iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62899

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Enhanced Detection of Credit Card Fraud Using Machine Learning Techniques

Nishant Doshi[1], Geetha Chillarge[2], Swati Shekapure[3]
[1]PG Student, [2]Assistant Professor, [3]Associate Professor, Marathwada Mitra Mandal's College of Engineering, Pune, Maharashtra, India

Abstract: ***As the number of credit card transactions continues to grow, they represent an increasing share of the global payment system. This growth has led to an increase in stolen account numbers and subsequent losses to banks. Machine Learning (ML) plays a crucial role in detecting credit card fraud in both online and offline transactions. Credit card fraud detection, which is a data mining problem, becomes challenging for two main reasons: first, the characteristics of normal and fraudulent behavior are continually changing, and second, the credit card fraud dataset is highly asymmetric. This study proposes an ensemble approach for accurately detecting credit card fraud transactions based on various ML algorithms and compares the performance of each algorithm with the proposed model. The results indicate that the proposed Ensemble model without SMOTE (Synthetic Minority Over-sampling Technique) outperforms all other methods, achieving a 99.94% accuracy rate..***
Keywords: Credit Card Transactions, Machine Learning, Credit Card Fraud Dataset, SMOTE, Ensemble Model

I. INTRODUCTION

A. Overview

Modern commerce is largely dependent on the use of e-commerce platforms and various electronic transactions conducted through them. Online banking has become more prevalent due to its numerous advantages, such as lower fees, better customer service, and faster processing times. However, security is a major concern for customers when it comes to online banking. The rise of fraudulent transactions has led many customers to fear for their financial security. Consequently, banks are compelled to develop fraud detection systems capable of identifying suspicious transactions.

B. Motivation

The motivation behind this research is to enhance the accuracy and reliability of fraud detection systems using advanced ML techniques. Given the increasing number of credit card fraud incidents, an efficient detection system is essential for minimizing financial losses and protecting consumers' trust.

C. Problem Definition

Credit card fraud detection is a challenging problem due to the dynamic nature of fraudulent behaviors and the highly imbalanced nature of fraud datasets. Traditional methods are often inadequate in identifying new and sophisticated fraudulent patterns. This research aims to develop a robust machine learning-based approach to effectively detect fraudulent transactions.

II. RELATED WORK

Previous studies have explored various ML techniques for fraud detection, including logistic regression, decision trees, random forests, and neural networks. While these methods have shown promise, there is a continuous need for improving detection accuracy and reducing false positives. Ensemble models and data balancing techniques like SMOTE have been identified as potential solutions to enhance the performance of fraud detection systems.

III. PROPOSED WORK

A. Objectives and Challenges

The primary objective of this research is to develop an ensemble ML model that improves the detection of fraudulent transactions with high accuracy and low false positive rates. Challenges include dealing with the imbalanced dataset and ensuring the model's adaptability to changing fraud patterns.

B. Project Scope

The scope of this project includes the development, training, and evaluation of various ML models, including logistic regression, decision trees, random forests, and ensemble methods. The performance of these models will be compared to determine the most effective approach for fraud detection.

C. Problem Statement

The problem statement for this research is to develop a machine learning-based system that can accurately detect fraudulent credit card transactions, thereby reducing financial losses and enhancing security measures for online and offline transactions.

IV. METHODOLOGY

A. Data Collection and Preprocessing

The dataset used for this study is the Credit Card Fraud Detection dataset from Kaggle, which contains transactions made by European cardholders in September 2013. The dataset consists of 284,807 transactions, of which 492 are fraudulent. The data is highly imbalanced, with fraudulent transactions representing only 0.172% of all transactions.

B. Model Development

The following models were developed and evaluated in this study:

1) Logistic Regression
2) Decision Tree
3) Random Forest
4) Gradient Boosting
5) Ensemble Model (Combination of Random Forest and Gradient Boosting)

C. Data Balancing Techniques

To address the imbalance in the dataset, SMOTE was used to oversample the minority class (fraudulent transactions). The impact of SMOTE on model performance was also evaluated.

D. Evaluation Metrics

The models were evaluated using the following metrics:

1) Accuracy
2) Precision
3) Recall
4) F1 Score
5) ROC-AUC Curve

V. RESULTS AND DISCUSSION

A. Model Performance

The performance of each model was evaluated based on the aforementioned metrics. The results are summarized in the following table:

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Logistic Regression	99.72%	87.23%	56.09%	68.51%	0.973
Decision Tree	99.92%	91.45%	81.21%	85.99%	0.999
Random Forest	99.94%	96.42%	82.74%	89.07%	0.999
Gradient Boosting	99.93%	94.29%	85.97%	89.93%	0.999

Ensemble (Without SMOTE)	99.94%	97.01%	83.92%	89.96%	0.999
Ensemble (With SMOTE)	99.28%	89.12%	86.78%	87.94%	0.994

The ensemble model without SMOTE achieved the highest accuracy of 99.94%, demonstrating its effectiveness in detecting fraudulent transactions. Figures 1 and 2 illustrate ROC curves and Confusion matrix for the proposed model.

B. Analysis and Comparison

From the results, it is evident that the ensemble model without SMOTE outperforms all other models, achieving the highest accuracy (99.94%) and a high ROC-AUC score (0.999). Although SMOTE improved recall, it slightly decreased the overall accuracy and precision of the ensemble model.

C. Graphical Representation

The following graphs illustrate the performance of the models:

ROC Curves

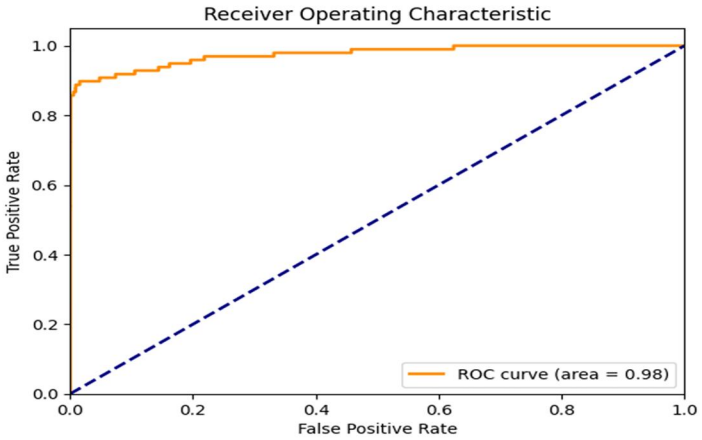


Fig. 1 A ROC curve of ensemble model

CONFUSION MATRIX FOR ENSEMBLE MODEL

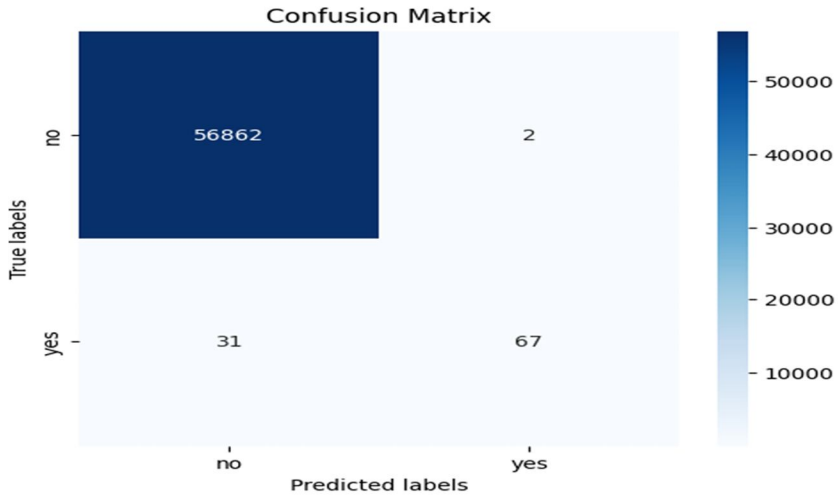


Fig. 2 Confusion Matrix of Ensemble Model

VI. CONCLUSION AND FUTURE WORK

A. *Conclusion*

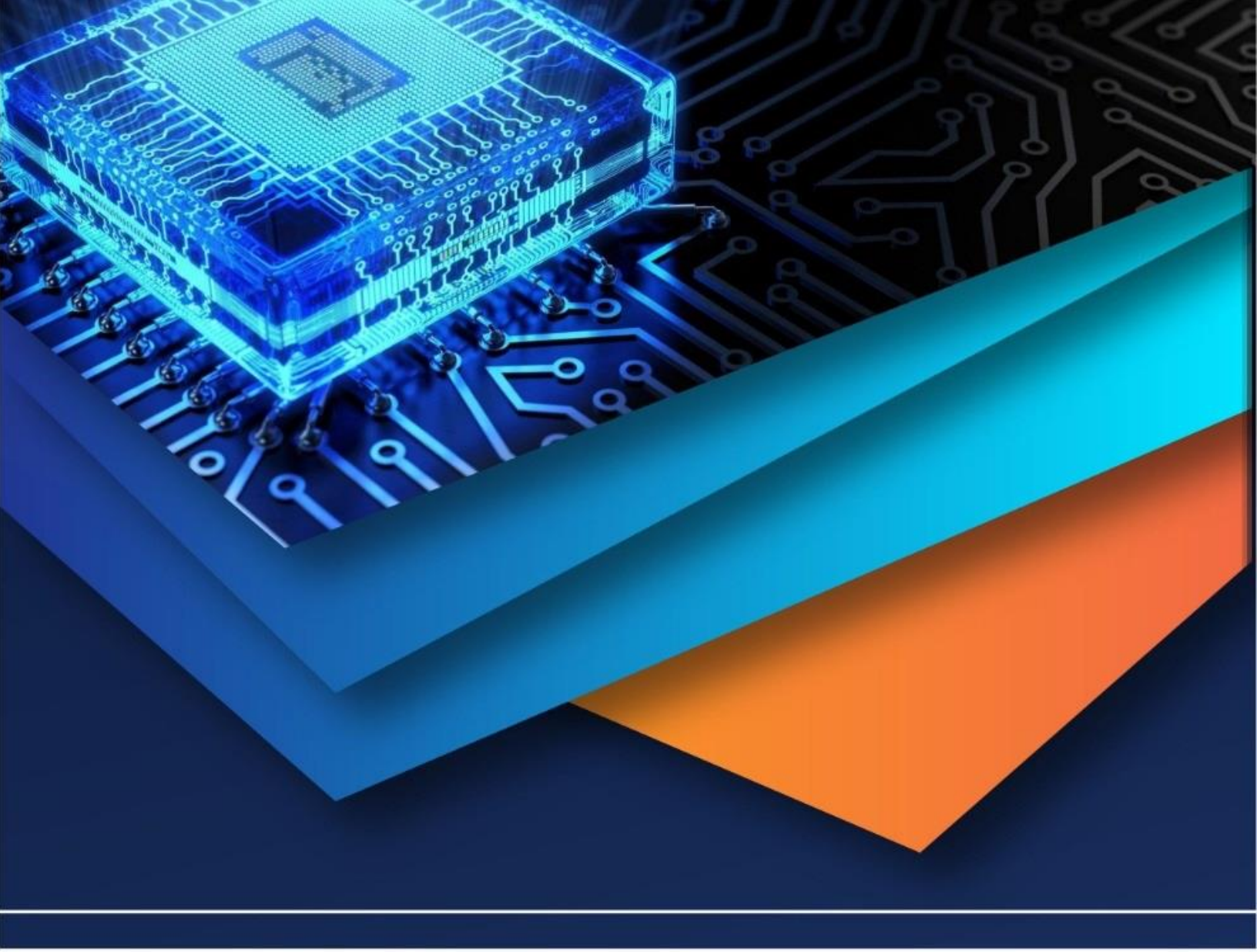
The study demonstrates that ensemble models, particularly without the application of SMOTE, provide superior performance in detecting credit card fraud. The results suggest that while data balancing techniques like SMOTE can enhance recall, they may not always lead to improvements in overall accuracy and precision.

B. *Future Work*

Future research should focus on exploring more advanced ensemble techniques and deep learning models to further enhance fraud detection accuracy. Additionally, incorporating real-time data and adaptive learning mechanisms can help in maintaining the model's effectiveness over time.

REFERENCES

[1] Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE transactions on neural networks and learning systems, 29(8), 3784-3797.
[2] Carcillo, F., Le Borgne, Y. A., Caelen, O., & Bontempi, G. (2017). Streaming active learning strategies for real-life credit card fraud detection: assessment and visualization. International Journal of Data Science and Analytics, 3(4), 293-308.
[3] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.
[4] Sahin, Y., & Duman, E. (2011). Detecting credit card fraud by ANN and logistic regression. Proceedings of the 2011 International Symposium on Innovations in Intelligent Systems and Applications (INISTA), 315-319.
[5] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. Expert Systems with Applications, 100, 234-245.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY