IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING: FRAUD DETECTION SYSTEM

[1]Ms. Gulnaj B. Sayyad, [2]Prof. Sushma Shinde,

[1]Student, Computer Engineering Dept., Siddhant College of Engineering, Sudumbre, Pune, India
[2]HoD, Department of Computer Engineering, Siddhant College of Engineering, Sudumbre, Pune, India

***Abstract*:** Financial services are being used everywhere and function with high complexity. With the increase in the usage of online modes for transacting throughout the world it is seen that the frauds too are increasing alarmingly in this sector. An automated Fraud Detection System is thus required to tackle this issue. Over the years, many techniques are being tried in order to efficiently tackle this issue. With millions of transaction taking place it is practically impossible to take care of this by manually checking for frauds.With that being said, speed and accuracy is needed while building such systems. Our system provides better accuracy rather than only works in these areas but also which in-turn will end up saving a lot of resources and the cost incurred. Our aim with this research is to provide a robust, cost effective, efficient yet accurate solution to find or detect frauds in both online payment transactions and payments that take place with credit cards. The proposed solution is a Machine Learning model that will serve the purpose of detecting 'Fraudulent' and 'Genuine' transactions in real time. This is very beneficent to all the sectors that are even mildly aligned to finance or make use of it. The solution will help them to analyse based on different factors if the ongoing transaction can be harmful and will prevent many unfortunate incidents.

Index Terms—fraud detection, credit card, online payments, fraudulent, genuine.

I. INTRODUCTION

Online Transactions Frauds are seen in abundance with the increase in the use of online methods within the government, private organizations and throughout various sectors in order to have faster payment nfrastructure. High reliance on technology has resulted in increased banking transactions in today's world. However, frauds in the banking sector have accelerated jointly as transactions online and offline. there is a lot of research being done in this area to combat against the ever increasing financial frauds online as transactions became a commonly used mode of payment, . The current systems which in place are not able to work efficiently to the pace at which the number of transactions is happening. Studies suggest 93% of the

merchants perform manual review anywhere between 1% - 10% of orders for fraud detection and this is costly, time consuming and it also leads to higher false negatives and human errors With online banking becoming an important & main platform for the whole sector there are a lot of challenges that need to be addressed. Users need the whole process to be secure but at the same time want it to be hassle free and want as few steps possible for authentication. The current systems with multilayer authentication are not able to combat the problem and identify if the transaction is authentic or not. The current systems in place may not be able to please the customer is comfort towards using the system but they provide a certain level of security and always have scope for improvement. The use of analytical models in order to identify malicious activities is a must with the increasing number of transactions and quantity of online banking services. We aim to analyze the data of transactions (Cash-In, Cash-Out etcetera) in both Online payments and credit card payments which consists both normal as well as fraud customer behavior. Using various machine learning techniques the classification model shall accurately classify transactions as genuine or fraudulent and provides speed, scalability and efficiency alongside accuracy. Businesses using the solution will in turn end up saving a lot of resources since they won't lose money in frauds as well as won't have to train employees in order to manually detect the same. Many machine learning chniques are commonly used in the banking sector at various occasions. We, with the proposed system aim to analyze the transaction data which comprises both normal as well as fraudulent transactions while giving us an insight about the customer behaviour. With the help of machine learning techniques, the proposed classification model shall accurately classify transactions as 'Genuine or Fraudulent'. The proposed system gives us a boost in terms of speed, scale and efficiency along with accuracy over the current systems in place. In Section II of the paper, we have discussed about the features in the existing applications and how they were beneficial. The implementations details and prototype are further discussed in Section III of the paper. The results have been further discussed in section IV. Section V and VI discuss the conclusion and well as future scope of our research.

II. LITERATURE REVIEW

With the increasing demand and need of Fraud detection systems in a world where most of the transactions happens online or via cards, a tremendous amount of work has been done with this domain. XGBoost model have have proposed by Yixuan Zhang et al. for the binary classification of transactions as fraud or not [1]. With Feature Engineering, initially the fraud transactions are generated with some time statistical, descriptive statistical and both combined features. To overcome the unbalanced target class distribution, SMOTE technique was used to oversample the minority class. Xgboost was found outperforming all the other models like SVM, Logistic Regression and Random Forest. By adding regularization to handle the sparse data.This work optimized the loss function To detecting a variety of frauds based on the historical data there is an another approach that discusses the variety of Data Mining techniques available for the Banking sector which can be implemented in real time [6]. Using classifiers like Bayesian, Bayesian, Support Vector, Decision Tree, Machines and Neural networks; activities are checked for being fraudulent or valid on a record-by-record basis. In Clustering; Model-based, Density based, Partitioning, Grid-based and Hierarchical methods are used for grouping together different bank transactions into one single cluster

to predict the percentage of clients contributing in profits. Then in Predication analysis technique; Linear Regression, Multivariate Linear Regression, Non-Linear Regression and Multivariate Non-Linear Regression are used prominently to determine and predict the frauds in the banking sector.

