Using Social Media Data to Monitor Natural Disaster: A Multi Dimension Convolutional Neural Network Approach with Word Embedding

Mohammad Reza Faisal[1], Irwan Budiman[2], Friska Abadi[3], Muhammad Haekal[4], Mera Kartika Delimayanti[5], Dodon Turianto Nugrahadi[6]

[1,2,3,4,6]Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat
[5]Program Studi D4 Teknik Informatika, Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta
[1]reza.faisal@ulm.ac.id*, [2]irwan.budiman@ulm.ac.id, [3]friska.abadi @ulm.ac.id, [4]muhammadhaekal77@gmail.com, [5]mera.kartika@tik.pnj.ac.id, [6]dodonturianto@ulm.ac.id

Abstract

Social media has a significant role in natural disaster management, namely as an early warning and monitoring when natural disasters occur. Artificial intelligence can maximize the use of natural disaster social media messages for natural disaster management. The artificial intelligence system will classify social media message texts into three categories: eyewitness, non-eyewitness and don't-know. Messages with the eyewitness category are essential because they can provide the time and location of natural disasters. A common problem in text classification research is that feature extraction techniques ignore word meanings, omit word order information and produce high-dimensional data. In this study, a feature extraction technique can maintain word order information and meaning by using three-word embedding techniques, namely word2vec, fastText, and Glove. The result is data with 1D, 2D, and 3D dimensions. This study also proposes a data formation technique with new features by combining data from all word embedding techniques. The classification model is made using three Convolutional Neural Network (CNN) techniques, namely 1D CNN, 2D CNN and 3D CNN. The best accuracy results in this study were in the case of earthquakes 78.33%, forest fires 81.97%, and floods 78.33%. The calculation of the average accuracy shows that the 2D and 3D v1 data formation techniques work better than other techniques. Other results show that the proposed technique produces better average accuracy.

Keywords: natural disaster, word embedding, convolutional neural network, twitter, social media