Fraud Detection and Authentication of Credit CardUsing ML

SHOAIB ATTAN KHAN[1], SHUBHAM ARORA[2], TIWARI SAHIL MUKESH[3], ZEESHAN KHAN[4], PRAKRUTHI M K[5]

[1, 2, 3, 4, 5] *Computer Science Department SJB Institute of Technology (VTU Aff.), Bangalore, India*

Abstract— This study aims to tackle the increasing issue of potential fraudulent activities within the sphere of credit card transactions. As the number of card transactions increases, there is a vital necessity for a strong and reliable system that could address and highlight such transactions as potentially fraudulent. The study suggests employing machine learning techniques and algorithms, more precisely decision trees and a random forest, for analyzing transactional data and selecting suspicious patterns indicative of such behavior. The system's training on a dataset containing historical information about fraudulent activities enables it to learn and recognize similar patterns, providing an advanced and proactive method of approaching to tackle credit card fraud. In addition to machine learning, the research implements application of varied authentication approaches, such as one-time passwords and security questions, to confirm the authenticity pertaining to credit card. By integrating these authentication measures with the analytical power of Naive Bayer's machine learning model, the system showed an accuracy of 98.75%. The expected outcome is not limited to finding and stopping fraud but also the enhanced authentication course, which leads to billions of dollars saved on an annual basis for consumers and financial organizations together. In the end, it is the goal of this study to generate a reliable solution to secure transactions involving credit cards relying on an algorithmic synergy for machine learning alongside strong authentication.

Index Terms— Fraud detection, deep learning, machine learn- ing, online fraud, credit card fraud, transaction data analysis.

I. INTRODUCTION

A. Background

Credit card fraud detection is an important topic for financialinstitutions and consumers. Many different algorithms andvarious methods have been devised for this purpose, like a Markov model of the CCFD system, a sensitive Decision Tree, a Support Vector Machine, and a Random Forest. Neuralnetworks were also suggested to be employed in this role,and the current paper specifically exploits the whale swarm optimization method to look for an optimal incentive valueand then corrects it utilizing the BP network. Nonetheless, these methods have numerous disadvantages, such as reduced accuracy, time-consuming processes, and inconveniences to the cardholder during the course of inquiry. It is important to keep a comprehensive log of all transactions, but that entails keeping a lot of data in retention. It is also impossible to identify the cardholder during online transactions that log the IP addresses for identification confirmation, meaning that cybercrime is an inevitable part of fraud investigation.

B. Problem Statement

The major problem with the existing system for detecting credit card fraud is that high-quality datasets are rare, andmost of the available ones are imbalanced and have largenumbers of unknown or absent data . To address this issue, overcome these drawbacks, an innovative system is suggested in this paper. This system relies on the whale algorithm and theSMOTE technique to improve the accuracy and efficiency of the algorithm and the level of misclassified normal transactionas fraud. The whale algorithm constitutes a type of swamintelligence system utilizing humpback whale performancewhich is rapidly becoming among the most efficient problem- solving areas . The SMOTE technique, which oversamplesthe minority class, is implemented to fix the problem with imbalanced datasets. This proposed system also seeks for enhancing the fraud detection in credit card transactions process's precision and productivity by minimizing instances of false positive and false negative outcomes. The suggested system will be able to learn from the past; by integrating the whale algorithm and SMOTE approach, it will recognize the new forms of fraudulent activity and, therefore, remainone step ahead of the fraudster. The intent is to establish a system that, using transaction data, can detect and identify instances of credit card fraud. In that case, it is planned to analyze previous consumer transaction data and outline trends in conduct. Finally, groupings were trained individually using different classifiers. After that, the findings obtained after applying accuracy, sensitivity, specificity, and precision as well as methods of authentication were considered to choose oneof the most efficient approaches in fraud prediction based on the classifier with a superior score rating. These approaches involve assessing just how well the predictors have performed.

II. RELATED WORK

Techniques from machine learning to detect fraudulent Master Card transactions were suggested by IFA in Parth Roy,Prateek Rao, Jay Gajre [1]. This approach aims to develop the current system for detecting fraud by reducing false alarmrates and improving the detection rate of scams. With data on card transactions continuing to be collected, the methods can be modified and implemented in the fraud detection system employed by banks to find and prevent fraud. For the protec- tion of both financial institutions and customers, an efficient system for detecting fraud remains essential, as European cardholders have reported 284,807 communications pertainingto fraud. Large-scale transactional data can be analyzed using techniques in machine learning, which can additionally beused to identify patterns that can point to fraudulent conduct. By identifying and stopping fraudulent transactions promptly, banks can mitigate financial losses while safeguarding customers against identity theft. Overall, implementing machine learning within the system for detecting credit card fraud isa decisive step towards addressing the intricate challengesposed by fraudulent activity. Ishika Sharma, Shivjyoti Dalai, Venktesh Tiwari, Ishwari Singh , Seema Kharb [2] presented that the problem in detecting credit card fraud can be addressed using various methods such as Naive Bayes, Random Forest, and Logistic Regression. Each transaction is evaluated individ- ually, and the most efficient technique for detecting fraud is applied. The primary goal is to identify and prevent fraud by improving the existing techniques to achieve optimal results. The approaches are utilized on credit card data to create a system capable of detecting frauds correctly identify frauds with the least amount of false positives. A real-time fraud detection technique should be able to identify frauds with the least possible effect on genuine transactions. In summary, the purpose of these techniques is to enhance the accuracyand effectiveness of fraud detection in credit card transactions discovery to ensure that real transactions are not mismarked asfalse and that data attempting to fake it is immediately blocked. Anuruddha Thennakoon, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, Nuwan Kuruwitaarachchi [3] offered an evaluation that offers a thorough manual for choosing theoptimal algorithm to address specific type of frauds, and we use an adequate performance metric to show the evaluation.In order to determine if a particular transaction is legitimateor fraudulent, they also created the application of predictive analytics performed by the implemented models from machine learning and an API module. H. Najadat, O. Altiti, A. A.Aqouleh, and M. Younes [4] conducted a study aimed at ad- dressing the challenge of imbalance classification and assessed methods from machine learning for fraud detection. There were several options analyzed and weak points, concerningthe machine learning fraud detection approach currently used.

M.W. S. Makki, Z. Assaghir, Y. Taher, R. Haque, M.-S. Hacid and H. Zeineddine provided several classification techniques based on machine learning techniques, such as Naive Bayes, Random Forest, and Logistic Regression, for the imbalanced datasets problem. The methods were tested using precision, accuracy, recall, f1 score, confusion matrix, and Roc-AUC score.

III. METHODOLOGY

A. Workflow Model Architecture

The architecture for fraud detection algorithm will involve several components, including data collection and analy- sis, machine learning-based classification algorithms, and a decision-making mechanism. The decision-making mechanism may include a rule-based system or a more sophisticated approach, such as a reinforcement learning algorithm. In con- clusion, such an architecture will have a feedback loop where the output of the decision-making mechanism is fed backfor enhancing the precision of the machine learning-basedclassification algorithms within the subsequent transactions. Such architecture will have to manage large amounts of data and perform the processing of transactions in real time for effective and efficient fraud detection.

The flowchart is illustrated in Figure. 1 depicts the process of credit card fraud detection and authentication that utilizes machine learning and a two-way authentication approach withOTP and face detection. It begins with receiving transac- tion data, followed by pre-processing and rule-based checksto identify suspicious patterns. An ML model analyses the data, assigning risk scores to transactions. Threshold checks compare the risk scores against predefined thresholds. Forflagged transactions, OTP verification is initiated, and simul- taneously, face detection captures the cardholder's face. Face recognition algorithms verify the cardholder's identity. Based on successful verification, a decision is made regarding the transaction. Reporting provides insights into fraud patterns andsystem performance. The flowchart concludes, representing theconclusion of the system.

B. Model Selection

The Naive Bayer's approach described in this report is grounded on the presumption that all features make equalcontributions to the classification target, and each possible attribute in the given record is statistically independent. Ina few practical instances, this assumption might not hold true. Therefore, the Bayes rule in this situation calculates the chance of an input record to be from each class of possibleand predict the output as a class with the largest probability result. This will enhance the prediction capability of the modelregardless of whether attributes are not independent as perthe assumption. The Bayes rule is a probabilistic model that considers prior knowledge of determining the likelihood of each class and the likelihood of the observed features to determine the likelihood of an input belonging to each class. By using this approach, the system can accurately classify inputs despite the existence of interdependencies among the attributes.

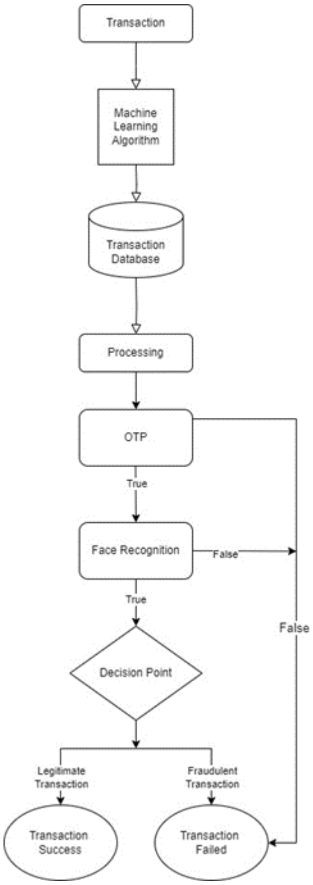


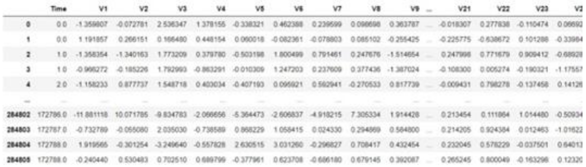
Fig. 1. Workflow Model Architecture.

C. Dataset

Data is extracted from the Kaggle Platform, also we have collected transactional data sourced from credit card com- panies, including cardholder information, transaction history, transaction location, and transaction amount. The data set contains details of more than 250,000 transactional data whichcontains both legitimate and fraudulent transactions. The data set also contains missing data which needs to be cleaned before processing it. At the same time, there presents an imbalanced data set that can be structured by normalizingand upsampling the least data set which in this case, hereis a fraudulent transaction. The data set examined has thefollowing dimensions:

Transaction Type	Count
Legitimate Transaction	284315
Fraud Transaction	492

The data-set that is being used to train our model shown in Figure.2 This dataset is also represented in a bar graph in Figure.3 as shown.



	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...	V21	V22	V23	V2
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...	-0.018307	0.277838	-0.110474	0.06692
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...	-0.225775	-0.638672	0.101288	-0.33984
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...	0.247998	0.771679	0.909412	-0.68928
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...	-0.108300	0.005274	-0.190321	-1.17557
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...	-0.009431	0.798278	-0.137458	0.14126
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
284802	172786.0	-11.881118	10.071785	-9.834783	-2.066656	-5.364473	-2.606837	-4.918215	7.305334	1.914428	...	0.213454	0.111864	1.014480	-0.50934
284803	172787.0	-0.732789	-0.055080	2.035030	-0.738589	0.868229	1.058415	0.024330	0.294869	0.584800	...	0.214205	0.924384	0.012463	-1.01622
284804	172788.0	1.919565	-0.301254	-3.249640	-0.557828	2.630515	3.031260	-0.296827	0.708417	0.432454	...	0.232045	0.578229	-0.037501	0.64013
284805	172788.0	-0.240440	0.530483	0.702510	0.689799	-0.377961	0.623708	-0.686180	0.679145	0.392087	...	0.265245	0.800049	-0.163296	0.12320

Fig. 2. Transactional Data-set

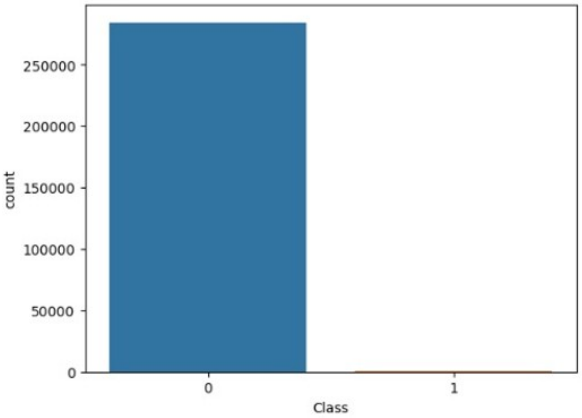


Fig. 3. Data-set graph Representation.

A. User Interface

The user interface the facilitated by an application. An API is being given for contacting bank servers when required.OTP (One-Time Password) verification is a key component for detecting credit card fraud programs. By integrating OTP verification into the system, an additional layer of security is added to authenticate transactions and prevent unauthorized card usage. When a suspicious transaction is flagged, the program can generate and send an OTP to the cardholder's registered mobile number or email address. The user thenenters this unique OTP within a specified time limit to validatethe transaction. If the OTP is entered correctly, the transactionmoves to the next step of authentication; otherwise, it is declined.

in Figure4 the verification still makes the transaction be one- way transaction and such that even if a fraudster gets holdof credit card details, it would not complete the transaction without OTP. OTP verification raises the security related to fraudulent credit card activity systems and systems by enablingan additional verification step even after supplying the card information to verify the user's possession of the smartphone. Ways in which security is guaranteed when two-way authentication fails within credit card transactions. This is accomplished while avoiding the transaction from happening before the criminal purchase occurs. It tightens security when facial recognition is successful in two-way authentication within credit card transactions. Once the user initiates a transaction, their face is captured and compared to a reference imagethrough facial recognition algorithms. If the facial recognition process is successful, along with a valid OTP verification, the transaction is approved.

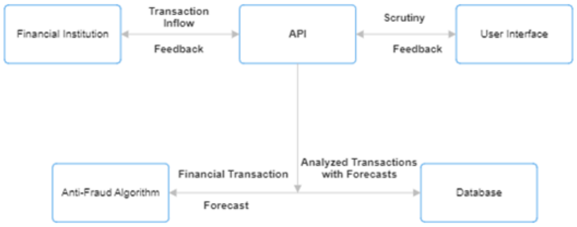


Fig. 4. User Interface.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

The Hardware Requirement is a computer system should have a Pentium IV 2.4 GHz processor, which is a relatively old processor by today's standards. However, this processor should be sufficient to run the proposed software application. The processor's clock speed plays a crucial role in determiningthe system's performance, and 2.4 GHz is considered an acceptable speed for most basic software applications. Forthe purpose of storing the data set that is used for trainingthe model a hard disk capacity of 500 GB is the quantity of storage space required, which is sufficient for most software applications, and this must be more than enough for the suggested model mentioned in our paper. Also, the computer system must have 4GB of RAM... The software application in this project requires a Windows 7 and above operating system.The choice of operating system is critical because it affects the software application’s compatibility and performance. The software application is developed using Python v 3.7 or above.

B. Results on Proposed Architecture

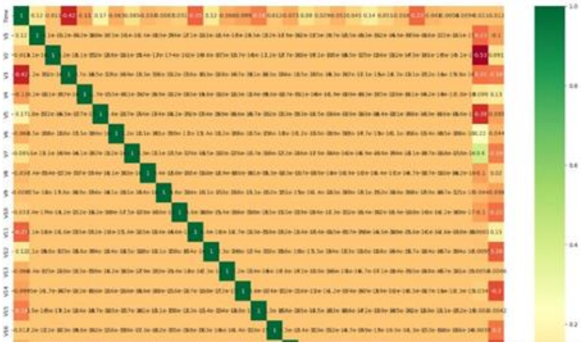


Fig. 5. Vizualizing the Data-set using HeatMap.

The HeatMap shown in Figure.5 is a visual representation of data that uses different colors to depict values. It can be used toanalyze various aspects of credit card usage and transactions. One common application is to examine the difference in the variable’s patterns, where the heatmap can represent different spending categories or time periods. This is mainly used in our project to help credit card companies identify unusual transaction patterns or outliers that could signal potential fraud. Furthermore, credit card issuers may employ heatmaps to assess risk by analyzing historical data, allowing them to identify high-risk areas or industries. They are a great approach to visualizing, which would make it simple and straightforward to analyze patterns, tracking, and exceptions infraudulent credit card transactions. This program may involve numerous verification channels such as facial identification verification, single-time passwords, and security questions to validate the identity of a credit card user. It ensures that a credit card’s legitimate owner completes all transactions, makingit almost impossible for unauthorized people to use a card. This architecture has provided us with an accuracy grade of 98.75%. The expense of commerce and payment for the end- user would be minimized drastically by implementing thisproject. Because the level of crime is reduced, fewer fraudulenttransactions will be executed with cost-free concessions and investigations and so financial institutes will save millions through offering checkups. Furthermore, consumers will not incur any associated losses due to fraudulent transactions.

The system under consideration has proven as a plausiblesolution to the problem of fraudulent credit card activity. Through the incorporation involving machine learning algo- rithms and multiple authentication processes, such systemswill also reduce credit transaction risks by identifying and authenticating any malicious cohorts prior to conducting anyin their actions that might result in expenditure reduction in consumers and financial institutions.

V. CONCLUSION AND SCOPE OF FUTURE WORK

Finally, fraudulent activity involving credit cards poses a significant challenge to the financial sector and it is impor-tant to adopt a dependable system capable of detecting and prevent such an action. This proposed system proposal will try to solve this problem by employing machine learning algorithms and authentication approaches. Including machine learning algorithms and decision trees, random forests, andso on, the installed system will behave depending on the transactional data for the identification of any suspicion. Lever- aging knowledge gained from previous fraud activities will beidentifying patterns and stop fraudulent activities. With this method, the system will be able to detect fraudulent activities as they are happening, stopping the further losses of credit card companies and customers. It will also utilize different authentication approaches for authenticating the credit card user. These approaches include facial detection authentication, one-time passwords, and security questions. The authenticationmethods ensure that the credit cards can only be accessed by authorized individuals. Using many authentication methods, the credit card accounts’ unauthorized access is minimized, reducing fraudulent activities. The assumed system is expectedto contribute greatly to the security of credit card transactions.It would prevent previously identified fraudulent activitiesbefore they happen and authenticate the customer accordingly, preventing the scenarios from happening to cause financial losses to both consumers and financial institutions. In addition, the system will aid in the reduction of costs related to fraud investigations and reimbursements. This will generate enor- mous savings for financial organizations. The creative system'sproduction is crucial in combating increasing fraudulent activi-ties involving credit cards. A range utilizing machine learning techniques, methods and verification techniques is enhances security measures during credit card transactions, avoiding fraudulent activity, and reducing the cost to consumers, and financial organizations.

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, Venkatesh 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. Zeined-dine, "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 Almusal- lam, 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 Computing and Control Systems, IEEE Xplore 2020.

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

[12] Subramanian R.R., Seshadri K. (2019) Design and Evaluation of a Hy- brid Hierarchical Feature Tree Based Authorship Inference Technique. In: Kolhe M., Trivedi M., T Tiwari S., Singh V. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, vol 39. Springer, Singapore.

[13] Joshva Devadas T ., Raja Subramanian R. (2020) Paradigms for Intelli- gent IOT Architecture. In: Peng SL., Pal S., Huang L. (eds) Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm. Intelligent Systems Reference Library, vol 174. Springer.

[14] R. R. Subramanian, B. R. Babu, K. Mamta and K. Manogna, "Design and Evaluation of a Hybrid Feature Descriptor based Handwritten Character Inference Technique," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-5.