iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51848

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Credit Card Fraud Detection Using Machine Learning

Achal T[1], Pratiksha B[2], Kalyani T[3], Vaishnavi M.[4], Rajvi D[5], Noorain S[6], Prof. Devchand Chaudhari Sir[7] (Guide)
[1, 2, 3, 4, 5, 6] *Student, Department of Computer Science Engineering, Government College of Engineering, Chandrapur, Maharashtra, India*

Abstract: ***Credit Card Fraud Detection is presently the most frequently occurring problem in the present. This is due to the rise in both online transactions and E-Commerce platforms. Credit Card Fraud Detection is mainly required by credit card users. The major goal of our project is to detect the given credit card transaction is real or fraud. This project aims to focus mainly on machine learning algorithm. We design the application to help the credit card users to be able to notify the transaction is real or fraud.***

Keywords: Digitalized Information, Authentication.

I. INTRODUCTION

In Today's present world scenario, most of our cash transactions have been switched to cashless transactions due to the rise of credit cards and the Internet. With the wide number of services available online, credit cards make it very simple for us to pay for them at the same time. With the ease in usability and convenience in executing transactions comes the threat of cybercrimes [1][3][16]. Every year the users and companies fall victim to fraudulent transactions and end up bearing huge losses. As of 2018, credit card owners in the U.S. created only 21.54% of $40.582 trillion in worldwide credit card volume in 2018, the companies in the United States suffered 33.99% of the total credit card fraud losses worldwide. To prevent this, we have proposed a unique credit card fraud detection system that will simply monitor the incoming transactions and then classify whether the transaction is fraudulent or not in real-time. '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 behavior 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 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 behavior 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 outnumber 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 analyze 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 feedback to the automated system which is used to train and update the algorithm to eventually improve the fraud-detection performance over time. Fraud detection methods are continuously developed to defend criminals in adapting to their fraudulent strategies.

These frauds are classified as:

1) Credit Card Frauds: Online and Off
2) Card Theft
3) Account Bankruptcy
4) Device Intrusion
5) Application Fraud
6) Counterfeit Card
7) Telecommunication Fraud

Some of the currently used approaches to detection of such fraud are:

a) Artificial Neural Network
b) Fuzzy Logic
c) Genetic Algorithm
d) Logistic Regression
e) Decision tree
f) Support Vector Machines
g) Bayesian Networks
h) Hidden Markov Model
i) K-Nearest Neighbor

- Credit card frauds are easy and friendly targets. E-commerce and many other online sites have increased the online payment modes, increasing the risk for online frauds.
- Increase in fraud rates, researchers started using different machine learning methods to detect and analyses frauds in online transactions.
- In year 2020 there were 1,387,615 cases of identity theft happens which causes to $40.85 Billions
- Fraud detection methods are continuously developed to defend criminals in adapting to their fraudulent strategies.

These frauds are classified as:

- Credit Card Frauds: Online and Offline
- Card Theft
- Account Bankruptcy
- Device Intrusion
- Application Fraud

A. Problem Statement

The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be a fraud. This model is then used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

Credit card frauds are increasing heavily because of fraud financial loss is increasing drastically. Every year due to fraud Billions of amounts lost. To analyze the fraud there is lack of research. Many machine learning algorithms are implemented to detect real world credit card fraud. KNN and hybrid algorithms are applied.

B. Objective

The key objective of any credit card fraud detection system is to identify suspicious events and report them to an analyst while letting normal transactions be automatically processed.

For years, financial institutions have been entrusting this task to rule-based systems that employ rule sets written by experts. But now they increasingly turn to a machine learning approach, as it can bring significant improvements to the process.

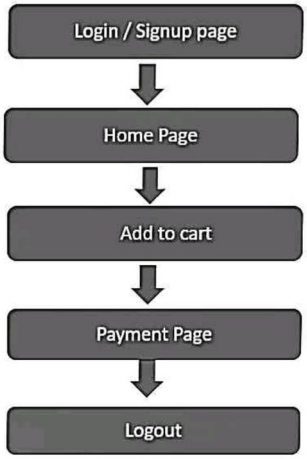
1) Higher accuracy of fraud detection. Compared to rule-based solutions, machine learning tools have higher precision and return more relevant results as they consider multiple additional factors. This is because ML technologies can consider many more data points, including the tiniest details of behaviour patterns associated with a particular account.
2) Less manual work needed for additional verification. Enhanced accuracy leads reduces the burden on analysts. “People are unable to check all transactions manually, even if we are talking about a small bank,” Alexander Kondor, data science competence leader at Attest, explains. “ML-driven systems filter out, roughly speaking, 99.9 percent of normal patterns leaving only 0.1 percent of events to be verified by experts.”
3) Fewer false declines. False declines or false positives happen when a system identifies a legitimate transaction as suspicious and wrongly cancels it.
4) Ability to identify new patterns and adapt to changes. Unlike rule-based systems, ML algorithms are aligned with a constantly changing environment and financial conditions. They enable analysts to identify new suspicious patterns and create new rules to prevent new types of scams

C. Methodology

The main approach of this paper is to design and develop a novel fraud detection method for Streaming Transaction Data, with an objective, to analyze the past transaction details of the customers and extract the behavioral patterns. Where cardholders are clustered into different groups based on their transaction amount. Some of the currently used approaches to detection of such fraud are:

- Artificial Neural Network
- Fuzzy Logic
- Genetic Algorithm
- KNN

DESIGNING OF PROJECT:



The approach that this paper proposes, uses the latest machine learning algorithms to detect anomalous activities, called outliers. The basic rough architecture diagram can be represented with the following figure:

When looked at in detail on a larger scale along with real life elements, the full architecture diagram can be represented as follows. First of all, we obtained our dataset from Kaggle, a data analysis website which provides datasets. Inside this dataset, there are 31 columns out of which 28 are named as v1-v28 to protect sensitive data. The other columns represent Time, Amount and Class. Time shows the time gap between the first transaction and the following one. Amount is the amount of money transacted. Class 0 represents a valid transaction and 1 represents a fraudulent one.

We plot different graphs to check for inconsistencies in the dataset and to visually comprehend it: This graph shows that the number of fraudulent transactions is much lower than the legitimate ones. This graph shows the times at which transactions were done within two days. It can be seen that the least number of transactions were made during night time and highest during the days. This graph represents the amount that was transacted. A majority of transactions are relatively small and only a handful of them come close to the maximum transacted amount. After checking this dataset, we plot a histogram for every column. This is done to get a graphical representation of the dataset which can be used to verify that there are no missing any values in the dataset. This is done to ensure that we do not require any missing value imputation and the machine learning algorithms can process the dataset smoothly. After this analysis, we plot a heatmap to get a colored representation of the data and to study the correlation between out predicting variables and the class variable. This heatmap is shown below:

The dataset is now formatted and processed. The time and amount column are standardized and the Class column is removed to ensure fairness of evaluation. The data is processed by a set of algorithms from modules. The following module diagram explains how these algorithms work together: This data is fit into a model and the following outlier detection modules are applied on it:

- Local Outlier Factor
- Isolation Forest Algorithm

These algorithms are a part of sclera. The ensemble module in the sclera package includes ensemble-based methods and functions for the classification, regression and outlier detection. This free and open-source Python library is built using NumPy, SciPy and matplotlib modules which provides a lot of simple and efficient tools which can be used for data analysis and machine learning. It features various classification, clustering and regression algorithms and is designed to interoperate with the numerical and scientific libraries.

We've used Jupiter Notebook platform to make a program in Python to demonstrate the approach that this paper suggests. This program can also be executed on the cloud using Google Collab platform which supports all python notebook files. Detailed explanations about the modules with pseudocodes for their algorithms and output graphs are given as follows:

1) Local Outlier Factor

It is an Unsupervised Outlier Detection algorithm. 'Local Outlier Factor' refers to the anomaly score of each sample. It measures the local deviation of the sample data with respect to its Neighbors. More precisely, locality is given by k-nearest neighbors, whose distance is used to estimate the local data. The pseudocode for this algorithm is written as:

On plotting the results of Local Outlier Factor algorithm, we get the following figure: By comparing the local values of a sample to that of its neighbors, one can identify samples that are substantially lower than their neighbors. These values are quite amanous and they are considered as outliers.

As the dataset is very large, we used only a fraction of it in out tests to reduce processing times. The result with the complete dataset processed is also determined and is given in the results section of this paper.

2) Isolation Forest Algorithm

The Isolation Forest isolates observations by arbitrarily selecting a feature and then randomly selecting a split value between the maximum and minimum values of the designated feature. Recursive partitioning can be represented by a tree, the number of splits required to isolate a sample is equivalent to the path length root node to terminating node.

The average of this path length gives a measure of normality and the decision function which we use. The pseudocode for this algorithm can be written as: On plotting the results of Isolation Forest algorithm, we get the following figure: Partitioning them randomly produces shorter paths for anomalies. When a forest of random trees mutually produces shorter path lengths for specific samples, they are extremely likely to be anomalies. Once the anomalies are detected, the system can be used to report them to the concerned authorities. For testing purposes, we are comparing the outputs of these algorithms to determine their accuracy and precision.

II. PROPOSED SYSTEM

We propose an end-to-end system that can be easily be deployed in any organization dealing with real-time online credit card transactions. The proposed architecture will consist of a training model known as the Artificial Neural Network to be able to correctly classify whether the incoming transaction is fraudulent or not in real-time. The model is then further integrated with Apache Airflow to automate the prediction model without any human intervention and also

added a simple functionality where if the predicted transaction in real-time has a Fraud Score ≥ 0.7 then the Fraud Manager will simply receive an alert on the dashboard or their authenticated Gmail inbox to review the transaction and act accordingly. We have also deployed a simple GUI using the FLASK application as the front end. The dataset used in our training and prediction was sourced from an open library on Kaggle. The data was collected in September 2013 over 48

hours of European Credit Card Users. The dataset consists of a total of 284,807 transactions where only 492 transactions are classified as fraud. Since the positive class only accounts for a total of 0.172% due to this the data is highly unbalanced or skewed [2][7]. For privacy reasons, most of the values in our dataset are numerical since they have undergone Principal Component Analysis (PCA). Only the attributes 'Time', 'Amount' and 'Class' are kept as it is. The PCA technique is used as an attempt to reduce the dimensionality of the dataset and then scaling it to be used with the learning model. The data is then oversampled and under sampled to create a balanced dataset. On oversampling, we use Synthetic Minority Oversampling Technique (SMOTE) to generate more fraud transactions to match the number of non-fraud transactions. This ends up increasing the total number of transactions present in the dataset. Similarly, on under sampling, we end up losing the number of non-fraud transactions to match it with the number of fraud transactions. This ends up in creating a dataset consisting of fewer transactions as compared to that of the original dataset.

III. RESULTS AND CONCLUSION

Our proposed system for an end-to-end fraud detection in real-time ensures the accuracy of 99.91% on its test results. The architecture of the system enables it to be easily scalable, dynamic, and elegant. It requires almost no human intervention until something suspicious is detected. The added functionality of Gmail alerts allows the fraud manager to focus only on those transactions which seem suspicious to fraud and then take an action to whether block, review, or pass the transaction. The K NEAREST NEIGHBOUR(KNN) model provides us the reliability of our results and Apache Airflow further helps us automate the entire job.

As per the main objective of the project is to classify and identify the credit card fraudsters and approval of cards for the users based on ML algorithms is being discussed throughout the project. Credit card fraud and approval is most common problem resulting in loss of lot money for people and loss for some banks and credit card company.

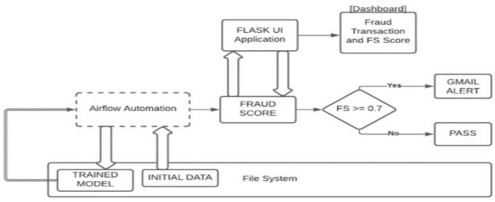
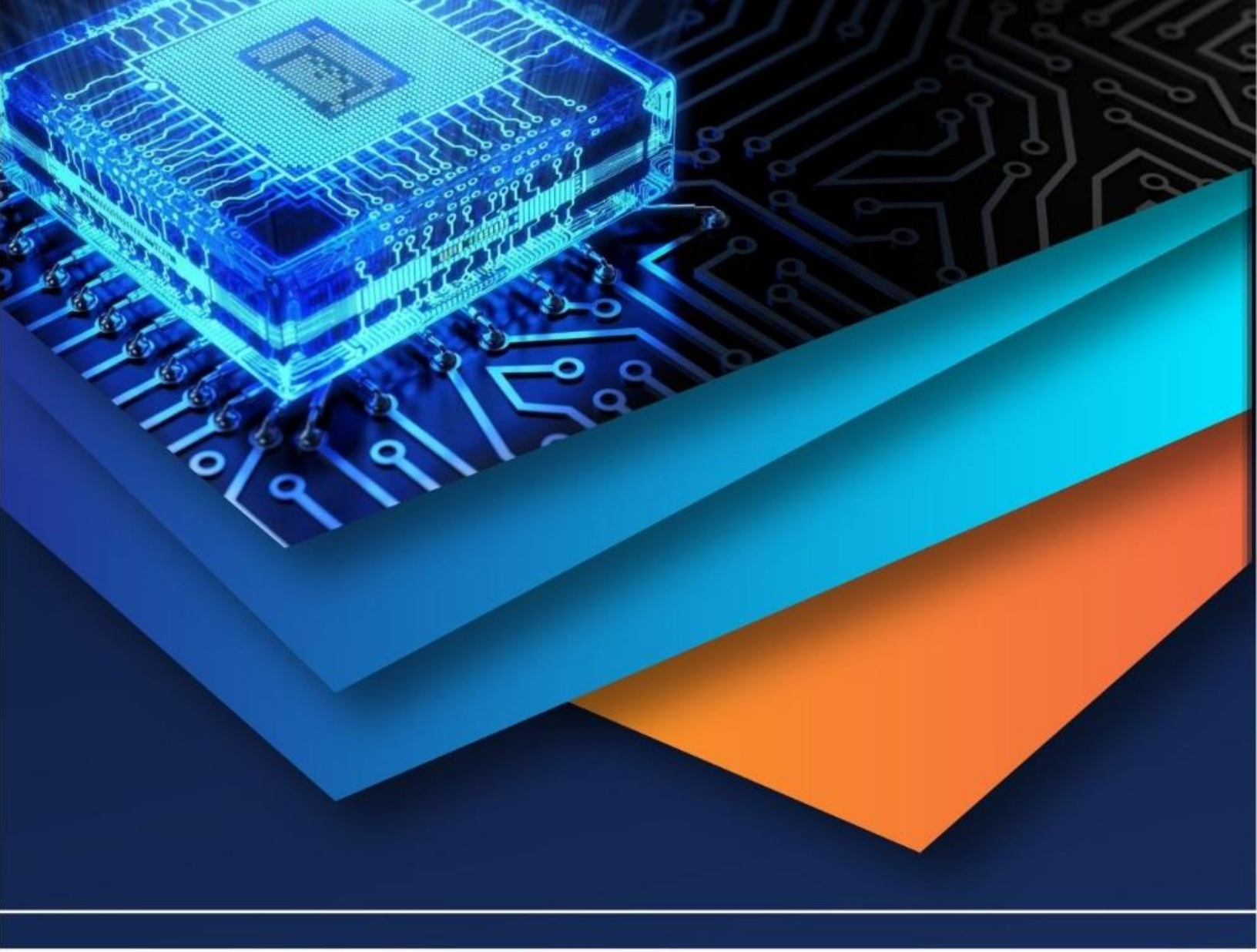


Figure 1. System Architecture

This project wants to help the peoples from their wealth loss and also for the banked company and trying to develop the model which more efficiently separate the fraud and fraudulent transaction by using the time and amount feature in dataset given in the KAGGLE. We built the model using some machine learning algorithm (KNN).

REFERENCES

[1] https://towardsdatascience.com/the-random-forestalgorithm-d457d499ffcd
[2] https://www.xoriant.com/blog/productengineering/decision-trees-machine-learningalgorithm.html
[3] Gupta, Shalini, and R. Johari. " A New Framework for Credit Card Transactions Involving Mutual Authentication between Cardholder and Merchant." International Conference on Communication Systems and Network Technologies IEEE, 2021:22-26.
[4] Y. Gumby and K. G. Co, "Global online payment methods: the Full year 2020," Tech. Rep., 3 2020.
[5] Bolton, Richard J., and J. H. David. "Unsupervised Profiling Methods for Fraud Detection." Proc Credit Scoring and Credit Control VII (2020): 5– 7.
[6] Drummond, C., and Holte, R. C. (2019). C4.5, class imbalance, and cost sensitivity: why under-sampling beats oversampling. Proc of the ICML Workshop on Learning from Imbalanced Datasets II, 1–8.
[7] Quah, J. T. S., and Sri Ganesh, M. (2020). Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications, 35(4), 1721-1732.
[8] "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
[9] 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
[10] "Survey Paper on Credit Card Fraud Detection by Suman" , Research Scholar, GJUS&T Hisar Hessonite published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014
[11] "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



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY