal, activities, and overall satisfaction. These results affirm that DSA techniques are effective in uncovering both surface-level linguistic structures and deeper semantic patterns, thereby adequately addressing **ROl** and thoroughly answering **RQ1.**

This research aimed to construct a hybrid approach integrating DSA in achieving Research Objective 2 **(RO** 2). This goal is directly related to Research Question 2 (RQ2), which explores how a hybrid approach integrating DSA can improve the accuracy of SA applied to YouTube reviews on Malaysia's island tourism. The initial sentiment polarity analysis using VADER indicated varied sentiment distributions among the islands, with Langkawi having the most positive sentiment, followed by Perhentian and Pangkor; however, Perhentian showed a higher neutral sentiment, indicating a more balanced range of opinions. These results underscore VADER's capability to capture subtle sentiment nuances in discussions related to tourism. Furthermore, adding VADER's compound sentiment score as an extra feature within the SVM classifier, along with TF-IDF and LDA features, led to improved classification performance. As a result, the learning curve analysis demonstrated SVM's ability to generalise effectively across different training sizes. Overall, these findings confirm the accuracy of the hybrid approach integrating DSA, thereby fulfilling **R02** and providing a clear answer to RQ2.

In order to meet Research Objective 3 **(RO** 3), the performance of the hybrid SA approach was evaluated using performance metrics and validated using the Area Under the Curve (AUC) metric, which addresses Research Question 3 **(RQ3),** which seeks to measure the accuracy of the hybrid SA approach that integrates DSA. Utilising stratified 9-fold cross-validation ensured balanced class representation and reliable performance evaluation across all datasets. The hybrid framework achieved impressive accuracy rates of 98.15% for Langkawi, 98.50% for Perhentian, and *91.19%* for Pangkor. Also, the AUC metric attained a perfect score of 100 for all three islands, underscoring the model's remarkable ability to differentiate between sentiment classes. Among the datasets, Perhentian emerged as the top performer, closely followed by Langkawi, while Pangkor also demonstrated strong and consistent performance. These results affirm the reliability, generalisability, and minimal overfitting of the hybrid framework, thereby addressing **R03** and conclusively answering **RQ3.**

To conclude, this research successfully this research effectively addresses all of the research questions by connecting DSA insights with a highly efficient hybrid SA approach. The integration of TF-IDF, LDA, VADER, and SVM establishes a comprehensive framework capable of accurately analysing large volumes of tourism-related user-generated content. From a practical standpoint, the findings provide valuable insights for tourism stakeholders, including policymakers and local businesses, enabling them to monitor public sentiment and refine strategic decision-making. Academically, the framework offers a scalable and adaptable foundation for future research in SA within tourism and other fields that require culturally and contextually sensitive interpretation of sentiment.

5.2 Research Limitations and Future Works

Despite the successful implementation of the hybrid SA approach integrating DSA techniques, several limitations emerged throughout this research that influence the overall reliability, generalisability, and computational efficiency of the findings. These limitations relate primarily to data source constraints, computational overhead, and methodological challenges associated with sentiment misclassification and interpretability. It is crucial to address these limitations to improve the robustness of SA practices, especially when dealing with large-scale, multilingual, and diverse platform tourism datasets. Hence, this section discusses the key research limitations observed in this research and provides suggestions for future research directions that could enhance the performance, scalability, and applicability of hybrid SA frameworks in tourism analytics.

5.2.1 Source Bias and Multilingual Issues

This research relies primarily on content from YouTube, which could contain biases associated with the different types of content and platform demographics. The emphasis on audio-visual formats, particularly in vlogs, may influence how opinions are conveyed and transcribed into written reviews, thereby affecting the dataset. Since the primary data source comprises reviews from the YouTube platform, it is recommended that future research expand the dataset by incorporating reviews from multiple platforms such as TripAdvisor, Google Reviews, Facebook, and X (formerly

Twitter). This broader data collection would provide a more comprehensive and balanced understanding of tourists' perceptions. When collecting data from multiple platforms, the dataset size can be increased for lessened island reviews, like Pangkor, in order to ensure a balanced SA across different island destinations, reducing destination-based sampling bias.

Besides that, Google Translate was implemented to process reviews in multiple languages, which could potentially affect the findings since translation tools may not fully capture idiomatic expressions and cultural nuances. Future studies should focus on enhancing multilingual data management beyond automatic translation tools like Google Translate, which may introduce cultural bias and semantic loss, given that SA relies on linguistic consistency. Hence, implementing more advanced multilingual SA techniques, such as transformer-based approaches, context-aware models, and language-specific lexicons, is recommended for further research to better preserve idiomatic expressions and cultural meanings across languages.

5.2.2 Computational Complexity

Another significant issue is processing time, as the hybrid approach incurs high processing time due to the integration of various techniques. For instance, a hybrid approach is computationally intensive due to the necessity to compute a huge number of predictions, which could lead to a significant amount of processing time and excessive workload on low-performance machines (Polignano et al., 2022). For instance, TF-IDF generates a document-term matrix, resulting in greater processing time and complexity due to n-gram features such as unigram, bigram and trigram. Although TF-IDF assigns higher values to frequently occurring terms, the presence of common words across documents complicates the feature extraction process (Barua et al., 2021). Additionally, SVM classification using TF-IDF features increases computational complexity as it handles a large number of features, resulting in higher memory usage and longer training times, particularly with large datasets, which can hinder the performance of the SVM (Febriansyah et al., 2022).

Future research could explore transformer architectures or deep learning models to enhance sentiment classification performance, such as Bidirectional Encoder Representations from Transformers (BERT) and Long Short-Term Memory (LSTM), which can enhance performance and generalisation while being optimised for large-

scale data processing. According to the study by Sinha et al. (2022), BERT-LSTM performs better in sentiment classification than conventional DL models because BERT-based techniques are extremely effective due to their enhanced accuracy and capacity to handle complex sentiment expressions. BERT is a transformer-based model that enhances understanding of multi-word expressions and sentiment nuances through bidirectional processing. In contrast, LSTM is a recurrent neural network that effectively identifies negatives and complex sentiment patterns in reviews by managing broad text associations. Overall, both models improve consistency and generalisation for opinion-rich data like YouTube reviews.

5.2.3 Sentiment Misclassification and Interpretability Difficulties

On the other hand, the effectiveness and reliability of the hybrid SA approach could potentially be influenced by a variety of limitations, such as sentiment misclassification and interpretability difficulties. This is because the lexicon-based approach faces limitations in accuracy and domain dependency, while the machine learning-based approach relies heavily on domain-specific knowledge (Islam et al., 2024). In addition, the distribution of sentiment classes within the dataset can influence SVM performance due to the bias on the dominant sentiment class (Anastasiou et al., 2023). These limitations underscore the necessity for methodological advancements aimed at enhancing the robustness, generalisability, and interpretability of the hybrid SA model in future applications.

