UNIVERSITI TEKNOLOGI MARA

AN ARABIC HADITH TEXT CLASSIFICATION MODEL USING CONVOLUTIONAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE

MOHD IRWAN BIN MAZLIN

Thesis submitted in fulfillment
of the requirements for the degree of
Master of Science
(Computer Science)

Faculty of Computer and Mathematical Sciences

September 2022

ABSTRACT

There is a lot of work which have been implemented to solve the problem of text classification, but there is only a little research doing Arabic text classification because of the difficulties in Arabic morphology and the limited public dataset. In order to construct the dataset, the dataset is validated by an expert from lecturer University Sains Islam Malaysia. The purpose validates the dataset is to maintain the authenticity of the content of the hadith. Convolution Neural networks and support vector machines are two different algorithms applied to text classification. CNN seems to be good in extracting the feature from input, and SVM is good for the classification task. This study is to introduce Hadith text classification using a Convolutional Neural Network and Support Vector Machine. There are 6 different ways of designing the experiment to evaluate the result of the study, which are an experiment with the model using different stemming techniques, an experiment with the model using three different algorithms, the result analysis of confusion matric of three algorithms, experiment the model using different SVM kernel, experiment the model using unseen data, produce precision, recall, F1-measure and accuracy result of the model and parameter. First, different model performances are being analysed to find which model gives higher accuracy for this study. CNN-SVM shows a promising result with 92% accuracy, while the CNN only and SVM only give lower accuracy than the proposed model with 82% and 74%. Second, parameter tuning is conducted to find the best parameter for CNN-SVM. Third, the model (CNN-SVM, CNN and SVM) is monitored to see if their performance predicts unseen data. In this study, the CNN-SVM model predicts all correct when using unseen data. Fourth, the model is being tested using different stemming techniques, and it found that the model using non-stemming techniques gives higher accuracy with 92%. Lastly, the different kernel of SVM kernels is being tested to investigate the model's performance for this study. The details about the other experiment can be seen in chapter five, Result and Discussion. The model (CNN-SVM) shows the potential in this study as the model shows better performance than other models. However, there are some limitation of this study, the dataset used were not applied to all categories. It only involved three classes which are prayer, fasting and zakat. So, the model not able to predict correctly if the model predict out of the selected classes. It might be better when the model learns more data and a more specific topic about the Hadith in Arabic. For future work, it is recommended to extend the dataset so that the model can predict the classes in more detail and combine the model with an optimization algorithm to improve the performance of the model.

ACKNOWLEDGEMENT

Alhamdulillah, praise to Allah because of his Almighty and His utmost blessing, I was able to finish this thesis within the period of time. Peace and blessing upon His Messenger Muhammad (SAW). While writing and developing for this thesis, there are a lot of challenges I have to face but I always put my trust to HIM to finish everything. I always believe there is always light at the end of the tunnel.

Firstly, during development process, I had a great chance for learning and professional development. I was very lucky person because ALLAH (SWT) sends me a person that is able to help me out for the development part. He also gives a great idea and a suitable technique to train the model so that the model able to predict the content of Hadith. The name given is Muhammad Danial Bin Ahmad Farid.

I am using this opportunity to express many thanks to my supervisor Dr Mohd Izani Bin Mohamed Rawi and Co-supervisor Dr Mohd Zaki Bin Zakaria who, in spite of being extraordinary busy with his duties, spent time out to hear and kept me on the correct path when writing this journal. I am grateful and indebted to them for their expertise, sincere and valuable guidance as well as the encouragement extended to me. They also have guided me on the ways to handle my temper when I am depressed.

I deeply appreciate, my parents, Mazlin Ibrahim and Aishah Abd Rahman who always pray for my success to finish to this thesis. To my lovely siblings, Izzat Mazlin, Izlin Mazlin and Izwan Mazlin, your attitude means something to me and teach me how to be a patient person. To be honest, I always say to myself I need to sacrifice my time for the sake of my future. May ALLAH always give HIS blessing to them who are involved during finishing this project

TABLE OF CONTENTS