IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

Credit Card Fraud Detection

Prof. Priyanka. B. Palve[1], Miss. Lashkare Utkarsha[2], Miss. Vedika Kolhe[3], Miss. Sontakke Tejal[4], Miss. Vaishnavi Hivarkar[5]

Prof. 1st Year Engineering Department, Adsul's Technical Campus, Ahilyanagar, India [1]
Students, Computer Engineering Department, Adsul's Technical Campus, Ahilyanagar, India [2,3]
Student, Information Technology Engineering Department, Adsul's Technical Campus, Ahilyanagar, India [4]
Student, E&TC Engineering Department, Adsul's Technical Campus, Ahilyanagar, India [5]

Abstract: *Credit card frauds are one of the most common frauds happening now. Many companies have been increasing their payment modes to online, raising the threat for online frauds. Many fraudsters started using different methods to steal the money used to make the online transactions. So, our aim is to use different machine learning algorithms to check whether the transactions made are fraud or genuine. So, we will be categorising the transactions into different groups so that we can apply different machine learning algorithms on them. Then different classifiers will be trained over the groups independently. Then the best classifier with a good accuracy score will be used to predict the fraud transactions. In this paper we will be using a dataset containing. The dataset is a collection of online transactions made by some anonymous people using their credit cards. This dataset is very unstable i.e., it has a large portion of genuine transactions and a very small number of fraud transactions*

Keywords: Credit card fraud, applications of machine learning, data science, isolation forest algorithm, local outlier factor, automated fraud detection

I. INTRODUCTION

'Fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behaviour of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future.In other words, Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card issuing authorities are unaware of the fact that the card is being used. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting. This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it is characterized by various factors such as class imbalance. The number of valid transactions far outnumbers fraudulent ones. Also, the transaction patterns often change their statistical properties over the course of time. These are not the only challenges in the implementation of a real-world fraud detection system, however. In real world examples, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize. Machine learning algorithms are employed to analyse all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent. The investigators provide a feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time.

Credit card fraud means unauthorized operation of an account that is used to make transactions without the actual owner of the account or the bank authority's knowledge. We need to take necessary precautions while doing these transactions to avoid these frauds. Also, the bank authorities need to use the latest technologies to predict these frauds

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

so that they can alert their customers beforehand. Fraud detection means (for our dataset) is to predict the transactions that are made by the account holders which are actually done by other people who has access to the account. This is a very complex problem that needs the attention of the account holder as well as the bank authorities so that their other customers need not suffer from the same problem. But this problem has a problem of class imbalance. The number of genuine transactions done by a customer will be far higher than the fraud transactions happened or even be zero. Also, the customer can do a transaction that deviates from his previous transactions that can be misinterpreted as a fraud transaction. Also, the payment requests sent are checked by automatic tools that confirms which request need to be confirmed. These algorithms check these requests and report suspicious requests to professionals who operate behind and they in turn investigate them by contacting the owners of the accounts whether the transactions are genuine or not.

Problem Statement:

Credit cards are an essential financial tool that enables its holders to make purchases and the luxury of paying back the amount later. Credit card holders have an advantage of paying the amount back later after a certain time. This makes the credit cards an easy target for the fraudsters. Without the owner's knowledge a good amount of money can be withdrawn by these fraudsters and they make it look like the actual owners of these cards made the withdrawal. The fraudsters make does this very carefully and anonymously that makes it difficult to stop and even catch them. In 2017, there were data breaches and approximately 179 million records among which Credit card frauds were the most common form. With many frauds happening all over the world with credit card frauds on the top, this makes this a serious issue to look after. Credit card dataset is largely imbalanced because there will be more valid data compared with a fraudulent bone. Banks are now moving to EMV cards, which store their data on integrated circuits making some card payments safer, but still leaving non-card payment frauds on advanced rates. According to 2017, the US Payments Forum report, felons have loosened their focus on conditioning related to CNP deals as the security of chip cards were increased.

II. LITERATURE SURVEY

There were different techniques that were used to predict the fraud transactions like Outlier detection, unsupervised outlier detection, Peer group analysis and breakpoint analysis.

Outlier detection detects the abnormal transactions made by the user which are different in scale, rage and type of transaction compared to the previous transactions, but these types of transactions can be actually done by the customer and so the prediction can be wrong.

Unsupervised outlier detection on the other hand does predict the required data, It just simply understands the behaviour of the customer transactions. Peer group analysis is another method that has been used which involves the comparison of entities that share similar characteristics.

A breakpoint is a structural change in data like an anomaly. Breakpoint analysis simply is analysing these breakpoints to better understand their existence and occurrence. Although supervised learning methods are used in fraud detection there is a possibility that they fail at some cases.

Fraud act as the unlawful or criminal deception intended to result in financial or personal benefit. It is a deliberate act that is against the law, rule or policy with an aim to attain unauthorized financial benefit. Numerous literatures pertaining to anomaly or fraud detection in this domain have been published already and are available for public usage. A comprehensive survey conducted by Clifton Phua and his associates have revealed that techniques employed in this domain include data mining applications, automated fraud detection, adversarial detection. In another paper, Suman, Research Scholar, GJUS&T at Hisar HCE presented techniques like Supervised and Unsupervised Learning for credit card fraud detection. Even though these methods and algorithms fetched an unexpected success in some areas, they failed to provide a permanent and consistent solution to fraud detection.

A similar research domain was presented by Wen-Fang YU and Na Wang where they used Outlier mining, Outlier detection mining and Distance sum algorithms to accurately predict fraudulent transaction in an emulation experiment

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

of credit card transaction data set of one certain commercial bank. Outlier mining is a field of data mining which is basically used in monetary and internet fields. It deals with detecting objects that are detached from the main system i.e. the transactions that aren't genuine. They have taken attributes of customer's behaviour and based on the value of those attributes they've calculated that distance between the observed value of that attribute and its predetermined value. Unconventional techniques such as hybrid data mining/complex network classification algorithm is able to perceive illegal instances in an actual card transaction data set, based on network reconstruction algorithm that allows creating representations of the deviation of one instance from a reference group have proved efficient typically on medium sized online transaction. There have also been efforts to progress from a completely new aspect. Attempts have been made to improve the alertfeedback interaction in case of fraudulent transaction.

In case of fraudulent transaction, the authorised system would be alerted and a feedback would be sent to deny the ongoing transaction. Artificial Genetic Algorithm, one of the approaches that shed new light in this domain, countered fraud from a different direction.

It proved accurate in finding out the fraudulent transactions and minimizing the number of false alerts. Even though, it was accompanied by classification problem with variable misclassification costs.

III. PROPOSED SYSTEM

Card payments are always different when compared to former payments made by the client. This creates a problem called conception drift. Concept drift can be said as a variable which changes over time and in unlooked-for ways. These variables create a high imbalance in data. The main agenda of our exploration is to overcome the problem of Concept drift to apply on real- world script. In our proposed system we will be using different machine learning algorithms like Decision trees, Random Forest and other algorithms and calculate their accuracy scores and then choose the best algorithm with the best accuracy score. We will also calculate the confusion matrix for each of the algorithm and take that into consideration along with the accuracy score to choose the best algorithm. Also, we need to consider the fact that our data set that we are about to look at is very much imbalanced.

A. Dataset

The dataset comprises transactions made by European credit cards Holders in September 2013. This dataset presents deals that passed in two days, where we've 492 frauds out of deals. The dataset is largely unstable, the positive class (frauds) account for 0.172 of all deals. Features v1,v2,v3…v28 are the key features achieved with PCA, the only features which haven't been converted with PCA are' Time 'and' Quantum'. Point' Time 'contains the seconds ceased between each sale and the first sale in the dataset. The point' Quantum's the sale Quantum, this point can be used as the amount. Point Class is the response variable and it takes values 1and 0 for fraud and genuine respectively.

B. Steps & Implementation

Steps to develop the Classifier in Machine Learning

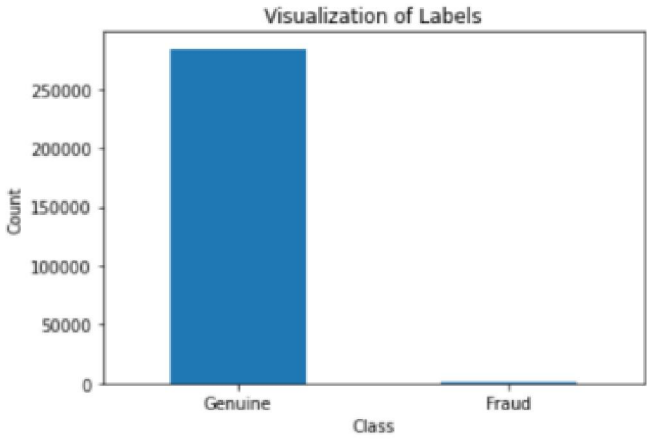
- Complete the Exploratory Data Analysis on the dataset,
- Apply different ML algorithms on our dataset,
- Train and evaluate the models to pick the best one

Step 1. Complete the Exploratory Data Analysis on the dataset

First we will import the required modules load the dataset and perform EDA on, it. Then we will make sure there are no null values in our dataset. The feature that we will be focusing is "Amount". Now, if we traverse the existence of each class tag and plot the data using matplotlib the plot will be as follows

We can observe from the above bar graph that the genuine transactions are over 99%. So, to avoid this problem we can apply the scaling techniques on the "Amount" feature to transform them to the range of values. We will remove the "Amount" column and add a new column with the scaled values in its place. We will also remove the "Time" column as it is not required.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2



Step 2: Use ML Algorithms to the Dataset

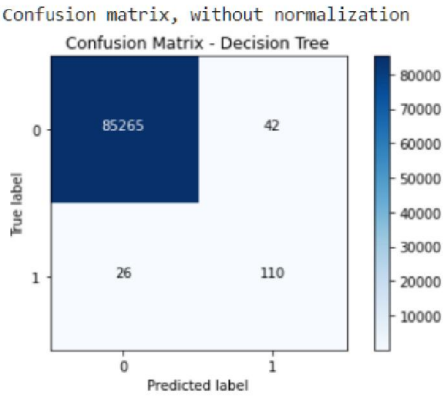
Let's use the Random Forest and Decision Tree Classifiers which are present in the sklearn package as RandomForestClassifier() and DecisionTreeClassifier() respectively.

```
# Decision Tree Classifier
from sklearn.tree import DecisionTreeClassifier
decision_tree = DecisionTreeClassifier()
decision_tree.fit(train_X, train_Y)

predictions_dt = decision_tree.predict(test_X)
decision_tree_score = decision_tree.score(test_X, test_Y) * 100
```

```
# Random Forest
from sklearn.ensemble import RandomForestClassifier
random_forest = RandomForestClassifier(n_estimators= 100)
random_forest.fit(train_X, train_Y)

predictions_rf = random_forest.predict(test_X)
random_forest_score = random_forest.score(test_X, test_Y) * 100
```

Step 3: Train and Evaluate the Models

Now, Let's train and evaluate the recently created models and pick the best one. Train the decision tree and random forest models using the fit() function. Note down the predictions made by the models using the predict () function and evaluate., Let's, visualize, the, scores, of, each, of, our, classifiers

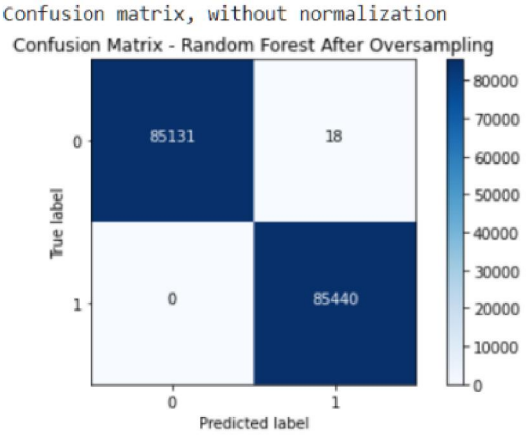


IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

The Random Forest classifier has somewhat an advantage over the Decision Tree classifier. Now we will calculate the accuracy,, precision,, recall,, and, f1-score for both of the classifiers by creating a function commonly used to calculate these values

IV. DISCUSSION & RESULT

Easily, Random Forest model works better than Decision Trees. But if we observe our dataset suffers a serious problem of class imbalance. The genuine (not fraud) deals are further than 99 with the fraud deals constituting of 0.17. With similar kind of distribution, if we train our model without taking care of the imbalance issues, it predicts the label with more significance given to genuine deals (as there are more data about them) and hence obtains further fragility. The class imbalance problem can be resolved by reasonable number of ways. Over slice is one of them. Finally, after oversampling the confusion matrix and the accuracy scores are calculated.



