PROJECT REPORT

COMPARISON BETWEEN RANDOM FOREST AND XGBOOST PERFORMANCE IN TEXT CLASSIFICATION FOR EMOTION DETECTION

JESSICA ANGELA LUKITO
21.K1.0029

Faculty of Computer Science
Soegijapranata Catholic University
2025

ABSTRACT

Humans can not read minds. In this era, where most people are using text-based communication through social media which are non-Face-to-Face interactions. A lot of misunderstandings happened during online conversations like texting because of unclear messages that leads to confusion. Unfortunately, the misunderstanding of a message can cause many negative things to happen such as fight, separation and many more. To resolve this issue, many research has been done by researchers. In some research, several researchers said that Random Forest is the best algorithm for text classification, while others said that XGBoost which is part of Decision Tree is the best. Moreover, there is a study that said Decision Tree is the worst performing algorithm for text classification. With this study, Random Forest and XGBoost as part of Decision Tree will be compared with several pre-processing scenarios and methods. Dataset used for this study is obtained from the Kaggle website which contains 416,809 unique values of sentences.

Keyword: Emotion Detection, Decision Tree, Random Forest, XGBoost