To address these limitations, future work may focus on fine-tuning VADER using domain-specific terms, slang, and contextual sentiment modifications to strengthen its ability to interpret ambiguous expressions. This is because VADER can be enhanced by implementing more slang terms and modifying sentiment scoring based on subtleties in the context (Barik & Misra, 2024), allowing it to more effectively comprehend ambiguous phrases and classify sentiment accurately across a range of industries. In contrast, SVM classification can be enhanced with kernel approximation approaches to provide more rapid training and inference times. Techniques such as the Nystroem Method, Random Fourier Features, Multiple Kernel Learning, Adaptive Kernel Approximation, and Low-Rank Kernel Approximation can be applied to reduce computing complexity while increasing scalability and applicability for large datasets (Du et al., 2024). To sum it up, these enhancements can assist the hybrid SA approach

in becoming more efficient and more interpretable, which ultimately becomes ideal for large-scale applications.

5.3 Summary

This chapter outlines the key contributions, limitations, and future directions of the research. It highlights the successful implementation of the TF-IDF, LDA, VADER, and SVM to analyse tourism-related sentiment from YouTube reviews of Malaysia's islands. The research demonstrated that integrating DSA techniques with a hybrid SA approach achieved high classification accuracy and provided valuable insights into tourist perceptions. The findings confirmed the framework's robustness, scalability, and generalisability, emphasising its importance for tourism stakeholders and future SA research.

However, some limitations were identified, including data source biases and multilingual issues, computational demands, and interpretability issues. These highlight the need for diverse datasets, better multilingual processing, and improved computational strategies to enhance hybrid SA models. The chapter also suggests future research directions, such as integrating data from various online platforms, utilising transformer-based multilingual models, implementing deep learning architectures and refining VADER with domain-specific lexicons, which further enhance sentiment detection and generalisability across extensive tourism datasets. In summary, this chapter reinforces the research's contributions while providing a roadmap for advancing hybrid SA techniques for sentiment-driven applications in tourism and beyond.

REFERENCES

Abdullah, N. A. S., & Rush, N. I. A. (2021). Multilingual sentiment analysis: A systematic literature review. *Pertanika Journal of Science and Technology, 29(1),* 445-470. https://doi.org/10.47836/pjst.29.L25

Abeysiriwardana, M., & Sumanathilaka, D. (2024). A survey on lexical ambiguity detection and word sense disambiguation. *ArXiv (Cornell University).*

Agrawal, R. (2020). Sentiment analysis of YouTube comments. Retrieved June 16, 2020, from https://www.analyticssteps.com/blogs/sentiment-analysis-youtube-comments

Akhtar, M. M. (2019). Sentiment analysis on YouTube: A brief survey. *Research Gate, 3,* 1250-1257. https://doi.org/10.48550/arXiv.1511.09142

Alabi, O., & Bukola, T. (2023). Introduction to descriptive statistics. *Recent Advances in Biostatistics [Working Title],* 1-12. https://doi.org/10.5772/intechopen.1002475

Alawi, A. B., & Bozkurt, F. (2024). A hybrid machine learning model for sentiment analysis and satisfaction assessment with Turkish universities using Twitter data. *Decision Analytics Journal,* //(April), 100473. https://doi.Org/10.1016/j.dajour.2024.100473

Alhujaili, R. F., & Yafooz, W. M. S. (2021). Sentiment analysis for Youtube videos with user comments: Review. *Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021,* 814-820. https://doi.org/10.1109/ICAIS50930.2021.9396049

Ali, T., Omar, B., & Soulaimane, K. (2022). Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. *MethodsX,* ^(November), 101894. https://doi.Org/10.1016/j.mex.2022.101894

Alqaryouti, O., Siyam, N., Abdel Monem, A., & Shaalan, K. (2024). Aspect-based sentiment analysis using smart government review data. *Applied Computing and Informatics, 20(1/2),* 142-161. https://doi.Org/10.1016/j.aci.2019.ll.003

Alsemaree, O., Alam, A. S., Gill, S. S., & Uhlig, S. (2024). Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions. *Heliyon, 10(9).* https://doi.Org/10.1016/j.heliyon.2024.e27863

Amalia, R. N., Sadik, K., & Notodiputro, K. A. (2023). A preliminary study of sentiment analysis on Covid-19 news: Lesson learned from data acquisition, pre-

processing, and descriptive analytics. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan,* 77(4), 1901-1914. https://doi.org/10.30598/barekengvoll7iss4ppl901-1914

Anastasiou, P., Tzafilkou, K., Karapiperis, D., & Tjortjis, C. (2023). YouTube sentiment analysis on healthcare product campaigns: Combining lexicons and machine learning models. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (USA)* (pp. 1-8). https://doi.org/10.1109/IISA59645.2023.10345900

Arora, S., Malhotra, A., Kumar, A., & Singh, N. (2024). Sentiment analysis unleashed: Leveraging VADER for evaluation of public opinion. *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2024,* 1-6. https://doi.org/10.1109/ICRITO61523.2024.10522172

Arslan, M., & Cruz, C. (2023). Leveraging NLP approaches to define and implement text relevance hierarchy framework for business news classification. *Procedia Computer Science, 225,* 317-326. https://doi.Org/https://doi.org/10.1016/j.procs.2023.10.016

Aye, Y. M., & Aung, S. S. (2020). Contextual lexicon based sentiment analysis in Myanmar text reviews. In *2020 23rd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)* (pp. 160-165). https://doi.org/10.1109/O-COCOSDA50338.2020.9295012

Azizan, A., Jamal, N. N., Abdullah, M. N., Mohamad, M., & Khairudin, N. (2019). Lexicon-based sentiment analysis for movie review tweets. In *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS)* (pp. 132-136). https://doi.org/10.1109/AiDAS47888.2019.8970722

Bagherzadeh, S., Shokouhyar, S., Jahani, H., & Sigala, M. (2021). A generalizable sentiment analysis method for creating a hotel dictionary: using big data on Trip Advisor hotel reviews. *Journal of Hospitality and Tourism Technology, 12(2),* 210-238. https://doi.Org/10.l 108/JHTT-02-2020-0034

Bakar, M. F. R. A., Idris, N., & Shuib, L. (2019). An enhancement of Malay social media text normalization for lexicon-based sentiment analysis. In *2019 International Conference on Asian Language Processing (IALP)* (pp. 211-215). https://doi.org/10.1109/IALP48816.2019.9037700

Baldha, T., Mungalpara, M., Goradia, P., & Bharti, S. (2021). Covid-19 vaccine tweets sentiment analysis and topic modelling for public opinion mining. In *2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)* (pp. 1-6). https://doi.org/10.1109/AIMV53313.2021.9671000

Balis, J. (2021). 10 truths about marketing after the pandemic. Retrieved March 10, 2021, from https://hbr.org/2021/03/10-truths-about-marketing-after-the-pandemic?ssp=l&setlang=en-MY&safesearch=moderate

Barik, K., & Misra, S. (2024). Analysis of customer reviews with an improved VADER lexicon classifier. *Journal of Big Data, 11(1).* https://doi.Org/10.l 186/s40537-023-00861-x

Barrera, R. (2023). Identifying the attributes of consumer experience in Michelin-starred restaurants: a text-mining analysis of online customer reviews. *British Food Journal*, 725(13), 579-598. https://doi.org/10.1108/BFJ-05-2023-0408

Barua, A., Sharif, O., & Hoque, M. M. (2021). Multi-class sports news categorization using machine learning techniques: Resource creation and evaluation. *Procedia Computer Science, 193,* 112-121. https://doi.0rg/https://doi.org/lO.lOl6/j.procs.202i.ll.OO2

Batta, V. M. B. (2024). Human language data processing using NLTK. *International Journal of Advanced Research in Science, Communication and Technology, 4(6),* 628-634. https://doi.org/10.48175/ijarsct-17685

Ben, T. L., N, R. R., Alia, P. C. R., G, K., & Mishra, K. (2023). Detecting sentiment polarities with comparative analysis of machine learning and deep learning algorithms. *2023 International Conference on Advancement in Computation and Computer Technologies, InCACCT 2023,* 186-190. https://doi.org/10.1109/InCACCT57535.2023.10141741

Birjali, M., Kasri, M., & Hssane, A. B. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems, 226,* 107134. https://doi.Org/10.1016/j.knosys.2021.107134

Bonthu, H. (2024). Rule-based sentiment analysis in python. Retrieved February 20, 2024, from https://www.analyticsvidhya.com/blog/2021/06/rule-based-sentiment-analysis-in-python/

Brandao, J. D. G., & Calixto, W. P. (2019). N-gram and TF-IDF for feature extraction on opinion mining of tweets with SVM classifier. *2019 International Conference on Artificial Intelligence and Data Processing Symposium, IDAP 2019.*

https://doi.org/10.1109/IDAP.2019.8875900

Cam, H., Cam, A. V., Demirel, U., & Ahmed, S. (2024). Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers. *Heliyon, 10(1),* e23784. https://doi.Org/10.1016/j.heliyon.2023.e23784

Cao, L. (2023). Sentiment analysis of social media text based on deep learning. In *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)* (pp. 1-5). https://doi.org/10.1109/ICMNWC60182.2023.10435901

Chakravarty, A., & Rengarajan, A. (2023). Analysis of deep learning models to identify uniqueness in user reviews and ratings. *2023 3rd International Conference on Advancement in Electronics and Communication Engineering, AECE 2023,* 698-702. https://doi.org/10.1109/AECE59614.2023.10428195

Chan, J. (2021). Domestic tourism as a pathway to revive the tourism industry and business post the covid-19 pandemic. *ERIA Discussion Paper Series,* (392), 1-35. Retrieved from https://www.eria.org/uploads/media/discussion-papers/ERIA-Research-on-COVID-19/Domestic-Tourism-as-a-Pathway-to-Revive-the-Tourism-Industry-and-Business-Post-the-COVID-19-Pandemic.pdf

Chang, C.-H., Wu, M.-L., & Hwang, S.-Y. (2019). An approach to cross-lingual sentiment lexicon construction. In *2019 IEEE International Congress on Big Data (BigDataCongress)* (pp. 129-131). https://doi.org/10.1109/BigDataCongress.2019.00030

Chathuranga, P. D. T., Lorensuhewa, S. A. S., & Kalyani, M. A. L. (2019). Sinhala sentiment analysis using corpus based sentiment lexicon. In *2019 19th International Conference on Advances in ICTfor Emerging Regions (ICTer)* (Vol. 250, pp. 1-7). https://doi.org/10.1109/ICTer48817.2019.9023671

Chauhan, G. S., Nahta, R., Meena, Y. K., & Gopalani, D. (2023). Aspect based sentiment analysis using deep learning approaches: A survey. *Computer Science Review, 49.* https://doi.org/10.1016/jxosrev.2023.100576

Chen, Y., Lin, L., Gong, Z., & Tan, W. (2023). A translanguaging turn in sentiment analysis: Exploring the role of language, culture, and linguistics in natural language processing. *International Journal of Engineering and Technology, 15(4),* 179-184. https://doi.org/10.7763/ijet.2023.vl5.1243

Datta, K. S. S., Naidu, G. N., Abhishek, S., & T, A. (2024). Enhancing veracity: Empirical evaluation of fake news detection techniques. *Procedia Computer Science, 233,* 97-107. https://doi.Org/https://doi.org/10.1016/j.procs.2024.03.199

Dornick, C, Kumar, A., Seidenberger, S., Seidle, E., & Mukherjee, P. (2021). Analysis of patterns and trends in COVID-19 research. *Procedia Computer Science, 185,* 302-310. https://doi.Org/https://doi.org/10.1016/j.procs.2021.05.032

Du, K. L., Jiang, B., Lu, J., Hua, J., & Swamy, M. N. S. (2024). Exploring kernel machines and support vector machines: Principles, techniques, and future directions. *Mathematics,* 72(24), 1-58. https://doi.org/10.3390/mathl2243935

Dutta, B. (2021). Semantic analysis: Working and techniques. Retrieved July 23, 2021, from https://www.analyticssteps.com/blogs/semantic-analysis-working-and-techniques

Fazrin, F., Pratiwi, O. N., & Andreswari, R. (2022). Comparison of k-nearest neighbor and logistic regression algorithms on sentiment analysis of Covid-19 vaccination on Twitter with vader And textblob labeling. *2022 International Conference of Science and Information Technology in Smart Administration, ICSINTESA 2022,* 39-44. https://doi.org/10.1109/ICSINTESA56431.2022.10041609

Febriansyah, M. R., Nicholas, Yunanda, R., & Suhartono, D. (2022). Stress detection system for social media users. *Procedia Computer Science, 216,* 672-681. https://doi.Org/10.1016/j.procs.2022.12.183

Finnigan, K. D. C, Anzum, F., Rokne, J., & Gavrilova, M. L. (2022). Weighted lexicon-based sentiment analysis for women career traits in information technology. In *2022 IEEE 21st International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)* (pp. 91-98). https://doi.org/10.1109/ICCICC57084.2022.10101520

Gkillas, A., Simos, M. A., & Makris, C. (2023). Popularity inference based on semantic sentiment analysis of YouTube video comments. In *14th International Conference on Information, Intelligence, Systems and Applications, USA 2023.* https://doi.org/10.1109/IISA59645.2023.10345963

Goyal, C. (2021). Part 9: Step by step guide to master NLP - semantic analysis. Retrieved June 23, 2021, from https://www.analyticsvidhya.com/blog/2021/06/part-9-step-by-step-guide-to-master-nlp-semantic-analysis/

Gronneberg, B. A., & Nygard, K. E. (2022). NLP text analysis on influential factors of persistence for female computer science undergraduates: A comparative study. In *2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)* (pp. 72-77).

https://doi.org/10.1109/ICT4DA56482.2022.9971229

Hall, D. (2023). What is descriptive analytics? Retrieved November 8, 2023, from https://technologyadvice.com/blog/information-technology/descriptive-analytics/

Harywanto, G. N., Siautama, R., A., A. C. I., & Suhartono, D. (2021). Extractive hotel review summarization based on TF/IDF and adjective-noun pairing by considering annual sentiment trends. *Procedia Computer Science, 179,* 558-565. https://doi.Org/https://doi.org/10.1016/j.procs.2021.01.040

Hasanati, N., Utami, T. S., & Kusumaningtyas, R. H. (2023). Sentiment analysis on news headlines of nation's capital relocation using CNN and SVM. *2023 11th International Conference on Cyber and IT Service Management, CITSM 2023,* (July), 1-6. https://doi.org/10.1109/CITSM60085.2023.10455228

Hashim, R., Omar, B., Saeed, N., Ba-Anqud, A., & Al-Samarraie, H. (2020). The application of sentiment analysis in tourism research: A brief review. *IJBTS International Journal of Business Tourism and Applied Sciences,* 5(1), 51-60. Retrieved from https://www.researchgate.net/publicati on/3 51866678THEAPPLIC ATION OF SENTFMENTANALYSISJNTOURISMRESEARCHABRIEFREVIE W

Herqutanto, M. F., Zatari, R. P., & Sutoyo, R. (2023). Topic modeling using LDA-based and machine learning for aspect sentiment analysis. In *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)* (pp. 142-148). https://doi.org/10.1109/ICFMCIS60089.2023.10349056

Hossain, M. S., Rahman, M. F., Uddin, M. K., & Hossain, M. K. (2023). Customer sentiment analysis and prediction of halal restaurants using machine learning approaches. *Journal of Islamic Marketing, 14(1),* 1859-1889. https://doi.org/10.1108/JIMA-04-2021-0125

Huang, B., Guo, R., Zhu, Y., Fang, Z., Zeng, G., Liu, J., ... Shi, Z. (2022). Aspect-level sentiment analysis with aspect-specific context position information. *Knowledge-Based Systems, 243,* 108473. https://doi.Org/10.1016/j.knosys.2022.108473

Imane, G., Kareem, D., & Faical, A. (2019). A set of parameters for automatically annotating a sentiment Arabic corpus. *International Journal of Web Information Systems, 15(5),* 594-615. https://doi.org/10.1108/IJWIS-03-2019-0008

Inam, G., Ullah, I., Singh, J., & Arumungam, T. (2020). Digital tourism: A possible

revival strategy for Malaysian tourism industry after covid-19 pandemic. *Electronic Journal of Business & Management,* 5(2), 1-17. Retrieved from https://www.researchgate.net/publicati on/3 57031698_Digital_Tourism_A_Possi ble_Revival_Strategy_for_Malaysian_Tourism_Industry_after_COVID-19_Pandemic

Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M., & Moni, M. A. (2024). *Challenges and future in deep learning for sentiment analysis: A comprehensive review and a proposed novel hybrid approach. Artificial Intelligence Review* (Vol. 57). Springer Netherlands. https://doi.org/10.1007/sl0462-023-10651-9

Janechek, J. (2023). Jargon. Retrieved from https://writingcommons.org/section/style/elements-of-style/diction/jargon/

Jardim, S., Mora, C, & Santana, T. (2021). A multilingual lexicon-based approach for sentiment analysis in social and cultural information system data. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)* (pp. 1-6). https://doi.org/10.23919/CISTI52073.2021.9476631

Jayaswal, V. (2020). Performance metrics: Confusion matrix, precision, recall, and Fl score. Retrieved September 14, 2020, from https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-fl-score-a8fe076a2262

Jehangir, B., Radhakrishnan, S., & Agarwal, R. (2023). A survey on named entity recognition — Datasets, tools, and methodologies. *Natural Language Processing Journal,* 3(10), 100017. https://doi.Org/10.1016/j.nlp.2023.100017

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,* <5(January), 100059. https://doi.Org/10.1016/j.nlp.2024.100059

Joshi, A., Dabre, R., Kanojia, D., Li, Z., Zhan, H., Haffari, G., & Dippold, D. (2024). Natural language processing for dialects of a language: A survey. *ArXiv (Cornell University),* 7(1), 1-36. https://doi.org/10.48550/arXiv.2401.05632

Jubair, F., Salim, N. A., Al-Karadsheh, O., Hassona, Y., Saifan, R., & Abdel-Majeed, M. (2021). Sentiment analysis for Twitter chatter during the early outbreak period of COVID-19. In *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 257-262).

https://doi.org/10.1109/ISRITI54043.2021.9702837

Kakde, S. T., Roychowdhury, S., Bhosale, A. T., & Maiti, J. (2024). CPAN chart: A novel customer perception analysis system using natural language processing and attribute control charting. *IEEE Transactions on Engineering Management, 71,* 7609-7622. https://doi.Org/10.l 109/TEM.2024.3375883

Kanade, V. (2022). What is semantic analysis? Definition, examples, and applications in 2022. Retrieved June 16, 2022, from https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-semantic-analysis/amp/

Kanavos, A., Antonopoulos, N., Mohasseb, A., & Mylonas, P. (2023). Analyzing public sentiment towards the Covid-19 pandemic: A Twitter-based sentiment analysis and machine learning approach. In *2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP 2023)* (pp. 1-6). https://doi.org/10.1109/SMAP59435.2023.10255176

Karam, M. W., Imran, A., Alzahrani, A., Alheeti, K. M. A., Abbas, A., & Al-Aloosy, A. A. N. (2022). Dramatic increase in fear-related discussion on Twitter during COVID-19: Analysis, topic modeling and tweets classification. In *2022 International Conference on Electrical Engineering and Sustainable Technologies (ICEEST)* (pp. 1-8). https://doi.org/10.1109/ICEEST56292.2022.10077859

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and research: A systematic literature review through text mining. *IEEE Access, 8,* 67698-67717. https://doi.org/10.1109/ACCESS.2020.2983656

Kartini, Darmawan, A. K., Syah, R. I., & Makruf, M. (2023). Sentiment analysis of social media for Indonesian m-Health PeduliLindungi Mobile-Apps (PLMA) with lexicon-based and support vector machine approach. *2023 IEEE 9th Information Technology International Seminar (ITIS),* 1-7. https://doi.org/10.1109/ITIS59651.2023.10419994

Kaur, G., & Sharma, A. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data, 10(\).* https://doi.org/10.1186/s40537-022-00680-6

Kavitha, K. M., Shetty, A., Abreo, B., D'Souza, A., & Kondana, A. (2020). Analysis and classification of user comments on YouTube videos. *Procedia Computer Science,* 777(2018), 593-598. https://doi.Org/10.1016/j.procs.2020.10.084

Khadija, M. A., Jayanti, I. S. D., Ni'Mah, F. U., & Kartikasari, H. (2023). Big data analysis of social media: A case study of citizens comments of integrated dynamic record information system application (SRIKANDI). In *2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)* (pp. 116-121). https://doi.org/10.1109/ICE3IS59323.2023.10335283

Kharisudin, I., & Masri'an, H. (2022). Topic modeling on WhatsApp user reviews using latent dirichlet allocation. *Scientific Journal of Informatics,* 9(1), 51-62. https://doi.org/10.15294/sji.v9il.34941

Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Information Sciences, 477,* 15-29. https://doi.Org/10.1016/j.ins.2018.10.006

Kim, S. W., & Gil, J. M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Human-Centric Computing and Information Sciences,* 9(1). https://doi.org/10.1186/sl3673-019-0192-7

Kleppen, E. (2023). What is sentiment analysis? Retrieved March 3, 2023, from https://builtin.com/machine-learning/sentiment-analysis

Kodicherla, P., Sathineni, S. R., & Sai, J. S. (2023). Comparative analysis of TextRank and latent semantic analysis algorithms for extractive news summarization. *2023 3rd Asian Conference on Innovation in Technology, ASIANCON 2023,* 1-6. https://doi.org/10.1109/ASIANCON58793.2023.10270050

Krishnan, A., & Kennedyraj. (2023). Exploring the power of topic modeling techniques in analyzing customer reviews: A comparative analysis. *ArXiv (Cornell University),* 13. Retrieved from https://doi.org/10.48550/arXiv.2308.11520

Kristiyanti, D. A., Aulianita, R., Putri, D. A., Utami, L. A., Agustini, F., & Alfianti, Z. I. (2022). Sentiment classification Twitter of LRT, MRT, and Transjakarta transportation using support vector machine. *2022 International Conference of Science and Information Technology in Smart Administration, ICSINTESA 2022,* 143-148. https://doi.org/10.1109/ICSINTESA56431.2022.10041651

Kumar, A. P., Nayak, A., & K, M. S. (2019). A statistical approach to evaluate the efficiency and effectiveness of the machine learning algorithms analyzing sentiments. *2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics, DISCOVER 2019 - Proceedings,* 1-6. https://doi.org/10.1109/DISCOVER47552.2019.9008028

Kurniawan, R., Lestari, F., Batubara, A. S., Nazri, M. Z. A., Rajab, K., & 2021, R. M. P. U.-. (2021). Indonesian lexicon-based sentiment analysis of online religious lectures review. *2021 International Congress of Advanced Technology and Engineering (ICO TEN).* https://doi.Org/https://doi.org/10.1109/icoten52080.2021.9493530

Laifa, M., & Mohdeb, D. (2023). Sentiment analysis of the Algerian social movement inception. *Data Technologies and Applications, 57(5),* 734-755. https://doi.org/10.1108/DTA-10-2022-0406

Lau, C.-Y., Yuen, M.-C, Cheng, W.-F., Lau, M.-Y., Chan, W.-F., & Wong, H.-C. (2022). Intellectual social scanning and analytics platform. In *2022 IEEE 16th International Conference on Big Data Science and Engineering (BigDataSE)* (pp. 13-20). https://doi.org/10.1109/BigDataSE56411.2022.00012

Li, H. (2023). A study of Chinese semantic analysis based on enhanced lexicon. In *2023 4th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)* (pp. 317-322). https://doi.org/10.1109/ICHCI58871.2023.10277988

Liang, M., & Niu, T. (2022). Research on text classification techniques based on improved TF-IDF algorithm and LSTM inputs. *Procedia Computer Science, 208,* 460-470. https://doi.Org/10.1016/j.procs.2022.10.064

Liaqat, M., & Rafique, F. (2024). Slang vs. colloquial — What's the difference? Retrieved March 14, 2024, from https://www.askdifference.com/slang-vs-colloquial/

Lim, M. K., Li, Y., & Song, X. (2021). Exploring customer satisfaction in cold chain logistics using a text mining approach. *Industrial Management & Data Systems, 121(12),* 2426-2449. https://doi.org/10.1108/FMDS-05-2021-0283

Liu, H., Chen, X., & Liu, X. (2022). A study of the application of weight distributing method combining sentiment dictionary and TF-IDF for text sentiment analysis. *IEEE Access, 10,* 32280-32289. https://doi.org/10.1109/ACCESS.2022.3160172

Luo, A. (2019). Content analysis: Guide, methods & examples. Retrieved July 18, 2019, from https://www.scribbr.com/methodology/content-analysis/

Mahajani, S. J., Srivastava, S., & Smeaton, A. F. (2023). A comparison of lexicon-based and ML-based sentiment analysis: Are there outlier words? In *2023 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS)* (pp. 1-4). https://doi.org/10.1109/AICS60730.2023.10470734

Mahavera, S. (2023, July 23). How tourism Malaysia is building back the sector in the post-covid era. *The Star*. Retrieved from https://www.thestar.com.my/news/focus/2023/07/23/how-tourism-malaysia-is-building-back-the-sector-in-the-post-covid-era

Manalu, B. U., Tulus, & Efendi, S. (2020). Deep learning performance in sentiment analysis. In *2020 4rdInternational Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)* (pp. 97-102). https://doi.org/10.1109/ELTICOM50775.2020.9230488

Mandal, S. K., & Bhardwaj, S. (2023). A hybrid classifier approach to analyze public sentiments in social domain. *2023 3rd International Conference on Advancement in Electronics and Communication Engineering, AECE 2023,* 723-726. https://doi.org/10.1109/AECE59614.2023.10428420

Manik, L. P., Susianto, H., Dinakaramani, A., Pramanik, N., & Suhardijanto, T. (2023). Can lexicon-based sentiment analysis boost performances of transformer-based models? In *2023 7th International Conference on New Media Studies (CONMEDIA)* (pp. 314-319). https://doi.org/10.1109/CONMEDIA60526.2023.10428401

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences, 36(4),* 102048. https://doi.org/10.1016/jjksuci.2024.102048

Mbooi, M., Rangata, M. R., & Sefara, T. J. (2024). Topic modelling of short texts in the health domain using LDA and bard. In *2024 Conference on Information Communications Technology and Society (ICTAS)* (pp. 82-87). https://doi.org/10.1109/ICTAS59620.2024.10507116

McCombes, S. (2019). Descriptive research: Definition, types, methods & examples. Retrieved May 15, 2019, from https://www.scribbr.com/methodology/descriptive-research/

Mishra, R. K., Urolagin, S., Jothi, J. A. A., Neogi, A. S., & Nawaz, N. (2021). Deep learning-based sentiment analysis and topic modeling on tourism during covid-19 pandemic. *Frontiers in Computer Science,* ^(November), 1-14. https://doi.org/10.3389/fcomp.2021.775368

Mondal, S., Srinivasan, P., & Ahammed, M. E. (2023). Fake news detection: A comparative study of machine learning techniques. In *2023 International*

Conference on Computational Intelligence, Networks and Security (ICCINS) (pp. 1-6). https://doi.org/10.1109/ICCINS58907.2023.10450148

Morision, N. S. I. M., Dams, N. F. I. M., Aziz, A. A., Zain, R. A., Adzmy, A., & Patah, M. O. R. A. (2023). Digital marketing impact on tourism destination in Malaysia: The role of social media reviews. *Journal of Tourism, Hospitality & Culinary Art (JTHCA),* 75(2), 135-153. Retrieved from https://ir.uitm.edu.my/id/eprint/94909

Motlagh, M. A., Shahhoseini, H. S., & Fatehi, N. (2023). A reliable sentiment analysis for classification of tweets in social networks. *Social Network Analysis and Mining,* 73(1), 1-11. https://doi.org/10.1007/sl3278-022-00998-2

Motz, A., Ranta, E., Calderon, A. S., Adam, Q., Alzhouri, F., & Ebrahimi, D. (2022). Live sentiment analysis using multiple machine learning and text processing algorithms. *Procedia Computer Science, 203,* 165-172. https://doi.Org/https://doi.org/10.1016/j.procs.2022.07.023

Moud, Z. A., Nejad, H. V., & Sadri, J. (2021). Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Systems with Applications,* 767(November 2020), 114324. https://doi.Org/10.1016/j.eswa.2020.114324

Mustika, H. F., Kartika, Y. A., Satya, I. A., Manik, L. P., Syafiandini, A. F., Setiawan, F. A., ... Saleh, D. R. (2020). Measuring the effect of elementary descriptive attributes on news recommender systems. *Proceeding - 2020 International Conference on Radar, Antenna, Microwave, Electronics and Telecommunications, ICRAMET 2020,* 302-307. https://doi.org/10.1109/ICRAMET51080.2020.9298686

Nanayakkara, A. C. (2024). Exploratory analysis of the black lives matter movement's visibility on YouTube. In *2024 4th International Conference on Advanced Research in Computing (ICARC)* (pp. 161-166). https://doi.org/10.1109/ICARC61713.2024.10499730

Nanyonga, A., Wasswa, H., & Wild, G. (2023). Topic modeling analysis of aviation accident reports: A comparative study between LDA and NMF models. *2023 3rd International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2023,* 1-2. https://doi.org/10.1109/SMARTGENCON60755.2023.10442471

Nugroho, D. K. (2021). US presidential election 2020 prediction based on Twitter data using lexicon-based sentiment analysis. In *202111th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 136-141).

https://doi.org/10.1109/Confluence51648.2021.9377201

Nurcahyawati, V., & Mustaffa, Z. (2023). Vader lexicon and support vector machine algorithm to detect customer sentiment orientation. *Journal of Information Systems Engineering and Business Intelligence,* 9(1), 108-118. https://doi.Org/10.20473/jisebi.9.l.108-118

Nurhaliza, A. C. A., Novita, R., Mustakim, & Rozanda, N. E. (2024). The implementation of TF-IDF and Word2Vec on booster vaccine sentiment analysis using support vector machine algorithm. *Procedia Computer Science, 234,* 156—163. https://doi.Org/10.1016/j.procs.2024.02.162

Octava, M. Q. H., Putri, D. G. P., Hilmy, F. M., Farooq, U., Nurhaliza, R. A., & Ganjar, A. (2023). Web-based sentiment analysis system using SVM and TF-IDF with statistical feature. *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2023,* 9-14. https://doi.org/10.1109/3ICT60104.2023.10391734

Ostheimer, S. (2019). 10 best islands for a Malaysia holiday. Retrieved October 6,2019, from https://edition.cnn.com/travel/article/malaysia-best-islands/index.html

Othman, A. K., Abas, M. K. M., Mat, A., Abujarad, I., & Yunus, N. M. (2024). Promoting and improving the tourism industry in Malaysia through tourist satisfaction. *AMH International - Information Management and Business Review, 16(Vo* 16No3S(I)a(2024)), 4-6. https://doi.org/10.22610/imbr.vl6i3S(I)a.4225

Park, S., & Lee, S. (2022). Online customer review platform categorization to identify market needs. In *2022 Portland International Conference on Management of Engineering and Technology (PICMET)* (pp. 1-8). https://doi.org/10.23919/PICMET53225.2022.9882804

Pattnaik, N., Li, S., & Nurse, J. R. C. (2023). Perspectives of non-expert users on cyber security and privacy: An analysis of online discussions on Twitter. *Computers & Security, 125,* 103008. https://doi.Org/https://doi.org/10.1016/j.cose.2022.103008

Polignano, M., Basile, V., Basile, P., Gabrieli, G, Vassallo, M., & Bosco, C. (2022). A hybrid lexicon-based and neural approach for explainable polarity detection. *Information Processing and Management, 59(5).* https://doi.Org/10.1016/j.ipm.2022.103058

Pradhan, R. (2021). Extracting sentiments from YouTube comments. In *2021 Sixth International Conference on Image Information Processing (ICIIP)* (Vol. 6, pp. 1-4). IEEE. https://doi.org/10.1109/ICIIP53038.2021.9702561

Pradhan, S. (2022). A guide to sentiment analysis - Part 1. Retrieved May 4, 2022, from https://www.nitorinfotech.com/blog/a-guide-to-sentiment-analysis-part-l/

Pranayteja. (2023). YouTube comments sentiment analysis using YouTube data API v3. Retrieved September 7, 2023, from https://medium.com/@pranayteja270/youtube-comments-sentiment-analysis-using-youtube-data-api-v3-bf4a2a041144

Prastiani, Fakhrurroja, H., & Hamami, F. (2023). Social media sentiment analysis for local water company customers using a support vector machine algorithm. *10th International Conference on ICTfor Smart Society, ICISS 2023 - Proceeding,* 1-6. https://doi.org/10.1109/ICISS59129.2023.10291991

Pratama, B. T., Utami, E., & Sunyoto, A. (2019). The impact of using domain specific features on lexicon based sentiment analysis on Indonesian app review. In *2019 International Conference on Information and Communications Technology (ICOIACT)* (pp. 474-479). https://doi.org/10.1109/ICOIACT46704.2019.8938419

Puh, K., & Babac, M. B. (2023). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights, 6(3),* 1188-1204. https://doi.org/10.1108/JHTI-02-2022-0078

Puri, A., Agrawal, G., Dukare, A., & Jawale, M. (2023). A survey and analysis of textual content based on exploratory data analysis technique and opinion analysis. In *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 1-6). https://doi.org/10.1109/ICCAKM58659.2023.10449608

Putra, M. M. D., Maki, W. F. Al, & Romadhony, A. (2021). Sentiment analysis on marketplace review using hybrid lexicon and SVM method. *2021 9th International Conference on Information and Communication Technology, ICoICT 2021,* (3), 66-70. https://doi.org/10.1109/ICoICT52021.2021.9527408

Raccuia, K. (2024). Langkawi vs Perhentian islands: An honest comparison to help you choose! Retrieved May 1, 2024, from https://sandinmycurls.com/short-getaway-in-malaysia-langkawi-tioman-perhentian/

Ramadhan, F. A., & Gunawan, P. H. (2023). Sentiment analysis of 2024 presidential candidates in Indonesia: Statistical descriptive and logistic regression approach. *2023 International Conference on Data Science and Its Applications, ICoDSA 2023,* 327-332. https://doi.org/10.1109/ICoDSA58501.2023.10276417

Raman, D. R., Jayalakshmi, S., Arumugam, K., Raj, A. V. A., Balaji, D., & Brightsingh, R. (2022). Implementation of data analysis and document summarization in social media data using R and Python. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1457-1464). https://doi.org/10.1109/ICIRCA54612.2022.9985479

Ramanathan, V., Hajri, H. Al, & Ruth, A. (2024). Conceptual level semantic sentiment analysis using Twitter data. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems, ADICS 2024,* 1-8. https://doi.org/10.1109/ADICS58448.2024.10533498

Ramdhani, Y., Mustofa, H., Topiq, S., Alamsyah, D. P., Setiawan, S., & Susanti, L. (2022). Sentiment analysis Twitter based lexicon and multilayer perceptron algorithm. In *2022 10th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-6). https://doi.org/10.1109/CITSM56380.2022.9936029

