

ABSTRACT

Classifying companies based on their industry sector is crucial for investment analysis. However, the current manual classification used by the Indonesia Stock Exchange (IDX) is not optimal in leveraging financial statement data for industry sector prediction. Therefore, this study develops a predictive model using Linear Discriminant Analysis (LDA) and compares it with Extreme Gradient Boosting (XGBoost). The dataset consists of financial reports from companies listed on IDX between 2010 and 2022, which undergo pre-processing, normalization, and oversampling using Borderline-SMOTE to address class imbalance. Model evaluation is conducted using accuracy, precision, recall, and F1-score, while Permutation Importance is applied to identify the most influential financial features. The results show that LDA achieves an accuracy of 27.51%, whereas XGBoost outperforms it with an accuracy of 63.87%, indicating that the non-linear XGBoost approach is more effective for industry sector classification. Additionally, total assets, total revenue, and inventory are identified as the most significant factors in predicting industry sectors. This study contributes to the development of an automated classification method based on financial reports, which can assist investors and analysts in making more accurate investment decisions.

Keywords: Indonesia stock exchange, linear discriminant analysis, xgboost, financial statements, industry sector prediction.