



Stock Industry Sector Prediction Based on Financial Reports using Random Forest Method

Kamil Elian Zhafran*, Deni Saepudin

School of Computing, Telkom University, Bandung, Indonesia

Email: ¹*zhafrankamil@student.telkomuniversity.ac.id, ²denisaepudin@telkomuniversity.ac.id

Correspondence Author Email: zhafrankamil@student.telkomuniversity.ac.id

Submitted: 99/99/9999; Accepted: 99/99/9999; Published: 99/99/9999

Abstract– This study aims to predict the stock industry sector on the Indonesia Stock Exchange (IDX) based on financial reports using the Random Forest method. Implementing this machine learning approach is crucial due to the complexity of financial data, which demands robust and adaptive methods for accurate predictions. The dataset comprises financial data from companies across 10 industrial sectors on the IDX, spanning 2010-2022, and includes 17 features from each financial report. Notably, there is an imbalance in the number of companies per sector, with sector B representing 14.76% and sector G only 1.98%. This imbalance introduces bias in data analysis, thus necessitating the application of the SMOTE oversampling method to address it. The research process involves data cleaning, splitting the data into 80% training and 20% testing sets, applying the SMOTE oversampling technique, and comparing predictions from imbalanced and balanced datasets. The Random Forest method is chosen for its capability to handle complex datasets for industrial sector classification. Evaluation results indicate that without oversampling, the model achieves an accuracy of 73.57%, precision of 74.29%, recall of 73.57%, and an F1-score of 73.51%. With oversampling, these metrics improve to an accuracy of 80.21%, precision of 81.34%, recall of 80.21%, and an F1-score of 80.45%.

Keywords: Industrial Sector Predictions; Indonesia stock exchange; Financial statements; Random Forest; SMOTE