CREDIT CARD FRAUD DETECTION AND AUTHENTICATION SYSTEM USING MACHINE LEARNING

Kavitha G L[1], S Harini Sree[2], Sakshi Nagarajarao Jadhav[3], Yuktha N[4]

[1]Assistant Professor, Dept. of Information Science and Engineering, Bangalore
[234]Student, Dept. of Information Science and Engineering, Bangalore, Karnataka, India

---***---

Abstract - *This project's objective is to create a reliable system that can detect fraudulent transactions and authenticate credit card users before the transaction gets completed. The increase in the use of credit cards for transactions has led to a rise in fraudulent activities, making it crucial to develop a system that can identify and prevent such activities. The system will use various machine learning algorithms, such as decision trees, random forests, KNN algorithm to analyze transactional data and detect any suspicious activities. It will be trained on a dataset containing information on past fraudulent activities to identify patterns and recognize similar fraudulent transactions. Furthermore, the system will implement diverse authentication methods, including face recognition detection and authentication and one-time passwords to verify the legitimacy of the credit card user. The implementation of this system is expected to increase the security of credit card transactions and prevent fraudulent activities by detecting and authenticating the fraud before it takes place, leading to significant savings for consumers and financial institutions.*

Key Words: *Fraudulent transactions, authentication, one-time passwords, face recognition, KNN Algorithm, Random Forest, Naïve Bayes.*

1. INTRODUCTION

This research work focuses on the identification of fraudulent transactions made through credit cards. The objective is to create a fraud detection algorithm that can accurately and efficiently classify transactions as either fraudulent or non-fraudulent by utilizing machine learning-based classification algorithms. With the increasing use of online payments and the decline of cash payments, fraudsters are taking advantage of the anonymity provided by these transactions. Online payments, in particular, require only the card number, expiration date, and CVV, making it easy for data to be lost or stolen without the cardholder's knowledge. In some cases, cardholders may not even be aware that their information has been compromised due to fraudulent purchases made through phishing techniques.

This highlights the importance of keeping card details private, though there are instances where this is not possible due to the prevalence of phishing sites and cases of lost or stolen cards.

One effective way to determine the legitimacy of a transaction is to analyze the spending patterns of the cardholder using existing data and applying machine learning algorithms. This can help identify anomalies in spending that may indicate fraudulent activity. There are various types of credit card fraud, including online and offline fraud, card theft, data phishing, application fraud, and telecommunication fraud. It is essential to address these different types of fraud to prevent fraudulent transactions and protect cardholders' data.

2. LITERATURE REVIEW

Parth Roy, Prateek Rao, Jay Gajre [1]. IFA suggested using machine learning to identify fraudulent Master Card transactions. The strategies are used to improve the best solution for issues with fraud detection. Methods for reducing false alarm rates and increasing the rate at which scams are discovered are still proven. Since European card holders have had 284,807 communications, data on card transactions continues to be collected. A modified version of these methods can be implemented to the bank's credit card scam detection system to help identify and stop fraud.

Ishika Sharma, Shivjyoti Dalai, Venktesh Tiwari, Ishwari Singh , Seema Kharb [2] presented various techniques such as Naive Bayes, Random Forest and Logistic Regression are utilized to tackle this problem. This transaction is evaluated individually, and whatever works best is carried out. The primary purpose is to detect fraud by filtering the aforementioned strategies in order to achieve a better outcome.

Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi [3offered an evaluation that offers a thorough manual for choosing the best algorithm for the kind of frauds, and we use an adequate performance metric to show the evaluation. In order to determine if a particular transaction is legitimate or fraudulent, they also created the use of predictive analytics performed by the implemented machine learning models and an API module

H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes [4] carried out a thorough experimental investigation using the answers to the imbalance classification issue. They investigated these options and the machine learning fraud detection methods, pinpointed their flaws, and summarised the findings using a credit card fraud labelled dataset.

S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine [5] presented various machine learning-based classification techniques, including Naive Bayes, Random Forest, and Logistic Regression, for handling the severely skewed dataset. Accuracy, precision, recall, f1 score, confusion matrix, and Roc-auc score will all be calculated as part of the research project.

D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith kumar, CH V N M Praneeth[6] proposed the usage of Random Forest, Gradient Boost, Support Vector Machine, and their combinations. Depending on the set of data and the application, the algorithms' efficacy vary. They demonstrate that, in spite of all computations, all algorithms exhibit some degree of imbalance at some point in the study. When learning curves were plotted, it was discovered that while logistic regression had a higher accuracy, KNN could only learn, while the majority of the algorithm was underfit. KNN is therefore a stronger classifier for identifying credit card fraud.

