Universiti Teknologi MARA

Visualizing the Reputation of Malaysian Communication Service Providers Through Twitter Sentiment Analysis Using Naïve Bayes

Aidil Amirul Safwan bin Abdullah Sani

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

January 2020

SUPERVISOR APPROVAL

VISUALIZING THE REPUTATION OF MALAYSIAN COMMUNICATION SERVICE PROVIDERS THROUGH TWITTER SENTIMENT ANALYSIS USING NAÏVE BAYES

By

AIDIL AMIRUL SAFWAN BIN ABDULLAH SANI
2016782377

This thesis was prepared under the supervision of the project supervisor, Dr Khyrina Airin Fariza binti Abu Samah. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science.

Approved by

……..
Dr. Khyrina Airin Fariza binti Abu Samah
Project Supervisor

JANUARY 17, 2020

STUDENT DECLARATION

I certify that this thesis and the project to which it refers is the product of my own work and that any idea or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

……
AIDIL AMIRUL SAFWAN BIN ABDULLAH SANI
2016782377

JANUARY 17, 2020

ABSTRACT

A text classifier model optimized for short snippets like tweets is developed to make bilingual sentiment analysis possible. The two languages explored are Bahasa Malaysia and English, since they are the two most commonly spoken languages in Malaysia. The classifier model is trained and tested on a huge multidomain dataset pre-labelled with the labels “0” and “1”, which resemble “positive” and “negative” respectively. Naïve Bayes ML technique is used as the core of the classifier model. The data are all pre-processed, and once the development of the classifier model is done, the model is run on real-time data, which are public tweets directly or indirectly mentioned to the three biggest CSP in Malaysia, which are Celcom, Maxis and Digi in the year of 2018. The result of the analysis is incorporated into a web application built on Bootstrap on top of Python’s Flask allowing interactive data visualization. Agile methodology is used throughout the development of the application to ensure that this project is done according to the guideline prepared in the design phase. Functionality testing is also done to ensure that there is no significant error that will render the application useless. In conclusion, the findings gathered show that Naïve Bayes is fairly suitable to be used in NLP problems. The future work that can be put into this project is to improve the corpus to include different slangs of Bahasa Malaysia and commonly used short forms as well as adding an extra class to represent texts that do not belong to either “positive” or “negative”.

TABLE OF CONTENTS