UNIVERSITI TEKNOLOGI MARA

MULTILABEL CLASSIFICATION OF NATURAL DISASTER NEWS EVENT USING SUPPORT VECTOR MACHINE

NUR FATINI BINTI MD NAZIR

2020482768

BACHELOR OF COMPUTER SCIENCE (HONS)

AUGUST 2023

ACKNOWLEDGEMENT

Alhamdulillah, praises and thanks to Allah because of His Almighty and His blessings, I was able to finish this research within the given time. I would like to acknowledge and give my warmest thanks to my supervisor, Madam Salehah Hamzah, who give me a support and made this work possible. Her guidance and advice carried me through all the stages of writing my thesis and working my project.

Special appreciation also goes to my beloved parents whose always give me moral support. I would also like to thanks them as whole for their continuous support and understanding when undertaking and writing my research. Your prayer for me was sustained me this far.

Finally, I would like to thanks and give my gratitude to my dearest friend who never miss encourage me and keep me in the tracks when I lose.

ABSTRACT

Nowadays, the ease of accessing online news through social media like Twitter is a concern for all generations. However, the problem is the process of verifying the accuracy of the information is time-consuming. Plus, the increment amount of online news stories makes it difficult to access disaster-relevant news promptly. Therefore, to tackle these issues, a solution is proposed: a multilabel classification system using Support Vector Machine (SVM) for natural disaster news event and using TF-IDF for feature extraction method. The chosen methodology for this project is the waterfall model, known for its systematic and linear software development approach. The results demonstrate the promising outcomes, with SVM model achieving 65.2% accuracy with value C=3 as the parameter of the SVM model. Future work aims to expand the system's capabilities to classify various types of natural disaster news event from social media, implement automatic classification with diverse techniques or models and extend the language support to include Malay. Incorporating techniques for spell-check, grammar correction and language quality assessment will enhance the system's ability to accurately classify and extract information from natural disaster news event on social media. In summary, the proposed solution employs SVM-based multilabel classification with TF-IDF feature extraction to address challenges in accessing accurate disaster-relevant news from social media. The system showcases high accuracy and presents possibilities for further enhancements and language expansion.

TABLE OF CONTENTS