The evaluation metrics for the Random Forest model (after oversampling) are as follows:

Accuracy: 0.99989

Precision: 0.99979

Recall: 1.00000

F1-score: 0.99989

As we can see the accuracy scores of the Random Forest model after the oversampling which is done to avoid the class imbalance issue, is quite good and better than the different algorithm approaches. So we can say that the Random Forest algorithm does a good job of predicting the anomalies in a huge imbalanced dataset.

V. CONCLUSION

Credit card fraud is the biggest frauds that are being happened right now around the whole ground. This paper has explained how credit card frauds have been happening and we studied these frauds using a dataset that consists of transactions made in the real world. We saw how different machine learning algorithms are used to predict the fraud transactions on our dataset and we also addressed the class imbalance issue of our dataset and used oversampling to finally use Random Forest classifier that got a good accuracy score.

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results. While the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. However, when the entire dataset is fed into the algorithm, the precision rises to 33%. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions Since the entire dataset

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

consists of only two days' transaction records, its only a fraction of data that can be made available if this project were to be used on a commercial scale. Being based on machine learning algorithms, the program will only increase its efficiency over time as more data is put into it.

Future Scope:

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.

REFERENCES

[1], Credit Card Fraud Detection Based on Transaction Behavior -by John Richard D., Kho Larry A. Vea" published by Proc of the 2017 IEEE Region 10 Conference (TENCON) Malaysia November 5-8 2017

[2] CLIFTON PHUA1VINCENT LEE1 KATE SMITH1 & ROSS GAYLER2 " A Comprehensive Survey of Data, Mining-based Fraud Detection Research" published by School of Business System Faculty of Information Technology Monash University Wellington Road Clayton Victoria 3800 Australia,

[3] "Survey Paper on Credit Card Fraud Detection by Suman" Research ScholarGJUS&T Hisar HCE Sonepat published by International Journal of Advanced, Research, in, Computer, Engineering, &, Technology, (IJARCET), Volume, 3, Issue, 3,, March, 2014

[4] "Research on Credit, Card, Fraud, Detection, Model, , Based, on, Distance, Sum, –, by, Wen-Fang, YU, and, Na, Wang", published, by, 2009, International Joint Conference on Artificial Intelligence,\ [6]

[5] "Credit, Card, Fraud, Detection: A Realistic Modelling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS,, VOL., 29,, NO., 8,, AUGUST, 2018.

[6] "Credit Card Fraud Detection Based on Transaction Behaviour -by John Richard D. Kho, Larry A. Vea" published by Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[7] CLIFTON PHUA1, VINCENT LEE1, KATE SMITH1 & ROSS GAYLER2 " A Comprehensive Survey of Data Mining-based Fraud Detection Research" published by School of Business Systems, Faculty of Information Technology, Monash University, Wellington Road, Clayton, Victoria 3800, Australia

[8] "Survey Paper on Credit Card Fraud Detection by Suman" , Research Scholar, GJUS&T Hisar HCE, Sonepat published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

[9] "Research on Credit Card Fraud Detection Model Based on Distance Sum – by Wen-Fang YU and Na Wang" published by 2009 International Joint Conference on Artificial Intelligence

[10] "Credit Card Fraud Detection through Parenclitic Network AnalysisBy Massimiliano Zanin, Miguel Romance, Regino Criado, and SantiagoMoral" published by Hindawi Complexity Volume 2018, Article ID 5764370, 9 pages

[11] "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy" published by IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 29, NO. 8, AUGUST 2018

[12] "Credit Card Fraud Detection-by Ishu Trivedi, Monika, Mrigya, Mridushi" published by International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 1, January 2016