Analysis of Optimization Techniques for Credit Card Fraud Detection

SHIKHAR AGRAHARI[1], TANU[2], ARUN KUMAR[3], ABHINAY PATEL[4], ABHINAV CHAUHAN[5]
[1, 2, 3, 4, 5] Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad

Abstract— In Today's scenario where we are promoting and going cashless to protect our cash currency from thefts. We are using credit cards and other online transaction methods. Now thefts and criminals find new ways to stole our money by doing fraud from our credit cards. To minimize these frauds we use computer programs that can learn from data to help catch fraud and aware the users from it. In our study, we looked at different ways to use these computer programs to find fraud and keep people's information safe. This review paper provides a comprehensive overview of the current state of credit card fraud detection techniques. It examines the various machine learning approaches used to identify and mitigate fraudulent credit card transactions, including supervised and unsupervised learning algorithms, as well as emerging trends and future research directions in this field.

Index Terms- Smart Computers, Credit Card Fraud, Working Together, Privacy, New Idea.

I. INTRODUCTION

In the 21st century, internet banking has made banking services accessible, with electronic payment methods becoming essential for modern transactions. Credit and debit cards, offered by financial institutions, simplify shopping and provide protection against loss or theft. However, this convenience has also attracted fraudsters who exploit global transactions, leading to significant security challenges. Credit card fraud, where unauthorized individuals use stolen card information for illicit gain, poses a substantial threat. Vulnerabilities in online platforms and identity theft schemes contribute to this issue. According to a study by "U.K. finance," credit and debit fraud cases reported in the U.K. totaled £574.2 million in 2020. As a result, there has been a significant increase in fraud detection efforts in recent years.

Credit card fraud presents a significant challenge, involving the illicit use of payment cards for unauthorized transactions. It's a method employed to acquire goods or funds unlawfully, posing risks to businesses and organizations. In practical terms, fraud detection systems (FDS) monitor transactions, flagging those deemed suspicious for further investigation. However, due to the sheer volume of transactions, investigators can only review a fraction of alerts daily, leaving many unchecked until customers report them. Moreover, fraud techniques evolve over time, leading to what's known as concept drift, making fraud identification even more challenging.

Machine learning stands out as an effective tool for fraud detection, employing classification and regression methods to recognize fraudulent credit card activities. Machine learning algorithms fall into two categories: supervised and unsupervised learning. Supervised learning utilizes labeled transactions for training classifiers, while unsupervised learning relies on peer group analysis to identify anomalies in spending behavior.

Various machine learning algorithms have been developed for credit card fraud detection, including neural networks, logistic regression, decision trees, Naive Bayes, support vector machines, k-nearest neighbors, and random forest. This paper evaluates the performance of these algorithms in classifying transactions as authorized or fraudulent, comparing them based on metrics such as accuracy, specificity, and precision.

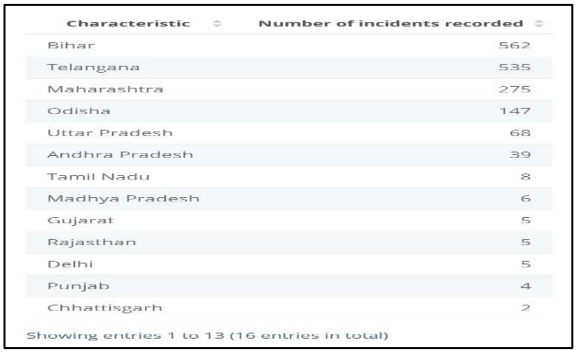
II. PROBLEM STATEMENT

Figure 1 shows that, Card frauds are increasing day by day which also leads to the financial losses are increasing very drastically. Every year a typical organization losses almost 5% of their yearly revenues, according to the some recent statistic reports on card fraud:

In 2022, the global financial losses due to card payments were 34 billion dollars, but the actual losses were 127 billon dollars after accounting for other costs.

In 2023, Card fraud cases are increased by 53% as compared to 2019.

In 2022, Over 1600 cases are reported for credit and debit card fraud, where Bihar having the highest number of cases.



Characteristic	Number of incidents recorded
Bihar	562
Telangana	535
Maharashtra	275
Odisha	147
Uttar Pradesh	68
Andhra Pradesh	39
Tamil Nadu	8
Madhya Pradesh	6
Gujarat	5
Rajasthan	5
Delhi	5
Punjab	4
Chhattisgarh	2

Showing entries 1 to 13 (16 entries in total)

Figure 1: Number Cases registered against card frauds.

In 2022-23, the number of frauds reported by banks in India increased from 7,263 in 2020-21 to 13,576, but the amount involved decreased from Rs 1,18,417 crore to Rs 26,632 crore.

These reports indicates the need of robust system that detect fraud and thefts due to card payments

III. LITERATURE REVIEW

[1] Noor Saleh Alfaiz and Suliman Mohamed Fati "Enhanced Credit Card Fraud Detection Model Using Machine Learning'' MDPI | Published: 21 February 2022

The COVID-19 pandemic has limited people's mobility to a certain extent, making it difficult to purchase goods and services offline, which has led the creation of a culture of increased dependence on online services. One of the crucial issues with using credit cards is fraud, which is a serious challenge in the realm of online transactions. Consequently, there is a huge need to develop the best approach possible to using machine learning in order to prevent almost all fraudulent credit card transactions. This paper studies a total of 66 machine learning models based on two stages of evaluation. A real-world credit card fraud detection dataset of European cardholders is used in each model along with stratified K-fold cross-validation. In the first stage, nine machine learning algorithms are tested to detect fraudulent transactions. The best three algorithms are nominated to be used again in the second stage, with 19 resampling techniques used with each one of the best three algorithms. Out of 330 evaluation metric values that took nearly one month to obtain, the All K-Nearest Neighbors (AllKNN) undersampling technique along with CatBoost (AllKNN-CatBoost) is considered to be the best proposed model. Accordingly, the AllKNN-CatBoost model is compared with related works. The results indicate that the proposed model outperforms previous models with an AUC value of 97.94%, a Recall value of 95.91%, and an *F*1-Score value of 87.40%.

[2] Abdullah Alharbi, Majid Alshammari , Ofonime Dominic Okon, Amerah Alabrah, Hafiz Tayyab Rauf 4, Hashem Alyami and Talha Meraj "A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach" MDPI | Published: 1 March 2022.

Online sales and purchases are increasing daily, and they generally involve credit card transactions. This not only provides convenience to the end-user but also increases the frequency of online credit card fraud. In the recent years, in some countries, this fraud increase has led to an exponential increase in credit card fraud detection, which has become increasingly important to address this security issue. Recent studies have proposed machine learning (ML)-based solutions for detecting fraudulent credit card transactions, but their detection scores still need improvement due to the imbalance of classes in any given dataset. Few approaches have achieved exceptional results on different datasets. In this study, the Kaggle dataset was used to develop a deep learning (DL)-based approach to solve the text data problem. A novel text2IMG conversion technique is proposed that generates small images. The images are fed into a CNN architecture with class weights using the inverse frequency method to resolve the class imbalance issue. DL and ML approaches were applied to verify the robustness and validity of the proposed system. An accuracy of 99.87% was achieved by Coarse-KNN using deep features of the proposed CNN.

[3] Rejwan Bin Sulaiman, Vitaly Schetinin, Paul Sant "Review of Machine Learning Approach on Credit Card Fraud Detection" Springer | Published : 5 May 2022

Massive usage of credit cards has caused an escalation of fraud. Usage of credit cards has resulted in the growth of online business advancement and ease of the e-payment system. The use of machine learning (methods) are adapted on a larger scale to detect and prevent fraud. ML algorithms play an essential role in analysing customer data. In this research article, we have conducted a comparative analysis of the literature review considering the ML techniques for credit card fraud detection (CCFD) and data confidentiality. In the end, we have proposed a hybrid solution, using the neural network (ANN) in a federated learning framework. It has been observed as an effective solution for achieving higher accuracy in CCFD while ensuring privacy.

[4] Eyad Abdel Latif Marazqah Btoush, Xujuan Zhou, Raj Gururajan, Ka Ching Chan, Rohan Genrich and Prema Sankaran "A systematic review of literature on credit card cyber fraud detection using machine and deep learning" PeerJ Computer Science | Published : 17 April 2023.

Credit card cyber fraud is a major security risk worldwide. Conventional anomaly detection and rule-based techniques are two of the most common utilized approaches for detecting cyber fraud, however, they are the most time-consuming, resource-intensive, and inaccurate. Machine learning is one of the techniques gaining popularity and playing a significant role in this field. This study examines and synthesizes previous studies on the credit card cyber fraud detection. This review focuses specifically on exploring machine learning/deep learning approaches. In our review, we identified 181 research articles, published from 2019 to 2021. For the benefit of researchers, review of machine learning/deep learning techniques and their relevance in credit card cyber fraud detection is presented. Our review provides direction for choosing the most suitable techniques. This review also discusses the major problems, gaps, and limits in detecting cyber fraud in credit card and recommend research directions for the future. This comprehensive review enables researchers and banking industry to conduct innovation projects for cyber fraud detection.

[5] S P Maniraj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed "Credit Card Fraud Detection using Machine Learning and Data Science" International Journal of Engineering Research & Technology (IJERT) | Vol. 8 Issue 09, September-2019.

It is vital that credit card companies are able to identify fraudulent credit card transactions so that customers are not charged for items that they did not purchase. Such problems can be tackled with Data Science and its importance, along with Machine Learning, cannot be overstated. This project intends to illustrate the modelling of a data set using machine learning with Credit Card Fraud Detection. The Credit Card Fraud Detection Problem includes modelling past credit card transactions with the data of the ones that turned out to be fraud. This model is then used to recognize whether a new transaction is fraudulent or not. Our objective here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications. Credit Card Fraud Detection is a typical sample of classification. In this process, we have focused on analysing and pre-processing data sets as well as the deployment of multiple anomaly detection algorithms such as Local Outlier Factor and Isolation Forest algorithm on the PCA transformed Credit Card Transaction data.

[6] Vaishnavi Nath Dornadula, Geetha Sa "Credit Card Fraud Detection using Machine Learning Algorithms" International Conference on Recent Trends in Advanced Computing(ICRTAC) | Published : 2019.

Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds. Increase in fraud rates, researchers started using different machine learning methods to detect and analyse frauds in online transactions. The main aim of the paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyse the past transaction details of the customers and extract the behavioural patterns. Where cardholders are clustered into different groups based on their transaction amount. Then using sliding window strategy [1], to aggregate the transaction made by the cardholders from different groups so that the behavioural pattern of the groups can be extracted respectively. Later different classifiers [3],[5],[6],[8] are trained over the groups separately. And then the classifier with better rating score can be chosen to be one of the best methods to predict frauds. Thus, followed by a feedback mechanism to solve the problem of concept drift [1]. In this paper, we worked with European credit card fraud dataset.

[7] Ameer Saleh Hussein, Rihab Salah Khairy, Shaima Miqdad Mohamed Najeeb, Haider Th. Salim ALRikabi "Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression" International Journal of Interactive Mobile Technologies (iJIM) – eISSN: 1865-7923 | Published : 2021.

The global online communication channel made possible with the internet has increased credit card fraud leading to huge loss of monetary fund in their billions annually for consumers and financial institutions. The fraudsters constantly devise new strategy to perpetrate illegal transactions. As such, innovative detection systems in combating fraud are imperative to curb these losses. This paper presents the combination of multiple classifiers through stacking ensemble technique for credit card fraud detection. The fuzzy-rough nearest neighbor and sequential minimal optimization are employed as base classifiers. Their combined prediction becomes data input for the meta-classifier, which is logistic regression resulting in a final predictive outcome for improved detection. Simulation results compared with seven other algorithms affirms that ensemble model can adequately detect credit card fraud with detection rates of 84.90% and 76.30%.

[8] K.Ratna Sree Valli , P.Jyothi , G.Varun Sai , R.Rohith Sai Subash "Credit card fraud detection using Machine learning algorithms" Journal of Research in Humanities and Social Science | Published : 2020.

Due to a rapid advancement in the electronic commerce technology, the use of credit cards has dramatically increased. Since credit card is the most popular mode of payment, the number of fraud cases associated with it is also rising.Thus, in order to stop these frauds we need a powerful fraud detection system that detects it in an accurate manner. In this paper we have explained the concept of frauds related to credit cards.Here we implement different machine learning algorithms on an imbalanced dataset such as logistic regression, naïvebayes,random forest with ensemble classifiers using boosting technique. An extensive review is done on the existing and proposed models for credit card fraud detection and has done a comparative study on these techniques. So Different classification models are applied to the data and the model performance is evaluated on the basis of quantitative measurements such as accuracy, precision, recall, f1 score, support, confusion matrix. The conclusion of our study explains the best classifier by training and testing using supervised techniques that provides better solution.

[9] Mr. Thirunavukkarasu.M1 ; Achutha Nimisha2 ; Adusumilli Jyothsna "CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING" International Journal of Computer Science and Mobile Computing | Published : 2021.

This Project is focused on credit card fraud detection in real world scenarios. Nowadays credit card frauds are drastically increasing in number as compared to earlier times. Criminals are using fake identity and various technologies to trap the users and get the money out of them. Therefore, it is very essential to find a solution to these types of frauds. In this proposed project we designed a model to detect the fraud activity in credit card transactions. This system can provide most of the important features required to detect illegal and illicit transactions. As technology changes constantly, it is becoming difficult to track the behavior and pattern of criminal transactions. To come up with the solution one can make use of technologies with the increase of machine learning, artificial intelligence and other relevant fields of information technology; it becomes feasible to automate this process and to save some of the intensive amounts of labor that is put into detecting credit card fraud. Initially, we will collect the credit card usage data-set by users and classify it as trained and testing dataset using a random forest algorithm and decision trees. Using this feasible algorithm, we can analyze the larger data-set and user provided current data-set. Then augment the accuracy of the result data. The performance of the techniques is gauged based on accuracy, sensitivity, and specificity, precision. The results is indicated concerning the best accuracy for Random Forest are unit 98.6% respectively. Keywords— Random forest algorithm, Criminal transactions, Credit card.

IV. METHODOLOGY

A. Explanation of the Approach:

As Shown in Figure 2, Credit card fraud detection using machine learning involves collecting transaction data, preprocessing for consistency, engineering features for pattern recognition, and training models like logistic regression or neural networks. Evaluation metrics guide model tuning, with ensemble methods boosting accuracy. Real-time monitoring employs the model to flag potentially fraudulent transactions, with continuous learning adapting to evolving fraud patterns. A feedback loop refines the system, and a combination of rule-based and ML approaches enhances accuracy while minimizing false positives.

The methodology adopted in this project to classify the non-fraudulent transactions and the fraudulent transactions. Figure 3.1 shows the steps used in this work. However, before we discuss the different steps of the methodology used in this work, we first discussed the dataset.

