

Title: Unmasking Banking Fraud: Unleashing the Power of Machine Learning and Explainable AI (XAI) on Imbalanced Data

Author Names: S. M. Nuruzzaman, Shirin Sultana, Sondip Poul Singha, Sudipto Chaki ¹ORCID, Md. Julkar Nayeem Mahi Tony Jan Alistair Barros Md Whaiduzzaman

Abstract: Recognizing fraudulent activity in the banking system is essential due to the significant risks involved. When fraudulent transactions are vastly outnumbered by non-fraudulent ones, dealing with imbalanced datasets can be difficult. This study aims to determine the best model for detecting fraud by comparing four commonly used machine learning algorithms: Support Vector Machine (SVM), XGBoost, Decision Tree, and Logistic Regression. Additionally, we utilized the Synthetic Minority Over-sampling Technique (SMOTE) to address the issue of class imbalance. The XGBoost Classifier proved to be the most successful model for fraud detection, with an accuracy of 99.88%. We utilized SHAP and LIME analyses to provide greater clarity into the decision-making process of the XGBoost model and improve overall comprehension. This research shows that the XGBoost Classifier is highly effective in detecting banking fraud on imbalanced datasets, with an impressive accuracy score. The interpretability of the XGBoost Classifier model was further enhanced by applying SHAP and LIME analysis, which shed light on the significant features that contribute to fraud detection. The insights and findings presented here are valuable contributions to the ongoing efforts aimed at developing effective fraud detection systems for the banking industry

Keywords: fraud detection; machine learning; logistic regression; decision tree; SVM; XGBoost; oversampling SMOTE; SHAP analysis; LIME analysis

<https://doi.org/10.3390/info15060298>

<https://www.mdpi.com/2078-2489/15/6/298>