iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.61180

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 IV Apr 2024- Available at www.ijraset.com

Identifying Credit Card Fraudulent Transactions through Machine Learning Analysis

Mr. G.Satya Mohan Chowdary[1] , Y Bala Subrahmanyam[2], Dulapalli Naga Naveen[3], Varasala Akshay[4], Vanapalli Venkat[5]
[1]*Assistant Professor* , [2, 3, 4, 5] *B.tech Students Department of Information Technology, Pragati Engineering College, Surampalem , Andhra Pradesh, India*

Abstract: ***In the last few years, there has been a sharp rise in credit card fraud transactions. For financial institutions, credit card theft is a severe problem, and reliably detecting fraud can be challenging. An annual study conducted in 2021 found that over 50% of Americans had experienced credit and debit card fraud, and over 13% of those who use these cards often also do so. This indicates that 127 million Americans have experienced credit card theft at some point in the past. Using the traditional way, it is very hard and time-consuming to find this form of fraud in a large database. An effective solution to handle this kind of issue is to create a fraud detection system that is fully automated to identify and categorize such cases using technologies like artificial intelligence (AI) and machine learning. SVM, Random Forest, KNN, XGBoost, and Naïve Bayes are the six supervised machine learning algorithms provided by this study. By using these techniques, a classification model that reliably identifies fraudulent transactions is created. It was discovered that Support Vector Machine was the most accurate model for differentiating between authentic and fraudulent transactions after each of these algorithms was examined.***
Keywords: Fraud Detection, AI, Credit Card, Data Imbalance, Machine learning Algorithms

I. INTRODUCTION

Ever since electronic commerce payment systems were introduced, there have always been people who would devise creative ways to steal money from other people without that person's knowledge or permission. Payment convenience has increased significantly with the introduction of online payment options. Payment fraud has also surged in tandem. Although it can happen with any payment method, credit card fraud is the most common type of online payment fraud. These days, this is a significant problem. There is a lot of credit card fraud. A significant sum of money can be quickly taken out without the owner's knowledge or danger. Without your knowledge, scammers use your credit card details to make unauthorized purchases on your account. Due to these conditions, credit card security is crucial. Since the goal of con artists is to make every fraudulent transaction seem authentic, it can be challenging to identify fraud. These days, customers use social media and online transaction technology to quickly transfer money from their bank accounts to other suppliers and clients for their enterprises. As a result, most business transactions are done online by entrepreneurs, which facilitates fraud. Historically, data analysts have been in charge of seeing and keeping an eye out for suspicious behaviors in order to uncover fraud. Card issuers may employ various countermeasures, such as fraud likelihood evaluation software, to detect and identify such behavior. For example, something that happens distance from the cardholders' home would seem suspect. But tracking it would also be challenging in the absence of a reliable system. This method is no longer successful due to the rapid increase of online transactions, which are often linked to mobile payments, online shopping, and other activities. Every day, millions of entries and hundreds of dimensions are added to a real-time transaction dataset that is updated online. Unfortunately, many techniques relying on cards to report transactional fraud have been ineffective. Most fraud protection systems work in the same way, monitoring incoming payments by recognizing suspect payment patterns among a huge number of payment records. Statistical learning, often known as machine learning, is a popular method because it is simple. Machine learning has shown to be quite effective in extracting these patterns. Depending on the application, it can be used as either an unsupervised (anomaly detection) or supervised (classification) model. It requires minimal upkeep because it may be automatically retrained to preserve its associations. This article will analyze and discuss the work of several researchers who have created models that can detect and classify fraudulent credit card transactions using a variety of machine learning techniques. It can be used as a supervised (classification) or unsupervised (anomaly detection) model, depending on the requirements of the application. Because it may be automatically retrained to keep its associations, it requires little maintenance.

This article will look at and discuss the work of several researchers who have created models that can detect and categorize fraudulent credit card transactions using a variety of machine learning techniques.

II. LITERATURE SURVEY

Md. Sufiyan et al. [6] employed a preprocessed technique to clean the dataset for their study, removing outliers and missing values. To mitigate the risks of overfitting, they balanced their dataset using the Smote technique, which involved balancing the ratio of minority class data against the percentage of majority class data. They then applied three machine learning algorithms: logistic regression, decision trees, and extreme learning machines. The results showed that, with a 99% accuracy rate, the extreme learning machine provides the most accurate outputs.

Vikrant et al. [7] used decision tree and statistical models to detect fraudulent transactions. After assessing three potential approaches: undersampling, oversampling, and rise sampling, they chose the one with the largest area under the curve. XGBoost, Decision Tree, Random Forest, and Logistic Regression were the learning methods used. With 99.96% accuracy and precision, the results show that XGBoost is the best classifier for detecting credit card fraud. The approach they utilized demonstrated that oversampling generated the highest AUC.

Nihar Ranjan et al. [1] highlighted the use of machine learning as a preventative method when discussing credit card theft. Overall, SVM and logistic regression performed well, with 94.84% accuracy and 97.58% precision, and 94.99% accuracy and 95.58% precision, respectively. They applied five machine learning techniques: SVM, Random Forest, XGBoost, Decision Tree, Naive Bayes, and Logistic Regression. The combination of machine learning algorithms and multiple sampling methodologies used by Rabiul Alam Bhuiyan et al. [5] was primarily focused on controlling class imbalance. Several sampling procedures, including adaptive synthetic, synthetic minority over-sampling, under-sampling, over-sampling, and synthetic over-sampling, were used to balance the imbalance dataset. Decision Tree, Random Forest, KNN, Logistic Regression, and Naive Bayes have also been used in various applications. The random forest algorithm and SMOTE sampling technique resulted in 85.71% recall, 91.30% precision, and 99.97% accuracy.

Shahnawaz Khan et al. [8] described how to create a reliable machine learning system for accurately detecting and classifying credit card thefts. Their methodology consisted of three models: support vector machines, artificial neural networks, and logistic regression, which had accuracy rates of 99.94%, 99.91%, and 99.91%, respectively. The models were thoroughly examined with a variety of matrices and performance assessment methodologies, such as the Matthews correlation coefficient, confusion matrix, and ROC curve. According on the accuracy calculation of the confusion matrix, all of the models got almost equal scores. However, additional research utilizing many performance metrics revealed that the SVM model outperformed the other models.

E. Esenogho et al. [4] developed a successful technique for detecting credit card fraud by combining an ensemble algorithm, a neural network, and a hybrid data resampling strategy. The early AdaBoost algorithm used the LSTM method. The latter technology, known as SMOTE-ENN, was created by combining nearest neighbor and the Synthetic Minority Oversampling process. A real-time dataset was used to compute the anticipated approach. The experiment results showed that the first algorithm outperformed the others. Furthermore, this approach offers 99.8% specificity and 99.6% sensitivity for detecting credit card fraud.

Noor Saleh Alfaiz et al. [3] conducted a two-level evaluation of 66 different machine learning models. The first round tested the ability of nine machine learning (ML) algorithms to detect fraudulent transactions. During the second round, which included 19 alternative resampling strategies, the best three algorithms were picked and used again. The All K-Nearest Neighbors undersampling strategy, paired with CatBoost, was shown to be the most effective suggested model out of 33 evaluation metric values. Their findings indicate that the proposed model outperforms the present ones by 97.94% AUC, 95.91% Recall value, and 87.40% F1-score value.

Esraa Faisal Malik et al. [2] created and evaluated seven hybrid machine learning models for detecting fraudulent activities. Credit card fraud was initially detected using modern machine learning (ML) techniques. Hybrid techniques were later developed using the most effective beginning phase algorithm. The hybrid strategy that combined LGBM with AdaBoost proved to be the most effective.

Manoj Kumar Reddy Mallidi et al. [11] used ten different strategies on the balanced and imbalanced datasets they proposed for their study. The dataset grew significantly after balancing, and numerous methods performed poorly. However, in both situations, RF produced the most accurate findings and maintained a steady F1-Score.

Tanouz et al. [10] cleansed the data before to processing, removing outliers that were causing the model to behave inconsistently.

Using a number of algorithms, including naive Bayes, logistic regression, Random Forest, and decision trees, they discovered that the RF classifier outperformed all others, with a 96.7741 percent accuracy rate. They also noticed an improvement in the other four techniques' performance.

III. SYSTEM ANALYSIS

A. *Existing System*

The current technology provides a machine learning and artificial intelligence-based approach to address the growing issue of credit card fraud. Because of the rise in credit card theft, traditional methods are ineffective and time-consuming. This project uses six supervised machine learning algorithms: SVM, Random Forest, KNN, XGBoost, Naïve Bayes, and Logistic Regression. Following extensive research, the Support Vector Machine was found to be the most accurate model for differentiating between fraudulent and non-fraudulent transactions. The proposed method uses technology to automatically identify and classify fraudulent events, making it a more effective and efficient alternative to manual detection techniques. Given the fact that a large portion of Americans have suffered credit and debit card theft, this method is critical. The work provides a potential paradigm shift in credit card fraud detection and is a proactive solution to the critical difficulties that financial institutions face.

DISADVANTAGES OF THE EXISTING SYSTEM

1) Restricted Feature Set: The current system's fraud detection functions are limited, which may cause complications. If crucial fraud indicators are not included in the feature set, the system's accuracy may suffer.
2) Dependence on Historical Data: When training machine learning models, historical data is commonly used. If the present system is not regularly updated with fresh data, its effectiveness may deteriorate over time due to its inability to adapt to evolving fraud tendencies.
3) imbalanced Class Distribution: Because credit card fraud is very rare compared to real transactions, the class distribution is imbalanced. As a result of this imbalance, the model may become biased in favor of the majority class, thus impairing its ability to detect fraud.
4) Lack of Explainability: Some machine learning algorithms, particularly advanced ones such as XGBoost and SVM, make decisions that are difficult to understand. This lack of explainability might make it difficult to interpret and act on the system's outputs, undermining confidence in the system's predictions.
5) Vulnerability to Adversarial Attacks: Adversarial attacks can harm machine learning models, notably those used to detect fraud. The model's ability to discriminate between genuine and fraudulent transactions may be jeopardized if fraudsters modify or inject fake data into the system, causing security concerns.

B. *Proposed System*

By incorporating cutting-edge innovations, the proposed solution aims to address the inadequacies of the present credit card fraud detection system. It ensures a comprehensive depiction of potential fraud indicators by augmenting the dataset with advanced feature engineering. To address the dynamic nature of fraud techniques, dynamic model retraining procedures are implemented. This allows the models to continuously adapt to new data while maintaining high accuracy over time. The proposed approach addresses class imbalance by increasing the model's sensitivity to occasional fraudulent cases using approaches such as oversampling and under sampling.

The integration of interpretable machine learning models and visualization tools improves explainability and provides clear insights into the decision-making process. Anomaly detection algorithms are also used to detect aberrant patterns that may suggest fraudulent behaviour, resulting in a more complete and sophisticated approach to fraud detection. To stay ahead of new risks, the suggested strategy emphasizes frequent updates. It also features feedback loops to enable continual development based on actual performance.

