

CREDIT CARD FRAUD DETECTION

Kalakuntla Sumanth¹, Gade Praveen Kumar², V.Rakesh³

¹ Student, Information Technology, B V Raju Institute of Technology, Telangana, India

² Student, Information Technology, B V Raju Institute of Technology, Telangana, India

³ Assistant Professor, Information Technology, B V Raju Institute of Technology, Telangana, India

ABSTRACT

Credit card plays a very important role in today's economy. Although using credit cards provides enormous benefits when used carefully and responsibly, significant credit and financial damages may be caused by fraudulent activities. Credit card fraud detection is when a business takes steps to prevent stolen money, products, or service obtained via an illegitimate credit card transaction. Credit card fraud can happen both by the cardholder themselves or by someone else. Due to the increase in electronic payments, the risk for credit card fraud is turning into a substantial challenge for financial institutions and service providers, which is forcing them to continuously improve their fraud detection systems. The credit card fraud detection problem includes modelling past credit card transaction with data of the ones that turned out to be fraud. There are many other credit card fraud detections using different algorithms giving out different accuracies. As these algorithms are not 100% accurate, we are trying to use different algorithm for better accuracy. So, in this credit card fraud detection project we are trying to detect the fraudulent transactions using better machine learning algorithms.

1. INTRODUCTION

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data. The field of study known as machine learning is concerned with the question of how to construct computer programs that automatically improve with experience

1. CLASSIFICATION

Classification uses an algorithm to accurately assign test data into specific categories. It recognizes specific entities within the dataset and attempts to draw some conclusions on how those entities should be labeled or defined. Common classification algorithms are linear classifiers, support vector machines (SVM), decision trees, k-nearest neighbor, and random forest, which are described in more detail below. □ Regression is used to understand the relationship between dependent and independent variables. It is commonly used to make projections, such as for sales revenue for a given business. Linear regression, logistical regression, and polynomial regression are popular regression algorithms.

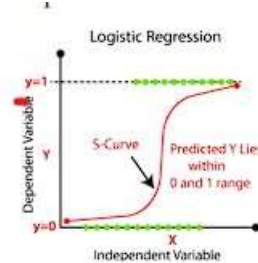
1.2 Objective of Project

In recent years, topics such as fraud detection and fraud prevention have received a lot of attention on the research front, in particular from payment card issuers. The reason for this increase in research activity can be attributed to the huge annual financial losses incurred by card issuers due to fraudulent use of their card products. A successful strategy for dealing with fraud can quite literally mean millions of dollars in savings per year on operational costs. Fraud prevention is interesting for financial institutions. The advent of new technologies as telephone, Automated Teller Machines (ATMs) and credit card systems have amplified the amount of fraud loss for many banks. Performing the analysis manually is literally impossible, while automation of this process might present a lot of practical difficulties. Analyzing every transaction is legitimate or not is very expensive. Moreover it is also time consuming, hence it is not practically possible. Confirming whether a transaction was done by a client or a fraudster is a better option, but by phoning all card holders is cost prohibitive if it is check in all transactions. Further it might

also lead to customer dissatisfaction. Fraud prevention by automatic fraud detections is where the well-known classification methods can be applied, where pattern recognition systems play a very important role.

2. Logistic Regression

A logistic regression is used for modelling the outcome probability of a class such as pass/fail, positive/negative and in our case—fraud/not fraud. Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible Values.



K-Nearest Neighbors The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. K-Nearest Neighbors is a Classification algorithm that counts similarities based on the distance in multi-dimensional space. The data point, therefore, will be assigned the class that the nearest neighbors have. This method is not vulnerable to noise and missing data points, which means composing larger datasets in less time. Moreover, it is quite accurate and requires less work from a developer in order to tune the model.

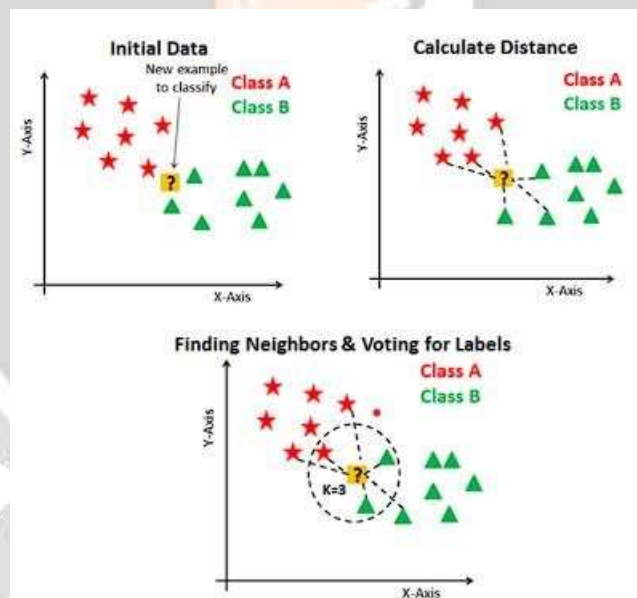
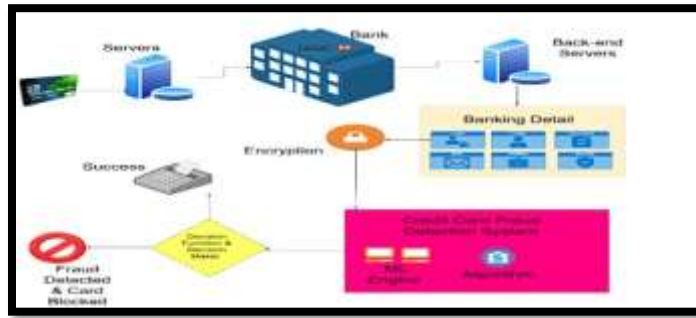


Fig -1 K NEAREST NEIGHBORS

2. PROPOSED SYSTEM

- 1)The proposed system will be able to detect frauds by considering a cardholder's spending habit without its significance.
 - 2)The proposed system will be an ideal choice for addressing this problem of current fraud detection system.
- Advantages of Proposed System
- 3)Easy accessing all transactions.
 - 4)Here we can know individual transactions to any fraud detection systems.
 - 5)Here legit transactions and fraud transactions are labeled as 0 & 1.



2.2 DESCRIPTION OF MODULES

Module I- Importing the needed Libraries:

1)The first thing we must do is import the necessary dependencies that we will use in the next part of our program. This project uses numpy, pandas, random, os, math, pickle, mfcc, scipy. io. wavefile.

Module II-Data Balancing:

- 1)Imbalanced classes are a general issue in ML based classification where there is an abnormal count of each class.
- 2)Due to these transactions it quite difficult to build a standard model with this much less number of fraudtransactions.



Module III-Feature Extraction:

- We use heatmap which provides a good visualization of the major and minor values in thematrix as different colored cells that define the values.
- Here, rows/columns of the matrix are clustered in sets.



MODULE IV: Outlier Detection:

- The outlier detection technique measures the distance of each data similar to the clusteringtechnique, but is used to find specific data and rules that are separated from the total data.
- Here our aim is to reduce the outliers to have a better trained model.

Module V- Feature Extraction:

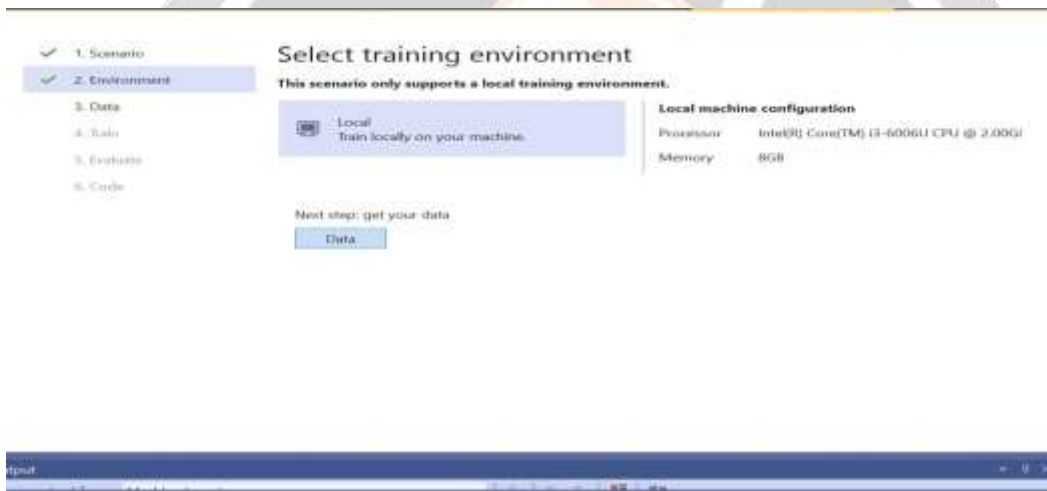
Feature extraction is a process to extract important features from data. Firstly, we use clustering method to divide the cardholders into different clusters/groups based on their transaction amount, i.e., high, medium and low using range partitioning.

Module VI- Training the Model and making predictions:

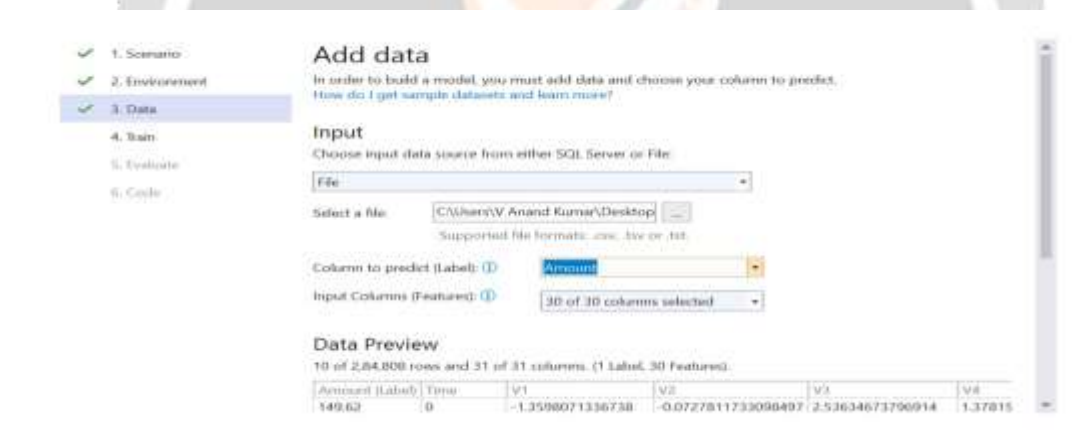
- The step has come you all were waiting to feed the data to KNN algorithms and make all predictions and receive accuracy on the test dataset. This step code seems to be large, but it is very small because we are following a functional programming approach in a stepwise manner, so we only need to call the functions. The first is to get neighbors, extract class and check the accuracy of the model.

3. RESULTS

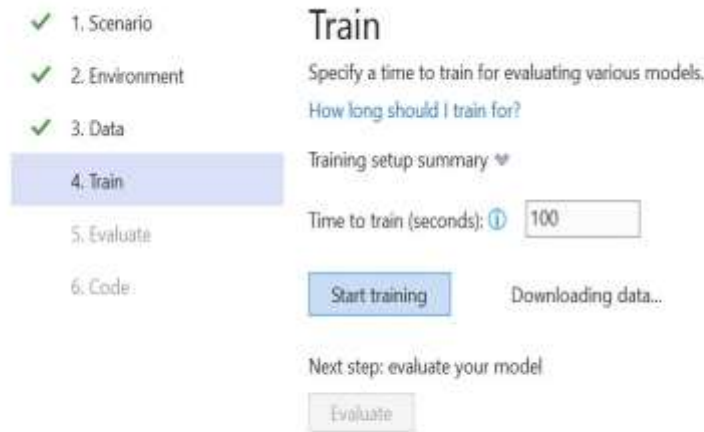
We will use the Credit Card Fraud Detection data set file in .csv format. Select the dataset from your local directories where you place the .csv file then Select the column that will be used to prediction label.



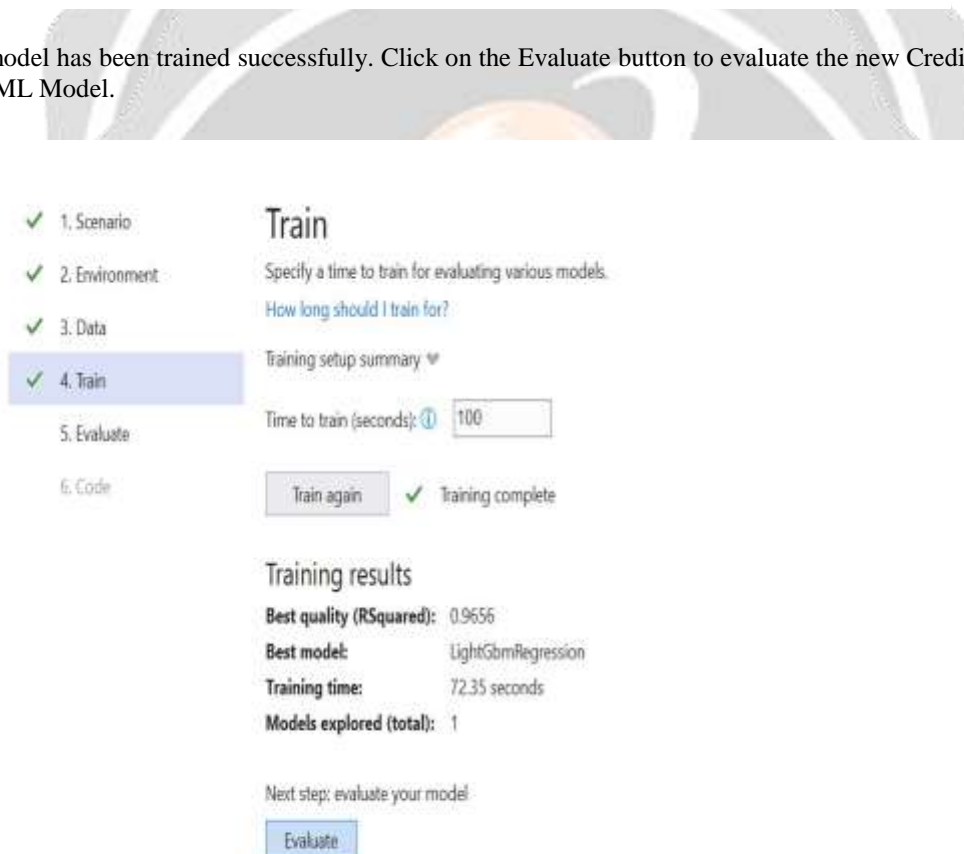
After training dataset selection, select the Machine learning task as binary-classification because this model will tell us only the transaction is fraud or not. Select the time to train the model. Click on start training.



ML.NET Model builder will start training our Machine Learning Model. I've selected only 2 minutes (100 seconds). You can choose the duration of training as you want.



Now our model has been trained successfully. Click on the Evaluate button to evaluate the new CreditCard Fraud Detection ML Model.



✓ 1. Scenario
 ✓ 2. Environment
 ✓ 3. Data
 ✓ 4. Train
 ✓ 5. Evaluate
 6. Code

Best model:
RSquared: 0.9656
Model: LightGbmRegression

Try your model

Sample data
 The following fields below are pre-filled by row 1 of your data.

Time

V1

V2

V3

V4

Results
 Amount: 124.64

The Evaluate window shows the performance of the best models on this dataset

4. CONCLUSIONS

Credit card fraud is without a doubt an act of criminal dishonesty. This article has listed out the most common methods of fraud along with their detection methods and reviewed recent findings in this field. This paper has also explained in detail, how machine learning can be applied to get better results in fraud detection along with the algorithm, pseudocode, explanation its implementation and experimentation results. One aim of this study is to spot the user model that best identifies fraud cases. There are numerous methods for detection of credit card fraud. If one in all these or combination of rule is applied into bank credit card fraud detection system, then the probability of fraud transactions can be predicted soon after credit card transactions by the banks. A series of anti-fraud ways is adopted to forestall the banks from nice losses sooner and scale back the risks

We can conclude that between logistic regression and KNN, logistic regression stood best by giving accuracy of 98%, whereas KNN gave accuracy of 94%. As we considered very small amount of data for analyzing and also as the data is imbalanced we couldn't get decent precision value. This high percentage of accuracy is to be expected due to the huge imbalance between the number of valid and number of genuine transactions.

5. REFERENCES

- [1]. <https://www.kaggle.com/mlg-ulb/creditcardfraud>
- [2]. <https://visualstudio.microsoft.com/downloads/>
- [3]. <https://dotnet.microsoft.com/apps/machinelearning-ai/ml-dotnet/model-builder>