Klasifikasi Multi-label pada Terjemahan Al-Quran Berbahasa Inggris Menggunakan *Convolutional Neural Networks*

Sarah Fauziah Lestari[1], Adiwijaya[2], Said Al Faraby[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]sarahfauziahlestari@students.telkomuniversity.ac.id, [2]adiwijaya@telkomuniversity.ac.id,
[3]saidalfaraby@telkomuniversity.ac.id

Abstract

Al-Quran is a Muslim holy book that was revealed to the Prophet Muhammad, consisting of 114 surahs and 6236 verses. The verses in each surah in the Quran may have one or more topics. In the previous study, a multi-label classification of the Al-Quran dataset was conducted. Therefore, in this study the author tested the performance of CNN in the case of the Al-Quran dataset. Because lately CNN which is part of *Deep Learning* has been used in task of natural language processing, in this English Al-Quran translation dataset is a relatively new dataset and has a high level of language that has an implicit meaning in several verses. In addition, pre-trained word embedding support the learning process of the CNN model which has a dataset of 6236 verses of the Quran that are considered too small to be used in *Deep Learning*. In this study used GloVe for the pre-trained word embedding. The results of this study show that from several scenarios CNN model that use GloVe has a better hamming loss value than those that do not use GloVe, and other algorithms namely Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM) with the value of hamming loss obtained is CNN GloVe with a dropout layer 0,0963 and execution time on test data 2.2662 seconds.

Keywords: Al-Quran, multi-label, convolutional neural networks, deep learning, GloVe