Strict security measures are used to protect the fraud detection system's integrity and dependability from adversarial attacks. In principle, the proposed strategy seeks to give a more advanced, adaptable, and transparent solution to the persistent issue of credit card fraud detection.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

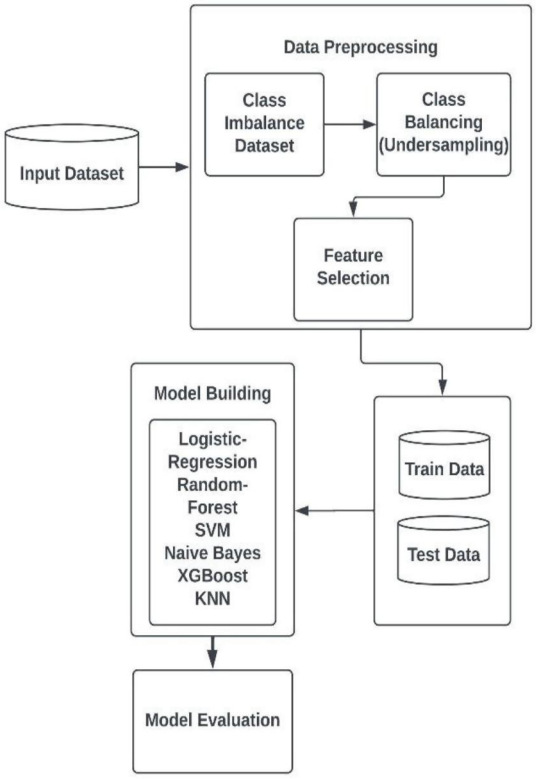


Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

MODULES

1) Data preprocessing: This module cleans, transforms, and enhances credit card transaction data. It encompasses scaling numerical qualities, encoding categorical variables, and dealing with missing data. Feature engineering approaches can be used to extract relevant data for machine learning systems.

2) Training Machine Learning Models: The system's core module trains many machine learning algorithms, such as XGBoost, Random Forest, SVM, KNN, and Naïve Bayes. Historical data is utilized to train the models to recognize legal and fraudulent transactions.

3) Anomaly Detection: The purpose of this module is to discover credit card transaction anomalies that may indicate fraud. It increases the complexity of fraud detection by employing specific algorithms designed to detect aberrant patterns or departures from the norm.

4) User Interface and Reporting: A user interface module is an integral aspect of the system that allows you to communicate with it, enter settings, and receive outputs. It may also generate detailed reports and visualizations to aid in the interpretation of the results. This module ensures that decision-makers, analysts, and other key stakeholders can simply navigate the system.

VI. RESULTS AND DISCUSSION

We used Naïve Bayes, SVM, Random Forest, KNN, Logistic Regression, and XGBoost algorithms to conduct our study. We used Jupyter Notebook to implement Python code, and the sklearn library was used. To achieve accurate findings and a well-fitting model, we trained our model using the techniques given in Section 3, verified it using KFold cross validation, and assessed it using the performance metrics listed in Table 1. After evaluating each algorithm's performance, we observed that SVM, Random Forest, and XGBoost had the highest accuracy levels.