MJ Madhury, H L Gururaj, B C Soundarya, K P Vidyashree ,B Rajendra [7] provided a variety of machine learning-based methods for credit card recognition, including XG Boost, Decision Tree, Random Forest, Support Vector Machine, and Extreme Learning Method. To get effective results, comparative examination of both machine learning and deep learning algorithms was done. Using the European card benchmark dataset for fraud detection, a thorough empirical investigation is conducted. The dataset was first subjected to a machine learning technique, which somewhat increased the accuracy of fraud detection.The evaluation of the research effort demonstrates the enhanced outcomes obtained, with optimised values for accuracy, f1-score, precision, and AUC Curves of 99.9%, 85.71%, 93%, and 98%, respectively. For situations involving credit card detection, the suggested model performs better than cutting-edge machine learning and deep learning techniques.

Fawaz Khaled Alarfaj , Iqra Malik , Hikmat Ullah Khan , Naif Almusallam , Muhammad Ramzan , And Muzamil Ahmed [8] Using both publicly available and actual transaction records, 13 statistical and machine learning models for payment card fraud detection were created. Analysis and comparison are done on the outcomes from both the original features and the aggregated features. To determine if the combined characteristics produced by a genetic algorithm have greater discriminative ability than the original features in detecting fraud, a statistical hypothesis test is performed. The results demonstrate that employing aggregated features to tackle real-world payment card fraud detection issues is effective.

Manjeevan Seera, Chee Peng Lim, Ajay Kumar, Lalitha Dhamotharan, Kim Hua Tan [9] proposed a method to determine whether transactions on the Kaggel-provided IEEE-CIS Fraud Detection dataset were genuine or fraudulent. Bidirectional Long Short-Term Memory (BiLSTM) and bidirectional Gated Recurrent Unit (BiGRU) are the foundations of the model, which is called BiLSTM-MaxPooling-BiGRUMaxPooling. They used the following six machine learning classifiers: Logistic Regression, Naive Base, Voting, Ada Boosting, Random Forest, and Decision Tree. When comparing the outcomes of machine learning classifiers and our model, it is clear that the model performed better because it received a 91.37% score.

Sailusha Ruttala; Gnaneswar V.; Ramesh R.; Rao, G. Ramakoteswara [10] The Adaboost algorithm and the random forest method are the algorithms used in the presentation. The two algorithms' outputs are based on F1-score, accuracy, precision, recall, and other metrics. On the basis of the confusion matrix, the ROC curve is plotted.Whthe algorithms from Random Forest and Adaboost are compared, the method with the highest accuracy, precision, recall, and F1-score is regarded as the best one for spotting fraud.

Xuan Shiyang Xuan,Guanjun Liu, Zhenchuan Li,Shuo Wang, Lutao Zheng, Changjun Jiang [11] discussed the use of two alternative random forest models to train the behavioural characteristics of typical and anomalous transactions, compare the two random forests, which differ in their underlying classifiers, and assess how well they detect credit fraud.

3.THE OBJECTIVE OF PROJECT

The specific objective of this project is to:

1.Develop a fraud detection system that can accurately identify and classify fraudulent and credit card transactions that are not fraudulent using machine learning techniques.

2. Implement a reliable authentication system that uses various authentication methods, such as face recognition authentication and one-time passwords verify the legitimacy of the credit card user.

3. Train the system on a dataset of past fraudulent activities to identify patterns and recognize similar fraudulent transactions, thereby increasing the accuracy of the fraud detection algorithm.

4. Increase the security of credit card transactions by detecting and preventing fraudulent activities before they occur, leading to significant savings for consumers and financial institutions.

5. Keep up with the advancements in technology and adapt the system to new fraudulent activities and authentication methods to maintain its effectiveness.

