Credit Card Fraud Detection

PROJECT REPORT

Machine Intelligence

BACHELOR OF TECHNOLOGY- V Sem CSE

Department of Computer Science & Engineering

SUBMITTED BY

Batch No: 1

Student Name 1: Satish
SRN: PES2UG20CS550

Student Name 2: Sharath
SRN: PES2UG20CS552

Student Name 3: Shreepathi
SRN: PES2UG20CS553

PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)

100 Feet Ring Road, BSK III Stage, Bengaluru-560085

Abstract and Scope

Due to a rapid improvement in the electronic commerce technology, the utilize of credit cards has augmented. As credit card becomes the trendiest style of payment for individually online as well as habitual acquisition, luggage of credit card fraud also growing. Economic fraud is increasing radically with the development of modern technology and the global super highways of communication, consequential in the loss of billions of dollars worldwide each year. The falsified transactions are sprinkled with genuine transactions and simple pattern corresponding techniques are not often sufficient to detect those frauds accurately. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. Many modern techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions.

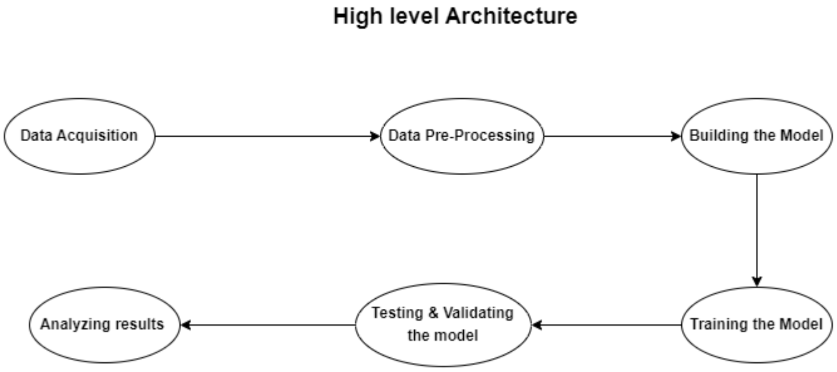
The ANN and KNN algorithm are an evolutionary search an optimization technique that Mimics natural evolution to find the best solution to a problem. Here the characteristics of credit card transactions undergo evolution to allow a modeled credit card fraud detection system to be tested. This method proves accurate in deducting fraudulent transaction and minimizing the number of false alerts. If this algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions.

Feasibility Study:

With the growth of e-commerce websites, people and financial companies rely on online services to carry out their transactions that have led to an exponential increase in the credit card frauds. Fraudulent credit card transactions lead to a loss of huge amount of money. The design of an effective fraud detection system is necessary in order to reduce the losses incurred by the customers and financial companies. Research has been done on many models and methods to prevent and detect credit card frauds. Some credit card fraud transaction datasets contain the problem of imbalance in datasets. A good fraud detection system should be able to identify the fraud transaction accurately and should make the detection possible in real-time transactions.

Design Approach/ Methodology/ Planning of work

High level Architecture:



Design Approach:

- We have done Exploratory Data Analysis on full data then we have removed outliers using "Local Outlier Factor", then finally we have used KNN and Ann technique to predict to train the data and to predict whether the transaction is Fraud or not. We have also applied T-SNE to visualize the Fraud and genuine transactions in 2-D.
- 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. Class 0 represents a valid transaction and 1 represents a fraudulent one.
- It contains only numerical (continuous) input variables which are as a result of a Principal Component Analysis (PCA) feature selection transformation resulting to 28 principal components.

Proposed Methodology:

- Behavioral characteristic of the card is shown by a variable of each profile usage representing the spending habits of the customers along with days of the month, hours of the day, geographical locations, or type of the merchant where the transaction takes place.
- Afterwards these variables are used to create a model which distinguish fraudulent activities. The details and background information of the features cannot be presented due to confidentiality issues.
- The time feature stores the seconds that has elapsed between each transaction along with first transaction in the dataset. The 'amount' feature is the transaction amount. Feature 'class' is the target class for the binary classification.
- Four basic metrics are used in evaluating the experiments, namely True positive (TPR), True Negative (TNR), False Positive (FPR) and False Negative (FNR) rates metric respectively.

