SENTIMENT ANALYSIS FOR SOCIAL MEDIA BY USING SVM

TANG LI PING

FAKULTI SENI, KOMPUTERAN & INDUSTRI KREATIF
UNIVERSITI PENDIDIKAN SULTAN IDRIS

2023

SENTIMENT ANALYSIS FOR SOCIAL MEDIA BY USING SVM

TANG LI PING

LAPORAN PROJEK TAHUN AKHIR DIKEMUKAKAN BAGI MEMENUHI
SYARAT UNTUK MEMPEROLEH IJAZAH SARJANA MUDA
KEJURUTERAAN PERISIAN (PERISIAN PENDIDIKAN) DENGAN KEPUJIAN

FAKULTI SENI, KOMPUTERAN DAN INDUSTRI KREATIF
UNIVERSITI PENDIDIKAN SULTAN IDRIS

2023

FAKULTI SENI, KOMPUTERAN DAN INDUSTRI KREATIF

PERAKUAN KEASLIAN PENULISAN

Nama Pelajar:	Tang Li Ping
No. Pendaftaran:	D20191086980
Nama Ijazah:	Sarjana Muda Kejuruteraan Perisian (Perisian Pendidikan) dengan Kepujian
Bidang Pengkhususan:	Teknologi Maklumat / Multimedia / Reka Bentuk Berkomputer
Tajuk Projek:	SENTIMENT ANALYSIS FOR SOCIAL MEDIA BY USING SVM

Saya sahkan bahawa segala bahan yang terkandung dalam laporan projek tahun akhir ini adalah hasil usaha saya sendiri. Sekiranya terdapat hasil kerja orang lain atau pihak lain sama ada diterbitkan atau tidak (seperti buku, artikel, kertas kerja, atau bahan dalam bentuk yang lain seperti rakaman audio dan video, penerbitan elektronik atau Internet) yang telah digunakan, saya telah pun merakamkan pengikhtirafan terhadap sumbangan mereka melalui konvensyen akademik yang bersesuaian. Saya juga mengakui bahawa bahan yang terkandung dalam laporan projek tahun akhir ini belum lagi diterbitkan atau diserahkan untuk program atau diploma/ijazah lain di mana-mana universiti.

22 February 2023

Tarikh

Tandatangan Pelajar
(TANG LI PING)

Perakuan Penyelia:

Saya akui bahawa saya telah membaca karya ini dan pada pandangan saya karya ini adalah memadai dari segi skop dan kualiti untuk tujuan penganugerahan Ijazah Sarjana Muda Pendidikan (Teknologi Maklumat / Multimedia / Reka Bentuk Berkomputer) dengan Kepujian.

25 February 2023

Tarikh

Tandatangan Penyelia

(Dr. Suliana Sulaiman)

ACKNOWLEDGEMENT

I would like to express my very great appreciation to several individuals for supporting me throughout this project. Firstly, I would like to express my sincere gratitude for the assistance given by my supervisor, Dr Suliana binti Sulaiman, for her insightful advice, constructive comments, valuable ideas and patience guidance that have helped me significantly at all times in my research and writing of this thesis. I am very appreciative of her willingness to spend her time so bountifully to discuss and review my work.

I also wish to express my special thanks to Madam Asma Hanee binti Ariffin and Sir Ahmad Nurzid bin Rosli, who are my coordinators in the final year project. Advices given by them during early preparation and planning for this project helped me a lot in completing this report. I would also extend my sincere thanks to my friends and coursemates, who support and help each other to get through all the challenges and difficulties. Finally, I am particularly grateful to my family members for their support and encouragement. I acknowledge my deepest sense of gratitude to them, their invaluable support has been more than sufficient to sustain me throughout my study.

ABSTRACT

This project attempts to assist educators in analysing the sentiment of Malay social media posts. The output from the sentiment can be used to enhance their teaching and learning activities. In this project, training and testing data was acquired from Husein in 2018, the Malay Stopwords List that used in data preprocessing stage was based on the research of Fatimah Ahmad (1995). All datasets need to be prepared using preprocessing, including tokenization, stop word removal, lower casing, removing numbers, and removing punctuations. Then the TF-IDF vectorization method was used. In this project, we implemented Support Vector Machine (SVM). The performance of trained models were evaluated using Confusion Matrix and Evaluation Matrix. From the experiment this project tends to produce 93% accuracy, 92% for prediction and 92% for recall.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Sentiment Analysis has become a popular topic in Web 2.0, a generation that enhances user-generated content, sharing of information, and participatory culture in social media platforms. This chapter will elaborate what Sentiment Analysis is, where we can use this technique, how Sentiment Analysis helps us, who will be benefited by this technique, and why it was chosen as a research topic. This project proposed a technique to analyse sentiment in UPSI confession by using Support Vector Machine (SVM).

1.2 Research Background

Nowadays, Social Media has become a huge data resource, millions of people sharing their life, expressing their opinions, and exchanging their ideas with countless persons through Social Media platforms, and creating gigantic data flow every day. Data that is created in daily operations have become valuable resources in 21 century; it contains unbounded potential in finding innovative services, new business opportunities, and market strategies. However, manually transferring such an enormous amount of unstructured data into useful information requires mass human effort, cost and time. To overcome the text data analysis problem, Semantic Analysis was applied.

Semantic Analysis is a technology that can automatically evaluate people's opinions toward different aspects, including services, products, organisations, and events. It is a branch of Artificial Intelligence (Natural Language Processing) and Machine Learning. Basically, Semantic Analysis takes text as input, analyzes and determines whether the input text is positive or negative. The analysed results of Semantic Analysis act as a significant indicator in decision-making and improving certain services or products. Hence, the Semantic Analysis technique is widely used by government and big organisations to understand sentiment of people and users toward certain topics.

Support Vector Machine (SVM) is a machine learning algorithm that is widely used in classification and regression. There are 3 main branches of Sentiment Analysis: machine-learning based, lexicon-based and corpus-based (Luo et al., 2016). As a machine-learning based method, performance of SVM is relying on quantity and quality of training dataset. Therefore, data processing is critical in machine learning based Semantic Analysis. SVM will transform the input data by using 'Kernel' mathematical functions, and draw a hyperplane (a decision boundary/surface) to perform classification. There are different types of Kernel, such as linear, non-linear, RBT, and polynomial. In this project, only linear classifiers are involved (positive and negative), hence it will be a linear SVM method.

1.3 Problem Statement

In the education field, educators are always seeking methods to discover the true feelings and feedback of learners. True and honest feedback from learners is significant for educators to adjust teaching plans, hence improve the quality of teaching and learning. However, to analyse students' feedback manually will cause heavy workload to educators (Zhai et al., 2020). Absence of specific tools in analysing students' tendency, behaviour and opinion result in low efficiency in analysing students' feelings.

Besides learning evaluation, social media is another essential platform to understand students' opinions. Posts and comments written by students in social media are useful in revealing issues or problems happening during the learning process. Such findings may help educators to detect problems and solve them as soon as possible, prevent social media crises and protect the school's reputation. Nevertheless, tracking and analysing big amounts of unstructured text data from social media require huge human effort and high administrative cost (Seki, 2016).

The Semantic Analysis system proposed in this work aims to reduce effort and time in social media data collection, and provide an automated way to analyse opinions of students in social media platforms.

1.4 Research Objective

- To analyse how students express their sentiment using social media.
- To develop sentiment analysis for social media dataset using SVM (Support Vector Machine) technique.
- To evaluate the sentiment analysis for social media dataset using SVM (Support Vector Machine) technique.

1.5 Research Questions

1. How students express their sentiment in social media in an educational environment.
2. How efficient SVM (Support Vector Machine) machine learning algorithm in classifying semantic polarity for social media content.
3. How to develop sentiment analysis for social media.

1.6 Research Scope

- Educators who want to understand more about students from time to time.
- Posts and comments retrieval from social media.
- Social media platform.
- Limit to text content only.

1.7 Significance of the Study

This study proposed a method to analyse sentiment polarity of students based on their expression in social media to enhance teaching and learning process, including the machine learning algorithm used and system required in carry out Sentiment Analysis. Semantic Analysis has become a new trend for teaching evaluation in the educational field (Zhai et al., 2020), it enables educators to have deeper understanding on current behaviour, opinion and idea of students efficiently, refining their way of teaching and communicating with students. Application of software tools and Semantic Analysis techniques in collecting students' posts and comments can reduce time, human effort and administrative cost.

1.8 Definitions, Acronyms and Abbreviations

Table 1: Definitions, Acronyms and Abbreviations

Term	**Definition**
NLP	Natural Language Processing. A computational technique to analyse, describe, and understand text naturally as human. (Fitri et al., 2019)
Semantics Analysis	Analysis of opinions from text data with the goal of classifying the degree of negative or positive of the text. (Hollander et al., 2016)
Machine Learning	A branch of Artificial Intelligence that imitates the human learning process, aims to gradually increase accuracy of prediction with the help of data and algorithms, without following explicit instructions.
SVM	Support Vector Machine, a machine learning algorithm that is widely used in classification and regression.

1.9 Summary

The proposed Semantics Analysis system focuses on the ability to analyse semantic polarity (negative or positive) of social media text data posted by students. This system aims to analyse the opinion and expression of students and provide reliable reference for educators in improving teaching quality. Traditional learning evaluation method is low efficient and costly, hence the application of Semantics Analysis technology is essential in increasing the performance of the analysis process.