



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: V Month of publication: May 2022

DOI: <https://doi.org/10.22214/ijraset.2022.43240>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Credit Card Fraud Detection System

Aakash¹, Akash kumar², Ravi Pratap Singh³

^{1, 2, 3}Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India

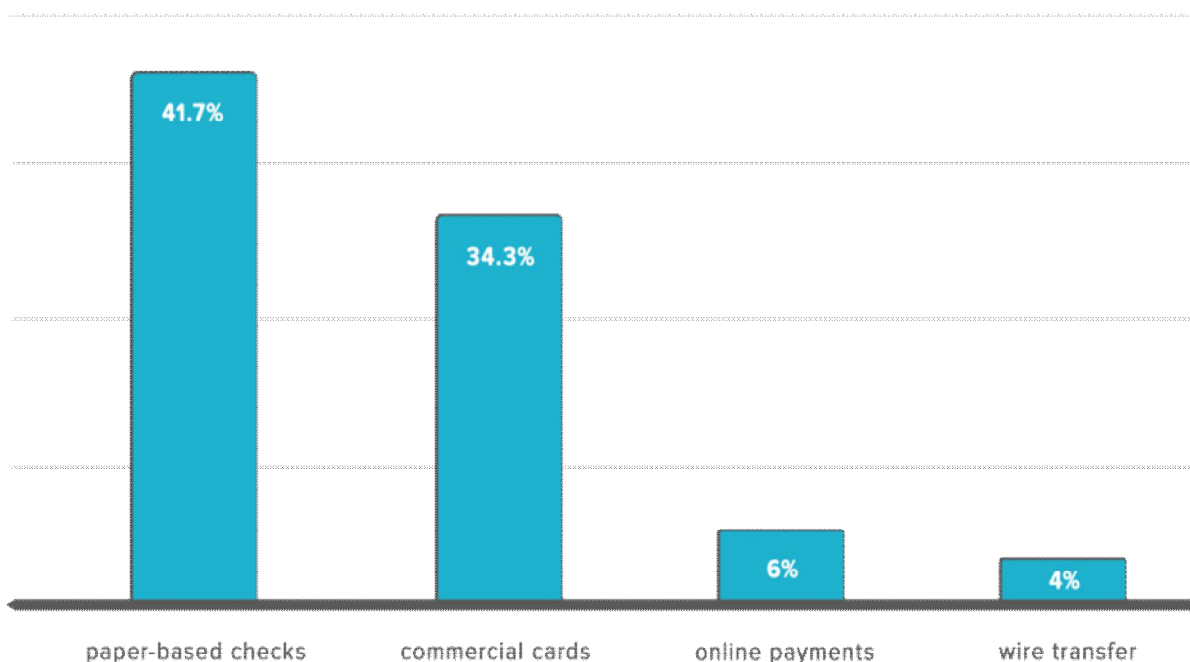
Abstract: In today's world everything is online which also increases the chances of fraud. There are approximately 36.4 percent fraud related to commercial cards which include credit card, debit card, etc. In 2022 there are 64 million people who uses credit card to initiate the transaction, therefore they are also prone to the card fraud. So, in this document we talk about the various machine learning algorithms to predict the credit card fraud based on the previous transaction pattern.

Keywords: K nearest neighbour, Machine learning, Credit card, decision tree, random forest

I. INTRODUCTION

The objective of this document is to find the fraudulent transaction over non-fraudulent transaction in credit card. Credit card is a thin hard plastic card which contains identification information of a card holder. Credit card contain twelve-digit number which has utmost important. This information is read by Automated Teller Machines (ATMs). The use of credit card is increasing year by year therefore the rate of fraudulent transactions is also increasing rapidly. Credit card fraud is a big term by theft and fraud committed using or involving time payment by using this card. The goal could be to buy goods without payment, or transfer of unauthorized funds from account. Credit card fraud is also an addition to ownership theft. As information from the United States Federal Commercial Commission, the level of identity theft was caught stabilized in the mid-2000s, but increased by 21percent in 2008. Although credit card fraud, that is a crime The majority of people associated with identity theft, dropped as percent of all ID theft complaints. In the year 2000, there were 13 billion produced annually, about 10 million or one of all 1300 transactions it appeared to be counterfeit. [1]

Payment fraud
Source: Aberdeen Group



II. LITERATURE REVIEW

Fraud acts as illegal or as criminal fraud aimed at financial or personal gain profit. It is an intentional act that is against the law, the law or policy for the purpose of obtaining unauthorized financial gain. Many books related to confusing discovery or fraud on this domain has been already published and available for public use. In-depth research by Clifton Phua and his participants indicated that strategies were used in this domain extracts applications for data mining, automated fraud discovery, discovery of enemies. Unusual strategies such as hybrid data mining / complex network the partition algorithm is able to detect illegal events is a data set of actual card activity, based on network a reconstruction algorithm that allows for creation single event diversion presentations in the reference team appeared to be generally successful in online shopping of medium size. Fraud detection is complex function and there is no system that accurately predicts any which is practiced as fraud. Properties of good fraud detection are:

- 1) The accuracy of system will be high.
- 2) It must be fast to detect fraud.
- 3) It should not declare honest transaction to fraud. [2]

III. PROBLEM STATEMENT

The credit card fraud detection system includes the understanding the previous pattern and predicting the correct output whether the transaction is fraudulent or non-fraudulent. Our aim is to build the model which gives the output that is 100 percent accurate and to minimize the fraudulent transaction in the society.

IV. PROBLEM SOLUTION

Hardware requirement for credit card fraud detection system

- RAM - 4GB and higher
- Processor - Intel i3 and above
- Hard disk – 500GB minimum

Software requirement for credit card fraud detection system

- OS - windows or Linux
- Python ide – python 2.7.x and above
- Language - python scripting

Tools and library required

- Python – 3.x
- NumPy – 1.19.2
- Scikit-learn – 0.24.1
- Matplotlib – 3.3.4
- Jupyter notebook

We will be creating and training our model for predicting whether a transaction is fraudulent or not. Since we are using the multiple algorithms, we will compare the accuracy scores after testing and pick the most accurate algorithm. Our solution is mainly divided into three steps, that are:

- 1) Pre-processing of data
- 2) Implementation of machine learning algorithms
- 3) Finding best model

Pre-processing of data: This step includes the processing of data set so that our model will give more accurate result. Pre-processing can be done by removing or treating null values, removal of dummy variables, treating outliers, etc.

Implementation of machine learning algorithms: We have split our data in 70 percent and 30 percent. Larger part of our dataset is used to train our model and rest of the part of our dataset is used for testing our model. To train our model we are using three machine learning algorithms which are decision tree, k nearest neighbour and random forest. After implementing these models, we compare the accuracy of different model with each other and algorithm with higher accuracy will be used.

- a) *Decision Tree*: Decision tree is a tree like structure where each node represents the decision taken on an attribute and each branch represent the outcome and leaf node represent the class label. It basically split the source set into the subsets. This step is repeated on each derived subset until it reaches the desired result. On implementing decision tree on our dataset, we attain the accuracy of 99.9288989494457 percent. [3]
- b) *K-nearest Neighbour*: The k-nearest neighbour (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression). On implementing k-nearest neighbour on our dataset, we attain the accuracy of 99.9506645771664 percent. [4]
- c) *Random Forest*: Random Forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It is a collection of decision tree It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. On implementing random forest on our dataset, we attain the accuracy of 99.95611109160493 percent.[5]

Finding best model: After applying the decision tree, k-nearest neighbour and random forest and on comparing it with each other we can clearly see that the accuracy of random forest is more than that of decision tree and k-nearest neighbour. So we will choose random forest algorithm for our model.

V. CONCLUSIONS

We can clearly see that random forest has highest accuracy. Therefore, we will use random forest algorithm for our model. Now we will be able to detect the credit card fraud by analyzing the pattern of spending on every card and to figure out any inconsistency with respect to the “usual” spending patterns. This will help to prevent losses in million. We will be able to minimize the fraudulent transaction using this model.

REFERENCES

- [1] <https://www.ijert.org/research/credit-card-fraud-detection-IJERTCONV9IS04018.pdf>
- [2] https://www.ripublication.com/ijaer18/ijaerv13n24_18.pdf
- [3] About decision tree <https://www.geeksforgeeks.org/decision-tree/>.
- [4] About K-nearest neighbour <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761#:~:text=Summary,that%20data%20in%20use%20grows.>
- [5] About random forest <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/#:~:text=Random%20forest%20is%20a%20Supervised,average%20in%20case%20of%20regression.>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)