Models used

1. ADA Sampling
2. Feed forward neural network

Tools Used:

1. **Numpy**

   NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices.

2. **Pandas**

   Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

3. **Matplotlib**

   Matplotlib is a low-level graph plotting library in python that serves as a visualization utility. A comprehensive library for creating static, animated, and interactive visualizations in Python

4. **Sklearn**

Scikitlearn (Sklearn) provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

5. **Seaborn**

Seaborn is a library that uses Matplotlib to plot graphs. It will be used to visualize random distributions.

Model Description and Results:

We proposed Artificial neural networks with different number hidden layer using keras making it deeper for training model to train it.

In the first MLP, we have the neural network with single hidden layer without drop and with drop.

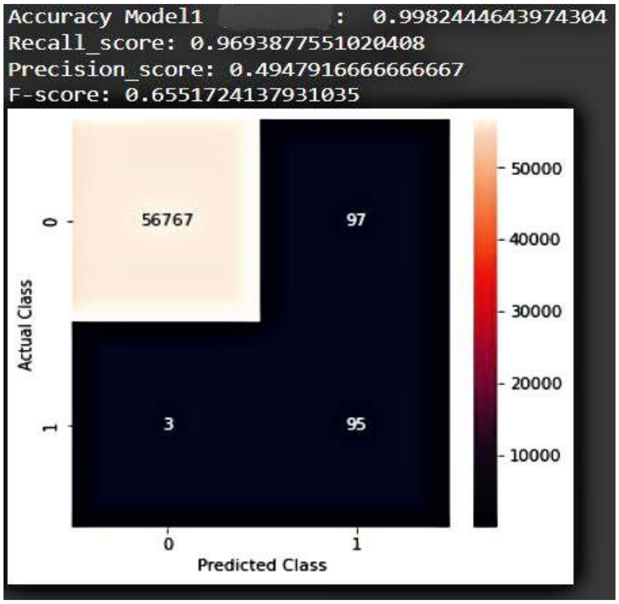
We have a neural network with two hidden layers with dropout in the second MLP

MLP with One Hidden Layer

- In the first MLP, there are three layers in the network: an input layer, a hidden layer, and an output layer.
- The input layer has 29 neurons, the hidden layer has 65 neurons, and the output layer has one neuron.
- A defined custom activation function was utilised, one without dropout and the sigmoid function was used in the output layer
- It was subsequently determined that our model was overfitting the data set and that dropout was necessary.

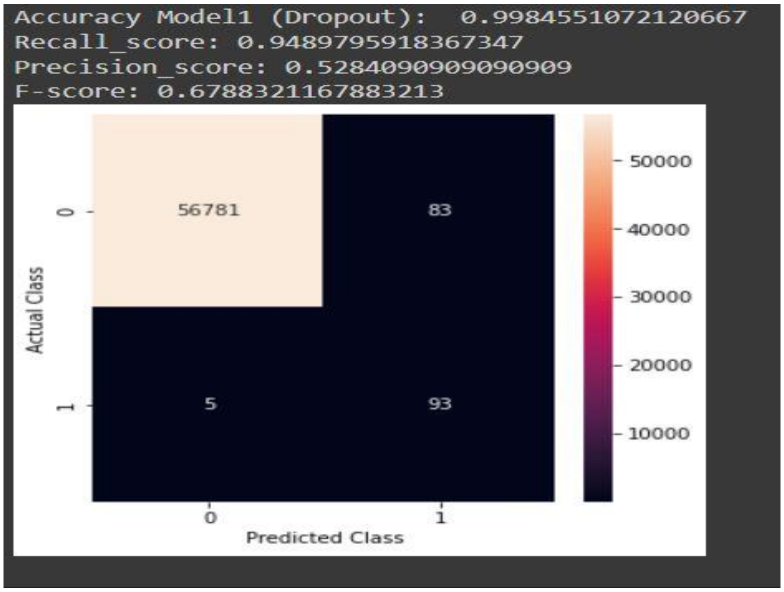
Result 1:

MLP without Dropout:



```
Accuracy Model1 [redacted]:  0.9982444643974304
Recall_score: 0.9693877551020408
Precision_score: 0.4947916666666667
F-score: 0.6551724137931035
```

MLP with Dropout:



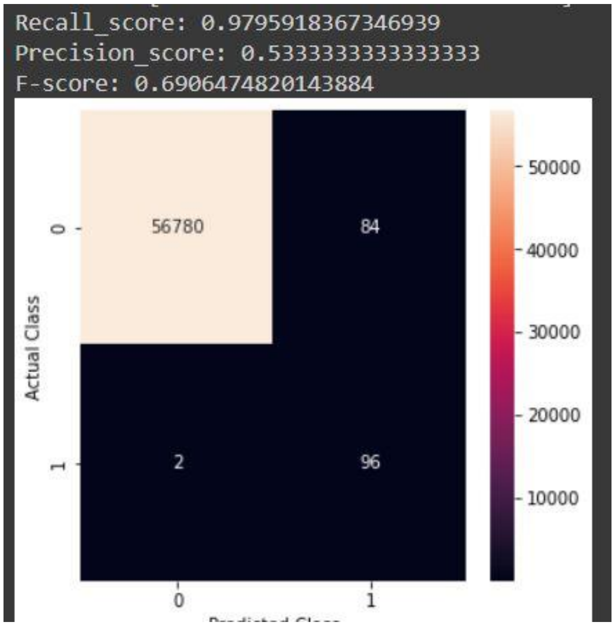
```
Accuracy Model1 (Dropout):  0.9984551072120667
Recall_score: 0.9489795918367347
Precision_score: 0.5284090909090909
F-score: 0.6788321167883213
```

MLP with Two Hidden layer with Dropout:

- In the second model, there are four layers in the network: an input layer, two hidden layer, and an output layer.
- The input layer has 29 neurons, the two hidden layer has 65 neurons each, and the output layer has one neuron.

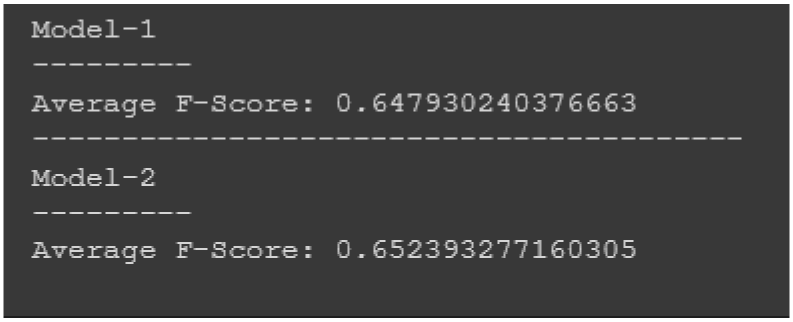
- For one of the hidden layers, a special activation function(custom) was used, while ReLU was used for the other.
- The sigmoid function was used in the output layer.
- However we had already determined that our model would overfit the data set without the use of dropout, So we continued using dropout.

Result 2:



```
Recall_score: 0.9795918367346939
Precision_score: 0.5333333333333333
F-score: 0.6906474820143884
```

Comparison Of Two models:



```
Model-1
---------
Average F-Score: 0.647930240376663
----------------------------------------
Model-2
---------
Average F-Score: 0.652393277160305
```

References

- Credit Card Fraud Detection Using Machine Learning | IEEE Conference Publication | IEEE Xplore
- Credit Card Fraud Detection Web Application using Streamlit and Machine Learning | IEEE Conference Publication | IEEE Xplore
- Analysis of Credit Card Fraud Detection using Machine Learning Techniques | IEEE Conference Publication | IEEE Xplore