

Universiti Teknologi MARA

**Sentiment Analysis on National Cultural Tourism
using Linear Support Vector Machine (LSVM)**

Nur Haida Hanna binti Samsuddin

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

August 2020

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty and His utmost blessings, I was able to finish this research within the time duration given. Firstly, my special thanks go to my supervisor, Puan Norizan Mohamad who has guided me from the start of my journey and helped me a lot about the project I had done. Special appreciation also goes to my beloved parents that put a lot of trust in me doing the Final Year Project, supporting me whenever I faced some problems. Last but not least, I would like to give my gratitude to my dearest friend that always supports me in completing this project and shared knowledge about related studies in this project.

ABSTRACT

Nowadays, sentiment analysis plays a big role for many industries especially it is something related with feedback or reviews from people in cyberspace. People reviewed some products, places and others by expressing their opinion or emotion into sentences. This leads to the problem of understanding the meaning behind the texts and difficult to discover the sentiment polarity of certain words. Cultural tourism in Malaysia is lacking in terms of their promotional activities and any related authorities in the tourism industry often overlooked the reviews from tourists about cultural heritage destinations. Moreover, negative reviews may impact the national tourism. This study will perform sentiment analysis on national cultural tourism of tourists reviews on TripAdvisor website. The study will identify sentiment analysis tasks based on classification model. A classifier will be designed and developed which is Linear Support Vector Machine (LSVM). Lastly, the accuracy of the proposed classifier will be tested. Therefore, the chosen technique is classification and the algorithm that will be applied in the classification process is Linear Support Vector Machines (LSVM). The output will be the accuracy of the LSVM model and the visualization of sentiment analysis of new data that user will choose in the prototype. The accuracy achieved from the project is 80%. The classifier is claimed to be bad classifier because AUC-ROC gained from the experiment is 0.5. In future, it is recommended to experiment with different algorithm or kernel of Support Vector Machine. The volume of data should be large as it can generate better result of classification method.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iv
ACKNOWLEDGEMENT	v
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	xi
LIST OF TABLES	xiii
LIST OF ABBREVIATIONS	xiv

CHAPTER ONE: INTRODUCTION

1.1	Background of Study	1
1.2	Problem Statement	3
1.3	Objectives	4
1.4	Scope	4
1.5	Significance of Study	4
1.6	Project Framework	5
1.7	Summary	6

CHAPTER TWO: LITERATURE REVIEW

2.2	National Cultural Tourism	7
2.2.1	An Overview of National Cultural Tourism Issues	7
2.2.2	Related Issues About Cultural Heritage Tourism	9
2.2.3	Related Works on National Cultural Tourism	10
2.3	Sentiment Analysis	12
2.3.1	An Overview of Sentiment Analysis	12
2.3.2	Sentiment Analysis Algorithm	14

2.3.3	Related Studies and Issues on Sentiment Analysis	15
2.4	Linear Support Vector Machines (LSVM)	17
2.4.1	An Overview of Support Vector Machines (SVM)	17
2.4.2	Basic Process Flow of SVM	20
2.4.3	Advantages of Linear Support Vector Machines	21
2.4.4	Related Studies of Support Vector Machine (SVM)	22
2.5	Applications of Support Vector Machines (SVM) in Sentiment Analysis	23
2.5.1	Aspect Term Extraction For Sentiment Analysis In Large Movie Reviews Using Gini Index Feature Selection Method And SVM Classifier (Manek et al., 2017)	23
2.5.2	Sentiment Analysis for Bangla Microblog Posts (Chowdhury & Chowdhury, 2013)	24
2.5.3	Sentiment Analysis on Punjabi News Articles Using SVM (Kaur & Kaur, 2015)	25
2.5.4	A Study using Support Vector Machines to Classify the Sentiments of Tweets (Wassim A. & Aziz M., 2017)	25
2.5.5	Sentiment analysis using Support Vector Machine (Zainuddin & Selamat, 2014)	26
2.5.6	Discussion of Comparison of Similar Studies	27
2.6	Conclusion	29

CHAPTER THREE: METHODOLOGY

3.1	Introduction	30
3.2	Project Methodology Overview	31
3.3	Analysis Phase	35
3.3.1	Data Collection and Pre-processing	35
3.4	Design Phase	40
3.4.1	Sentiment Analysis on National Cultural Tourism using LSVM Logical Design	40
3.4.2	Linear Support Vector Machines (LSVM)	42
3.4.3	Proposed System Layout	46
3.5	Implementation Phase	48