Universiti Teknologi MARA

Depression Prediction System From Twitter's Tweet by Using Sentiment Analysis

Nur Amalina Binti Kamaruddin

Final Year Project
Bachelor of Computer Science (Hons.)
Faculty of Computer and Mathematical Sciences

January 2020

SUPERVISOR APPROVAL

DEPRESSION PREDICTION SYSTEM FROM TWITTER'S TWEET BY USING SENTIMENT ANALYSIS

By

NUR AMALINA BINTI KAMARUDDIN
2016734703

This thesis was prepared under the supervision of the project supervisor, Madam Nurazian Binti Mior Dahalan. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted as part of the bachelor's degree requirements.

Approved by

……..
Madam Nurazian Binti Mior Dahalan
Project Supervisor

JANUARY 6, 2020

STUDENT DECLARATION

I declare that this study and the project to which it relates is the result of my own research and that any concept or quote from other people's work, written or otherwise, is fully recognized in keeping with the discipline's standard reference practices.

……
NUR AMALINA BINTI KAMARUDDIN
2016734703

JANUARY 6, 2020

ABSTRACT

According to the research conducted by the World Health Organization (WHO) in 2015, approximately 300 million of people around the globe are suffering with depression. The research also shows that there is an increase of 18% in the number depression cases diagnosed between 2007 and 2015. Depression is defined as a mental disorder that leads to constant feeling of sadness and also disintegration of interest in an activity that an individual used to enjoy. It also contributes to the inability to carry out daily activities (WHO, 2015). Thus, a Depression Prediction System was developed to predict depression from tweets. The main function of this system is to classify tweet into "depressed" and "not depressed". The classification model was built using Naïve Bayes algorithm. The number of data used in this project is 15952 with 1 independent variable and 1 dependent variables. These data in term of tweets need to go through data cleaning and data transformation before it can be processed by the classification model. Once the data has been transformed, it is divided into 80% to be used training data and the remaining 20% as testing data.

CONTENTS **PAGE**