ID 23

Detecting AI-generated and Human-Written Documents Using an Ensemble Learning Approach

HMDT Herath[1#] and RAHM Rupasingha[1]

[1]Department of Economics and Statistics, Faculty of Social Sciences and Languages, Sabaragamuwa University of Sri Lanka

[#]<hmrupasingha@gmail.com>

Artificial Intelligence (AI) has revolutionized many parts of modern life including written content. Because of this reason, it is challenging to identify separate AI-generated documents and human-written documents. There are different positive and negative effects of AI-generated documents in different fields including Education. This study aimed to develop a detective mechanism to identify. AI-generated documents and human-written documents automatically using machine learning (ML) algorithms. The acquired AI-generated and human-written documents were pre-processed by cleaning the data set and Term Frequency-Inverse Document Frequency (TF-IDF) was used for feature extraction. Then the study continued utilizing five classification methods such as Naïve Bayes, Random Forest, Decision Tree, Support Vector Machine (SVM), and ensemble learning algorithm that combined the four individual algorithms listed above. The Random Forest individual algorithm shows the best testing accuracy with 65% training and 35% testing dataset for the classification. Ensemble learning outperformed the outcomes in the precision, accuracy, recall, f-measure, and error values. Results demonstrate that this mechanism can successfully detect AI-generated documents and human-written documents separately using an ensemble learning approach.

Keywords: *Artificial intelligence, AI-generated, human-written, machine learning, classification, natural language processing.*