International Journal of Engineering & Scientific Research
Vol.11 Issue 05, May 2023
ISSN: 2347-6532 Impact Factor: 6.660
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Using ML models to determine best approach to detect credit card fraud

Dewansh Chaudhary*1, Diksha Singht*2, Ajay Singh*3, Faraz Ahmad Siddqui*4,Prof.Jain Singh*5

*1,2,3,4AKTU, Department of Computer Science & Engineering, IIMT College Of Engineering,GreaterNoida,UttarPradesh, India.

*5Department Of Computer Science & Engineering, IIMT College Of Engineering,GreaterNoida,UttarPradesh, India.

ABSTRACT

With the growth in e-commerce, credit card and debit card use has also increased. This also comes with the increase in credit card frauds. Credit card fraud occurs when unauthorized users attain an individual's credit card information and misuse this information to make purchases, open accounts, etc. The aim of our project is to develop a ML model which learns from Fraud-Detection Dataset to detect Credit card fraud in real-time.

Keywords: Credit card, Fraud Detection, e-commerce, ML model, real-time

I. INTRODUCTION

The risk of online fraud also increased with the rapid increase in online merchants, card users, and card issuers in the last few years.

Statistics show that the growth of online banking in India has become an important part of everyone's lives. However, this also increased the risk of online fraud in India. It is also important for online users to be aware of risks and implement the necessary precautions to protect their credit card details.

In the 21st century, various financial institutions made Internet banking and E-payment methods available to the public for purchasing goods and services very convenient and provide credit cards to make their life easy without carrying cash.

Credit cards allow consumers to purchase goods and services using the funds in their bank accounts. It also offers the consumers protection in case of damage, lost, or even stolen of their purchased goods

Overall, the availability of credit cards has made online transactions more secure and convenient for consumers and become a fundamental part of their daily uses.

Aggregate of credit cards issued in India from the year 2013 to the year 2021 (in millions)

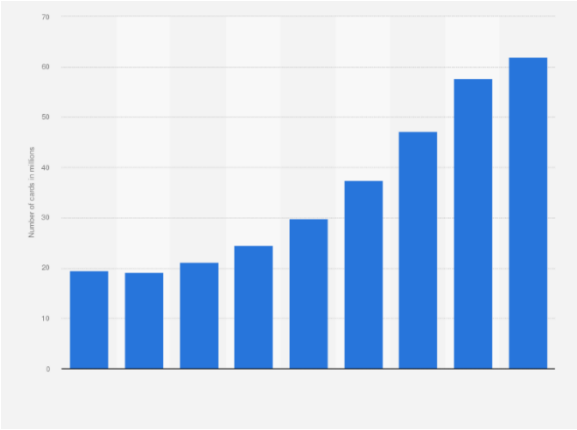

Fig. 1 Bar Chart of number of credit cards in India

India has about 77 million active credit cards. Credit card growth has been increasing significantly in India. HDFC bank is the largest issuer of credit card in India.

Machine learning is a collection of multiple algorithms, working on historical data to suggest risk rules.

Using of Machine learning for Fraud detection

Machine learning model can recognise unusual credit card transactions and fraud. They can detect thousands of patterns from huge data files. Following are the benefits of credit card fraud detection using machine learning:

-Rapid observation

-Higher accuracy

-Scalable

Working of Machine Learning in fraud detection

-Input Data:

Machine learning algorithms work on historical datasets. In this framework, it will be transaction data. Firstly, the model has to collect data, as much data it collects the better the model work.

-Extract features:

Features extraction is a process of modifying source data into binary data. Features extraction remove the redundant and unnecessary data. In machine learning features extraction boost the model reliability.

-Training and algorithm:

Giving feedback data is the key to clarifying and getting finer accuracy. In machine learning models it is important to trained it before using it. By training algorithm, it boosts the model reliability.

-Create model:

Lastly, when the all steps are completed, our fraud detection machine leaning model is completed. This model can detect fraud in a high speed with higher reliability.

II. LITERATURESURVEY

In our project (Credit Card fraud Detection) we help to find the more efficient way to detect the fraud and suspicious activity with the help of different algorithms and alert the costumer or user who is used the credit or debit card for several need. We will discuss on these algorithms which is used in our project mentioned above.

Decision Tree:

Decision tree is a machine learning algorithm which is very helpful in machine learning project for classification and regression analysis purpose and achieve a specific goal. Basically, the main goal of this algorithm to train the modal for predict the value from the previous data. It works like a similarity tree by using its own analysis. The analyzed data are defining in the terms of transaction and satisfy certain rules and condition. Here we used the decision tree for analyze the individually checked the transaction activity.

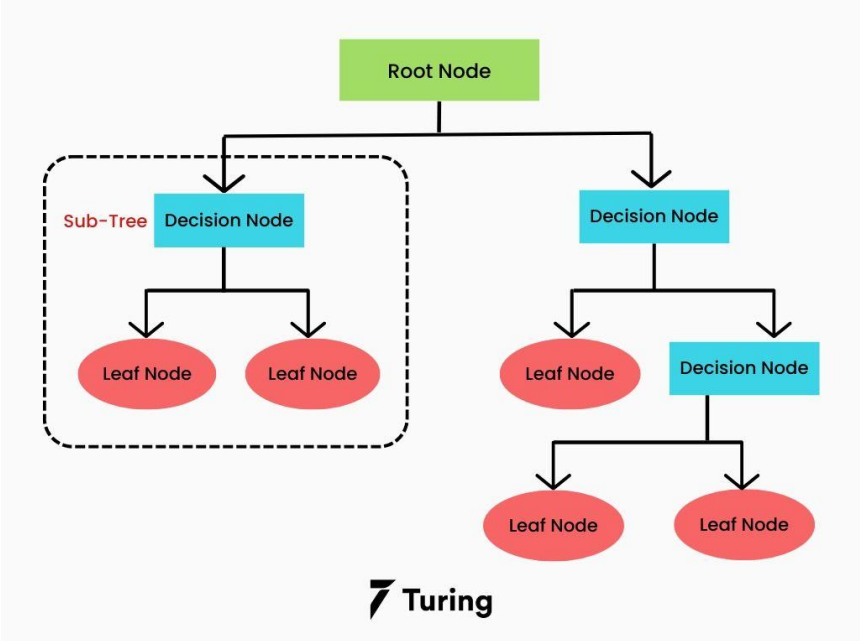

Fig. 2 Decision Tree Representation

Support Vector Machines (SVMs):

Support Vector machine (SVMs) is a machine learning algorithm which is generally used in classification problem in 1960's. The SVM's modal is basically represent different classes or object for data in hyperplane in a multidimensional space. The main goal of this algorithm minimized the error according to margin of the classes and object and find minimum marginal hyperplane (MMH).

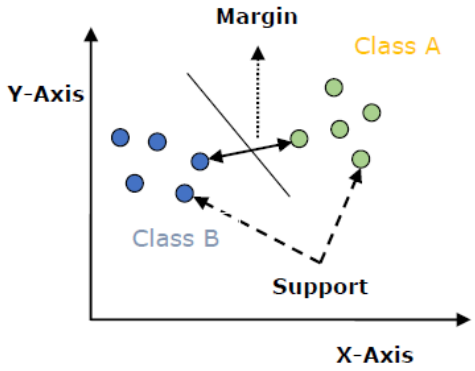

Fig. 3 Graph of SVMs

Random Forest:

Random Forest is a supervised machine learning algorithm developed by Tin Kam Ho in 1995. Random Forest is most used algorithm due to its accuracy and simplicity. It can be used in regression and classification tasks. Here we used Random Forest for predict the data according the dataset. First the dataset is divided into the tree nodes and after the similar nodes merged together and find the accurate prediction.

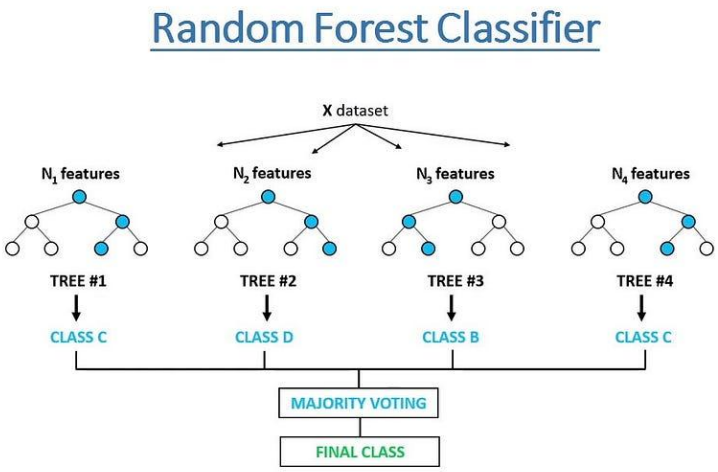

Fig. 4 Diagram of Random Forest

K-Nearest Neighbor Algorithms:

The K-Nearest Neighbor algorithm is a machine learning algorithm based on supervised technique. It is mostly used to predict the data value to classify the similar feature of the data. KNN algorithm is helped to identify the category, feature or class of a particular dataset.

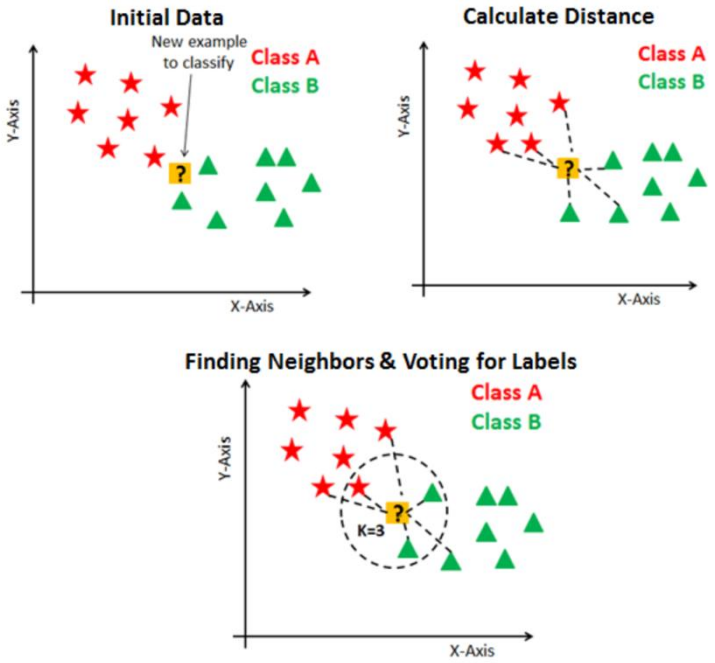

Fig. 5 Process of KNN

III. METHODOLOGY

Dataset

We will be using Python 3.9 and its libraries in a Jupyter Notebook to Determine Dataset used: Credit Card Fraud Detection on kaggle.com.

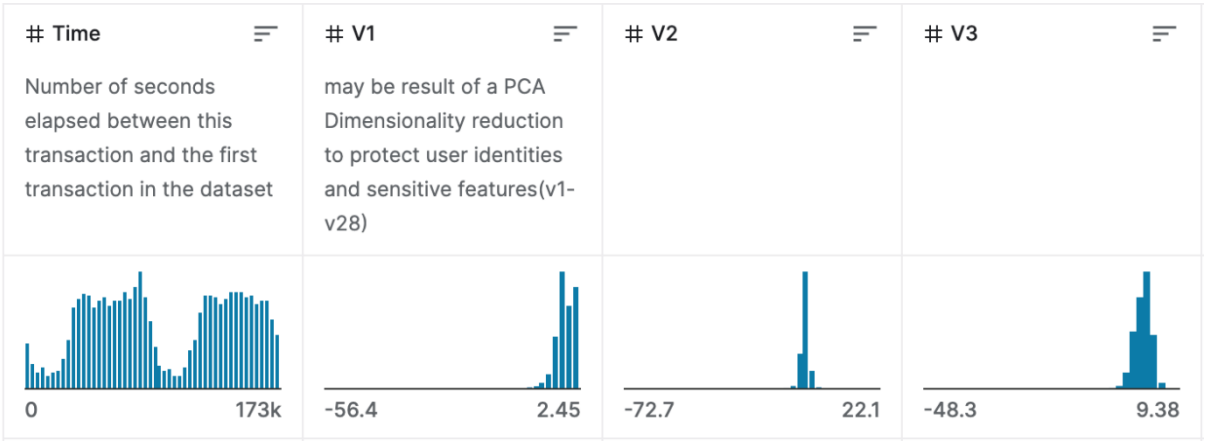

Fig. 6Snapshat of Dataset

The transactions made with credit cards completed by European cardholders within the month of September 2013 have been included in the dataset. It comprises of transactions over two successive days, with 492 out of 284,807 transactions classified as fraudulent.

The dataset was skewed, with just 0.172% of all transactions being labelled as fraudulent. The dataset only comprises numerical input parameters resulting from a Principal Component Analysis transformation, with attributes V1 - V28 indicating the PCA principle components. PCA was not used to alter the 'Time' and 'Amount' characteristics. 'Time'

refers to the number of seconds that have passed between each one of the transactions and the first transactional data in the dataset, whereas 'Amount' refers to the transaction amount. The answer variable is indicated by the 'Class' feature, with a value of 1 representing fraud and 0 representing no fraud. For example-dependent cost-sensitive learning, the 'Amount' function might be beneficial.

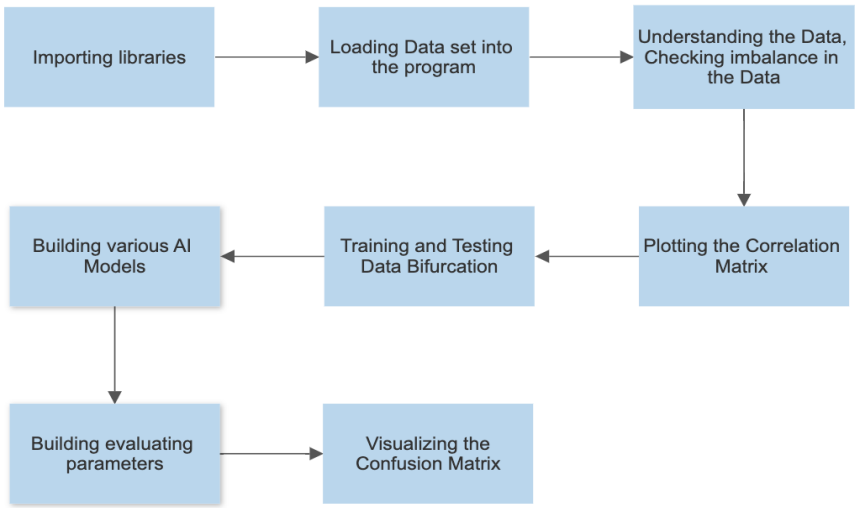

Fig. 7 Flow Chart of Project

Importing Libraries

In our project, we utilize several Python libraries for data manipulation and visualization. Numpy is one such library, which we use for working with arrays and for manipulating pandas Dataframes. Pandas, on the other hand, provides various data structures and operations for working with numerical and relational data. For creating static, animated, and interactive visualizations in Python, we rely on the comprehensive library, Matplotlib. Additionally, we make use of Seaborn, a visualization library that is based on Matplotlib.

Loading dataset into the program

Using Pandas methods, we read the dataset and convert the csv file into a Pandas Dataframes.

Understanding the Data

We read the various attributes of the dataset and determine the optimal features in it. We also detect the imbalance in the data.

Plotting the correlation matrix

To predict the most relevant features, we plot the correlation matrix, which provides a graphical representation of the correlations between different features. By dividing the data

into input parameters, we gain a better understanding of how the features are correlated with each other.

Splitting Training and Testing data

Our dataset will be separated into two primary categories: one for training the model and the other for evaluating the performance of our trained model.

Building various AI models

We will build and test different models; random forest model, decision tree model, linear regression.

Building the evaluating parameters

We will check the accuracy, the precision, the recall, the F1 – score of all the AI models have built.

Visualizing the Confusion Matrix

Visualizing the confusion matrix through a plot enables us to observe the true positives, true negatives, false positives, and false negatives. This offers a more comprehensive analysis compared to solely examining the accuracy of classifications.

IV. RESULT

In our experimentation of trying to find the best algorithm for credit card fraud detection, we found the following results:

Random Forest – 99.95%

Logistic Regression – 97%

K-Nearest Neighbours – 99.86%

Support Vector Classifier – 99.89%

* These percentages were calculated on the accuracy of each model to correctly detect credit card fraud.

While the difference between the accuracy scores may seem minute between the models, they are significant when considering huge number of data. So, these differences are significant for us to use only the best model for credit card fraud detection.

V. CONCLUSION

From this research we can conclude that Random Forest Model is the most accurate model for credit card fraud detection. Using this algorithm for monitoring will give many benefits on the industry.

As credit card fraud is detected more accurately then it will generate more sales and retain customers. It will also build trust in the system while also improving customer satisfaction. It also adds a layer of security to banking.

VI. REFERENCE

[1] John RichardD. Kho,LarryA. Vea,"Credit CardFraud Detection Basedon Transaction Behaviour", Proc.ofthe2017IEEERegion10Conference(TENCON),Malaysia, November 5-8, 2017

[2] CLIFTONPHUA1,VINCENTLEE1,KATESMITH1&ROSS GAYLER2"A ComprehensiveSurveyofDataMining-basedFraudDetectionResearch",SchoolofBusinessSystems,FacultyofInformationTechnology,MonashUniversity,WellingtonRoad,

Clayton,Victoria3800,Australia

[3] Suman, Research Scholar, GJUS&T Hisar HCE, Sonepat, "SurveyPaperonCreditCardFraudDetection", International JournalofAdvancedResearchinComputerEngineering&Technology(IJARCET)Volume3Issue3,March2014

[4] Wen-Fang YU and Na Wang, "Research on Credit Card Fraud Detection Model Based on DistanceSum",InternationalJointConferenceon Artificial Intelligence, 2009

[5] MassimilianoZanin,MiguelRomance, Regino Criado, and SantiagoMoral, "CreditCardFraudDetectionthroughParencliticNetworkAnalysis",Hindawi Complexity Volume 2018, Article ID5764370,9 pages

[6] "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy",IEEETRANSACTIONSONNEURALNETWORKSANDLEARNINGSYSTEMS,VOL.29,NO.8,AUGUST2018

[7] David J.Wetson,DavidJ.Hand,MAdams,Whitrow and Piotr Jusczak "Plastic Card Fraud Detection using Peer Group Analysis" Springer, Issue 2008.