For Association rule mining; Quantitative, Multilevel and Multidimensional association rules are used to find the various sets of binary values that occur frequently together in the transactions so as to find various such groups who are related to one another containing such variables. Neural networks firstly detects the frauds by learning and then predict the transactions as fraudulent or genuine by category labeling to the input transactions.

There are few other works that have been done extensively for the credit card frauds. In [2], N. Malini et al. Have tried to optimize the Fraud detection process for credit card transactions by reducing the false positives and has gotten an increase in the fraud detection rate using KNN classification and outlier detection technique. It simply works on finding out the unusual transaction activities that are fraudulent. Here the closest or nearest point of any incoming transaction is calculated to the upcoming new transaction; now suppose the incomingtransaction is fraudulent in nature, then the model will show it as fraudulent and for continuous attributes, the Euclidean distance is calculated. While for categorical attributes, an easy matching coefficient has been used. To detect frauds, Outlier detection, unsupervised learning has been used . An ensemble method comprising of Random forest with Neural network is proposed in [16], out of which Random forest is able to detect the normal transaction instances and Neural Networks is used to detect the core fraudulent transactions. In [5], The credit card fraud detection technology based on Whale-algorithm optimized Back propagation(BP) neural network and is aimed at solving the problems of network defects, which is easy to fall into local optimum, poor system stability and slow convergence rate derived from Back propagation neural network is proposed by Chunzhi Wang et al. This research work goes through various methods andstudies the permutation and combination between algorithms like Back propagation, Genetic Algorithm and Particle Swarm Optimisation. Using Whale Swarm Optimization Algorithm to best optimise the weight of back propagation network, the Whale Optimization algorithm is used to get an optimal initial value and then BP network algorithm to correct the error value to finally obtain the optimal value. Based on the same lines of using Artificial Intelligence techniques, Ibtissam Benchaji et al. have proposed a method to enhance results which is already classified of the frauds in credit card transactions in the imbalanced dataset [9]. A sampling method is used which is on the basis of the K-means clustering algorithm that clusters and groups the fraudulent transaction samples. And the genetic algorithm is used in each cluster obtained to obtain the new transaction samples aims to build a most accurate classifier to detect frauds. Some of the works done in this domain also suggest the best AI techniques that can be used for detection of Fraud in different varieties of credit card transaction as seen in [15]. Another such approach is presented in [7] where a Multilayer perceptron (MLP) model is used for fraud detection in Parenclitic Networks analysis. Such networks generally highlight the difference between the standard instance and a particular new instance. Here they have checked the features of the historical data and the structure of correlations between all the instances. A transaction is represented by a node and the links between the nodes represent how the correlation is between various transactions. To find these links, the Euclidean distance is calculated and compared with a threshold value. After this entire process, this network is transformed to a set of features to

finally apply the MLP model. A live system that detects credit card frauds has also been illlustrated in [17] which is based on a Deep Neural Network methodology. It uses an auto-encoder to classify the transactions as fraudulent or genuine in real time. A real time approach for frauds in credit card transactions using Machine Learning is proposed in [3]. Four major types of frauds like bankruptcy fraud application fraud, , behavioral fraud and theft/counterfeit fraud that fall under Card-not-present (CNP) frauds category have been identified in this and each one is addressed using a series of ML algorithms to find Support vector Machine outperforming all other algorithms. For real time analysis, predictive analysis of the ML models is done along with a API module to decide whether the transactions are genuine or fraudulent. Support vector Machine outperformed all the ML algorithms used here. Another ML approach is depicted in [14] wherein many different algorithms outlier detection techniques are reviewed and finally integrated to establish a best financial fraud detection system. The accuracy of detection is based on the balance in the classes in the dataset [4]. Data mining techniques are used here in order to solve the issue of undersampling and versampling. The emergence of hybrid models is able to produce higher accuracy compared to supervised or unsupervised machine learning algorithms. Hinderance in performance of machine learning algorithms is reduce with usage of balanced dataset. In [8], a comparative study of various different credit card fraud detection techniques is made. Techniques like Breadth first depth first greedy approaches, fuzzy logic, a probabilistic model using Bayesian Network, etc. to classify the transactions as genuine and fraudulent are compared to find the best one suitable to variable environments. Many a times the user behavior (noise) isn't considered and the two classes of transactions are balanced based on the data's volume In credit card fraud detection systems, . Thus in [18], an imbalanced classification technique has been proposed based on a variety of user behaviors and clustering. Here the different user behaviors are grouped to remove the noise and hierarchical sampling is finally used to find out the fraudulent transactions. An approach which is similar to [18] has been proposed in [20] but with a hyper sphere model, the user's behaviour analysed with the history of his transactions and then the risk threshold of every user is calculated individually using the algorithm for risk threshold. The hyper sphere model is hence created based on the user behavior which as we have seen is in-turn determined from the transactional behavior and optimal risk threshold. In [19], a recurrent neural network is used to bring out a method that is inexpensive for detection of frauds. The history of card ayments is considered as an interleaved sequence to be used for further identification of frauds in transactions. Some of the other recent works in this domain include detecting malware using Machine Learning [12], detecting illegal entities in bitcoin data [13], detecting financial fraud using credit card data, etc.