4.PROPOSED METHODOLOGY

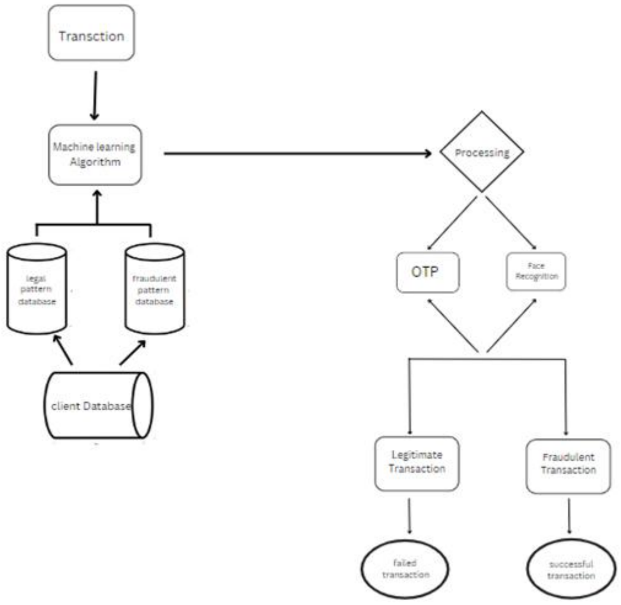


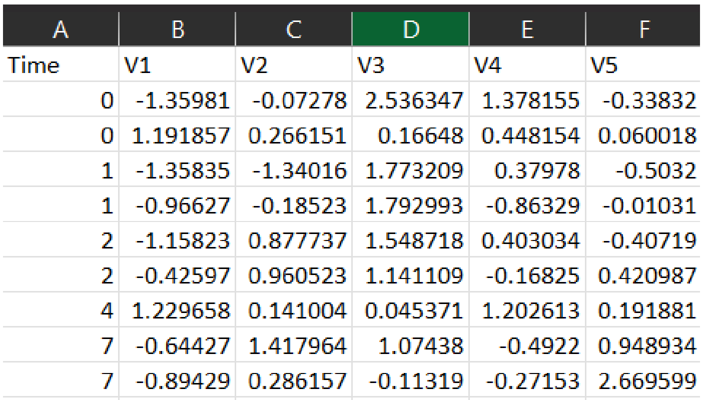
Figure 4.1-Flowchart of the system

The figure shows a sequence of critical events that must occur in the creation of the proposed model. The task of classifying the data is completed after procedures like data processing, decision making and face detection and OTP authentication.

The system receives data from various sources, such as credit card transactions, user account information, and device information. The input data is pre-processed to normalize and standardize the data. This may involve converting data into a specific format, identifying missing values, and removing any irrelevant information.

In order to find patterns and anomalies in the data, the pertinent characteristics are retrieved from the pre-processed data. To find potential fraudulent transactions, the collected features are analyzed using a variety of methods, including rule-based systems, anomaly detection, and machine learning algorithms. The system authenticates legitimate users using various techniques such as face detection, two-factor authentication, and device recognition. Then based on the fraud detection and authentication results, the system makes a decision on whether to approve or reject a transaction.

The system informs the user and the company of the transaction's status, including whether it was accepted or refused. In order to increase the accuracy and efficiency of the fraud detection and authentication system, the system generates reports and analyses the data to find trends and patterns.



A	B	C	D	E	F
Time	V1	V2	V3	V4	V5
0	-1.35981	-0.07278	2.536347	1.378155	-0.33832
0	1.191857	0.266151	0.16648	0.448154	0.060018
1	-1.35835	-1.34016	1.773209	0.37978	-0.5032
1	-0.96627	-0.18523	1.792993	-0.86329	-0.01031
2	-1.15823	0.877737	1.548718	0.403034	-0.40719
2	-0.42597	0.960523	1.141109	-0.16825	0.420987
4	1.229658	0.141004	0.045371	1.202613	0.191881
7	-0.64427	1.417964	1.07438	-0.4922	0.948934
7	-0.89429	0.286157	-0.11319	-0.27153	2.669599

Figure 4.2. The dataset used for the research

All Our study suggests a novel use of sophisticated machine learning algorithms to identify outliers, or false or unusual events. In order to achieve this, we made use of a Kaggle dataset made up of transaction records by using it as our training data. The dataset consists of v1-v28 PCA feature concerns with secrecy, time, and amount as our variable factors and classes with 0 and 1 where 0 indicates no fraud and 1 means fraud, respectively.

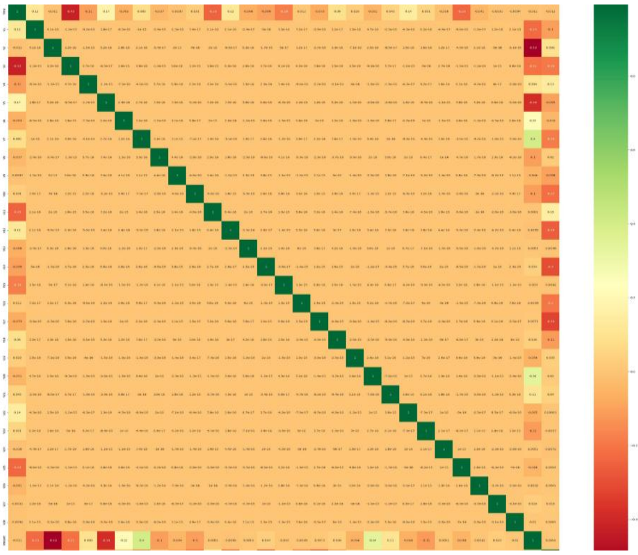


Figure 4.3 The heatmap drawn for finding correlation.

5.ALGORITHMS USED TO DETECT THE FRAUDULENT TRANSACTIONS

In order to extract the data and find trends, such as fraudulent transactions, we will train our system using a variety of machine learning approaches after building the heatmap. To forecast a fraudulent transaction in the future, we will leverage the pattern discovered between the legitimate transactions and the fraudulent transactions that are currently present in the dataset. We then employed a variety of algorithms, including those listed below, to construct and train a model aimed at identifying fraudulent activities in credit card transactions:

5.1.RANDOM FOREST CLASSIFIER:

A classification system called Random Forest uses a number of decision trees, each of which is constructed using a different subset of the dataset. The algorithm then aggregates the outcomes of each decision tree's predictions to produce a final outcome. When compared to employing a single decision tree, this method is intended to increase the predictability of results. The final output is predicted by the algorithm using majority voting, which means that the forecast that received the greatest support from the decision trees is chosen as the result. More trees in the forest can be used to improve accuracy and lessen the problem of overfitting.

5.2. Naïve Bayes algorithm:

5.3. Decision tree algorithm:

A decision tree is a type of hierarchical structure that resembles a tree, where every internal node, including the root node, represents a "test" run on a dataset's instances' attributes. The results of each test are represented by the corresponding branches, which connect to further internal nodes or leaf nodes, which stand in for the class labels. For the execution of this form of tree, the following three factors are necessary: Target Attribute is the attribute that represents the class label. Instances is a collection of instances for which the class label is already known. Attributes List is a collection of predictor attributes used to create the decision tree.

5.4.Logical Regression:

A categorical dependent variable is analyzed in relation to one or more independent variables using the statistical procedure known as logistic regression. It is a common use of regression analysis in machine learning and predictive modelling. The dependent variable in logistic regression is a binary or dichotomous variable, meaning it can only have two possible outcomes, such as yes or no, true or false, success or failure, etc. The independent variables may be categorical or continuous. Based on the values of the independent variables, logistic regression seeks to determine the likelihood that the dependent variable will fall into one of the two groups. A probability value between 0 and 1 is the result of logistic regression, which can be transformed The theory presented here relies on two fundamental assumptions. Firstly, it assumes that each feature in a given input carries equal weight towards the objective of classification. Secondly, it assumes that all attributes provided are statistically independent, meaning that the values of each attribute do not provide any information about the values of other attributes. While these assumptions may hold true in some scenarios, they may not always be applicable in practice. To address such cases, the Bayes rule is employed to determine whether a given input is fraudulent or legitimate. The rule computes the probability of an input belonging to each possible class and assigns the predicted class based on the one with the highest probability. By utilizing this approach, the system can make more accurate predictions even when the independence assumption between attributes does not hold.

6.ADVANTAGES

- Enhanced Security: A credit card fraud detection and authentication system can provide an extra layer of security to protect against fraudulent transactions. It can use various techniques such as machine learning, artificial intelligence, and facial recognition to identify and authenticate legitimate users while blocking unauthorized access.

- Real-time Monitoring: Such systems can monitor credit card transactions in real-time, which enables quick identification and prevention of fraudulent activities. This is especially important for e-commerce transactions where speed is critical.

- Lessened Financial Losses: Both consumers and businesses can suffer large financial losses as a result of credit card fraud. By recognising and preventing fraudulent transactions from taking place, an efficient fraud detection and authentication system can aid in the prevention of these losses.

- Improved Customer Trust: When customers know that their credit card information is secure and protected, they are more likely to trust the company they are dealing with. This can result in improved customer loyalty and repeat business.

- Regulation Compliance: Credit card fraud detection and authentication tools can assist companies in meeting security standards like the Payment Card Industry Data Security Standard (PCI DSS) and other legal obligations. Penalties and other legal problems may be avoided in this way.

- Regulation Compliance: Credit card fraud detection and authentication tools can assist companies in meeting security standards like the Payment Card Industry Data Security Standard (PCI DSS) and other legal obligations. Penalties and other legal problems may be avoided in this way.

7.CONCLUSION

Developing a reliable system that can identify fraudulent transactions and authenticate credit card users is the goal of this project, to sum up. The need to develop a system that can stop such activities in their tracks by identifying them and preventing them has increased due to the rising prevalence of credit card theft. The suggested system will examine transactional data and look for any suspicious behaviour using a variety of machine learning algorithms, including decision trees, random forests, and more.. To identify trends and identify similar fraudulent transactions, it will be trained on a dataset of prior fraudulent acts. In order to confirm the legality of the credit card user, the system will also use a variety of authentication techniques, such as facial detection authentication, one-time passwords, and security questions. The introduction of this system is anticipated to result in significant cost savings for both customers and financial institutions while also enhancing the security of credit card transactions, preventing fraud by identifying and validating it beforehand, and preventing fraudulent activity.

8.FUTURE ENHANCEMENTS

In the future, the proposed system could be enhanced by incorporating advanced machine learning algorithms and techniques such as deep learning and natural language processing to better analyses and interpret transactional data. Additionally, the system could be integrated with real-time fraud detection mechanisms that could monitor transactions as they occur and identify suspicious activities in real-time. Furthermore, the system could leverage blockchain technology to enhance security and prevent fraudulent activities by creating an immutable and transparent record of all transactions. Finally, the system could be extended to support more diverse payment methods and authentication mechanisms to accommodate the evolving needs of consumers and businesses. These enhancements are expected to significantly improve the effectiveness and efficiency of the system, leading to even greater savings for consumers and financial institutions while further reducing the risk of credit card fraud.

9. REFERENCES

[1] Parth Roy, Prateek Rao, Jay Gajre, "*Comprehensive Analysis for Fraud Detection of Credit Card through Machine Learning*" ,2021 International Conference on Emerging Smart Computing and Informatics (ESCI) , March 2021.

[2] Ishika Sharma, Shivjyoti Dalai, Venktesh Tiwari, Ishwari Singh , Seema Kharb, "*Credit Card Fraud Detection Using Machine Learning & Data Science*" , International Research Journal of Engineering and Technology (IRJET) Vol. 09, Issue 06, Jun 2022.

[3] Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi ,"*Real-time Credit Card Fraud Detection Using Machine Learning*",IEEE 9th International Conference on Cloud Computing, Data Science & Engineering, 2019.

[4] H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, "*Credit card fraud detection based on machine and deep learning*," in Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS), Apr. 2020.

[5] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine, '*'An experimental study with imbalanced classification approaches for credit card fraud detection*,'' IEEE Access.

[6] D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith kumar, CH V N M praneeth, "*Credit Card Fraud Detection Using Machine Learning* ", Fifth International Conference on Intelligent Computing and Control Systems , 2021.

[7] MJ Madhury, H L Gururaj, B C Soundarya, K P Vidyashree ,B Rajendra , "*Exploratory analysis of credit card fraud detection using machine learning techniques*",Elsevier B.V. 2022 .

[8] Fawaz Khaled Alarfaj , Iqra Malik , Hikmat Ullah Khan , Naif Almusallam , Muhammad Ramzan , And Muzamil Ahmed ,"Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms", Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University through the Research 2022.

[9] Manjeevan Seera, Chee Peng Lim, Ajay Kumar, Lalitha Dhamotharan, Kim Hua Tan ,"*An intelligent payment card fraud detection system*", Springer Science+Business Media, LLC, part of Springer Nature 2021.

[10] Sailusha Ruttala; Gnaneswar V.; Ramesh R.; Rao, G. Ramakoteswara, "Credit Card Fraud Detection Using Machine Learning", International Conference on Intelligent Computing and Control Systems,IEEE Explore 2020

[11] Shiyang Xuan,Guanjun Liu, Zhenchuan Li,Shuo Wang, Lutao Zheng, Changjun Jiang,"*Random forest for credit card fraud detection*" ,IEEE, 2018.