"CREDIT CARD FRAUD ANALYSIS AND DETECTION METHOD"

Ms. Aksa Biju

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

aksabiju003@gmail.com

Ms. Clarin Joe

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

clarinamalajoe@gmail.com

Mr. P M Wajid Rehman

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

pmwr.bca2124@saintgits.org

Mr. Aswin A

BCA Student

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

aswin.bca2124@saintgits.org

Dr. Meenu Suresh

Assistant Professor

PG Department of CA & AI

Saintgits College Of Applied Sciences Kottayam, Kerala

Mahatma Gandhi University - Kerala

meenupillai1988@gmail.com

Abstract

Electronic payment (e-payment) has transformed financial transactions, offering speed and convenience. However, it has also brought about significant challenges, notably in credit card security. This paper explores the landscape of e-payment, focusing on credit card fraud detection—a crucial aspect of financial security systems. Advanced algorithms and machine learning techniques are employed to analyse transaction data for patterns indicative of fraudulent activity. Despite these advancements, credit cards face various security threats, including card skimming, phishing scams, and data breaches. Cybercriminals continuously adapt their tactics, necessitating ongoing advancements in credit card protection. This paper highlights the importance of evolving security measures to safeguard users' financial information in the ever-changing digital landscape.

I.Introduction

Electronic payment, or e-payment, refers to the digital exchange of money between parties for goods or services, conducted through electronic devices and online platforms. This method of financial transaction eliminates the need for physical currency or traditional paper-based checks, offering a convenient and efficient means of conducting business in the modern digital era[2]. E-payment systems encompass various forms, including credit/debit card transactions, mobile payments, online banking transfers, and digital wallets[4,7,8]. The widespread adoption of e-payment solutions has revolutionised the way individuals and businesses manage financial transactions, offering speed, security, and accessibility.

Credit card fraud detection is a critical aspect of financial security systems aimed at identifying and preventing unauthorised or fraudulent transactions. Advanced algorithms and machine learning techniques play a pivotal role in analysing vast amounts of transaction data to identify patterns and anomalies indicative of fraudulent activity[9]. These systems assess various factors, including transaction frequency, location, purchase amounts, and unusual spending patterns.

Credit cards face various security and fraud challenges, ranging from unauthorised transactions and identity theft to sophisticated cyberattacks[10-12]. One prevalent issue is card skimming, where criminals install devices on ATMs or point-of-sale terminals to capture card information[13]. Phishing scams also target cardholders through deceptive emails or websites, aiming to obtain sensitive details. Additionally, data breaches at retailers or financial

institutions can compromise large volumes of credit card information. Cybercriminals constantly evolve their tactics, employing malware and other advanced techniques to exploit vulnerabilities in online transactions[14,15]. The ongoing battle between security measures and fraudulent activities underscores the need for continuous advancements in credit card protection to safeguard users' financial information.

II.Related Work

Sadgali et al[6] focuses on addressing the increasing challenge of credit card fraud in the digital age. The authors use a consistent dataset to evaluate different methods, aiming to select the best technique for implementation in future work. The paper underscores the potential of machine learning in enhancing the accuracy and efficiency of fraud detection systems, thereby providing a significant contribution to the security of financial transactions in the digital era. Zhang et.al [5] explores the application of the Xgboost model for detecting fraud in customer transactions. The study uses a dataset from the IEEE-CIS Fraud Detection Competition on Kaggle, involving data mining techniques like feature engineering, visualization, and the use of SMOTE (Synthetic Minority Oversampling Technique) for addressing class imbalance. The authors compare Xgboost with other machine learning methods like Support Vector Machine, Random Forest, and Logistic Regression, demonstrating that the Xgboost-based model outperforms these in terms of ROC AUC score and accuracy. They also highlight the importance of feature selection in improving model performance. Liu et.al [16] presents a comprehensive study on the application of machine learning techniques in detecting online transaction fraud. The paper introduces two fraud detection algorithms based on Fully Connected Neural Networks and XGBoost, respectively, and highlights the design of an interactive online transaction fraud detection system that utilizes the XGBoost model. The study includes extensive experiments and comparisons and also discusses the system's capacity to analyze transaction data automatically and provide users with fraud detection results, contributing significantly to the field of online financial security. Sailusha et.al [1] explores the application of machine learning algorithms for detecting credit card fraud. The study focuses on using Random Forest and Adaboost algorithms to analyze credit card transaction data. The effectiveness of these algorithms is evaluated based on accuracy, precision, recall, and F1-score. The researchers also employ a Receiver Operating Characteristic (ROC) curve based on the confusion matrix for further analysis. The results show that while both algorithms perform well, the Random Forest algorithm exhibits a higher performance. GPT

The author Jain et.al [3] examines the effectiveness of various machine learning algorithms in detecting credit card fraud. The study specifically focuses on three algorithms: Decision Tree, Random Forest, and XGBoost. It compares their performance in terms of prediction accuracy and uses a dataset of over one lakh credit card transactions for testing. The results reveal that XGBoost has the highest prediction accuracy (99.962%), followed by Random Forest (99.957%) and Decision Tree (99.923%).

III.Proposed Method

The research methodology comprises three main sections: data pre-processing, addressing imbalanced classifiers, and providing descriptions of the models used in the study.

3.1 Dataset Description

The dataset utilized in this study to evaluate students' adaptability to online learning was sourced from Kaggle, a well-known machine learning repository widely used for sharing and assessing datasets. Kaggle serves as a platform where individuals, organizations, and scholars contribute datasets spanning various sectors. Ensuring alignment with the goals and parameters of our investigation was crucial in selecting an appropriate dataset from Kaggle's diverse offerings. Researchers conducted a thorough evaluation of the dataset to ascertain its dependability and quality. This evaluation included a meticulous assessment for completeness, correctness, and relevance to ensure that the dataset adequately met the requirements of our study.

A. Data Preprocessing

In this research, the data underwent thorough cleaning, transformation, and organization to meet quality standards for analysis. Subsequently, normalization was performed, applying rules to categorise data into low, moderate, and high adaptability levels. The Interquartile Range [71,72] was utilized to eliminate any inaccurate data points, ensuring data accuracy.

B. Balancing the Dataset using SMOTE

Addressing the inherent data imbalance, the dataset comprises 625 instances of Moderate values, 480 instances of Low values, and 100 instances of High values. To rectify this imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) [73,74] is employed through oversampling and is depicted in Fig.1. This technique involves generating synthetic instances to augment the minority class, ensuring a more balanced representation in the dataset. Subsequently, the balanced dataset is fed into each classifier for effective classification.

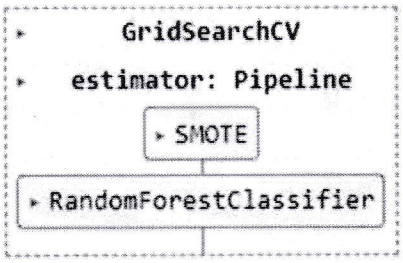

Figure 1. Balancing the Dataset using SMOTE

C. Description of Models

Numerous machine learning algorithms have been employed to forecast potential outcomes regarding student adaptability levels in this study. The dataset underwent training and testing with a diverse set of ML algorithm - RandomForest. Given the dataset's multi-labeled nature, a One Vs. Rest approach was integrated into binary class classification algorithms to tackle the challenges of multi-class classification. In this methodology, one class is designated as the positive class, while all other classes are collectively treated as the negative class. Each class undergoes individual testing, and results are generated by evaluating one class against all the other classes using this approach.

1) Random Forest

The Random Forest Classifier is a machine learning method specifically designed for addressing classification problems. This ensemble learning technique combines multiple decision trees to generate predictions. The key feature of a random forest lies in training each tree with a different subset of the training data and features, resulting in a diverse set of decision trees. As incoming data is fed into the ensemble, each tree independently categorises it, and the final prediction is determined by a majority vote among the trees. Renowned for its ability to handle high-dimensional data, mitigate overfitting, and deliver robust and accurate classification results, the Random Forest Classifier finds widespread application in various industries such as banking, healthcare, and image categorization, thanks to its efficiency and adaptability.

$$\text{norm}fi_i = \frac{fi_i}{\sum j \in features fi_i}$$

IV.Evaluation and Result Analysis

As shown in Table 1, Random forest with recall, precision, F1 Score and accuracy with SMOTE oversampling are 0.765957,0.186528,0.300000,0.994079 respectively and with no oversampling are 0.723404,0.918919,0.809524,0.999436 is depicted in figure 2.

Table 1: Comparison of the various metrics

Random Forest with	Recall	Precision	F1 Score	Accuracy
SMOTE Oversampling	0.765957	0.186528	0.300000	0.994079
No under/oversampling	0.723404	0.918919	0.809524	0.999436

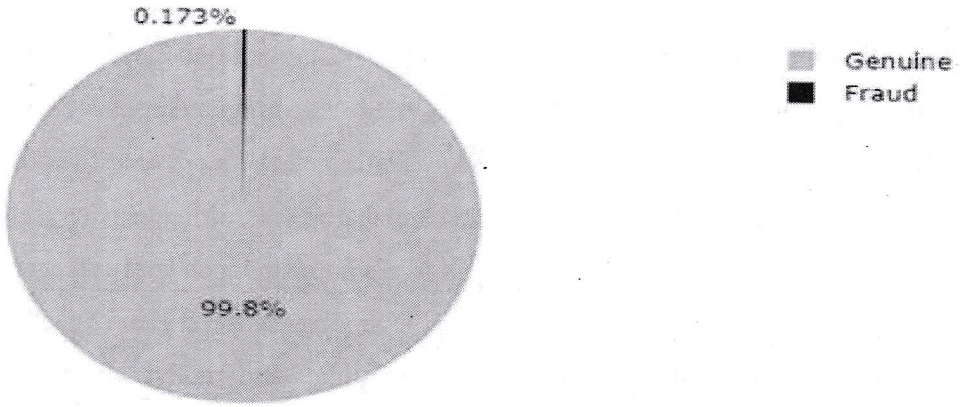

Figure 2. Fraud and Genuine transaction

V.Conclusion

Even though there are many fraud detection techniques we can't say that this particular algorithm detects the fraud completely. From our analysis, we can conclude that the accuracy for the Random Forest with SMOTE Oversampling is 99.4079 and Random Forest with No Under/Oversampling is 99.9436. When we consider the precision, recall, and the F1-score the Random Forest algorithm has the highest value . Hence we conclude that the Random Forest Algorithm works with best accuracy to detect credit card fraud.

REFERENCES

1) Ruttala Sailusha,V. Gnaneswar,R. Ramesh,G. Ramakoteswara Rao,"Credit Card Fraud Detection Using Machine Learning",Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020).
2) Kanika,Dr Jimmy Singla,"A Survey of Deep Learning based Online Transactions Fraud Detection Systems",2020 International Conference on Intelligent Engineering and Management (ICIEM).
3) Vinod Jain,Mayank Agrawal,Anuj Kumar,"Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection",2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).
4) Parth Roy,Prateek Rao,Jay Gajre,Kanchan Katake,Arvind Jagtap,Yogesh Gajmal,"Comprehensive Analysis for Fraud Detection of Credit Card through Machine Learning",2021 International Conference on Emerging Smart Computing and Informatics (ESCI).
5) Yixuan Zhang,Jialiang Tong,Ziyi Wang,Fengqiang Gao,"Customer Transaction Fraud Detection Using Xgboost Model",2020 International Conference on Computer Engineering and Application (ICCEA).
6) Imane SADGALI,Nawal SAEL,Fouzia BENABBOU,"Fraud detection in credit card transaction using machine learning techniques",2019 978-1-7281-4368-2/19/$31.00 ©2019 IEEE.
7) V.Kamesh, M.Karthick,K.Kavin,M.Velusamy, R.Vidhya,"Real-Time Fraud Anamaly Detection in E-banking Using Data Mining Algorithm",2017 South Asian Journel of engineering and technology
8) I. Sadgali, N. Sael, F. Benabbou, "Detection of credit card fraud:State of art",International Journal of computer science and network security, Vol.18, No.11, pp.76-83, 2018.
9) S. Askari, A. Hussain, "Credit Card Fraud Detection Using Fuzzy ID3", Proc. of International Conf.On Computing, Communication and Automation (ICCCA), Greater Noida, India, pp.446-452, 2017.
10) V. Sapovadia, "Financial Inclusion, Digital Currency, and Mobile Technology," in Handbook of Blockchain, Digital Finance, and Inclusion, Volume 2, Elsevier, 2018, pp. 361–385.

11) Arvind M Jagtap, Prof.Dr.Gomathi N, "Meta-Heuristic based Trained Deep Convolutional Neural Network for Crop Classification" ,International Journal of Emerging Trends in Engineering Research (IJETER) Volume 8. No. 7, July 2020.

12) ULB (2018), Kaggle, "Machine Learning Group-Credit Card Fraud Detection"

13) Minastireanu, E. A., & Mesnita, G. (2019). Light gbm machine learning algorithm to online click fraud detection. J. Inform. Assur. Cybersecur, 2019.

14) Fabrizio Carcillo, Yann-Aël Le Borgne, Olivier Caelen, Yacine Kessaci, Frédéric Oblé, Gianluca Bontempi, Combining unsupervised and supervised learning in credit card fraud detection, Information Sciences, 2019, ISSN 0020-0255.

15) Ugo Fiore, Alfredo De Santis, Francesca Perla, Paolo Zanetti, Francesco Palmieri, Using generative adversarial networks for improving classification effectiveness in credit card fraud detection, Information Sciences, Volume 479, 2019, Pages 448-455, ISSN 0020-0255.

16) ARMEL and D. ZAIDOUNI, "Fraud Detection Using Apache Spark," 2019 5th International Conference on Optimization and Applications (ICOA), Kenitra, Morocco, 2019, pp. 1-6. doi: 10.1109/ICOA.2019.8727610.

17) Bocheng Liu, Xiang Chen* and Kaizhi Yu ,"Online Transaction Fraud Detection System Based on Machine Learning",ICCTPE 2021.