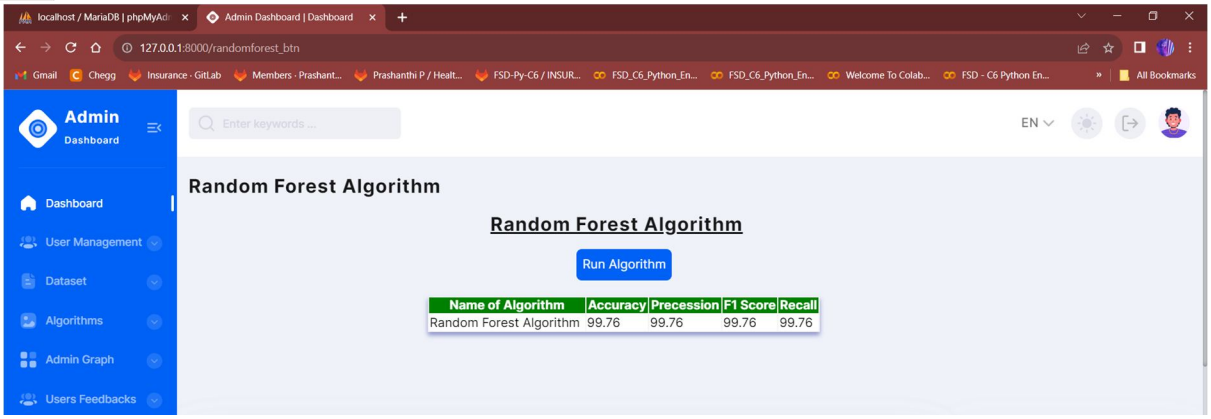


Fig 2. Proposed Algorithm Accuracy

Fig 3. Fraud Detection Based on the Transactions data

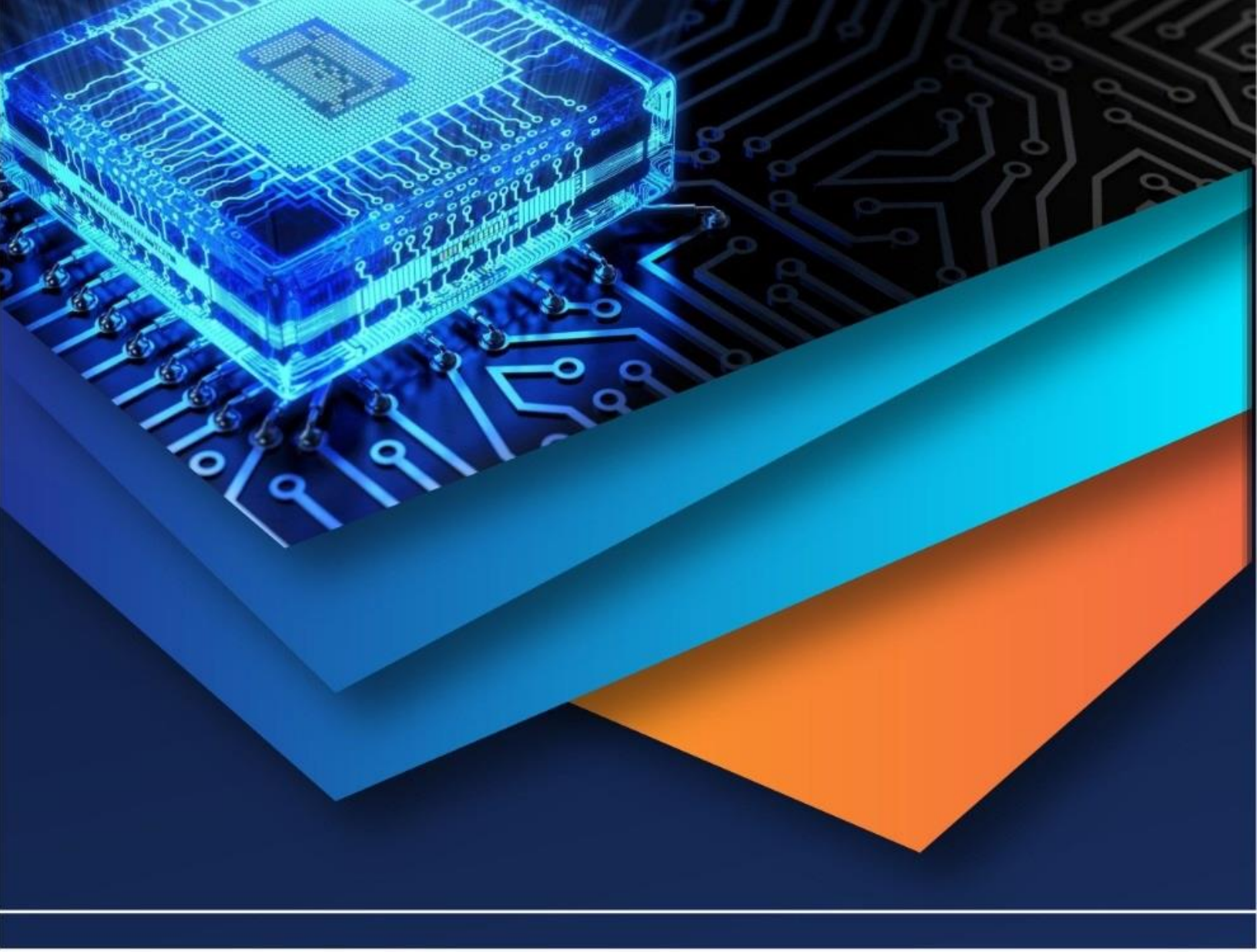
VII. CONCLUSION AD FUTURE WORK

The primary goal is to create a dependable and accurate machine learning model that can detect and classify credit card fraud. This paper utilized various techniques, including Naïve Bayes, SVM, Random Forest, KNN, Logistic Regression, and XGBoost. XGBoost, Random Forest, and SVM all achieved the highest accuracy, with SVM having the best model fit. As a result, we discovered that SVM produces more accurate classification results for both fraudulent and valid credit card transactions. Ensemble approaches, deep learning tactics, and improved data pretreatment through improved data cleaning and data balancing procedures may be used to improve this research and produce more exact results in future investigations.

REFERENCES

[1] Nihar Ranjan, Sneha George, Pallavi Pathade, Rakshita Nikhindi, Sneha Kamble, “Implementation of Machine Learning Algorithm to Detect Credit Card Frauds,” International Journal of Computer Applications (0975 – 8887) Volume 184 – No.1, March 2022.
[2] Malik, E.F.; Khaw, K.W.; Belaton, B.; Wong, W.P.; Chew, X, “Credit Card Fraud Detection Using a New Hybrid Machine Learning Architecture,” Mathematics 2022, 10, 1480.
[3] Noor Saleh Alfaiz * and Suliman Mohamed, “Enhanced Credit Card Fraud Detection Model Using Machine Learning,” Electronics 2022,11, 662.
[4] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba and G. Obaido, “A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection,” in IEEE Access, vol. 10, pp. 16400- 16407, 2022
[5] Rabiul Alam Bhuiyan , Mst. Shimu Khatun , Md. Taslim , Md.Alam Hossain, “American Journal of Multidisciplinary Research and Innovation (AJMRI)”, October 03, 2022.

[6] Md Shufyan, Dr. Prashant Prashun, "An Optimized Machine Learning Algorithms for Solving Class Imbalance Problem in Credit Card Fraud Detection," International Journal of Scientific Research & Engineering Trends, July-Aug-2022.
[7] Vikrant Chole, Ayan Mukherjee, Kshitij Gaikwad, Pradhnya Gawai, Pradhnya Bagde, Rati Mahule and Puja Pawar, "Revelation of Credit Card Fraud using Machine Learning Algorithm," International Journal for Modern Trends in Science and Technology, June 2022
[8] Shahnawaz Khan, Abdullah Alourani, Bharavi Mishra, Bharavi Mishra, Ashraf Ali, Mustafa Kamal, "Developing a Credit Card Fraud Detection Model using Machine Learning Approaches," International Journal of Advanced Computer Science and Applications, Vol. 13, No. 3, April 2022
[9] Undersampling and oversampling an old and a new approach [online] Available at: https://medium.com/analytics-vidhya/undersamplingand- oversampling-an-old-and-a-new-approach-4f984a0e8392
[10] D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith kumar, CH V N M praneeth, "Credit Card Fraud Detection Using Machine Learning," 2021 Fifth International Conference on ISMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)
[11] Manoj Kumar Reddy Mallidi, Yeshwanth Zagabathuni, "Analysis of Credit Card Fraud detection using Machine Learning models on balanced and imbalanced datasets," International Journal of Emerging Trends in Engineering Research, July 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY