CREDIT CARD FRAUD DETECTION

AP18110010468, AP18110010491
AP18110010492,
AP18110010508, AP18110010518

Abstract

One of the main and popular payment methods is the use of credit cards. This advent of free transactions has made it easier and easier for users to pay. The growing growth in its use directly affected counterfeit transactions which increased the frequency of illegal activities. Statistics suggest that the losses incurred by these illegal activities amount to millions of dollars each year. The scam is so well done that a normal eye looks like a real transaction. As a result, many banking and economic sectors have begun to rely on technology to combat these illegal agreements. Many machine learning applications are designed to combat deception, but with the improvement of these systems, the techniques used by these technicians are also improved. In this project, we are trying to use electronic learning algorithms such as Logistic Regression, SVM, Decision Tree, KNN, XGboost and Random Forest which in these applications are best suited for the purpose of solving the problem of credit card fraud.

Introduction

Financial fraud is a growing concern with far-reaching consequences for governments, corporations, the financial industry. As credit card transactions become a widespread payment method, the focus is on the latest computerized accounting systems to deal with the issue of credit card fraud. There are many fraudulent solutions and software to prevent fraud in businesses such as credit card, retail, e-commerce, insurance and industry. The data mining process is one of the most notable and popular methods used to solve the problem of obtaining credit fraud. It is impossible to be absolutely sure of the real purpose and suitability behind a plan or transaction. In fact, looking for existing evidence of fraud from data obtained using mathematical techniques is the best way to work. Credit card fraud is a real process of identifying fraudulent transactions into two phases of legal and fraudulent transactions, a number of strategies designed and used to resolve credit card fraud such as genetic algorithm, neural network architecture, neural algorithm -A migratory bird algorithm, comparative analysis of asset retrieval, SVM, decision tree and random forest is performed. Credit card fraud is a very common problem but it is also a difficult problem to solve.

LITERARTURE SURVEY:

Sahil Dhankhad and three other writers [3] have provided a solution to the perplexing issue of financial services that cost billions of dollars each year. Using surveillance machine learning systems and real-world databases, they have used a number of credit card fraud techniques. In this article, they have used these algorithms to deliver good divisions using aggregation methods. This article provides the differences between the 10 classification algorithms and their accuracy comparisons. The authors used model testing using Accuracy, F1-Score, Recall, Precision, G-Mean, FPR, TRP techniques. The authors identified the most important factors in the fraudulent discovery of a credit card that could lead to greater testing accuracy models.

Samuel A. Oluwadare and three other authors [4] introduced a model for using and improving a credit card fraud model as transaction size and information size grow. In this paper, the database is obtained from European card holders containing 284,807 transactions. Credit card fraud and providing accuracy, sensitivity, clarity, Matthews equity and rating, three Naïve Bayes, a neighborhood of 4 k and structured models are used. In terms of results, they say KNN works better than other methods in terms of results.

N.Malini et al [1] author uses the K-Nearest Neighbor Algorithm (KNN) and the unsupervised outsourcing process to identify fraudulent credit card transactions on a banking basis. The author uses the distance metric to calculate the nearest point of any incoming credit card transactions in other transactions. Fraudulent transactions could be inconsistent with the KNN approach. The author has opted for an unsupervised method of detection of inaccuracies in order to capture the invisible type of illegal accuracy with limited memory.

Linda Delamaire and two other authors [5] have identified various types of credit card fraud and revised methods used for fraudulent detection. In this paper, various findings have been published published on detecting credit card fraud and analyzed 3. Depending on the type of fraud involved, different proposed measures may be changed. The proposed methods are expensive and effective over time. The main purpose of these measures is to reduce credit card fraud, but authors still face the problem when what is actually done is also classified as fraud.

Dataset and data pre-processing:

The data used for credit card acquisition analysis in this project contains data from European credit card holders containing lines of credit card transactions. The data analysis result revealed 284,807 records and 31 most prominent features were selected. Features include 28 intentionally hidden hidden features by data source, as well as 'time' and 'value' of activity.

DATA CLEANSING:

➢ Data can have many inactive and non-functional components. To handle this part, data cleaning is done. It involves managing lost data.
After analyzing the data we can see that of the 284,807 samples, there are only 492 cases of fraud which is only 0.17 percent of the total samples. Therefore, we can say that the data we are dealing with is very unequal data and needs to be carefully managed when modeling and testing. After examining the statistics, we have noticed that the ‘Amount’ variable values are very different compared to other variables. To reduce its variety, we can get used to using the 'StandardScaler' method in python.

Data mining Models:

Below supervised as well as unsupervised models have implemented for the classification of fraud and nonfraud transactions.

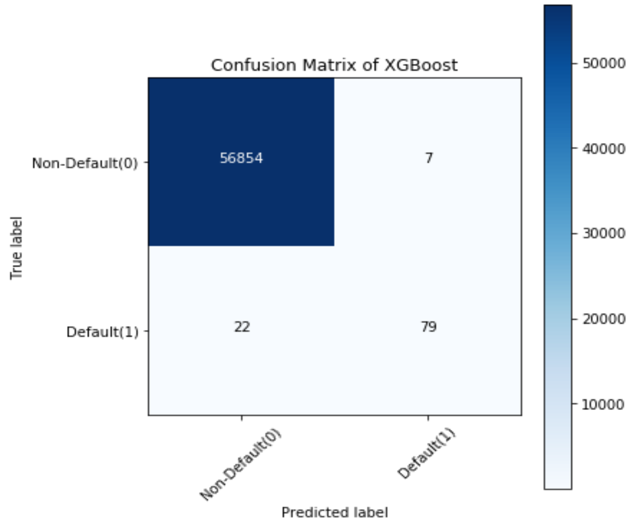
Supervised

1. Logistic Regression
2. Support Vector Machines
3. Decision Tree
4. K Nearest neighbor
5.Random Forest
6. XGboost

Performance of models:

Xgboost:

Accuracy score obtained after training and testing the model **0.9994908886626171**

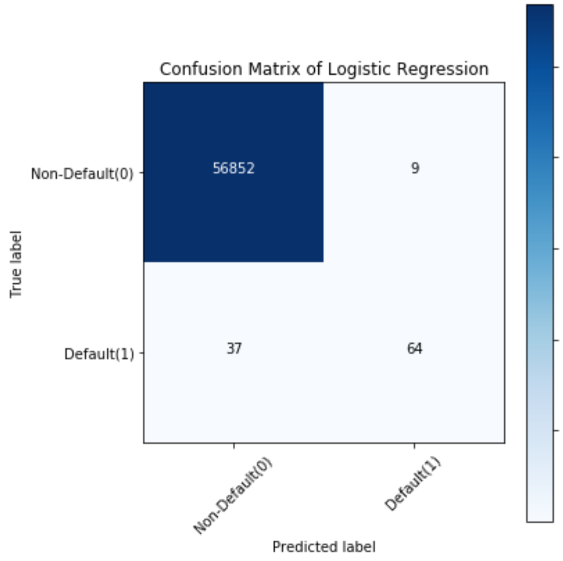


Look at the first line. The first line for the real fake value in the 0 test set. As you can count, the fraud rate of 56861 is 0. And of these 56861 non-fraudulent transactions, the editor correctly represented their 56854 as 0 and their 7 as 1 It means that, in 56854 non-fraudulent activities, the actual value of churn was 0 on the test set, and the editor correctly guessed. those as 0s. We can say that our model has distinguished the non-fraudulent transaction.

Let's look at the second line. Apparently there were 101 transactions whose fraud rate was 1. A well-divided editor predicted 79 of them as 1, and 22 of them as negative as 0. Incorrectly predicted values can be considered a model error.

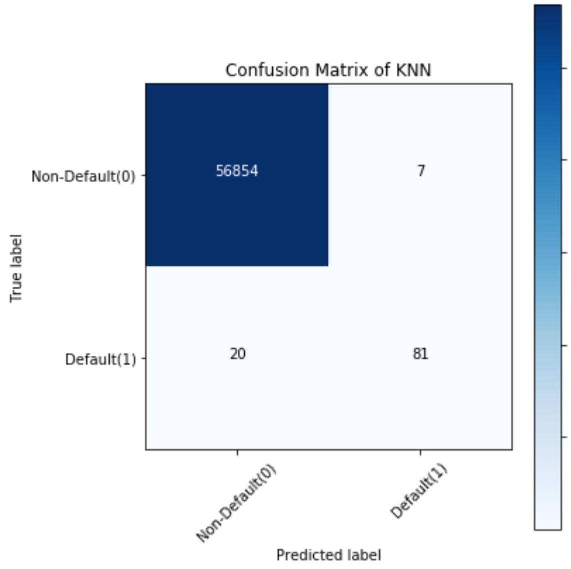
Logistic Regression

Accuracy score obtained after training and testing the model **0.9991924440855307**



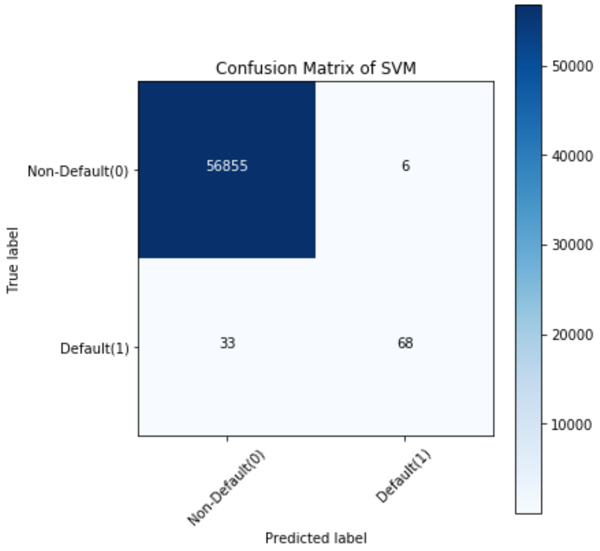
K-Neighbors Classifier

Accuracy score obtained after training and testing the model **0.9995259997893332**



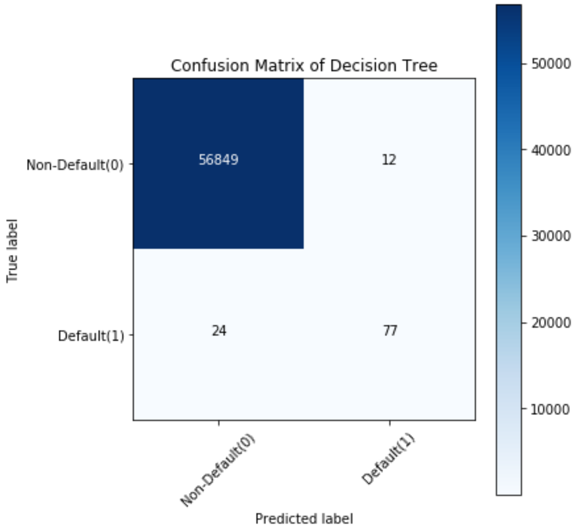
SVM

Accuracy score obtained after training and testing the model **0.9993153330290369**



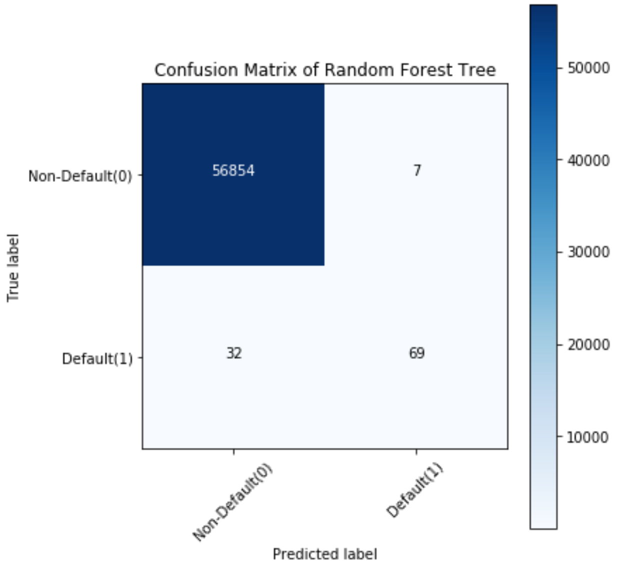
Decision Tree Classifier

Accuracy score obtained after training and testing the model **0.9993679997191109**



Random Forest

Accuracy score obtained after training and testing the model **0.9993153330290369**



CONCLUSION:

In terms of metrics for accuracy testing, the KNN model proves to be the most accurate model and the Logistic regression model to be the most accurate model. However, when we compile the results for each model, it shows 0.99 (99% accurate) which is a very good score. When comparing the matrix of confusion of all kinds, it can be seen that the K-Nearest Neighbors model did an excellent job of distinguishing fraudulent transactions from non-fraudulent items followed by the XGBoost model. We can therefore conclude that the most suitable model that can be used in our case is the K-Nearest Neighbors model and the most neglected model for the Logistic regression model.

References

[1] N. Malini, M. Pushpa “Analysis on credit card fraud identification techniques based on KNN and outlier detection” in 2017 Third International Conference on Advances in Electrical, Electronics, Information, 10 Communication and Bio-Informatics (AEEICB),27-28 Feb. 2017 IEEE

[3] Sahil Dhankhad ; Emad Mohammed ; Behrouz Far “Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study” 2018 IEEE International Conference on Information Reuse and Integration (IRI)

[4] John O. Awoyemi ; Adebayo O. Adetunmbi ; Samuel A. Oluwadare “Credit card fraud detection using machine learning techniques: A comparative analysis” 2017 International Conference on Computing Networking and Informatics (ICCNI)

[5] Delamaire, Linda & Abdou, Hussein & Pointon, John. (2009). “Credit card fraud and detection techniques: A review.” Banks and Bank Systems. Volume 4. Issue 2.2009

Group members:

CSE-H

AP18110010468- B. Rajya Lakshmi
AP18110010491- T. Vani
AP18110010492- B. Aditya
AP18110010508- Saurabh Poonia
AP18110010518- V. Mounika