III. FRAUDETECT

Frauds in transactions can be considered broadly in two types - Online Payments Transaction Frauds and Credit Card Transaction Frauds.

A. Online Payment Transaction Frauds The "Paysim" dataset from Kaggle [10], a well known online community has been used for analysis and detection of frauds. This is a synthetic dataset of Online payment transactions consisting of nearly 6.4 million rows of data. The entire 6.4 million rows data has been segregated into five majorly identified transaction types : CASH-IN, CASHOUT, DEBIT, PAYMENT and TRANSFER. The task was to find out the types in which frauds mainly occur. The transactions are classfied as "fraudulent" and "genuine" using binary classification analysis. The overall flow of this module can be visualized in Fig. 1. The dataset was found to be quite skewed, thus for proper analysis data pre-processing is done.

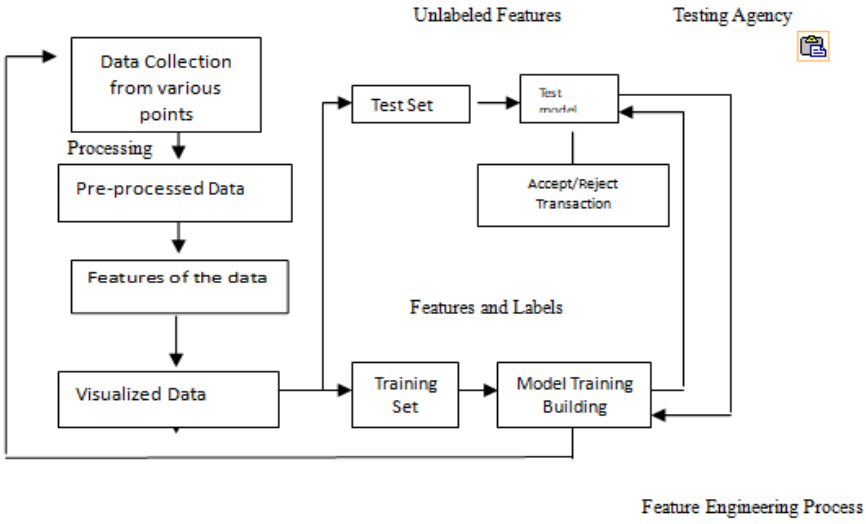


Fig. 1 Online Payment Transaction Frauds Flow diagram

The procedural steps for the detection of fraudulent transactionsare as follows : 1) Adjusting Data Types and removing missing values: The data is preprocessed as per the algorithm in order to be able to make calculations with them. Then missing values if any were tested to fill them and make the data consistent. 2) Fraudulent transactions: The different types of frauds and their respective count is found out. These become the explanatory and dependent variables for analysis. Binary encoding is then used by labelling the two cases as 'Fraudulent' and 'Genuine' where 0 represents genuine & 1 represents fraudulent. From the 5 majorly identified transaction types; CASH IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER, it is observed that Fraud occurs mainly in only two of the five typesa) TRANSFER: here, the money is sent/transferred to a user/fraudster/customer. b) CASH OUT: here, the money is sent/transferred to a merchant who usually pays to the customer/user/fraudster in cash. Thus TRANSFER and CASH OUT became the only independent variables to be looked on for further analysis. Among all the transactions, the fraudulent transactions were to be segregated from the genuine ones. 3) Analyzing the Data Pattern: For this