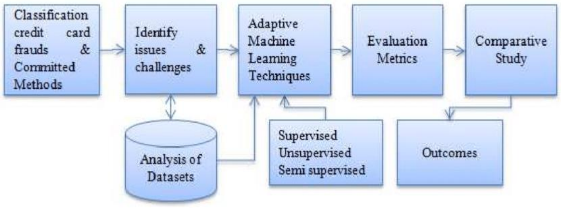

Classification credit card frauds & Committed Methods
Identify issues & challenges
Adaptive Machine Learning Techniques
Evaluation Metrics
Comparative Study
Analysis of Datasets
Supervised Unsupervised Semi supervised
Outcomes

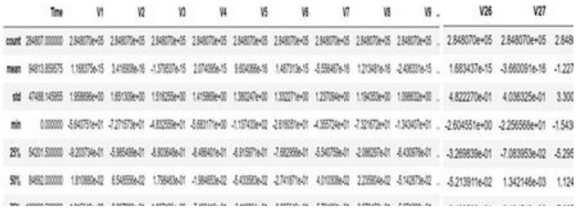
Figure. 2: Methodology for card fraud detection system

B. Dataset:
As Shown figure 3, the dataset for this project work is obtained from Kaggle.

The dataset contains transactions made by credit cards in September 2019 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have notbeen transformed with PCA are 'Time' and 'Amount'. Feature 'Time' contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Given the class imbalance ratio, we recommend measuring the accuracy using the Area Under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not meaningful for unbalanced classification.



	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V26	V27	
count	284807.000000	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	...	2.848070e+05	2.848070e+05	2.848
mean	94813.859575	1.168375e-15	3.416908e-16	-1.379537e-15	2.074095e-15	9.604066e-16	1.487313e-15	-5.556467e-16	1.213481e-16	-2.406331e-15	...	1.683437e-15	-3.660091e-16	-1.227
std	47488.145955	1.958696e+00	1.651309e+00	1.516255e+00	1.415869e+00	1.380247e+00	1.332271e+00	1.237094e+00	1.194353e+00	1.098632e+00	...	4.822270e-01	4.036325e-01	3.300
min	0.000000	-5.640751e+01	-7.271573e+01	-4.832559e+01	-5.683171e+00	-1.137433e+02	-2.616051e+01	-4.355724e+01	-7.321672e+01	-1.343407e+01	...	-2.604551e+00	-2.256568e+01	-1.543
25%	54201.500000	-9.203734e-01	-5.985499e-01	-8.903648e-01	-8.486401e-01	-6.915971e-01	-7.682956e-01	-5.540759e-01	-2.086297e-01	-6.430976e-01	...	-3.269839e-01	-7.083953e-02	-5.295
50%	84692.000000	1.810880e-02	6.548556e-02	1.798463e-01	-1.984653e-02	-5.433583e-02	-2.741871e-01	4.010308e-02	2.235804e-02	-5.142873e-02	...	-5.213911e-02	1.342146e-03	1.124
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]

Figure. 3: Features of data

C. Machine Learning Technique Used:
Credit Card Fraud Detection using Machine Learning can be done using –

➢ Unsupervised Learning :

a. K-Means Clustering:
K-Means can identify clusters of normal and potentially fraudulent transactions based on features like transaction amount, location, and time.

b. Isolation Forest:
Isolation Forest is an anomaly detection algorithm that isolates instances in a dataset by randomly selecting features to create isolation trees.

It is efficient and effective in detecting outliers.

➢ Supervised Learning:

a. Logistic Regression:
It is a widely used algorithm for binary classification problems.
Logistic regression models the probability that a given instance belongs to a particularclass.

b. Decision Trees:
Decision trees can be used for both classification and regression tasks.
They are interpretable and can capture complex decision boundaries.

c. Random Forest:

Random Forest is an ensemble method that builds multiple decision trees and combines their Predictions.

It improves accuracy and helps reduce over fitting.

V. DESIGN AND IMPLEMENTATION OF ALGORITHMS

As shown in figure 4, In the process followed to detect the fraud with the help of models which has been trained and tested and after testing the statistical analysis of accuracy, recall and F-1 score is been done to find which machine learning algorithm is best for prediting the fraud.

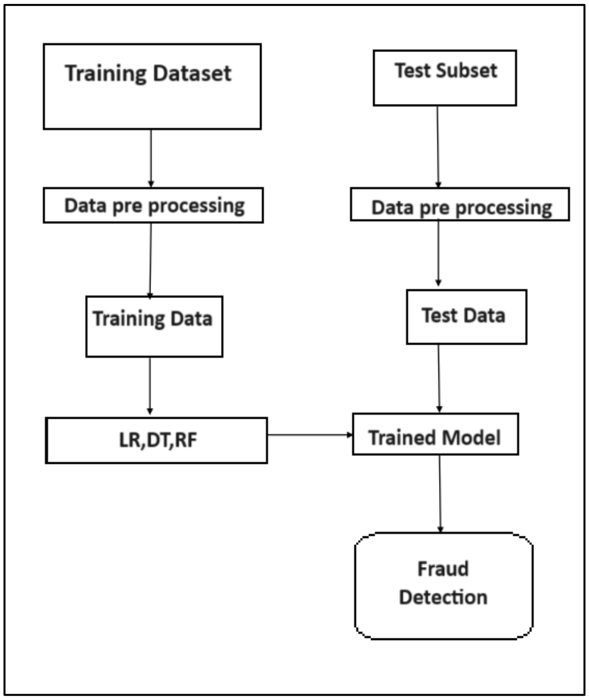
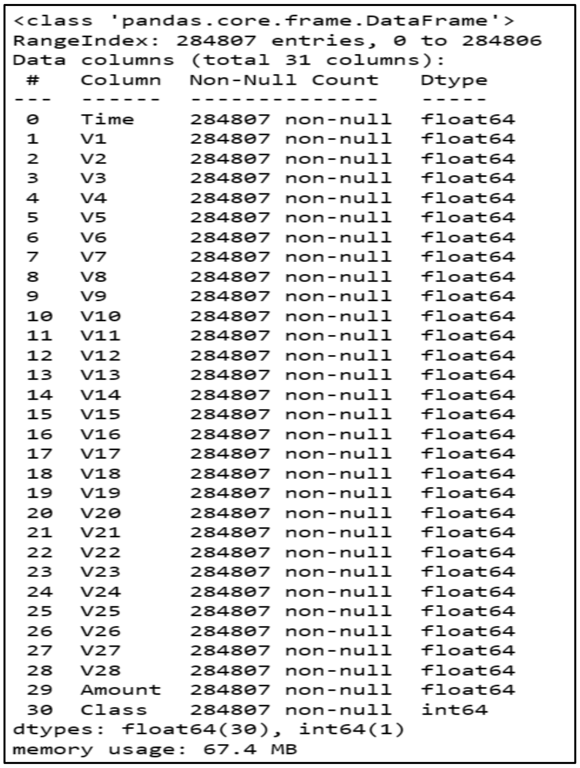


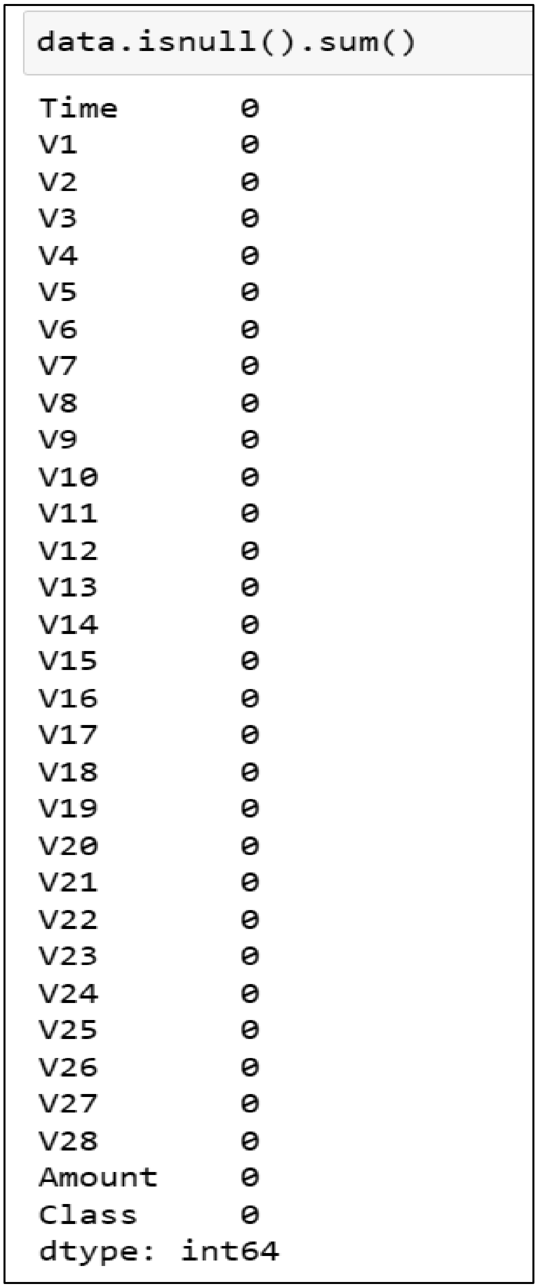
Figure 4: Architecture of the proposed methedology

Before performing the training and testing it is needed to be cleaned, balanced and optimized, as shown in figure 5



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
 #   Column  Non-Null Count   Dtype
---  ------  --------------   -----
 0   Time    284807 non-null  float64
 1   V1      284807 non-null  float64
 2   V2      284807 non-null  float64
 3   V3      284807 non-null  float64
 4   V4      284807 non-null  float64
 5   V5      284807 non-null  float64
 6   V6      284807 non-null  float64
 7   V7      284807 non-null  float64
 8   V8      284807 non-null  float64
 9   V9      284807 non-null  float64
 10  V10     284807 non-null  float64
 11  V11     284807 non-null  float64
 12  V12     284807 non-null  float64
 13  V13     284807 non-null  float64
 14  V14     284807 non-null  float64
 15  V15     284807 non-null  float64
 16  V16     284807 non-null  float64
 17  V17     284807 non-null  float64
 18  V18     284807 non-null  float64
 19  V19     284807 non-null  float64
 20  V20     284807 non-null  float64
 21  V21     284807 non-null  float64
 22  V22     284807 non-null  float64
 23  V23     284807 non-null  float64
 24  V24     284807 non-null  float64
 25  V25     284807 non-null  float64
 26  V26     284807 non-null  float64
 27  V27     284807 non-null  float64
 28  V28     284807 non-null  float64
 29  Amount  284807 non-null  float64
 30  Class   284807 non-null  int64
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

Figure 5: Dataset Structure

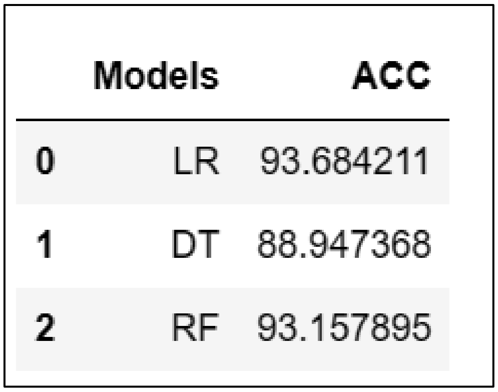


```
data.isnull().sum()
```

```
Time      0
V1        0
V2        0
V3        0
V4        0
V5        0
V6        0
V7        0
V8        0
V9        0
V10       0
V11       0
V12       0
V13       0
V14       0
V15       0
V16       0
V17       0
V18       0
V19       0
V20       0
V21       0
V22       0
V23       0
V24       0
V25       0
V26       0
V27       0
V28       0
Amount    0
Class     0
dtype: int64
```

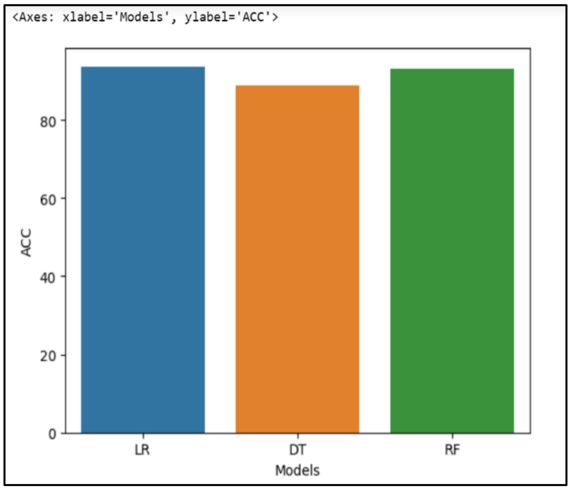
Figure 6: Null Attributes In Dataset

First the dataset is to be cleaned by removing the null rows and excluding the duplicate entries. After this it needed to balance the imbalanced data in dataset so first the undersampling is used to balance the data were the data of legitimate payment are reduced as equal to the fraud payment data. After undersampling all the machine learning algorithms (Logistic Regression, Decision Tree Classifier, Random Forest classifier) are been applied.



	Models	ACC
0	LR	93.684211
1	DT	88.947368
2	RF	93.157895

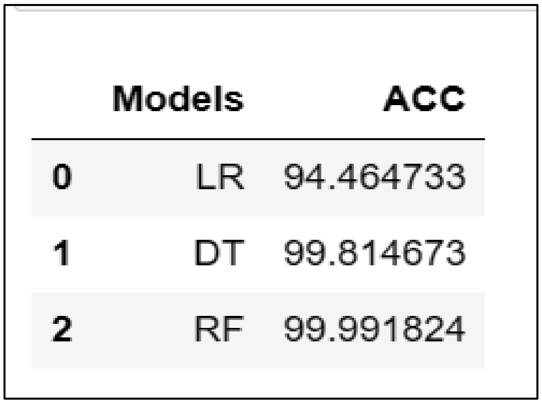
Figure 7: Accuracy Of ML Algorithm After Undersampling



```
<Axes: xlabel='Models', ylabel='ACC'>
```

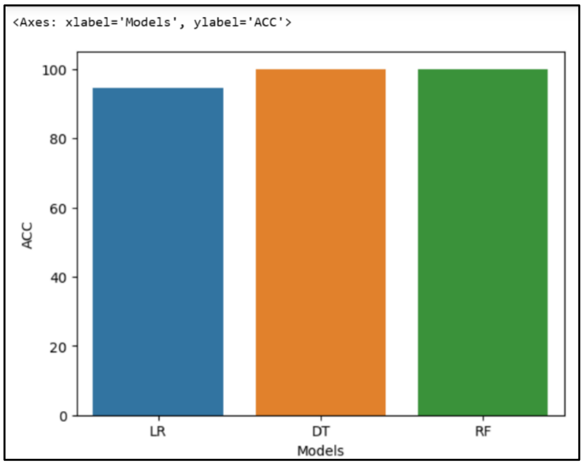
Figure 8: Statistical representation of accuracy of ML algorithms.

After this oversampling is done were the data of fraud payments are made equal to the legitimate payment data using standardization and normalization. Afte oversampling all the machine learning algorithms (Logistic Regression, Decision Tree Classifier, Random Forest classifier) are been applied. After performing all the above operations a statistics analysis is done to find which machine learning algorithm along with optimization is giving best result.



	Models	ACC
0	LR	94.464733
1	DT	99.814673
2	RF	99.991824

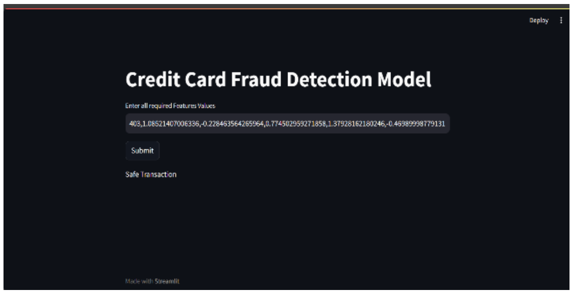
Figure 9: Accuracy of ML algorithm after oversampling



```
<Axes: xlabel='Models', ylabel='ACC'>
```

Figure 10 : Statistical representation of accuracy of ML algorithms.

Anaconda navigator is used as it is having several IDE's installed in it python programming language is used to implement machine learning algorithms as it is easy to learn and implement. In this project Jupyter notebook is used to process the complete code where the code can be viewed as block of codes and running each section and identifying the errors is easier. A streamlit web app is developed for predicting the result in user have to provide all the feature of payments.



Credit Card Fraud Detection Model

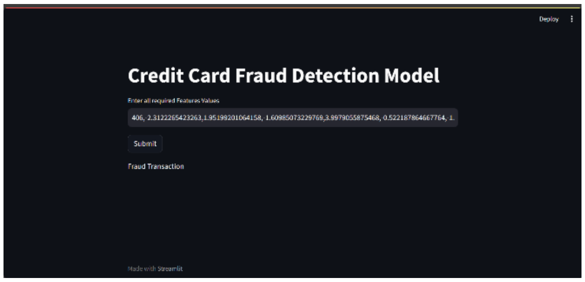
Enter all required Features Values

403,1.08521407006336,-0.228463564265964,0.774502959271858,1.37928162180246,-0.46989998779131

Safe Transaction

Made with Streamlit

Figure 11: representation of web app showing safe transaction



Credit Card Fraud Detection Model

Enter all required Features Values

406,-2.3122265423263,1.95199201064158,-1.60985073229769,3.9979055875468,-0.522187864667764,-1.

Fraud Transaction

Made with Streamlit

Figure 12: representation of web app showing fraud transaction

VI. RESULTS AND DISCUSSION

Referring to Table 1, the first stage provides an initial assessment of how well each machine learning algorithm performs. These algorithms are Logistic Regression, Decision Tree, Random Forest.

Classifier	Metrics Accuracy Precision
Logistic Regression	0.947 0.996
Decision Tree	0.998 0.913
Random Forest	0.999 0.997

It's evident from the Table 1 that Random Forest outperforms significantly in other terms of algorithms accuracy. Additionally, Random Forest exhibits the highest precision and accuracy among all algorithms, with Logistic Regression and Decision Tree following suit. Consequently, implementing the proposed system with Random Forest is expected to yield superior accuracy, particularly when dealing with a larger volume of training data.

CONCLUSION

The technology change influenced several improvements. We are talking about online transactions done through credit cards, which leads to credit card frauds, and this study is about improving machine learning algorithms for fraud detection. In this study, we put forth fraud detection methods based on supervised learning such as Random Forest, Decision Tree, and logistic regression, We compared all the algorithms with different datasets by first using the original dataset itself; we then use resampling techniques such as undersampling and oversampling because our dataset is highly imbalanced. Finally, we concluded that Random Forest would be the perfect fit for our model. It can be inferred that oversampling works better because the smaller number of observations helps in training our model efficiently. Oversampling will be an ideal sampling technique in the real-world scenario as the information containing a pattern is not lost.

The synergy between advanced machine learning models and optimization methods contributes to the development of more robust and efficient fraud detection systems.Each model in both stages is evaluated based on the Accuracy, Recall, Precision, and F1-Score. In the first stage, algorithms applied are Logistic Regression (LR), Decision Tree (DT), Random Forest (RF).In the second stage, the 5 resampling techniques are divided as follows: 1 undersampling, 2 oversampling, and 2 combinations of both undersampling and oversampling techniques. The best model out of all these is Ramdom forest classifier along Data preprocessing, Class balancing, Feature selection, Emsembling. Finally, Ramdom forest classifier is compared with previous works with the same dataset and similar approaches. Indeed, Ramdom forest classifier outperforms previous models in terms of Accuracy (0.999918238308078), Recall (1.0), and F1-Score (0.9999181929736854).
Our approach involves training the Random Forest classifier using feedback and delayed supervised samples. Subsequently, we aggregate the probabilities generated by the classifier to identify potential alerts. Moreover, we propose the utilization of a learning to rank approach, whereby alerts are ranked according to priority. The suggested method offers solutions to challenges such as class imbalance and concept drift in fraud detection. In future research endeavors, we plan to explore the application of semisupervised learning methods for classifying alerts within Fraud Detection Systems (FDS), further enhancing the efficacy and robustness of our approach.

VII. FUTURE SCOPE

Future work in credit card fraud detection should explore advanced optimization algorithms (e.g., Monarch Butterfly, Earthworm, Elephant Herding), anomaly detection techniques (including deep learning and unsupervised methods), real-time detection, and response systems. Additionally, researchers can focus on improving model interpretability, experimenting with ensemble methods, integrating behavioral biometrics, addressing adversarial robustness, leveraging blockchain technology, and implementing cross-channel analysis. Continuous model evaluation and updating strategies are crucial for adapting to evolving fraud patterns. Exploring these areas will contribute to the ongoing enhancement of credit card fraud detection systems in the dynamic landscape of digital transactions.

While we couldn’t reach out goal of 100% accuracy in fraud detection, we did end up creating a system that can, with enough time and data, get very close to that goal. As with any such project, there is some room for improvement here. The very nature of this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result. This model can further be improved with the addition of more algorithms into it. However, the output of these algorithms needs to be in the same format as the others. Once that condition is satisfied, the modules are easy to add as done in the code. This provides a great degree of modularity and versatility to the project. More room for improvement can be found in the dataset. As demonstrated before, the precision of the algorithms increases when the size of dataset is increased. Hence, more data will surely make the model more accurate in detecting frauds and reduce the number of false positives. However, this requires official support from the banks themselves.

VIII. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who have contributed to the completion of this research paper.

First and foremost, we are deeply indebted to our advisor, Mrs. Meenu Sharma, whose expertise, understanding, and patience added considerably to our graduate experience. We appreciate her vast knowledge and skill in many areas, and her assistance in writing reports and code.

Special thanks are due to IMS Engineering College Ghaziabad and AKTU University for providing the necessary resources and environment for conducting our research.

Lastly, We wish to thank all the participants of our study, without whom this research would not have been possible.

Thank you all for your invaluable contributions.

REFERENCES

[1] Noor Saleh Alfaiz and Suliman Mohamed Fati "Enhanced Credit Card Fraud Detection Model Using Machine Learning'' MDPI , 21 February 2022.

[2] S P Maniraj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed "Credit Card Fraud Detection using Machine Learning and Data Science" International Journal of Engineering Research & Technology(IJERT) | Vol. 8 Issue 09, September-2019.

[3] Abdullah Alharbi, Majid Alshammari , Ofonime Dominic Okon, Amerah Alabrah, Hafiz Tayyab Rauf 4, Hashem Alyami and Talha Meraj "A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach" MDPI, 1 March 2022.

[4] Vaishnavi Nath Dornadula, Geetha Sa "Credit Card Fraud Detection using Machine Learning Algorithms" International Conference on Recent Trends in Advanced Computing(ICRTAC) , 2019.

[5] Rejwan Bin Sulaiman, Vitaly Schetinin, Paul Sant "Review of Machine Learning Approach on Credit Card Fraud Detection"Springer, 5 May 2022.

[6] Eyad Abdel Latif Marazqah Btoush, Xujuan Zhou, Raj Gururajan, Ka Ching Chan, Rohan Genrich and Prema Sankaran "A systematic review of literature on credit card cyber fraud detection using machine and deep learning" PeerJ Computer Science, 17 April 2023.

[7] Andrea Dal Pozzolo, Giacomo Boracchi, Olivier Caelen, Cesare Alippi and Gianluca Botempi, |Credit card Fraud Detection : A realistic Modeling and a Novel Learning Strategy|, IEEE Trans. on Neural Network and Learning system,vol.29,No.8, August 2018.

[8] Shiyang Xuan,Guanjun Liu,Zhenchuan Li,Lutao Zheng,Shuo Wang, Jiang,| Random Forest for credit card fraud detection|,Int.conf.on Networking,Sensing and control,2018.

[9] Y. Sahin , and Duman,E.,(2018) ―Detecting credit card fraud by ANN and logistic regression.| In Innovations in Intelligent Systems and Applications(INISTA),2018 international Symposium (pp.315-319).IEEE

[10] Google Developers. Classification: ROC Curve and AUC. Available online: https://developers.google.com/machine-learning/ crash-course/classification/roc-and-auc (accessed on 22 January 2022).

[11] Masís, S. Interpretable Machine Learning with Python: Learn to Build Interpretable High-Performance Models with Hands-On Real-World Examples; Packt Publishing Ltd.: Birmingham, UK, 2021; p. 81.

[12] Prusti, D.; Rath, S.K. Fraudulent transaction detection in credit card by applying ensemble machine learning techniques. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Rourkela, India, 6–8 July 2019; pp. 1–6. [CrossRef]

[13] Wang,G.-G.; Deb, S.; Cui, Z. Monarch butterfly optimization. Neural Comput. Appl. 2019, 31.[CrossRef]

[14] Ghosh, I.; Roy, P.K. Application of earthworm optimization algorithm for solution of optimal power flow. In Proceedings of the 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), Kolkata, India, 18–20 March 2019; pp. 1–6. [CrossRef]

[15] BORA MEHAR SRI SATYA TEJA1, BOOMIREDDY MUNENDRA2, Mr. S. GOKULKRISHNAN "A Research Paper on Credit Card Fraud Detection" International

Research Journal of Engineering and Technology (IRJET) | Published : 2022.

[16] Ameer Saleh Hussein, Rihab Salah Khairy, Shaima Miqdad Mohamed Najeeb, Haider Th. Salim ALRikabi "Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression" International Journal of Interactive Mobile Technologies (iJIM) – eISSN: 1865-7923 | Published : 2021