Ravisankar, S., Marimuthu, A., & Praveen, K. (2023). Elevating movie reviews with intelligent recommendation systems. *2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2023,* 1-6. https://doi.org/10.1109/ICDSAAI59313.2023.10452615

Rawat, A. S. (2021). An overview of descriptive analysis. Retrieved March 31, 2021, from https://www.analyticssteps.com/blogs/overview-descriptive-analysis

Rawat, C, Sarkar, A., Singh, S., Alvarado, R., & Rasberry, L. (2019). Automatic detection of online abuse and analysis of problematic users in Wikipedia. In *2019 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 1-6). https://doi.org/10.1109/SIEDS.2019.8735592

Ridzuan, A. R., Ann, L. C, Razak, M. I. M., Akhwan, M. H., Ridzuan, M., & Zarin, N. I. (2024). Langkawi's ecological and economic renaissance: A study of blue and green opportunities. *ASEAN Journal on Hospitality and Tourism,* 22(1), 73-87. https://doi.org/10.5614/ajht.2024.22.L05

Rouhani, S., & Abedin, E. (2020). Crypto-currencies narrated on tweets: a sentiment analysis approach. *International Journal of Ethics and Systems, 36(1),* 58-72. https://doi.org/10.1108/IJOES-12-2018-0185

Sadiq, M. T., & Saji, K. M. (2022). The disaster of misinformation: A review of research in social media. *International Journal of Data Science and Analytics, 13(4),* 271-285. https://doi.org/10.1007/s41060-022-00311-6

Sahu, A. K. (2020). YouTube sentiment analysis using Python. Retrieved October 9, 2020, from https://medium.com/analytics-vidhya/youtube-sentiment-analysis-using-python-7b4dcb6907b0

Saini, A. (2024). Guide on support vector machine (SVM) algorithm. Retrieved May 22, 2024, from https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/

Salim, S., & Tan, C. (2023). Malaysia's 2024 tourist arrivals expected to surpass pre-Covid-19 levels - Tourism Malaysia. Retrieved December 6, 2023, from https://theedgemalaysia.com/node/692906

Sano, A. V. D., Stefanus, A. A., Madyatmadja, E. D., Nindito, H., Purnomo, A., & Sianipar, C. P. M. (2023). Proposing a visualized comparative review analysis model on tourism domain using Naive Bayes classifier. *Procedia Computer Science, 227,* 482-489. https://doi.Org/10.1016/j.procs.2023.10.549

Satrya, R. N., Pratiwi, O. N., Farifah, R. Y., & Abawajy, J. (2022). Cryptocurrency sentiment analysis on the Twitter platform using support vector machine (SVM) algorithm. *Proceedings - International Conference Advancement in Data Science, E-Learning and Information Systems, ICADEIS 2022,* 2-6. https://doi.org/10.1109/ICADEIS56544.2022.10037413

Schoene, A., & Mel, G. de. (2019). Pooling Tweets by Fine-Grained Emotions to Uncover Topic Trends in Social Media. In *2019 22th International Conference on Information Fusion (FUSION)* (pp. l-7). https://doi.org/10.23919/FUSION43075.2019.9011265

Segarra, E., Guapi, F., Brito, L. M. Y., & Lopez, C. (2024). The impact of tourism on islands: A sustainability and conservation analysis. *Green World Journal,* 7(2), 147-147. https://doi.org/10.53313/gwj72147

Semary, N. A., Ahmed, W., Amin, K., Plawiak, P., & Hammad, M. (2024). Enhancing machine learning-based sentiment analysis through feature extraction techniques. *PLoSONE, 19(1* February), https://doi.org/10.1371/journal.pone.0294968

Shah, P. K., & Bhuvana, J. (2023). Performing semantic analysis of short descriptive answers using NLP. *2023 International Conference on Advances in Computation, Communication and Information Technology, ICAICCIT 2023,* 118-122. https://doi.org/10.1109/ICAICCIT60255.2023.10465750

Sham, N. M., & Mohamed, A. (2022). Climate change sentiment analysis using lexicon, machine learning and hybrid approaches. *Sustainability (Switzerland), 14(%),* 1-

28. https://doi.org/10.3390/sul4084723

Shamisavi, M., & Jahanshahi, A. (2022). Forecasting tehran stock exchange trend with time series analysis, fundamental data, and sentiment analysis in news. *2022 30th International Conference on Electrical Engineering, ICEE 2022,* 1-7. https://doi.org/10.1109/ICEE55646.2022.9827232

Sharma, A., & Ghose, U. (2021). Lexicon a linguistic approach for sentiment classification. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 887-893). https://doi.org/10.1109/Confluence51648.2021.9377057

Sharma, R., Lowe, D., Somani, K., & Galhotra, B. (2022). A study on lexicon based techniques of Twitter sentiment analysis. *8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022, 1,* 562-566. https://doi.org/10.1109/ICACCS54159.2022.9785231

Shekhar, S., Akanksha, & Saini, A. (2021). Utilizing topic modelling to identify abusive comments on YouTube. In *2021 International Conference on Intelligent Technologies (CONIT)* (pp. 1-5). https://doi.org/10.1109/CONIT51480.2021.9498368

Shi, X. (Crystal), & Chen, Z. (2021). Listening to your employees: analyzing opinions from online reviews of hotel companies. *International Journal of Contemporary Hospitality Management, 33(6),* 2091-2116. https://doi.org/10.1108/IJCHM-06-2020-0576

Silviana, L., Nababan, E. B., & Zarlis, M. (2023). Long short-term memory and word embedding for sentiment analysis of user review. In *2023 1st IEEE International Conference on Smart Technology (ICE-SMARTec)* (pp. 66-71). https://doi.org/10.1109/ICE-SMARTECH59237.2023.10461949

Singh, H., & Srivastava, D. (2023). Sentiment analysis: Quantitative evaluation of machine learning algorithms. *Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023,* (Icssit), 946-951. https://doi.org/10.1109/ICSSIT55814.2023.10061130

Singh, P. K., Sharma, S., & Paul, S. (2020). Identifying hidden sentiment in text using deep neural network. In *2nd International Conference on Data, Engineering and Applications (IDEA)* (pp. 1-5). https://doi.org/10.1109/IDEA49133.2020.9170726

Singh, R., & Tiwari, A. (2021). Youtube comments sentiment analysis. *International*

Journal of Scientific Research in Engineering and Management (IJSREM, 5(May), 5. Retrieved from https://www.researchgate.net/publication/351351202

Sinha, S., Jayan, A., & Kumar, R. (2022). An analysis and comparison of deep-learning techniques and hybrid model for sentiment analysis for movie review. *2022 3rd International Conference for Emerging Technology, INCET 2022,* 1-5. https://doi.org/10.1109/INCET54531.2022.9824630

Song, C, Zheng, L., & Shan, X. (Gene). (2022). An analysis of public opinions regarding Internet-famous food: a 2016-2019 case study on Dianping. *British Food Journal, 124(12),* 4462-4476. https://doi.org/10.1108/BFJ-05-2021-0510

Srivastava, R., Bharti, P. K., & Verma, P. (2022). Comparative analysis of lexicon and machine learning approach for sentiment analysis. *International Journal of Advanced Computer Science and Applications, 13(3),* 71-77. https://doi.org/10.14569/IJACSA.2022.0130312

Statista. (2023). Tourism in Malaysia - statistics & facts. Retrieved December 21, 2023, from https://www. statista.com/topics/5741/travel-and-tourism-in-malaysia/#topicOverview

Su, V., & Thakur, N. (2025). COVID-19 on YouTube: A data-driven analysis of sentiment , toxicity , and content recommendations. *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC),* 715-723. https://doi.org/10.1109/CCWC62904.2025.10903713

Sudiro, R. V. O. I., Prasetiyowati, S. S., & Sibaroni, Y. (2021). Aspect based sentiment analysis with combination feature extraction LDA and word2vec. *2021 9th International Conference on Information and Communication Technology, ICoICT2021,* 611-615. https://doi.org/10.1109/ICoICT52021.2021.9527506

T, R., Pradeep, S. D., & R, V. K. (2023). Japanese language review mining using translators, word embedding and ML techniques. In *2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS)* (pp. 171-176). https://doi.org/10.1109/ICRAIS59684.2023.10367133

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences (Switzerland), 13(1).* https://doi.org/10.3390/appl3074550

Tetteh, M., & Thushara, M. (2023). Sentiment analysis tools for movie review evaluation - A survey. In *2023 7th International Conference on Intelligent*

Computing and Control Systems (ICICCS) (pp. 816-823). https://doi.org/10.1109/ICICCS56967.2023.10142834

Thakkar, A., & Chaudhari, K. (2020). Predicting stock trend using an integrated term frequency-inverse document frequency-based feature weight matrix with neural networks. *Applied Soft Computing Journal, 96,* 106684. https://doi.Org/10.1016/j.asoc.2020.106684

Tharim, A. H. A., Sayuthy, A. I. A., & Wahi, N. (2024). An assessment on the impact of physical development of Perhentian island, Terengganu, Malaysia toward the quality of life of its residents. *Planning Malaysia, 22(2),* 428-441. https://doi.org/10.21837/pm.v22i31.1480

Tho, C, Heryadi, Y., Kartowisastro, I. H., & Budiharto, W. (2021). A comparison of lexicon-based and transformer-based sentiment analysis on code-mixed of low-resource languages. In *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)* (Vol. 1, pp. 81-85). https://doi.org/10.1109/ICCSAI53272.2021.9609781

Trivedi, S. K., & Singh, A. (2021). Twitter sentiment analysis of app based online food delivery companies. *Global Knowledge, Memory and Communication,* 70(8-9), 891-910. https://doi.org/10.1108/GKMC-04-2020-0056

Uddin, M. R., Prova, N. S., Jawad, A., Rayhan, M. R., Arnab, S. M. R., & Iqbal, S. M. H. S. (2023). Qualitative sentiment analysis of YouTube contents based on user reviews. In *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 1099-1102). https://doi.org/10.1109/ICAAIC56838.2023.10141517

Verma, R., & Thakur, B. (2023). Machine learning techniques for the prediction of cyber-attacks. *Proceedings - 4th IEEE 2023 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2023,* 978-985. https://doi.org/10.1109/ICCCIS60361.2023.10425542

Vipin, C, Harshit, S., Kumar, K. L. S., Vijay, K., & Zaid, M. (2023). Checking the truthfulness of news channels using NLP techniques. In *2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM)* (pp. 1-5). https://doi.org/10.1109/IC-RVITM60032.2023.10435241

Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model. *IEEE Access, 8,*

138162-138169. https://doi.org/10.1109/ACCESS.2020.3012595

Waspodo, B., Aini, Q., Singgih, F. R., Kusumaningtyas, R. H., & Fetrina, E. (2022). Support vector machine and lexicon based sentiment analysis on kartu prakerja (Indonesia pre-employment cards government initiatives). In *2022 10th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-6). https://doi.org/10.1109/CITSM56380.2022.9935990

Widiantoro, A. D., Wibowo, A., & Harnadi, B. (2021). User sentiment analysis in the Fintech OVO review based on the lexicon method. In *2021 Sixth International Conference on Informatics and Computing (ICIC)* (pp. 1-4). https://doi.org/10.1109/ICIC54025.2021.9632909

Yenkikar, A., Bali, M., Patil, R. R., Mirajkar, R., & Ara, T. (2024). Biomedical named entity recognition through spaCy: A visual exploration. In *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)* (pp. 17-22). https://doi.org/10.1109/InCACCT61598.2024.10551087

Yusoh, M. P., Hanafi, N., Latip, N. A., Steaphen, J., Abdullah, M. F., & Hua, A. K. (2023). The relevance of scuba diving activities as a tourist attraction on Pangkor island. *Planning Malaysia, 21(4),* 217-232. https://doi.org/10.21837/pm.v21i28.1328

Zaman, Q. A. B. K., Yusoff, W. N. S. B. W., & Shah, Q. B. B. A. (2024). Sentiment analysis on the place of interest in Malaysia. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 43(1),* 54-65. https://doi.org/10.37934/araset.43.L5465

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2023). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering,* 35(11), 11019-11038. https://doi.org/10.1109/TKDE.2022.3230975

Zhao, F., Yao, Z., Luan, L, & Liu, H. (2020). Inducing stock market lexicons from disparate Chinese texts. *Industrial Management & Data Systems, 120(3),* 508-525. https://doi.org/10.1108/FMDS-04-2019-0254

Zhu, Z. (2024). Systematic optimization of overfitting problem in machine learning. *Highlights in Science, Engineering and Technology, 111,* 353-359. https://doi.org/10.54097/3tkzrj84

AUTHOR'S PROFILE

Nurain Batrisyia Binti Bidin will complete her Master's studies in Computer Science at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA.

LIST OF PUBLICATIONS

Bidin, N. B. B., Samah, K. A. F. A., & Aminuddin, R. B. (2024). Conceptualizing hybrid descriptive-semantic techniques to enhance sentiment analysis of YouTube reviews: Malaysian island contexts. In *Proceedings of the IEEE Conference on Systems, Process and Control, ICSPC* (pp. 417-422). IEEE. https://doi.org/10.1109/ICSPC63060.2024.10862383

Bidin, N. B., Samah, K. A. F. A., Abidin, N. A. B. Z., & Riza, L. S. (2025). Exploring enhancement in sentiment analysis using descriptive-semantic techniques: A systematic literature review. *Journal of Theoretical and Applied Information Technology*, 703(13), 4819-4837. Retrieved from https://www.j atit.org/volumes/Vol 103No 13/26 Vol 103No 13 .pdf

Bidin, N. B. B., Samah, K. A. F. A., & Kamarudin, N. 'Azwa. (2025). Hybrid sentiment analysis of Malaysian islands : A descriptive-semantic approach using TF-IDF and LDA. In *2025 12th IEEE Symposium on Industrial Electronics & Applications.* IEEE Malaysia Industrial Electronics & Industrial Applications Joint Chapter. https://doi.org/10.1109/ISIEA65768.2025.11138181