step, comparison of the recipients balances after and before a transaction were considered. From these, 2 different cases came out - a) Both the old and new balance in the recipient's account were zero, but transferred amount was not zero: Funds are transferred, but still the recipient's balance amount before and after is zero. It was recognize that almost half of such kind of transactions were fraudulent. For genuine transactions, the recipient's balance before and after was zero only in few of the cases. b) Both of the new and old balance in the sender's account were zero, but transferred amount was not zero: Funds are transferred, but still the sender's balance after & before is zero. Very few of such transactions were fraudulent. For genuine transactions, the sender's balance before and after waszero in almost half of the cases. We observed that zero balances in both sender's and recipient's accounts are strong indicators of fraud, when the transaction is non-zero. The balance values of zero are replaced by -1 and -2 in the sender's and recipient's accounts respectively, For explicitly detecting such frauds. These will make fraudulent transactions stand out, since keeping their values as zeros would make them all appear as Genuine.4) Feature Engineering on data: In order to utilize the observation of zero-balances serving to differentiate between fraudulent and genuine transactions, two new features or columns are created to record errors in the senders' and receivers' accounts for each transaction for better analysis. It turned out in the further analysis that these new features became important in highlighting the fraudulent transactions. 5) Machine Learning to Detect Fraud in Skewed Data: For the detection of the Frauds transactions, two algorithms of Machine Learning have been worked on - Decision tree classifier and XGBoost classifier. Both of the algorithms have a great capacity to work with skewed data well and give more accurate results and hence have been chosen to be worked on. 6) Investigating the AUPRC (Area under Precision-RecallCurve): Calculation of the skewness of the data gives a clear evidence of the highly imbalanced nature of the data.Since the data is highly skewed, the AUPRC(Area under the precision-recall curve) is used rather than the conventionalmethod of using AUROC(Area under the receiver operatingcharacteristic). The purpose of using AUPRC over AUROChere is that it is more conscious about finding the differences between different algorithms used and also setting up their parameters more than that of AUROC. To deal with the large skew in the data, an appropriate metric turned out to be AUPRC and the XGBoost Machine Learning algorithm for classification which seemed to work well with the strongly imbalanced classes. Few other metrics like Confusion matrix, Precision Recall and F1 score were also calculated to confirm the results from AUPRC. 7) Bias-variance trade off: The XGBClassifier model had a degree of bias and was slightly underfit. To optimize the model's performance, the most convenient way was to increase the max depth parameter of the XGBClassifier at the expense of the longer time spent learning the model. B. Credit Card Transaction Frauds To detect the frauds that occur with credit card transactions, "creditcard" dataset from Kaggle [11] has been taken for experimentation purpose. It includes the transactions which made by the European cardholders in transactions of credit cards for the month of September, 2013. The transactions that occurred in particular two days of the month have been considered here for implementation. The nature of the data is highly imbalanced. The Principal Component Analysis(PCA) transformation resulted into numerical input variables, thus making the data highly imbalanced. Feature "Class" in the dataset is the dependent variable. It has a binary nature to take value 1 if there is fraud & 0 otherwise. A simple architecture of the credit card transactions fraud detection module can be visualized in Fig. 2. The optimal solution can be evaluated by

using different algorithms to do the predictions. The highest predictive power model is further used for getting the prediction results. Different ML algorithms like XGB Classifier, KNN Classifier, Logistic Regression, Random Forest Classifier, Support Vector Classifier, Decision Tree Classifier and LGB Classifier have been worked upon to find out the most suitable model for fraud detection in transactions made using credit cards.

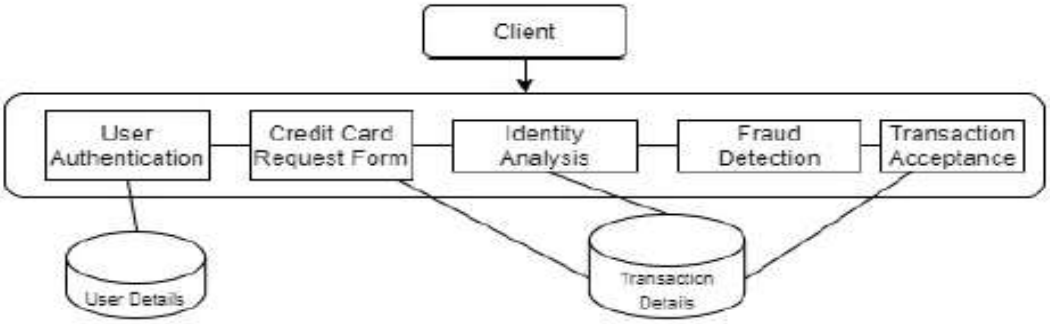


Fig. 2. Credit Card Transaction Frauds Architecture

The working of Fraud detection in credit card transactions can be can be visualized in Fig 3.

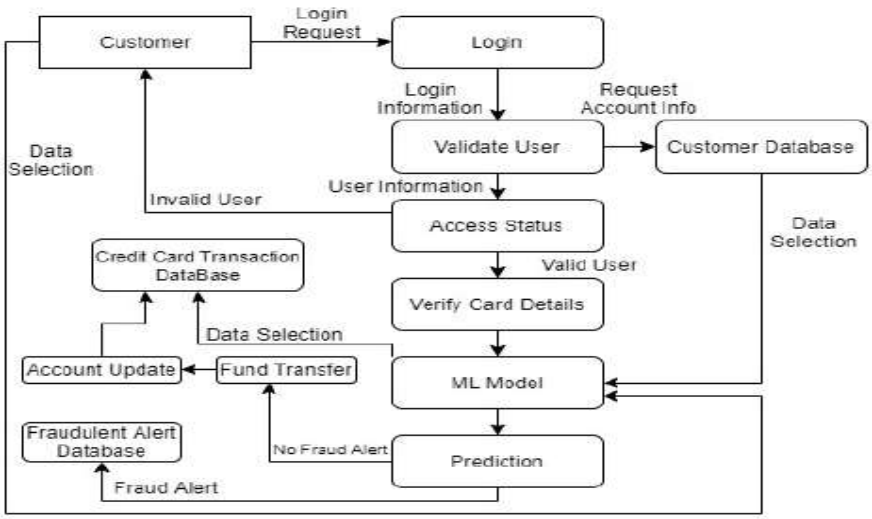


Fig. 3 Credit Card Transaction Frauds Flow diagram

The steps involved in this process are as follows –

1) Highlighting the fraudulent transactions from the data: Feature “Class” in the dataset is the dependent variable. It has a binary tendency to take value 1 if fraud occurs in the transactions and 0 otherwise. Thus the Class 0 transactions are “Genuine” while Class 1 are “Fraudulent”. Then such fraudulent and genuine transactions were pointed out from the data by plotting the amount of money over time in such transactions in the form of plots for visualizing them. A new dataframe of the shuffled class 0 and 1 transactions is created to be put together into a new csv file. Now using this new dataframe the Class 0 and 1 transactions are separated out widely and visualized. 2) Synthetic Minority Over-sampling Technique(SMOTE): SMOTE is one of the most commonly used oversampling methods to solve the imbalance problem. It is used here to generate synthetic samples. Here, we need the training set to be a balanced one, more

specificallly a class-balanced set or even a nearly class-balanced one. Thus to meet the requirements for the training set, SMOTE is used here. This balanced training set is then used to train the classifier 3) Feature Scaling: This technique is used here to standardize the independent features present in the data in a fixed range. 4) Dimensionality Reduction: As there are a total of 31 columns in the dataset, it is necessary to reduce a few of them. This will thus lead to a decrease in the no. of input variables from the dataset because more input variables can increase the complexity to model a predictive model like this. Such a concept is usually referred to as the "Curse of dimensionality". This step is only limiting the data to the most important features only and adding the relevant features as the data is highly unbalanced. Thus the new dataframe now contains only the relevant 4000 rows and 8 columns. The new data ready for training the model is now placed in a different csv file to be worked on for further analysis. 5) Hyperparameter tuning: A hyperparameter is a parameter whose value is used to control the learning process and they tend to define the architecture of the model. Here as different classification algorithms are to be used to build the model, there is a need of a process of searching for the ideal model architecture. Hyperparameter tuning is a step that can get this job done. 6) ML Algorithms: Our approach to the solution is finding out the best Machine Learning algorithm that will give the most accurate result for detection of frauds in transactions. a) Logistic Regression: Logistic regression model is selected here as the dependent variable is of binary nature. This model is a statistical one and to model a binary target variable"Class" here in the data, a logistic function has been used. In the regression analysis, to estimate the parameters of the created logistic model, logistic regression has been applied. For this regression model, a predictive accuracy of 96.08% is obtained which is quite a good one. b) Support Vector Classifier: Support Vector classifieris used to fit the model to the data, returning a "best fit" hyperplane that divides, or categorizes the data. Here it is observed that SVC gives an accuracy of around 96.08% as per the classification report.c) Decision Tree Classifier: This algorithm of classification is used to create the model by constructing a decision tree. Every node in the decision tree specifies a test on the attributes. Each node has descending branches which gives one of the possible values for that particular attribute. This is found to be giving an accuracy of 95.58%. d) Random Forest Classifier: Now to get more accurate predictions, the use of multiple decision trees is made with the help of the Random Forest classifier algorithm. It aggregates the votes from the various decision trees made to finally decide upon the type of the transaction as either fraudulent or genuine of the "Class" test object from the data. As expected, RandomForest appears to give an improved accuracy of 96.25% than the Decision Tree classifier. e) K Nearest Neighbor Classifier: Each data point (object) in the k closest neighbors casts a vote and the category with the most votes wins (k is a positive integer, typically small). As per the k value, the object is simply assigned to the class of that single nearest neighbor. Here we have tried giving two different values for k ie. 5 and 2. When k=5, we are getting an accuracy score of 96.91% whereas for k=2 which is a much smaller value for k, we are getting more accurate predictions as the accuracy comes out to be 97.08%. f) XGBoost Classifier: Now XGBClassifer which is an improved implementation of gradient boosted decision trees designed for speed and performance that is, dominative competitive machine learning, frauds are to be detected is been used. Our model needs to be the best fit for detection and ,prediction of frauds. The accuracy for prediction using XGBClassifieris 96.91% which is better performing than the decision tree and random forest approach. g) LGB Classifier: LGB Classifier is a gradient boosting framework. It uses

tree based learning technique as like decision trees, Random forest and XGBClassifer. LGB algorithm has a nature of growing the tree leaf-wise unlike other algorithms that grow level-wise, thus it is said to have a vertically growing tree unlike other algorithms who have a horizontally growing tree. Such an algorithm which is also called a leaf-wise algorithm can reduce loss as compared to a level-wise algorithm when growing the same leaf. Hence a model using this algorithm is created. It is giving a prediction accuracy of 97.41%. 7) ROC (Receiver operating characteristic) Curve: An ROC curve is a graph showing the performance of lassification model at all classification thresholds. This curve plots two parameters: True Positive Rate & False Positive Rate. It plots both of these parameters against each other at various threshold settings for every classifier used here. This is done just to build the prediction models for detection of Frauds in Credit card payments transactions.

IV.RESULTS

A. Online Payment Transaction Frauds

It can be observed that in the Online payment transactions, fraud has occurred only in TRANSFER and CASHOUT as seen in Fig. 4.

```
The types of fraudulent transactions are ['TRANSFER', 'CASH_OUT']

The number of fraudulent TRANSFERs = 4097

The number of fraudulent CASH_OUTs = 4116
```

Fig. 4. Fraud occurring transaction types

s

While analysing the data pattern, it is seen that 49.55% transactions are fraudulent while only 0.06% are genuine in case of “Both the new and old balance in the recipient’s account were zero, but transferred amount was not zero” which can be seen in Fig. 5. Also, Fig. 6 shows that for “Both the new and old balance in the sender’s account were zero, but transferred amount was not zero”, only 0.30% transactions are fraudulent and 47.37% are genuine.

```
The percentage of 'fraudulent' transactions where both the old and new balance in the recipient´s account were zero,
but the transacted amount was not zero: 49.5556%

The percentage of 'genuine' transactions where both the old and new balance in the recipient´s account were zero,
but the transacted amount was zero: 0.0618%
```

Fig. 5. Both the old and new balance in the recipient’s account were zero, but transferred amount was not zero

```
The percentage of 'fraudulent' transactions where both the old and new balance in the sender´s account were zero,
but the transacted amount was not zero: 0.3044%

The percentage of 'genuine' transactions where both the old and new balance in the sender´s account were zero,
but the transacted amount was zero: 47.3732%
```

Fig. 6. Both the old and new balance in the sender’s account were zero, but transferred amount was not zero

Altogether 27,70,409 transactions of TRANSFERs and CASH OUTs were detected from the data. The 3D plot in Fig. 7 distinguishes best between fraud and genuine data by musing both of the engineered error-based features ie. Error in Balance at Destination and Origin. The fraudulent transactions can be seen standing out clearly in the plot.

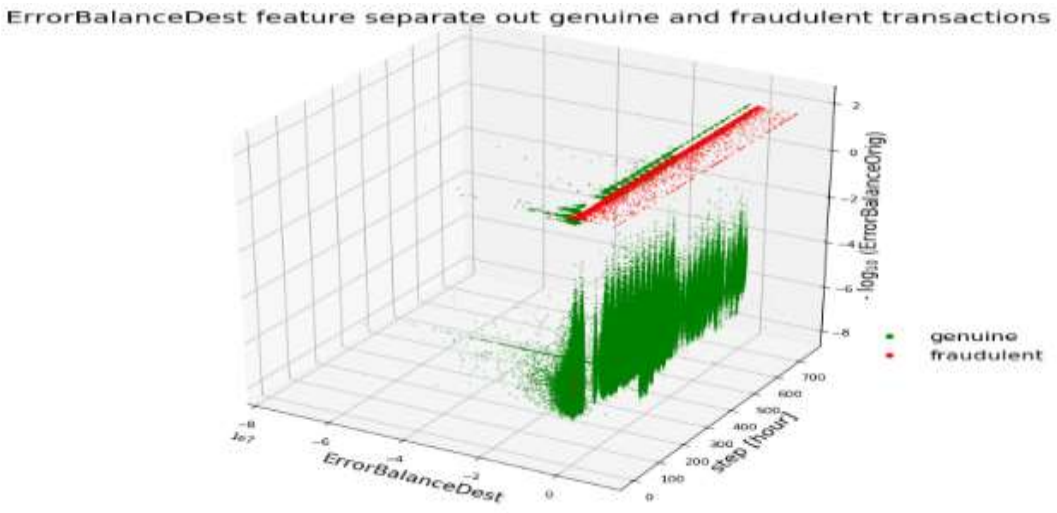


Fig. 7. Separating out genuine from fraudulent transactions

The accuracies of XGBoost and Decision Tree classifiers can be seen in Table I. Table II displays the confusion matrix obtained by the XGBoost Classifier. The method used here should therefore be broadly applicable to a range of such problems.

B. Credit Card Transaction Frauds

For Credit Card transactions, the Class 0 ie. genuine and 1 ie. fraudulent transactions can be separated out widely and visualized as in Fig. 8. Since the optimal solution can be evaluated by using different algorithms to do the predictions,

TABLE -I

ML ALGORITHMS AND THEIR PERFORMANCE MEASURES

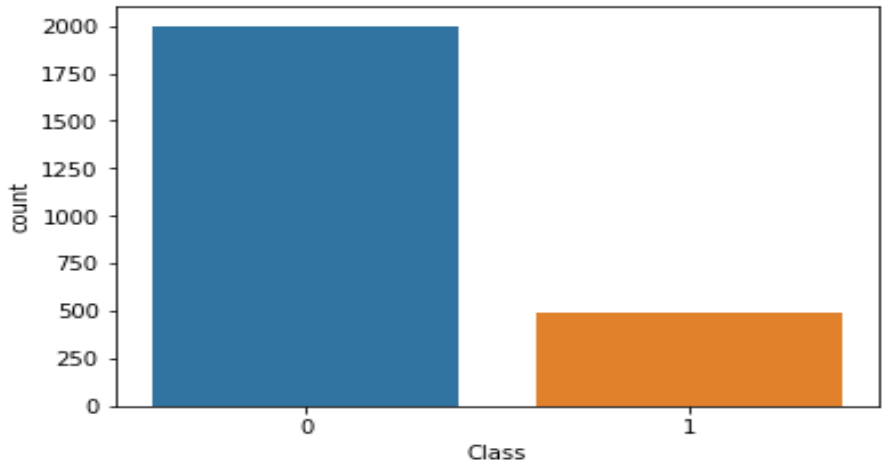
Algorithm Name	Performance Metrics	
	Accuracy	AUPRC
XGBooster	99.99%	0.9967
Decision Tree Classifier	99.96%	0.7450

TABLE II

CONFUSION MATRIX OF XGBOOST CLASSIFIER

	Genuine	Fraudulent	Total
Genuine	2209196	564	2209760
Fraudulent	21	6446	6567
Total	2209217	7110	2216327

the highest predictive power model is further used for getting the prediction results.



It is found that the algorithm that best suits for prediction is Light Gradient Boosting Classifier with an predictive accuracy of 97.41%. The overall performance of all the ML algorithms can be seen in Table III.

TABLE III

ML ALGORITHMS AND THEIR ACCURACY SCORES

Algorithm Name	Accuracy
Logistic Regression	96.08%
Support Vector Classifier	96.08%
Decision Tree Classifier	95.58%
Random Forest Classifier	96.25%
KNN Classifier	(k==5) 96.91% (k==2) 97.08%
XGBoost Classifier	96.91%
LGB Classifier	97.41%

The ROC curve at various threshold settings for every classifier used here to build the prediction models for detection of Frauds in Credit card payment transactions can be analyzed in Fig. 9. Crucially, these results were obtained without artificial balancing of the data, such as SMOTE (Synthetic Minority Fig. 9. ROC (receiver operating characteristic) curve Over-sampling Technique), making this approach suitable to real-world applications.

V. CONCLUSION

Fraud detection in current trending payment methods is a need of the hour. The solution proposed in this paper

is a robust, quick and accurate way to detect frauds that occur in both Online payments and payments through credit cards. As the solutions are based on high performing Machine Learning algorithms they provide some cost effective and quick predictions is reducing the risk of occurrence of frauds. This approach can be suitable to real-world applications where the financial services companies, banks, financial institutions, etc. can deploy it and achieve maximum customer satisfaction by preventing frauds that happens in transactions made by their users. Hence, FrauDetect will fulfill the real time demand of frauds detection before they even occur.

VI. FUTURE SCOPE

The project demonstrates a system which is able to decide a unique approach amalgamate various features in order to perform fraud detection. Having achieved a significant improvement in classification accuracies, the future scope of this study can be identified as various secure methods of transacting and achieving an overall better accuracy with using a unique combination of features.

REFERENCES

[1] Nerurkar, P., Busnel, Y., Ludinard, R., Shah, K., Bhirud, S. and Patel, D., 2020, August. Detecting Illicit Entities in Bitcoin Using Supervised Learning of Ensemble Decision Trees. In Proceedings of the 2020 10thInternational conference on Inform Ation Communication and Management (pp. 25-30). DOI:https://doi.org/10.1145/3418981. 3418984

[2] Thushara Amarasinghe, Achala Aponso, and Naomi Krishnarajah. 2018. Critical Analysis of Machine Learning Based Approaches for Fraud Detection in Financial Transactions. In Proceedings of the 2018 International Conference on Machine Learning Technologies (ICMLT '18). Association for Computing Machinery, New York, NY, USA, 12– 17.DOI:https://doi.org/ 10.1145/3231884.3231894.

[3] Imane Sadgali, Nawal Sael, and Faouzia Benabbou. 2019. Fraud detection in credit card transaction using neural networks. In Proceedings of the 4th International Conference on Smart City Applications (SCA '19). Association for Computing Machinery, New York, NY, USA, Article 95, 1–4. DOI:https://doi.org/10.1145/3368756.3369082.

[4] Ishan Sohony, Rameshwar Pratap, and Ullas Nambiar. 2018. Ensemble learning for credit card fraud detection. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAD '18). Association for Computing Machinery, New York, NY, USA, 289–294. DOI:https://doi.org/10.1145 / 3152494.3156815.

[5] Y. Zhang, J. Tong, Z. Wang and F. Gao, "Customer Transaction Fraud Detection Using Xgboost Model," 2020 International Conference on Computer Engineering and Application (ICCEA), Guangzhou, China, 2020, pp. 554-558, doi: 10.1109/ICCEA50009.2020.00122.

[6] N. Malini and M. Pushpa, "Analysis on credit card fraud identification techniques based on KNN and outlier etection," 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp. 255-258, doi: 10.1109/AEEICB.2017.7972424.

[7] A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," 2019 9th International Conference on Cloud Computing, Data Science

Engineering (Confluence), Noida, India, 2019, pp. 488- 493, doi: 10.1109/CONFLUENCE.2019.8776942.

[8] S. Khatri, A. Arora and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection:

A Comparison," 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence),

Noida, India,2020, pp. 680-683, doi: 10.1109/Confluence47617.2020.9057851.

[9] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai and S. Pan, "Credit Card Fraud Detection Based on Whale Algorithm

Optimized BP Neural Network," 2018 13th International Conference on Computer Science Education (ICCSE),

Colombo, 2018, pp. 1-4, doi: 10.1109/ICCSE.2018.8468855.

[10] R. Rambola, P. Varshney and P. Vishwakarma, "Data Mining Techniques for Fraud Detection in Banking Sector,"

2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida,

India, 2018, pp. 1-5, doi: 10.1109/CCAA.2018.8777535

[11] Massimiliano Zanin, Miguel Romance, Santiago Moral, Regino Criado, "Credit Card Fraud Detection through

Parenclitic Network Analysis", Complexity, vol. 2018, Article ID 5764370, 9 pages, 2018. https://doi.org/

10.1155/2018/5764370.

[12] Jain, Y. Tiwari, N. Dubey, S. Jain, Sarika. (2019). A comparative analysis of various credit card fraud detection

techniques. International Journal of Recent Technology and Engineering. 7. 402-407.

[13] I. Benchaji, S. Douzi and B. ElOuahidi, "Using Genetic Algorithm to Improve Classification of Imbalanced

Datasets for Credit Card Fraud Detection," 2018 2nd Cyber Security in Networking Conference (CSNet), Paris,

2018, pp. 1-5, doi: 10.1109/CSNET.2018.8602972.

[15] Dal Pozzolo, Andrea Adaptive Machine learning for credit card fraud detection ULB MLG PhD thesis (supervised

by G. Bontempi).

[16] Kumar A. et al. (2020) Malware Detection Using Machine Learning. In: Villaz´on-Terrazas B., Ortiz-Rodr´ıguez

F., Tiwari S.M., Shandilya S.K. (eds) Knowledge Graphs and Semantic Web. KGSWC 2020. Communications in

Computer and Information Science, vol 1232. Springer, Cham. https://doi.org/10.1007/978-3-030-65384-2 5

[17] Youness Abakarim, Mohamed Lahby, and Abdelbaki Attioui. 2018. An Efficient Real Time Model For Credit

Card Fraud Detection Based On Deep Learning. In Proceedings of the 12th International Conference on

Intelligent Systems: Theories and Applications (SITA'18). Association for Computing Machinery, New York,

NY, USA, Article 30, 1–7. DOI:https://doi.org/10.1145/3289402.3289530

[18] Qi Li and Yu Xie. 2019. A Behavior-cluster Based Imbalanced Classification Method for Credit Card Fraud

Detection. In Proceedings of the 2019 2nd International Conference on Data Science and Information Technology (DSIT 2019). Association for Computing Machinery, New York, NY, USA, 134–139. DOI:https://doi.org/10.1145/3352411.3352433

[19] Bernardo Branco, Pedro Abreu, Ana Sofia Gomes, Mariana S. C. Almeida, Jo˜ao Tiago Ascens˜ao, and Pedro

Bizarro. 2020. Interleaved Sequence RNNs for Fraud Detection. Proceedings of the 26th ACM SIGKDD

International Conference on Knowledge Discovery Data Mining. Association for Computing Machinery, New
York, NY, USA, 3101–3109. DOI:https://doi.org/10.1145/3394486.3403361

[20] Ligong Chen, Zhaohui Zhang, Qiuwen Liu, Lijun Yang, Ying Meng, and Pengwei Wang. 2019. A method for
online transaction fraud detection based on individual behavior. Proceedings of the ACM Turing Celebration
Conference - China. Association for Computing Machinery, New York, NY, USA, Article 119, 1–8.
DOI:https://doi.org/10.1145/3321408.332664