

Universiti Teknologi MARA

**Sentiment Analysis of Hotel Reviews
Using Convolutional Neural Network**

Sofea Aini Bt. Mohd Sufian

**Thesis submitted in fulfilment of the requirements
for Bachelor of Computer Science (Hons.)**

February 2021

ACKNOWLEDGMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this Final Year Project (FYP) within the time duration given. Firstly, my special thanks go to my supervisor, Associate Professor Ts. Dr Hamidah bt Jantan, who help me a lot of things in completing projects and gives ideas improving the quality of my project.

Special appreciation also goes to my beloved parents, Mohd Sufian bin Abdul Kadir and Mazlida Aini bt Ghazali, who are always support and give me inspire to finishing my final year project. Not forgetting to UiTM Kuala Terengganu especially Faculty of Science Computer and Mathematics for giving me chance to study here.

Last but not least, I would like to give my gratitude to my dearest friend and lecturers who involved directly or indirectly support in completing this project especially my friends that under same supervision group with me. I hope this project help me to improve my knowledge in computer science and algorithm and practice as a programmer.

ABSTRACT

Tourism industry is one of the largest sectors and also one of the reasons to the development of hotels around the world. Most of them also use online booking system. Many online reviews are increasing every day because travellers are willing to share their experiences. This report presents the worked progress of sentiment analysis on hotel reviews using convolutional neural network. It is implemented to help users which is hotel's staff that want to know more about their hotels from the reviews. The users face problems while reading the reviews because it consumes a lot of times and expressions of the sentences that are hard to understand whether it is positive or negative opinions since human language is complex. The traditional algorithm that been used a lot also reach a plateau of performance, with no room for improvement. So, the aim of this paper is to develop a prototype for hotel reviews using CNN Algorithm, to identify the requirements of CNN technique for text classification and also to evaluate the accuracy of CNN algorithm. The hotel reviews will be analysed using CNN text classification whether it is positive or negative sentiment. The experiment was doing by replacing the batch size, filter size, dropout and also the split of the data between training and testing numbers. It also used word2vec as word embedding. The result shows that CNN algorithm method can have high accuracy with more than 90% accuracy. The prototype also been implemented using CNN model to predict the sentiment on hotel review. In conclusion, the CNN algorithm can be used as text classification as it gives a high accuracy.

Keywords

Hotel prediction, Convolutional Neural Network (CNN), hotel reviews, sentiment analysis

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	x

CHAPTER ONE: INTRODUCTION

1.1 Background of Study	1
1.2 Problem Statement	3
1.3 Objective	5
1.4 Project Scope	5
1.5 Project Significance	6
1.6 Overview of Research Framework	7
1.7 Conclusion	8

CHAPTER TWO: LITERATURE REVIEW

2.1. Tourism Industries in Malaysia	9
2.2. Sentiment Analysis	10
2.2.1. The Use of Sentiment Analysis	11
2.2.2. Phases of Sentiment Analysis	12
2.2.2.1. Information Collection Process	12
2.2.2.2. Learning Phase	12
2.2.2.3. Prediction/Recommendation Phase	13
2.2.3. Sentiment Analysis Approaches & Techniques	14
2.2.3.1. Supervised Learning	14

2.2.3.2.	Unsupervised Learning	15
2.2.3.3.	Semi-Supervised Learning	15
2.2.3.4.	Lexicon-Based Approach	16
2.2.3.5.	Hybrid Approach	16
2.2.4.	Advantages and Disadvantaged of Sentiment Analysis	16
2.3.	Hotel Prediction	18
2.3.1.	The Prediction on Hotel Reviews	18
2.3.2.	Problems with Hotel Prediction	19
2.3.3.	Benefit of Hotel Prediction Existence	19
2.4.	Convolutional Neural Network Algorithm	19
2.4.1.	The Methods for CNN	20
2.4.2.	Implementation CNN Algorithm in Various Problems	21
2.5.	Application of Sentiment Analysis in Tourism Industries	23
2.6.	The Implication of Literature Reviews	25
2.7.	Conclusion	26

CHAPTER THREE: METHODOLOGY

3.1.	Overview of Research	27
3.1.1.	Detailed of Research Methodology Framework	27
3.2.	Preliminary Study	28
3.2.1.	Literature Analysis	28
3.3.	Data Collection and Understanding	30
3.3.1.	Data Collection	29
3.3.2.	Data Preparation	30
3.3.2.1.	Lowercase	30
3.3.2.2.	Stopwords	31
3.3.2.3.	Tokenization	32
3.3.2.4.	Lemmatization	32
3.3.2.5.	Stemming	33
3.4.	Design Phase	34
3.4.1.	System Architecture	34
3.4.2.	Flow Chart for CNN algorithm on Data	35