ABSTRACT

Felicia Adeline Setiawan

Undergraduate Thesis

A Comparative Performance Assessment Of Single Classifier And Ensemble Learning For Credit Card Default Prediction

The credit card debt crisis has become a significant concern, impacting card-issuing institutions, despite the continuous global rise in credit card customers since 2018. The rise of e-commerce platforms encourages consumerist behavior, with credit cards becoming a preferred and convenient payment method, leading to increased transactions and affecting the risk of customer default. Consequently, banks, as issuers, should avoid this behavior to prevent costly defaults. Machine learning is a recent tool used to predict credit card defaults due to its ability to handle large datasets, explored in previous research. The prevalence of imbalanced data is a common challenge in practice and can significantly impact prediction performance if the existence of imbalanced data is neglected. Hence, this study aims to investigate the impact of imbalanced data by comparing several classification algorithms and identifying features with significant contributions to predictions. Results indicate that Random Forest stands out as the most effective algorithm due to the highest F1-Score compared to others. Additionally, implementing SMOTE to address data imbalances enhances model performance across various imbalance ratios. Certain features, such as payment status in the most recent one to two months and credit card limit balances, play crucial roles in predicting default payments.

Keywords : Credit card debt crisis, machine learning, imbalanced data, classification algorithms, Random Forest, SMOTE, F1-Score, payment status, credit card limit balances, customer default.

TABLE OF CONTENT

LIST OF FIGURE

LIST OF TABLE

LIST OF FORMULA

LIST OF APPENDIX