e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING WITH PYTHON

Pratik Bawangade[*1], Mayur Gedam[*2], Faisal Khan[*3], Shadab Sheikh[*4], Akash Bhilawe[*5], Prof. Muzaffar Khan[*6]

[*1,2,3,4,5]Student, Department Of Electronic And Telecommunication Engineering, Anjuman College Of Engineering And Technology, Nagpur, Maharashtra, India.

[*6]Guide, Department Of Electronic And Telecommunication Engineering, Anjuman College Of Engineering And Technology, Nagpur, Maharashtra, India.

ABSTRACT

Online transactions have become a significant and crucial aspect of our lives in recent years. It's critical for credit card firms to be able to spot fraudulent credit card transactions so that customers aren't charged for things they didn't buy. The number of fraudulent transactions is rapidly increasing as the frequency of transactions increases. Machine Learning and its algorithms can be used to solve such issues. With Credit Card Fraud Detection, this project aims to demonstrate the modelling of a data set using machine learning. Modeling prior credit card transactions with data from those that turned out to be fraudulent is part of the Credit Card Fraud Detection Problem. The model is then used to determine whether or not a new transaction is fraudulent. Our goal is to detect 100% of fraudulent transactions while reducing the number of inaccurate fraud classifications. Credit Card Fraud Detection is an example of a common classification sample. On the PCA converted Credit Card Transaction data, we concentrated on evaluating and pre-processing data sets, as well as deploying different anomaly detection techniques such as the Local Outlier Factor and Isolation Forest algorithm.

Keywords: Credit Card Fraud Detection, Python, Fraud Detection Analysis, Detection Using Python, Fraud Detection Using Machine Learning, Detecting Fraud Transactions.

I. INTRODUCTION

Credit Card Fraud is described as when a person uses another person's credit card for personal gain while the owner and card issuing authorities are uninformed.

Credit card fraud is a simple and inviting target. E-commerce and many other online sites have increased the number of payment options available online, raising the risk of online fraud. Due to the rise in fraud rates, researchers began employing various machine learning approaches to detect and analyze online transaction fraud.

Credit card fraud is a common occurrence. Without posing any risks, a large sum of money can be withdrawn in a short period of time without the owner's knowledge. Fraudsters strive to make every fraudulent transaction appear legal, making fraud detection a tough and time-consuming task.

Credit Card Fraud is described as when a person uses another person's credit card for personal gain while the owner and card issuing authorities are uninformed.

Due to the rise and acceleration of E-Commerce, there has been a massive increase in the use of credit cards for online purchasing, resulting in a high number of credit card frauds. The necessity to identify credit card frauds has become increasingly important in the digital age. Fraud detection entails tracking and analyzing the behavior of a variety of users in order to detect or avoid fraudulent activity. We need to grasp the many technologies, algorithms, and types involved in identifying credit card frauds in order to properly detect credit card frauds. An algorithm can tell whether a transaction is fraudulent or not. To detect fraud, they must have access to a dataset as well as information of fraudulent transactions. They categorize all transactions after analyzing the dataset.

Fraud detection is tracking the behaviors of large groups of people in order to predict, detect, or minimize unacceptable activity such as fraud, intrusion, or defaulting. To analyze all permitted transactions and report suspect ones, machine learning techniques like Logistics Regression are used. Professionals evaluate these

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

reports and call cardholders to establish whether the transaction was legitimate or fraudulent. The investigators submit feedback to the automated system, which is utilized to train and update the algorithm over time in order to improve fraud detection effectiveness.

II. METHODOLOGY

Proposed System

Fraud detection is a problem that applies to a variety of businesses, including banking and finance, insurance, government agencies, and low enforcement, among others. Millions of transactions can be searched using advanced data mining algorithms to find patterns and detect fraudulent transactions.

Credit card fraud detection is a technique for detecting fraudulent credit card transactions and preventing customers from being charged for products they did not purchase.

The major goal of our project is to make Credit Card Fraud Detection more accessible to people who are victims of credit card online fraud. The primary goal of a credit card fraud detection system is to keep our transactions and security safe. Fraudsters won't be able to make repeated transactions on a stolen or counterfeit card before the cardholder notices the fraudulent activity with this approach. The model is then used to determine whether or not a new transaction is fraudulent. Our goal is to detect 100% of fraudulent transactions while reducing the number of inaccurate fraud classifications.

The following figure shows the complete Fraud Detection Process

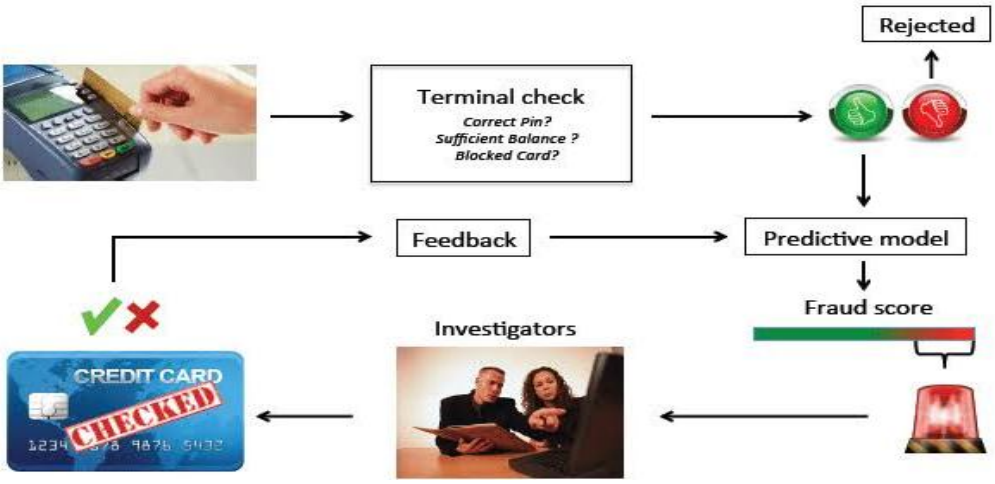

Figure: Fraud Detection Process

Flow Diagram

Basically, there are five steps in Credit Card Fraud Detection process

Step 1: Dataset (Credit Card Data)
Step 2: Data Pre-Processing
Step 3: Data Analysis
Step 4: Train Test split
Step 5: Logistics Regression Model
Step 6: Evaluation

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com

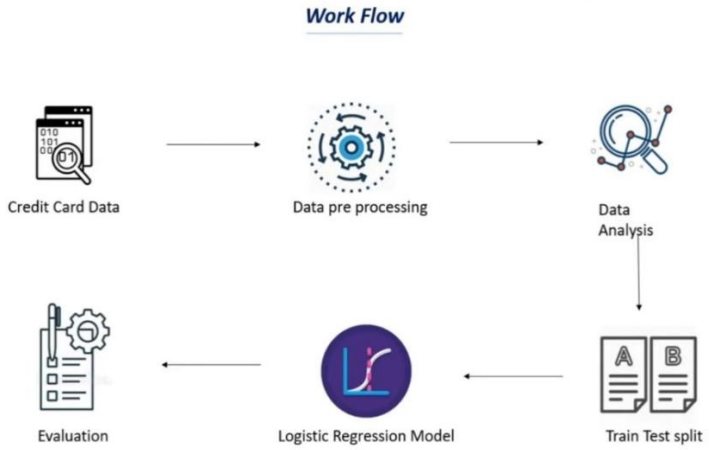

Figure: Work Flow of System

Step 1: The first step is to collect data. In order to do so, we must first determine where we may gather data and what attributes are present in the dataset. Data Preprocessing is the next step after we receive the data.

Stage 2: Because the dataset we have is quite unbalanced, this is a difficult step in data preprocessing. This is a crucial aspect of the process. We'll also do some data pre-processing here.

Step 3: In this section of the data analysis, we will learn about the numerous properties of the dataset and the relationships between them. It will provide us with information about the dataset we have in order to help us find a better model for this specific goal.

Step 4: Next, divide the data into Training and Test categories. We feed this training data into our Machine Learning Model, and then we test or find the accuracy of our model after it has been trained.

Step 5: After splitting the data, we'll feed the training data into a Logistic Regression Model, which is the model we'll employ because this is a Binary Classification problem. In this scenario, we'll determine whether a transaction is legitimate or fraudulent.

Step 6: The final step is to assess our model's performance; in this case, we'll use Testing data to do so

III. RESULT

We recommend testing accuracy using the Area Under the Precision-Recall Curve because of the class imbalance ratio (AUPRC). For unbalanced categorization, the accuracy of the confusion matrix is meaningless.

The code prints the number of false positives it found and compares it to the real numbers. This is used to compute the algorithm's accuracy score and precision.

The percentage of data we used for speedier testing was 10% of the total dataset. At the end, the entire dataset is used, and both results are printed.

These results, as well as the classification report for each algorithm, are included in the output, where class 0 indicates that the transaction was considered to be valid and class 1 indicates that the transaction was determined to be fraudulent.

IV. CONCLUSION

Fraud detection is a complicated problem that necessitates extensive planning before applying machine learning techniques. Nonetheless, it is a solid use of data science and machine learning, as it ensures that the customer's money is secure and not readily tampered with.

The Random Forest algorithm, as I mentioned earlier, will be fine-tuned in the future. Having a data set with non-anonymized features would make this even more fascinating, as displaying feature importance would allow one to identify which individual characteristics are most significant for detecting fraudulent transactions. Please do not hesitate to contact me if you have any questions or discover any errors. The introduction of this article includes a link to the notebook containing my code.

V. REFERENCES

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

[2] L.J.P. van der Maaten and G.E. Hinton, Visualizing High-Dimensional Data Using t-SNE (2014), Journal of Machine Learning Research.

[3] Machine Learning Group — ULB, Credit Card Fraud Detection (2018), Kaggle.

[4] Nathalie Japkowicz, Learning from Imbalanced Data Sets: A Comparison of Various Strategies (2000), AAAI Technical Report WS-00–05.