PROJECT REPORT

Comparing Specific Preprocessing Scenarios on Recurrent Neural Network, Gradient Boosting Machines, Naive Bayes, and K-Nearest Neighbors for Predicting Fake News

Christopher Matthew Soeharto
21.K1.0060

Faculty of Computer Science
Soegijapranata Catholic University
2024

ABSTRACT

One major issue in human life is finding valid information to make decisions. With improvement in technology, information is very accessible anywhere and anytime, but this comes with a downside. As information is easy to access, the chance of fake news is also higher. Fortunately for us humans, we have AI to help us with our problems and in this case AI will be used to identify fake news. As AI is still considered new and require more development, The author will try to address one common issue found in most paper talking about detecting fake news using AI, which is the lack of research about pre-processing effect on machine learning algorithm such as RNN, GBM, Naive Bayes, and KNN at identifying fake news. The author will test multiple specific pre-processing scenarios on each algorithm to see the effect of the pre-processing. Dataset from kaggle consisting of 72.134 labeled news articles.

Keywords : Fake news, RNN, GBM, Naive Bayes, KNN, Pre-processing.

5