ZKG INTERNATIONAL

ISSN: 2366-1313

Fraud Detection in Credit Card Using Machine Learning

[1]KIRAN SINGH BONDILI, [2]CHANDU DELHI POLICE

[1] M. Tech Scholar,[2]Assistant Professor, Department of CSE,

Priyadarshini Institute of Technology & Science, Chintalapudi - 522306

Abstract:

The use of credit cards to make purchases on the internet and for regular use is increasing exponentially as is the risk that is associated to it. An increasing number of fraudulent transactions occur every single day. There are a variety of modern methods, including artificial neural network (ANNN) and various algorithmic methods for machine learning are compared which include Logistic Regression and decision Trees, Random Forest, Artificial Neural Networks, Logistic Regression, K-Nearest Neighbors and K-means clustering etc. These are utilized to identify fraudulent transactions. This paper makes use of genetic algorithms and a neural network that includes methods for determining the an optimal solution to the problem and, implicitly, generating the outcome that the transaction is fraudulent. The goal is to find out the fraudulent transaction as well as come up with a method to generate tests data. This is a heuristic technique that is used to solve complex computational issues. Implementing a reliable method of detecting fraudulent transactions is crucial in all credit card issuer businesses and their customers to limit the losses they suffer.

Keywords: Machine learning, Credit card, Electronic commerce, Fraud detection.

.I. INTRODUCTION

Credit cards are small, handy plastic card which contains specific information about the card's owner, like an image or signature and permits the person who is named on it to debit charges for services or purchases to their account - charges for which he'll be charged on a regular basis. Nowadays, the data of the card are processed by automated ATMs (ATMs) as well as stores, banks and also in the online banking systems. They are issued a unique card number, which is of vital importance. Security is based upon the security physical of the card, as well as the security associated with the card's number. There has been a significant increase in the volume of credit card transactions that has led to a dramatic increase in the number of fraudulent transactions. Fraud on credit cards is broad term that refers to theft and fraud committed by using a credit or debit card as a source of funds for the course of a transaction. In general, the majority of fraud detection systems for credit cards are built around Artificial Intelligence, Meta Learning and pattern matching..

OBJECTIVE

The primary goal is to identify online fraud when making an online financial transactions. Presently, the threat of security breaches in networks is growing in both quantity and danger. The most popular method used by hackers is to target end-to-end technology and exploit human weaknesses..

II. LITERATURE SURVEY

1]Vimala Devi. J et al. To find fraudulent transactions, 3 machine learning algorithms were developed and implemented. There are numerous measures to assess the effectiveness of predictors or classifiers, like those of the Vector Machine, Random Forest or Decision Tree. These metrics are either prevalence- dependent or prevalence-independent. In addition, these algorithms are utilized in fraud detection techniques, and the results of these algorithms are examined.

2]Popat and Chaudhary. They presented supervised algorithms: The algorithms included Logistic Regression, Deep Learning Nave Bayesian, Support Vector Machine (SVM), Neural Network, Artificial Immune System, K Nearest Neighbour, Data Mining, Decision Tree, Fuzzy Logic based System as well as Genetic Algorithm are some of the methods used. We evaluated machine-learning algorithms against prediction as well as clustering and outlier detection.

3]Deepa and Akila . To detect fraud, various methods such as Anomaly Detection Algorithm, K-Nearest Neighbor, Random Forest, K-Means and Decision Tree were used. Based on a specific scenario, we presented various methods and suggested the most effective method for detecting fraudulent transactions. To identify the fraud outcome the system employed a variety of techniques and procedures to calculate the score of fraud for that particular transaction.

4]Kibria and Sevkli. Utilizing the grid search method and a deep learning model. The performance of the model is compared to that of two other machine-learning algorithms such as logistic regression (LR) and support vector machine (SVM). The model developed is used to analyze the data of credit cards set , and the results are then compared to the logistic regression model and models for support vector machines.

5]Borse Suhas as well as Dhotre Machine Learning's Naive Bayes classification employed to identify fraudulent or common transactions.

6]Asha R B et al. have suggested a deep learning-based method to detect fraud in the transactions made with credit cards. Making use of machine-learning algorithms like support vector machines, knearest neighbor and artificial neural network to anticipate the likelihood of fraud.

4.PROBLEM DEFINITION

There are a myriad of problems which make this process challenging to carry out and one of the most significant issues with the detection of fraud is the inability to find the research literature that provides experiments and real-world data that academic researchers can use to test their theories on. The reason for this is because of the confidential financial information associated with fraud that must be kept private in order to ensure customer privacy. In this article, we'll look at the various properties that the fraud detection system must be able to provide the best results. The system must be able to deal with irregular distributions, as only a tiny percentage of credit card transactions are fraudulent. It is essential to have a way to handle the background noise. Noise refers to the error which are in the data, such as for instance, inaccurate dates. This noise in real data reduces the precision of generalization that is accomplished, no matter how large the training set. Another issue in this field is the fact that data is often overlapping. A lot of transactions can appear to be fraudulent, but they're genuine.

ALGORITHM

Random Forest

Classifier has fined decision trees for an area of data, and then combines their data to obtain the complete dataset's predictive capacity. Instead of relying on one decision tree.

The RF analyzes the forecasts from each tree and predicts the final outcome by analyzing the votes of the majority of forecasts. Utilizing a large number of trees within the forest increases precision and reduces the risk of over fitting. It forecasts output with great accuracy and is able to run efficiently even when large data sets. It is also able to maintain the accuracy even when a significant portion of data goes missing.

Random Forest can handle both the tasks of classification and regression. It is capable of handling large datasets with high dimension. It increases the accuracy of the model and eliminates the over-fitting problem. Two-step techniques for training are employed to train trees-based Random Forest: First, we build randomly the forest, mixing the N trees, and then we estimate each tree we build in the first stage.

Naive Bayes

The non-naive Bayes groups of statistical algorithms are among the most commonly used algorithms in the field of text classification and analysis of tests. Naive Bayes Algorithm is extremely efficient in comparison to other algorithms for classification.

Naive Bayes classifier is a set of classification algorithms based on Bayes Theorem.

It's not just a single algorithm, but rather a group of algorithms in which each algorithm shares the same fundamental.

5.EXPERIMENTAL SETUP

The following section discuss our research study we conducted using a selection of algorithms for machine learning and techniques for imbalance classification. We begin by providing details of the research design then we present the results and an analysis. We also discuss the critical flaws that we observed during our tests.

A]Workflow of Experiments

Our research study is structured in the following manner. The research is described as well as discussed over two stages. In the initial phase the eight classification techniques are evaluated. The evaluation was conducted in relation to three parameters, including factors: sensitivity, accuracy and the Area Under Precision-Recall Curve (AUPRC). This analysis results in the selection of the most appropriate algorithms, including those of SVM as well as the ANN.

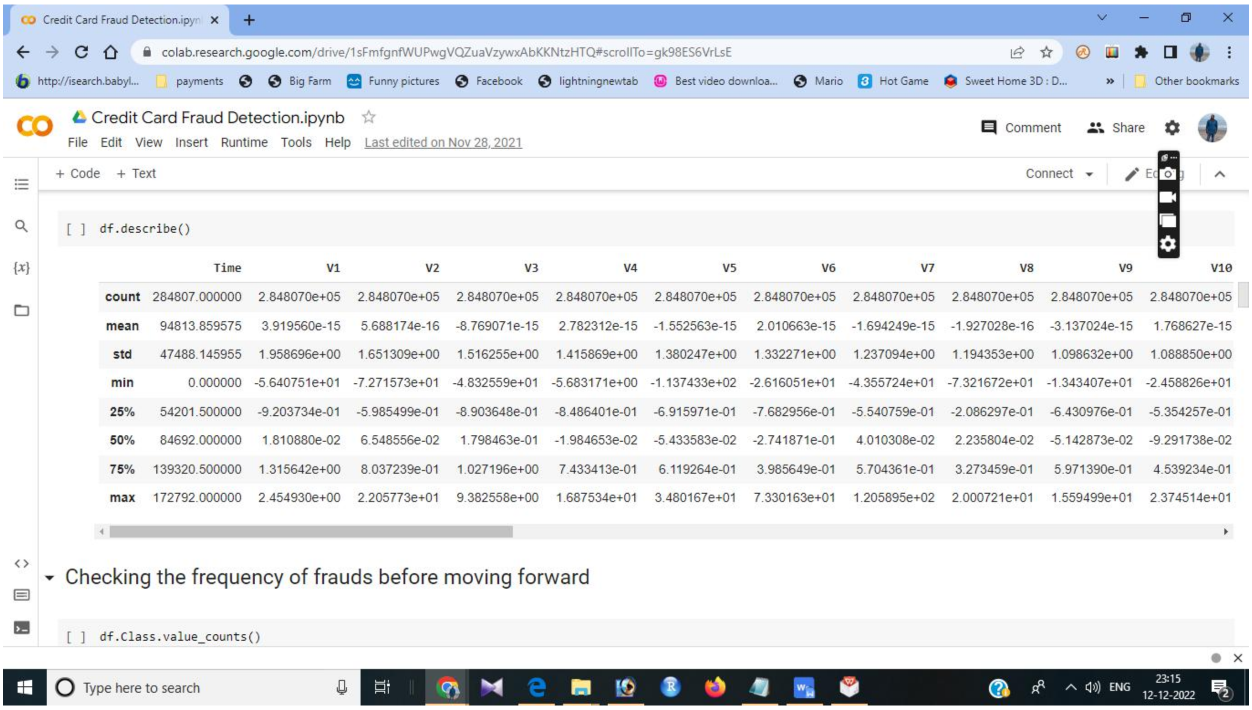
In the next phase in the second phase, the algorithms chosen are used to compare the various imbalance classification methods like Random Oversampling One-Class Classification, and Cost Sensitive. Then the SVM is utilized as a binary tool for classification and is compared with one-class classification SVM along with the Cost Sensitive SVM. Additionally the ANN is is compared with it's counterpart, the Auto-Associative Neural Network.

B]Dataset and Variable Selection

The data used in our study contains the data that has been labeled as fraudulent by credit card. It includes ten million transactions involving credit cards described by eight variables that are listed:

Cust ID can be described as an automatic incrementing integer value that represents the customer's identification number: The variable is later removed since it is not relevant in detecting fraud. Gender refers to the gender of the client. State: is the state where the customer is located and is located in the United States. The card holder is the number of credit cards the user has (maximum 2.). Balance : shows the amount of the credit card's balance in USD. Num Trans is a discrete variable which indicates the amount of transactions that have been made up to date. NumInt Trans is a discrete number that represents the amount of international transactions that have been completed up to date. Credit Line: indicates the limit on credit for the customer. Fraud Risk: The binary target variable, which takes the values of 0 for legitimate transactions and 1 for a fraudulent transactions.

6. RESULTS & EVALUATION


Credit Card Fraud Detection.ipynb
colab.research.google.com/drive/1sFmfgnfWUPwgVQZuaVzywxAbKKNtzHTQ#scrollTo=gk98ES6VrLsE
File Edit View Insert Runtime Tools Help Last edited on Nov 28, 2021
Comment
Share
+ Code + Text
Connect
Type here to search
ENG
23:15
12-12-2022

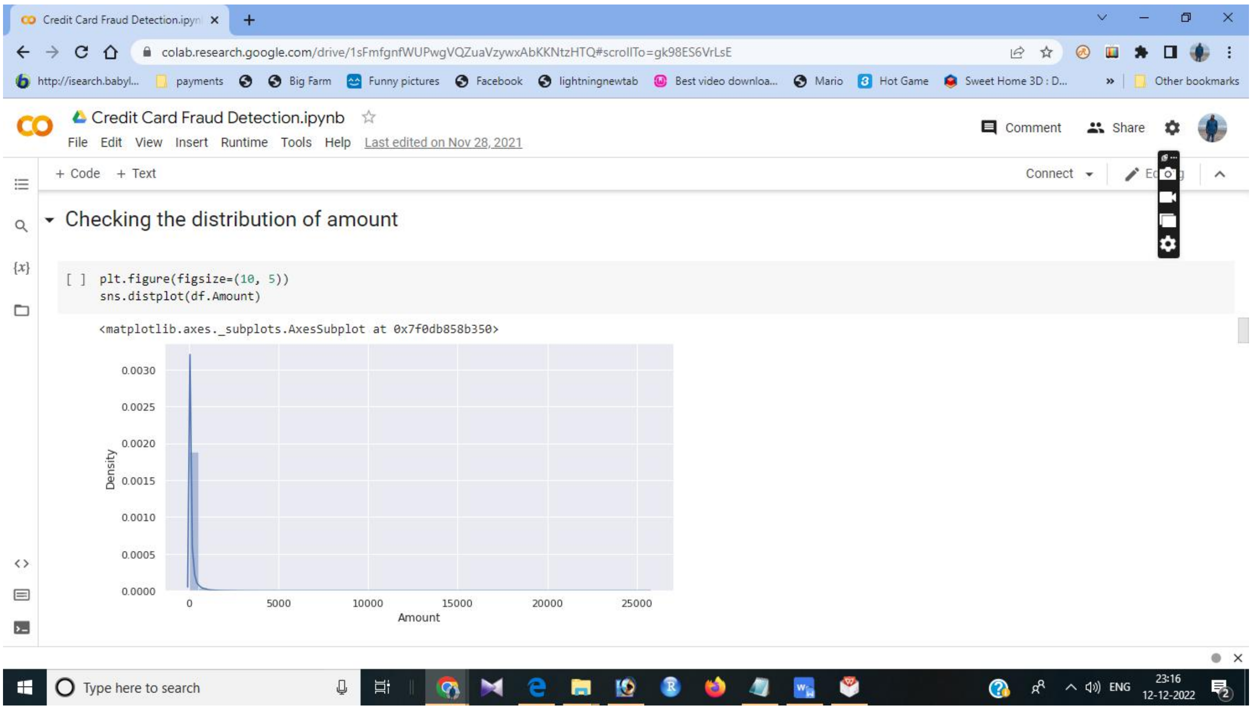
```
[ ] df.describe()
```

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
count	284807.000000	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05	2.848070e+05
mean	94813.859575	3.919560e-15	5.688174e-16	-8.769071e-15	2.782312e-15	-1.552563e-15	2.010663e-15	-1.694249e-15	-1.927028e-16	-3.137024e-15	1.768627e-15
std	47488.145955	1.958696e+00	1.651309e+00	1.516255e+00	1.415869e+00	1.380247e+00	1.332271e+00	1.237094e+00	1.194353e+00	1.098632e+00	1.088850e+00
min	0.000000	-5.640751e+01	-7.271573e+01	-4.832559e+01	-5.683171e+00	-1.137433e+02	-2.616051e+01	-4.355724e+01	-7.321672e+01	-1.343407e+01	-2.458826e+01
25%	54201.500000	-9.203734e-01	-5.985499e-01	-8.903648e-01	-8.486401e-01	-6.915971e-01	-7.682956e-01	-5.540759e-01	-2.086297e-01	-6.430976e-01	-5.354257e-01
50%	84692.000000	1.810880e-02	6.548556e-02	1.798463e-01	-1.984653e-02	-5.433583e-02	-2.741871e-01	4.010308e-02	2.235804e-02	-5.142873e-02	-9.291738e-02
75%	139320.500000	1.315642e+00	8.037239e-01	1.027196e+00	7.433413e-01	6.119264e-01	3.985649e-01	5.704361e-01	3.273459e-01	5.971390e-01	4.539234e-01
max	172792.000000	2.454930e+00	2.205773e+01	9.382558e+00	1.687534e+01	3.480167e+01	7.330163e+01	1.205895e+02	2.000721e+01	1.559499e+01	2.374514e+01

Checking the frequency of frauds before moving forward

```
[ ] df.Class.value_counts()
```

Fig 1: Fraud and non fraud cases


Credit Card Fraud Detection.ipynb
colab.research.google.com/drive/1sFmfgnfWUPwgVQZuaVzywxAbKKNtzHTQ#scrollTo=gk98ES6VrLsE
File Edit View Insert Runtime Tools Help Last edited on Nov 28, 2021
Comment
Share
+ Code + Text
Connect
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
Density
0
5000
10000
15000
20000
25000
Amount
Type here to search
ENG
23:16
12-12-2022

Checking the distribution of amount

```
[ ] plt.figure(figsize=(10, 5))
    sns.distplot(df.Amount)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0db858b350>
```

Fig 2: Word represntation of the data

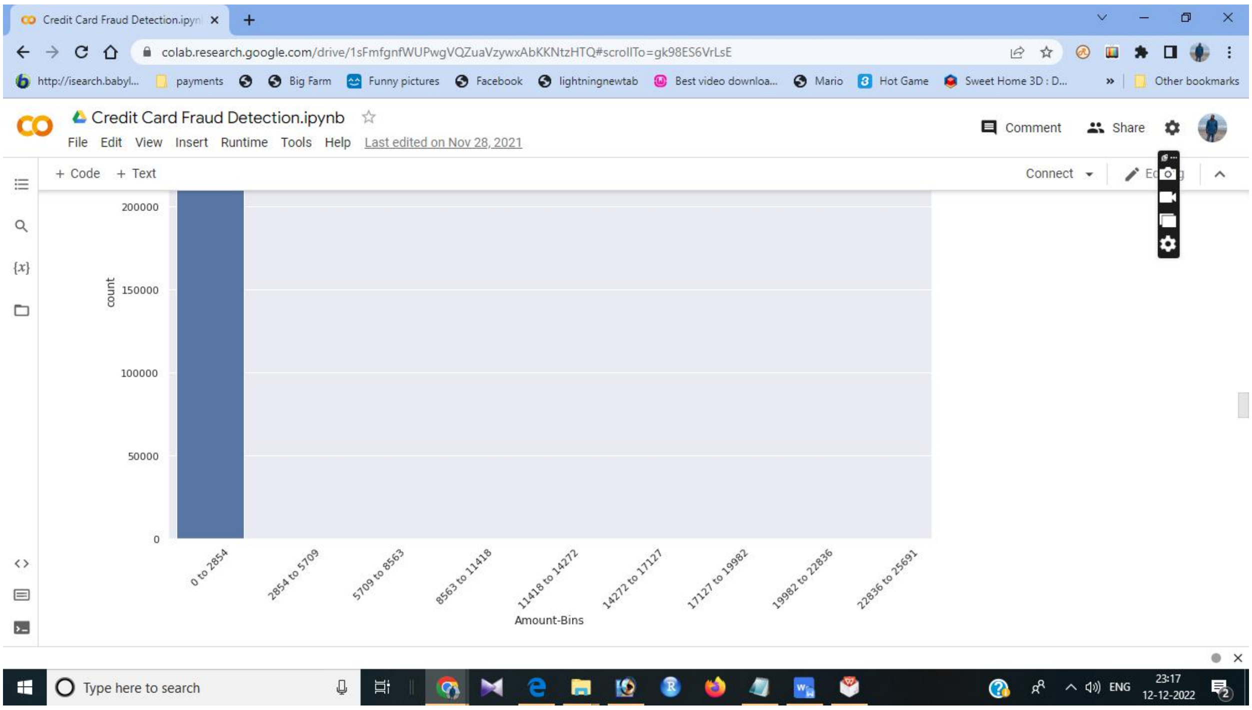
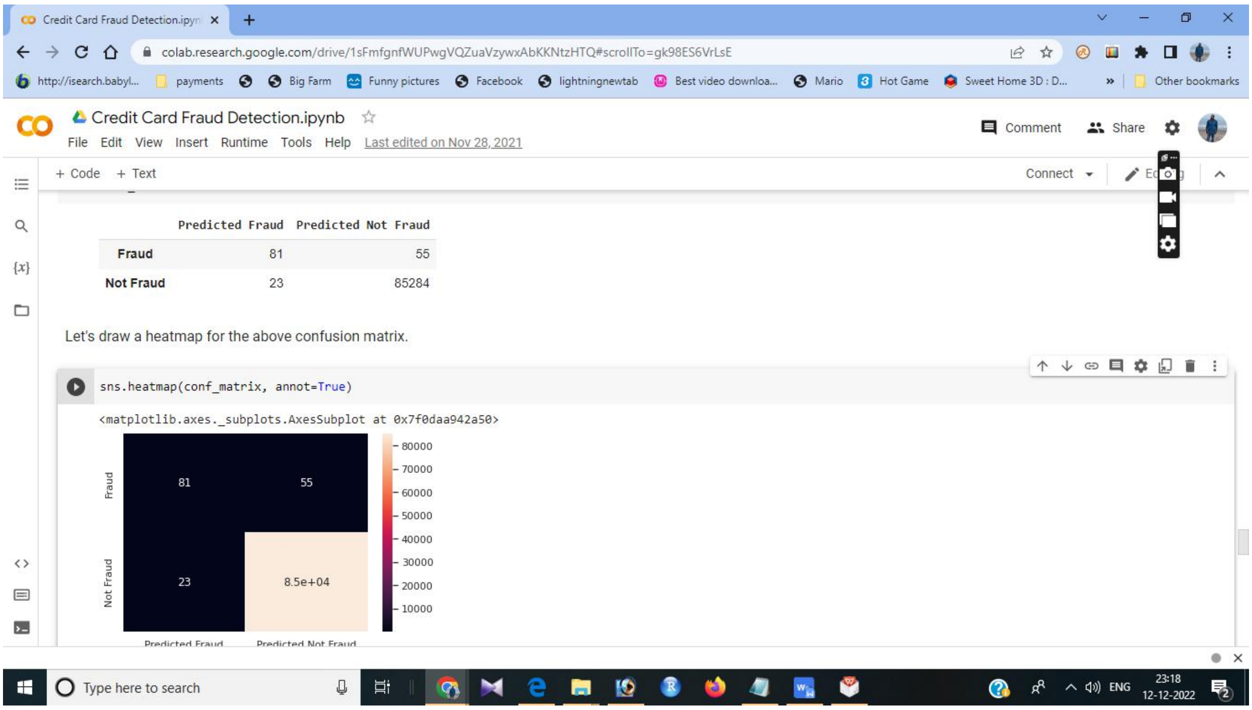


Fig 3: One hot encoding on the data set


Credit Card Fraud Detection.ipynb
colab.research.google.com/drive/1sFmfgnfWUPwgVQZuaVzywxAbKKNtzHTQ#scrollTo=gk98ES6VrLsE
Credit Card Fraud Detection.ipynb
File Edit View Insert Runtime Tools Help Last edited on Nov 28, 2021
Comment Share
+ Code + Text
Connect

	Predicted Fraud	Predicted Not Fraud
Fraud	81	55
Not Fraud	23	85284

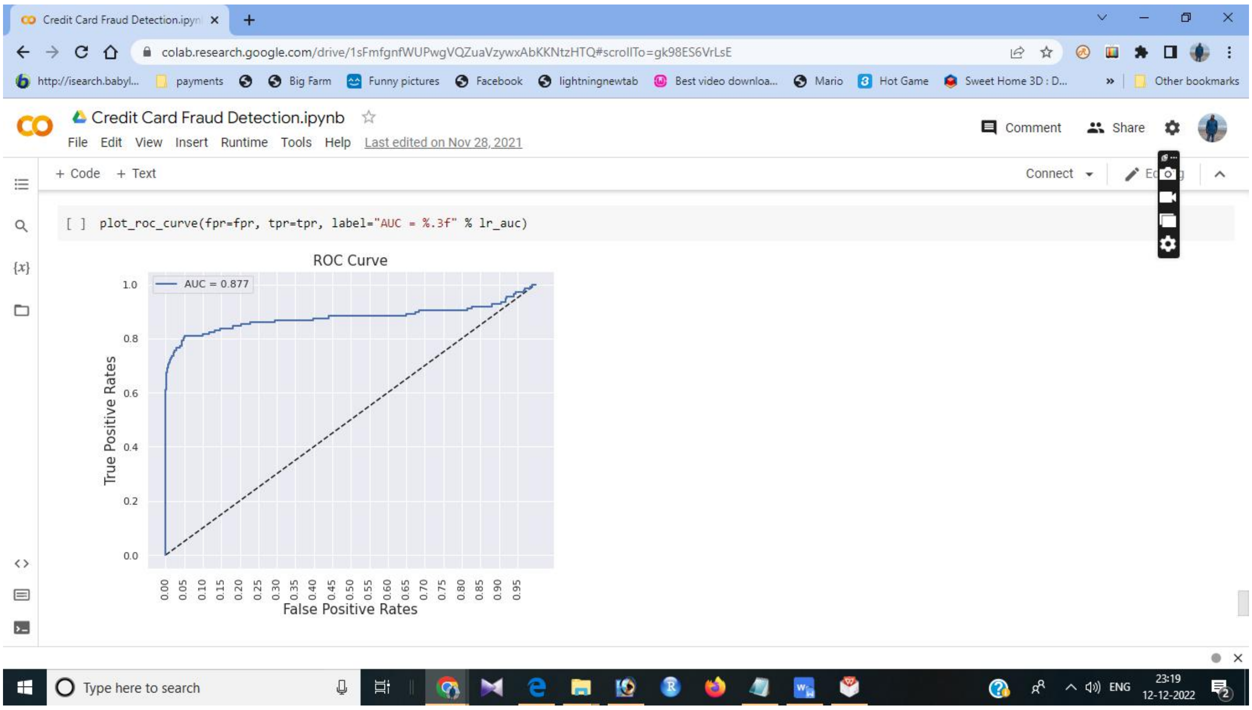
Let's draw a heatmap for the above confusion matrix.

```
sns.heatmap(conf_matrix, annot=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0daa942a50>
```

Fraud
Not Fraud
81
55
23
8.5e+04
Predicted Fraud
Predicted Not Fraud
80000
70000
60000
50000
40000
30000
20000
10000
Type here to search
23:18
12-12-2022

Fig 4: Resuls Fraud and Non Fraud evaluations



```
[ ] plot_roc_curve(fpr=fpr, tpr=tpr, label="AUC = %.3f" % lr_auc)
```

Fig 5: ROC curve plotting word represntation of the data

7. CONCLUSION AND FUTURE SCOPE

The theft of credit cards becomes an issue for the entire world. Fraud causes huge financial losses around the globe. The credit card industry have invested funds to develop and create methods to identify and limit fraud. The primary objective of this research is to establish algorithms that provide the correct ability to be used to credit card firms to assist in better identifying fraudulent transactions while consuming less time and at a lower cost. A variety of machine learning techniques are examined to determine which one is the most effective, including Logistic Regression, Decision Trees, Random Forest, Artificial Neural Networks, Logistic Regression K-Nearest Neighbors and Kmeans clustering. Since not all scenarios are identical, a situation-based algorithm can be utilized to identify the most suitable scenario suitable for the scenario.

8. REFERENCES

[1] A. Srivastava, A. Kundu, S. Sural and A. Majumdar, “Credit card fraud detection using hidden markov model,” In:IEEE transactions on dependable and secure computing, vol. 5, no. 1, Jan.-March 2008, pp. 37-48.

[2] Statista the statistic portal, https://www.statista.com/topics/871/online-shopping/, March 14, 2017.

[3] S. Yusuf, E. Duman, "Detecting credit card fraud by decision trees and support vector machines," IMECS 2011- International multiconference of Engineers and Computer Scientists 2011, 1, 442-447, 2011.

[4] Y. Wang, S. Adams, P. Beling, S. Greenspan, S. Rajagopalan, M. Velez-Rojas, S. Mankovski, S. Boker, D. Brown, "Privacy preserving distributed deep learning and its application in credit card Fraud detection," 1070-1078, 2018.

[5] S. Bhattacharyya, S. Jha, K. Tharakunnel, J. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, 50, 602-613, 2011.

[6] Y. Sahin, S. Bulkan, E. Duman, "A cost-sensitive decision tree approach for fraud detection" Expert Syst. Appl., 40, 5916-5923, 2013.

[7] S. Panigrahi, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detection: a fusion approach using dempster–shafer theory and bayesian learning," Information Fusion, 10, 354-363, 2009.

[8] S. Maes, K. Tuyls, B. Vanschoenwinkel, B. Manderick, "Credit card fraud detection using bayesian and neural networks, 2002.

[9] M. Zareapoor, P. Shamsolmoali, "Application of credit card fraud detection: based on bagging ensemble classifier," Procedia Computer Science, 48, 679-686, 2015.

[10] L. Zheng, et al, "A new credit card fraud detecting method based on behavior certificate," IEEE 15th international conference on Networking, Sensing and Control (ICNSC), Zhuhai, pp. 1-6, 2018.

[11] E. Duman, M. Özçelik, "Detecting credit card fraud by genetic algorithm and scatter search," Expert Syst. Appl., 38, 13057-13063, 2011.

[12] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, P. Beling, "Deep learning detecting fraud in credit card transactions," Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, 2018, pp. 129-134, 2018

[13] F. Kang, C. Dawei, T. Yi, Z. Liqing, "Credit card fraud detection using convolutional neural networks," 483-490, 2016.

[14] P. Suraj, N. Varsha and S. P. Kumar, "Predictive modelling for credit card fraud detection using data analytics," Procedia Computer Science, 132, 385-395, 2018.

[15] J. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.E. Portier, L. He, O. Caelen, "Sequence classification for credit-card fraud detection," 10.1016/j.procs.2015.04.201, 2018.

[16] A. De Sá, A. Pereira, G. Pappa, "A customized classification algorithm for credit card fraud detection," 2018.

[17] J. S. Bayer, München, Technische Universität München, Diss., "Learning Sequence Representations," 2015.

[18] T. K. Behera, S. Panigrahi, "Credit card fraud detection: a hybrid approach using fuzzy clustering, neural network," Second International Conference on Advances in Computing and Communication Engineering, Dehradun, pp. 494-499, 2015.

[19] T. R. C. Sudha, "Credit Card Fraud Detection in Internet using K Nearest Neighbour Algorithm," IPASJ international journal of computer science, vol. 5, no. 11, 2017.

[20] M. Sanaz, S. Mehdi, "Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors.," 2018.