APPENDIX 1

REVEALING & PREVENTING FRAUDLENT TRANSACTION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

DESAVATH REVANTH NAIK	196301033
BUSAPPAGARI SHANMUKHA	196301028
DONAPATI SUNIL KUMAR REDDY	196301038
BURLE VINEETH	196301027

In partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE ENGINEERING

FACULTY OF ENGINEERING & TECHNOLOGY

GURUKUL KANGRI (DEEMED TO BE UNIVERSITY)

MAY & 2023

APPENDIX 2

GURUKUL KANGRI
(DEEMED TO BE UNIVERSITY)

BONAFIDE CERTIFICATE

Certificate that this project report "**REVEALING & PREVENTING FRAUDLENT TRANSACTION USING MACHINE LEARNING**" is the bonafide work of "**DESAVATH REVANTH NAIK, BUSAPPAGARI SHANMUKHA, DONAPATI SUNIL KUMAR REDDY, BURLE VINEETH**" who carried out the project work under my supervision.

SIGNATURE
Dr. Mayank Agarwal
HEAD OF THE DEPARTMENT
Department of CSE,
FET, GKV, Haridwar.

SIGNATURE
Dr. Suyash Bhardwaj
SUPERVISOR
Department of CSE,
FET, GKV, Haridwar.

APPENDIX 3

CONTENTS

ABSTRACT III

LIST OF FIGURES V

LIST OF TABLES V

CHAPTER 1: INTRODUCTION 1

1.1 INTRODUCTION 2

1.2 PROJECT GOALS 2

1.3 RESEARCH METHODOLOGY 3

CHAPTER 2: LITERATURE REVIEW 5

2.1 INTRODUCTION 6

2.2 LITERATURE REVIEW 6

2.3 LITERATURE REVIEW CONCLUSION 11

CHAPTER 3: PROJECT DESCRIPTION 12

3.1 INTRODUCTION 13

3.2 DATA SOURCE 13

CHAPTER 4: DATA ANALYSIS 14

4.1 DATA PREPARATION 15

4.1.1 CORRELATION BETWEEN ATTRIBUTES 16

4.2 DATA PREPROCESSING 18

4.3 DATA MODELING 18

4.3.1 K-NEAREST NEIGHBOR (KNN) 18

4.3.2 NAIVE BAYES 19

4.3.3 SUPPORT VECTOR MACHINE(SVM) 20

4.3.4 LOGISTIC REGRESSION 21

4.4 EVALUATION AND DEPLOYMENT 23

4.4.1 TRAINING & TESTING AND VALIDATION 25

CHAPTER 5: CONCLUSION 26

5.1 CONCLUSION & RECOMMENDATIONS 27

CHAPTER 6: BIBLIOGRAPHY 28

APPENDIX 4

DISCLAIMER

This project, Revealing & Preventing Fraudulent Transaction using Machine learning report/dissertation and completely functional project has been prepared by the students under the Major Project of the Faculty of Engineering & Technology, Gurukul Kangri Vishwavidyalaya, for academic purposes only. The views expressed in the report are personal to the students and do not necessarily reflect the view of the FET, GKV or any of its staff or personnel and do not bind the Faculty of Engineering & Technology, Gurukul kangri Vishwavidyalaya in any manner. This report and functional project is the intellectual property of the Faculty of Engineering & Technology, Gurukul Kangri Vishwavidyalaya and the same or any part of may not be used in any manner whatsoever, without express permission of the Faculty of Engineering & Technology, Gurukul Kangri Vishwavidyalaya, in writing.

Yours student

DESAVATH REVANTH NAIK

BUSAPPAGARI SHANMUKHA

DONAPATI SUNIL KUMAR REDDY

BURLE VINEETH

REVEALING & PREVENTING FRAUDLENT TRANSACTIONS USING MACHINE LEARNING

ABSTRACT

The purpose of this project is to detect the fraudulent transactions made by credit cards using machine learning techniques, to stop fraudsters from the unauthorized usage of customers' accounts. The increase of credit card fraud is growing rapidly worldwide, which is the reason actions should be taken to stop fraudsters. Putting a limitfor those actions would have a positive impact on the customers as their money would berecovered and retrieved back into their accounts and they will not be charged for items or services that were not purchased by them which is the main goal of the project.

Detectionof the fraudulent transactions will be made by using three machine learning techniques KNN, SVM and Logistic Regression, those models will be used on a credit card transaction dataset.

Keywords*: Credit Card Fraud Detection, Fraud Detection, Fraudulent Transactions, K- Nearest Neighbors, Support Vector Machine, Logistic Regression, Naive Bayes*

LIST OF FIGURES

Figure 1 - Dataset Structure 15
Figure 2 - Class Distribution 16
Figure 3 - Correlations 16
Figure 4 – Variable 18 17
Figure 5 - Variable 28 17
Figure 6 – Accuracy of KNN Model 19
Figure 7 - Accuracy of NAIVE Model 19
Figure 8 - Accuracy of SVM Model 20
Figure 9 - Accuracy of Logistic Regression Model 21
Figure 10 – Accuracy comparison graph of all Models 23

LIST OF TABLES

Table 1 - Confusion Matrix 24
Table 2 - Table of Accuracies 24

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

With the increase of people using credit cards in their daily lives, credit card companies should take exceptional care in the security and safety of the customers. According to (Creditcard statistics 2021) the number of people using credit cards around the world was 2.8 billion in 2019, in addition 70% of those users own a single card at least.

Reports of Credit card fraud in the US rose by 44.7% from 271,927 in 2019 to 393,207 reports in 2020. There are two kinds of credit card fraud, the first one is by having a creditcard account opened under your name by an identity thief, reports of this fraudulent behavior increased 48% from 2019 to 2020. The second type is by an identity thief usesan existing account that you created, and it is usually done by stealing the information of the credit card, reports on this type of fraud increased 9% from 2019 to 2020 (Daly, 2021).Those statistics caught my attention as the numbers are increasing drastically and rapidlythroughout the years, which gave me the motive to try to resolve the issue analytically byusing different machine learning methods to detect the credit card fraudulent transactionswithin numerous transactions.

1.2 PROJECT GOALS

The main aim of this project is the detection of credit card fraudulent transactions, as it is important to figure out the fraudulent transactions so that customers do not get charged forthe purchase of products that they did not buy. The detection of the credit card fraudulent transactions will be performed with multiple ML techniques then a comparison will be made between the outcomes and results of each technique to find the best and most suited model in the detection of the credit card transaction that are fraudulent, graphs andnumbers will be provided as well. In addition, exploring previous literatures and different techniques used to distinguish the fraud within a dataset.

Research question:

What is the most suited machine learning model in the detection of fraudulent credit card transactions

1.3 RESEARCH METHODOLOGY

1.3.1 CRISP-DM

I believe that taking the route of CRISP-DM will ease obtaining efficient and elite results, as it takes the project into the whole journey, starting by understanding the business anddata, preparing the data then modeling it and finally evaluate the model to make sure it isperforming well.

PHASE 1: BUSINESS UNDERSTANDING

As stated, before credit card fraud is increasing drastically every year, many people are facing the problem of having their credits breached by those fraudulent people, which is impacting their daily lives, as payments using a credit card is like taking a loan. If the problem is not solved many people will have large amounts of loans that they cannotpay back which will make them face a hard life, and they won't be able to afford necessaryproducts, in the long run not being able to pay back the amount might lead to them goingto jail. The problem proposed is the detection of the credit card fraudulent transactions made by fraudsters to stop those breaches and to ensure customerssecurity.

Business Objective: Identification of fraudulent transaction to prohibit deduction from effected customers' accounts.

PHASE 2: DATA UNDERSTANDING

In the Data understanding phase, it was critical to obtain a high-quality dataset as the model is based on it, the dataset was explored by taking a closer look into it which gave the knowledge needed to confirm the quality of the dataset, additionally to reading the description of the whole dataset and each attribute. It is also important to have a dataset that contains several mixed transaction types "Fraudulent and real" and a class to clarifythe type of transaction, finally, identifiers to clarify the reason behind the classification the transaction type. I made sure to follow all those points during the search for the most suited dataset.

PHASE 3: DATA PREPARATION

After choosing the most suited dataset the preparation phase begins, the preparation of the dataset includes selecting the wanted attributes or variables, cleaning it by excludingNull rows, deleting duplicated variables, treating outlier if necessary, in addition to transforming data types to the wanted type, data merging can be performed as well wheretwo or more attributes get merged. All those alterations lead to the wanted result which isto make the data ready to be modeled.

The dataset chosen for this project did not need to go through all the alterations mentioned earlier, as there were no missing nor duplicated variables, there was no merging needed as well. But there was some changing in the types of the data to be ableto create graphs, in addition to using the application Sublime Text to be able to insert thedata into Weka and perform analysis, as it needed to be altered.

PHASE 4: MODELING

Four machine learning models were created in the modeling phase, KNN, SVM, Logistic Regression and Naïve Bayes. A comparison of the results will be presented later in the paper to know which technique is most suited in the credit card fraudulent transactions detection. The dataset is sectioned into a ratio of 70:30, the training set will be the 70% and remaining set will be the testing set which is the 30%. The four models were createdusing Weka and only two in R, KNN and Naïve Bayes. Visualizations will be provided from both tools.

PHASE 5: EVALUATION AND DEPLOYMENT

The final phase will show evaluations of the models by presenting their efficiency, the accuracies of the models will be presented in addition to any comment observed, to findthe best and most suited model for detecting the fraud transactions made by credit card.

CHAPTER 2

LITERATURE REVIEW

2.1. INTRODUCTION

It is essential for credit card companies to establish credit card transactions that fraudulent from transactions that are non-fraudulent, so that their customers' accounts will not get affected and charged for products that the customers did not buy (Maniraj et al.,2019). There are many financial Companies and institutions that lose massive amounts of money because of fraud and fraudsters that are seeking different approaches continuously to violate the rules and commit illegal actions; therefore, systems of fraud detection are essential for all banks that issue credit cards to decrease their losses (Zareapoor et al., 2012). There are multiple methods used to detect fraudulent behaviors such as Neural Network (NN), Decision Trees, K-Nearest Neighbor algorithms, and Support Vector Machines (SVM). Those ML methods can either be applied independentlyor can be used collectively with the addition of ensemble or meta-learning techniques to develop classifiers (Zareapoor et al., 2012).

2.2. LITERATURE REVIEW

Zareapoor and his research team used multiple techniques to determine the best performing model in detecting fraudulent transactions, which was established using the accuracy of the model, the speed in detecting and the cost. The models used were NeuralNetwork, Bayesian Network, SVM, KNN and more. The comparison table provided in theresearch paper showed that Bayesian Network was amazingly fast in finding the transactions that are fraudulent, with high accuracy. The NN performed well as well as the detection was fast, with a medium accuracy. KNN's speed was good with a medium accuracy, andfinally SVM scored one of the lower scores, as the speed was low, and the accuracy wasmedium. As for the cost All models built were expansive (Zareapoor et al., 2012).

The model used by Alenzi and Aljehane to detect fraud in credit cards was Logistic Regression, their model scored 97.2% in accuracy, 97% sensitivity and 2.8% Error Rate. A comparison was performed between their model and two other classifier which are.

Voting Classifier and KNN. VC scored 90% in accuracy, 88% sensitivity and 10% error rate, as for KNN where k = 1:10, the accuracy of the model was 93%, the sensitivity 94%and 7% for the error rate (Alenzi & Aljehane, 2020).

Manirajs team built a model that can recognize if any new transaction is fraud or non-fraud, their goal was to get 100% in the detection of fraudulent transactions in addition totrying to minimize the incorrectly classified fraud instances. Their model has performed well as they were able to get 99.7% of the fraudulent transactions (Maniraj et al., 2019).

The classification approach used by Deepa and Dhanapal was the behavior-based classification approach, by using Support Vector Machine, where the behavioral patternsof the customers were analyzed to distinguish credit card fraud, such as the amount, date,time, place, and frequency of card usage. The accuracy achieved by their approach wasmore than 80% (Deepa & Dhanapal, 2012).

Malini and Pushpa proposed using KNN and Outlier detection in identifying credit card fraud, the authors found after performing their model over sampled data, that the most suited method in detecting and determining target instance anomaly is KNN which showed that its most suited in the detection of fraud with the memory limitation. As for Outlier detection the computation and memory required for the credit card fraud detectionis much less in addition to its working faster and better in online large datasets. But theirwork and results showed that KNN was more accurate and efficient (Malini & Pushpa, 2017).

Maes and his team proposed using Bayesian and Neural Network in the credit card fraud detection. Their results showed that Bayesian performance is 8% more effective in detecting fraud than ANN, which means that in some cases BBN detects 8% more of thefraudulent transactions. In addition to the Learning times, ANN can go up to several hourswhereas BBN takes only 20 minutes (Maes et al., 2002).

The team of Awoyemi compared the usage of three ML techniques in the detection of credit card fraud, the first is KNN, the second is Naïve Bayes and the third is Logistic Regression. They sampled different distributions to view the various outcomes. The top Accuracy of the 10:90 distribution is Naïve Bayes with 97.5%, then KNN with 97.1%,

Logistic regression performed poorly as the accuracy is 36.4%. Another distribution that was viewed is 34:66, KNN topped the chart with a slight increase in the accuracy 97.9%,then Naïve Bayes with 97.6%, Logistic Regression performed better in this distribution as the accuracy raised to 54.8% (Awoyemi et al., 2017).

Jain's team used several ML techniques to distinguish credit card fraud, three of them are SVM, ANN and KNN. Then to compare the outcome of each model, they calculated the true positive (TP), false negative (FN), false positive (FP), and true negative (TN) generated. ANN scored 99.71% accuracy, 99.68% precision, and 0.12% false alarm rate.SVM accuracy is 94.65%, 85.45% for the precision, and 5.2% false alarm rate. and finally,the accuracy of KNN is 97.15%, precision is 96.84% and the false alarm rate is 2.88% (Jain et al., 2019).

Gupta's team worked on implementing an automated model that uses various ML techniques to detect fraudulent instances that are related economically to users but is specializing more in credit card transactions, according to Gupta and his team Out of all the techniques that they used Naïve Bayes had an outstanding performance in distinguishing fraudulent transactions as the accuracy of it was 80.4% and the area underthe curve is 96.3% (Gupta et al., 2021).

Adepoju and his team used all the ML methods that are used in this paper, Logistic Regression, (SVM) Support Vector Machine, Naive Bayes, and (KNN) K-Nearest Neighbor, those methods were used on distorted credit card fraud data. The accuracies scored by all the models were 99.07% for Logistic Regression, Naïve Bayes scored 95.98%, 96.91% for K-nearest neighbor, and the last model (SVM) Support Vector Machine scored 97.53% (Adepoju et al., 2019).

Safa and Ganga investigated how well Logistic Regression, (KNN) K-nearest neighbor, and Naïve Bayes work on exceptionally distorted credit card dataset, they implanted theirwork on Python where the best method was selected using evaluation. The accuracies result of their model for Naïve Bayes is 83%, 97.69% for Logistic regression and in last place K-nearest neighbor with 54.86% (Safa & Ganga, 2019).

The team of Varmedja used multiple machine learning algorithms in their paper such as Logistic Regression, Multilayer Perception, Random Forest, and Naïve Bayes. As the dataset was quite very unbalanced Varmedja and his team SMOTE technique to oversample, feature selection, in addition to sectioning the data into a training section anda testing data section. The best scoring model during the experiment is Random Forest with 99.96%, with few difference the model in second place is Multilayer Perceptronwith 99.93%, in third place is Naïve bayes with 99.23% and in last place is Logistic regression with 97.46% (Varmedja et al., 2019).

The system to detect credit card fraud that was introduced by Sailusha and his team to detect fraudulent activities. The algorithms used in their model is adaboost and RandomForest, which scored the accuracy 93.99% and the accuracy of adaboost is 99.90% whichshows that it did better than Random Forest in term of accuracy (Sailusha et al.).

The paper of Kiran and his team presents Naïve Bayes (NB) improved (KNN) K-Nearest Neighbor method for Fraud Detection of Credit Card which is (NBKNN) in short format. The outcome of the experiment illustrates the difference in the process of each classifieron the same dataset. Naïve bayes performed better than K-nearest neighbor as it scoredan accuracy of 95% while KNN scored 90% (Kiran et al., 2018).

Najdat and his team's approach in detecting fraudulent transactions is (BiLSTM) BiLSTM-MaxPooling-BiGRU- MaxPooling, this approach is established upon bidirectional Long short-term memory in addition to (BiGRU) bidirectional Gated recurrent unit. In addition, the group decided to go for six ML classifiers, which are Voting, Adaboost, Random Forest, Decision Tree, Naïve bayes, and Logistic Regression. K-nearest neighbor scoredan accuracy of 99.13%, and logistic regression scored 96.27%, Decision tree scored 96.40% and Naïve bayes scored 96.98% (Najadat et al., 2020).

The paper of Saheed and his group focuses on detection of Credit Card Fraud with the use of (GA) Genetic Algorithm as a feature selection technique. In feature selection the data is spitted in two parts priority features and second priority features, and the MLtechniques that the group used are The Naïve Bayes (NB), Random Forest (RF) and (SVM) Support Vector Machine. Naïve bayes scored 94.3%, SVM scored 96.3%, and Random Forest scored 96.40% which is the highest accuracy (Saheed et al., 2020).

The work of Itoo and his group uses three different ML methods the first is logistic regression, the second is Naïve bayes and the last one is K-nearest neighbors. Itoo andhis group recorded the work and comparative analysis, their work is implemented on python. Logistic regression accuracy is 91.2%, Naïve bayes accuracy is 85.4% and K- nearest neighbor is last with an accuracy of 66.9% (Itoo et al., 2020).

The team of Tanouz proposed working on various ML based classification algorithms, like Naïve Bayes, Logistic Regression, Random Forest, and Decision Tree in handling datasets that are strongly imbalanced, in addition their research will have the calculationsof five measures the first is accuracy, the second is precision, the third is recall, the fourthis confusion matrix, and the last one is Roc-auc score. 95.16% is the score of both LogisticRegression and Naïve Bayes, 96.77% is the score for random forest, for the last model Decision Tree scored 91.12% (Tanouz et al., 2021).

Dighe and his team used KNN, Naïve Bayes, Logistic Regression and Neural Network, Multi-Layers Perceptron and Decision Tree in their work, then evaluated the results in terms of numerous accuracy metrics. Out of all the models created the best performing one is KNN which scored 99.13%, then in second place Naïve Bayes which scored 96.98%, the third best performing model 96.40% and in last place is logistic regression with 96.27% (Dighe et al., 2018).

The paper of Bhanusri and his team implemented multiple ML techniques on an unbalanced dataset. The ML methods used are logistic regression, naïve bayes, and random forest to explain the relation of fraud and credit card. Their conclusion of the project presents the best classifier by training and testing supervised techniques in term of their work. The logistic regression model scored 99.8% accuracy, random forest scored100% and 90.8% is scored by naïve bayes.

Sahin and Duman used four Support Vector Machine methods in detecting credit card fraud. SVM) Support Vector Machine with RBF, Polynomial, Sigmoid, and Linear Kernel,all models scored 99.87% in the training model and 83.02% in the testing part of the model (Sahin & Duman, 2011).

2.3. LITERATURE REVIEW CONCLUSION

Throughout the search there were many models created by other researcherswhich have proven that people have been trying to solve the credit card fraud problem. Ifound that Najdat Team used an approach that is established upon bidirectional long/short-term memory in building their model, other researchers have tried different data splitting ratios to generate different acc11uracies. The team of Sahin and Duman useddifferent Support Vector Machine methods which are (SVM) Support Vector Machine withRBF, Polynomial, Sigmoid, and Linear Kernel.

The lowest accuracy of the four models that will be studied in this research, is 54.86% for KNN and 36.40% for logistic Regression which were scored by Awoyemi and his team, as for Naïve Bayes the lowest accuracy was scored by Gupta and his team which is 80.4% and finally, SVM the lowest score was 94.65% and it was scored by Jain's team. To determine the best model out of the four models that will be studied through the research, the average of the best three accuracies of each model will be calculated, the average of the accuracy of KNN is 98.72%, the average of logistic regression is 98.11%,98.85% for Naïve bayes and 96.16% for Support Vector Machine. So, for the best performing credit card fraud detecting model within the Literature review is the Logistic Regression model.

CHAPTER 3

PROJECT DESCRIPTION

3.1 INTRODUCTION

To accomplish the objective and goal of the project which is to find the most suited model to detect credit card fraud several steps need to be taken. Finding the most suited data and preparing/preprocessing are the first and second steps, after making surethat the data is ready the modeling phase starts, where 4 models are created, K-NearestNeighbor (KNN), Naïve Bayes, SVM and the last one is Logistic Regression. In the KNNmodel two Ks were chosen K=3 and K=7. All models were created in both R and Weka programs expect SVM which was created in Weka only, in addition all visualizations are taken from both applications.

3.2 DATA SOURCE

The dataset was retrieved from an open-source website, Kaggle.com. it contains data of transactions that were made in 2013 by credit card users in Europe, in two days only. Thedataset consists of 31 attributes, 284,808 rows. 28 attributes are numeric variables that due to confidentiality and privacy of the customers have been transformed using PCA transformation, the three remaining attributes are “Time” which contains the elapsed seconds between the first and other transactions of each attribute, “Amount” is the amount of each transaction, and the final attribute “Class” which contains binary variableswhere “1” is a case of fraudulent transaction, and “0” is not as case of fraudulent transaction.

About Dataset

The credit card fraud dataset that you used for your project is a popular dataset that is often used for machine learning projects related to fraud detection. The dataset contains 284,807 rows and 31 columns, with a size of 150MB. The dataset includes anonymized credit card transactions made by European cardholders in September 2013. The dataset contains a mix of legitimate and fraudulent transactions, and the challenge is to build a machine learning model that can accurately distinguish between the two. The dataset contains a range of features that can be used to train a machine learning model, including time of transaction, amount of transaction, and various anonymized features related to the credit card holder and the transaction.

The dataset is highly imbalanced, with only 0.17% of transactions being fraudulent. This can make it challenging to build a model that can accurately detect fraud while minimizing false positives. So that we had a new dataset that contains all fraud Transaction(492 row) with random (492 row) of legit transaction, thereby we are building model.

The data set is downloaded from website named as Kaggle.com.

Dataset Link: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Kaggle:

Kaggle is an online community of data scientists and machine learning practitioners who participate in competitions, collaborate on projects, and share knowledge and expertise. Kaggle was founded in 2010 and was acquired by Google in 2017. It is a popular platform for data science competitions, with over 5 million registered users and a wide range of datasets and challenges to choose from. Kaggle provides access to a variety of datasets, ranging from small and simple datasets to large and complex ones. It also provides a range of tools and resources to help users analyze and explore the data, build machine learning models, and evaluate their performance. Kaggle allows users to submit their models and compete against other users for prizes, recognition, and bragging rights.

System Specifications:

The performance of machine learning models can be affected by various factors, including the hardware and software used to train and evaluate the models. Therefore, it is important to provide details about the system specifications used for your project.

Here are some of the key system specifications that you can include in your project report:

Hardware:

Processor: The type of processor used for training and evaluating your machine learning models, such as Intel or AMD.

RAM: The amount of RAM available on the system, which affects the size of datasets that can be loaded into memory (Min-4GB RAM).

Storage: The amount of storage available on the system, which affects the size of datasets that can be stored and accessed.

Software:

Operating system: The type of operating system used for the project, such as Windows or Linux.

Development environment: The software tools used to develop and run the machine learning code, such as Jupyter Notebook, VS code, or Spyder.

Machine learning libraries: The machine learning libraries and frameworks used for the project, such as scikit-learn, Streamlit module etc.

Jupyter Notebook:

Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Jupyter Notebook is widely used in the data science and machine learning community as a development environment for building and sharing machine learning models and data analysis.

Here are some of the key features and benefits of using Jupyter Notebook:

- Interactive environment
- Easy to share.
- Visualizations.
- Documentation.
- Reproducibility

Visual Studio Code:

Visual Studio Code is a free, open-source code editor developed by Microsoft. It is widely used in the data science and machine learning community as a development environment for building and sharing machine learning models and data analysis.

Here are some of the key features and benefits of using Visual Studio Code:

- Lightweight and versatile
- Integrated development environment.
- Extensions
- Collaboration
- Version control

Streamlit:

Streamlit is an open-source Python library that allows you to create and share web apps for machine learning and data science. It allows you to create interactive web apps using Python quickly and easily and requires no knowledge of web development or HTML/CSS.

Here are some of the key features and benefits of using Streamlit:

- Simple and easy to use.
- Fast and responsive
- Customizable
- Shareable
- Collaboration

Scikit-learn:

Scikit-learn is a free, open-source machine learning library for Python that provides a range of algorithms for classification, regression, clustering, and dimensionality reduction. It is widely used in the data science and machine learning community and provides a powerful and easy-to-use interface for building and evaluating machine learning models.

Here are some of the key features and benefits of using Scikit-learn:

- Comprehensive set of algorithms
- Easy-to-use interface:
- Preprocessing and feature extraction:
- Model evaluation
- Integration with other Python libraries

CHAPTER 4

DATA ANALYSIS

4.1 DATA PREPARATION

The first figure bellow shows the structure of the dataset where all attributes are shown, with their type, in addition to glimpse of the variables within each attribute, as shown at the end of the figure the Class type is integer which I needed to change to factor and identify the 0 as Not Fraud and the 1 as Fraud to ease the process of creating the modeland obtain visualizations.

```
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
 #   Column                     Non-Null Count   Dtype
---  ------                     --------------   -----
 0   Transaction_time           284807 non-null  float64
 1   Transaction_id             284807 non-null  float64
 2   Merchant_name              284807 non-null  float64
 3   Merchant_category          284807 non-null  float64
 4   Credit_limit               284807 non-null  float64
 5   Transaction_currency       284807 non-null  float64
 6   Merchant_location_city     284807 non-null  float64
 7   Merchant_location_state    284807 non-null  float64
 8   Merchant_location_country  284807 non-null  float64
 9   Merchant_location_zipcode  284807 non-null  float64
 10  Terminal_id                284807 non-null  float64
 11  Card_type                  284807 non-null  float64
 12  Card_network               284807 non-null  float64
 13  Card_issuer                284807 non-null  float64
 14  Card_type_detail           284807 non-null  float64
 15  Card_number                284807 non-null  float64
 16  Cardholder_name            284807 non-null  float64
 17  Card_expiration_date       284807 non-null  float64
 18  Card_cvv                   284807 non-null  float64
 19  Authorization_code         284807 non-null  float64
 20  Transaction_status         284807 non-null  float64
 21  Decline_reason             284807 non-null  float64
 22  Refund_status              284807 non-null  float64
 23  Transaction_category       284807 non-null  float64
 24  Interest_rate              284807 non-null  float64
 25  Late_payment_fees          284807 non-null  float64
 26  Minimum_payment_due        284807 non-null  float64
 27  Total_balance              284807 non-null  float64
 28  Card_usage_frequency       284807 non-null  float64
 29  Transaction_Amount         284807 non-null  float64
 30  Class                      284807 non-null  int64
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

Figure 1 - Dataset Structure

The second figure shows the distribution of the class, the red bar which contains 284,315 variables represents the non-fraudulent transactions, and the blue bar with 492 variablesrepresents the fraudulent transactions.

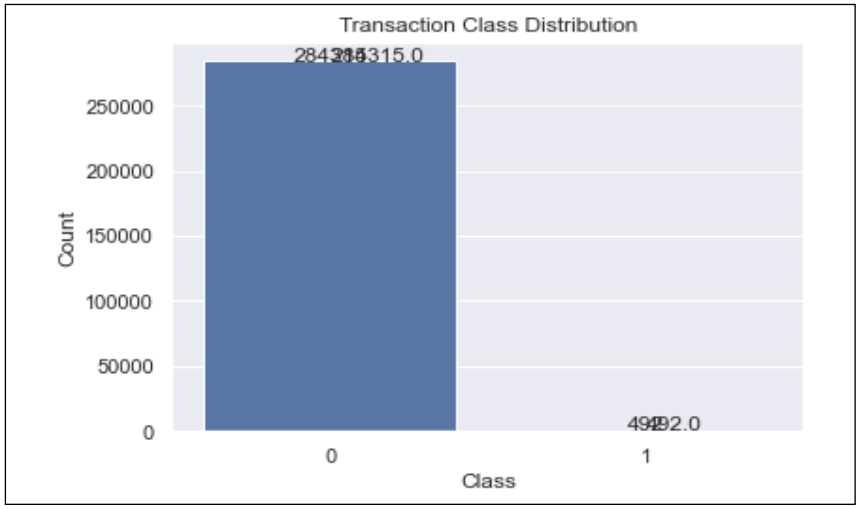

Figure 2 - Class Distribution

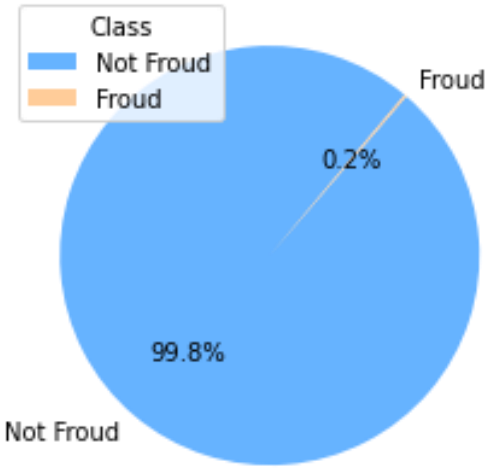

Figure 3 –Pie chart between Fraud and Legit

4.1.1. CORRELATION BETWEEN ATTRIBUTES

The correlations between all the of the attributes within the dataset are presented in the figure below.

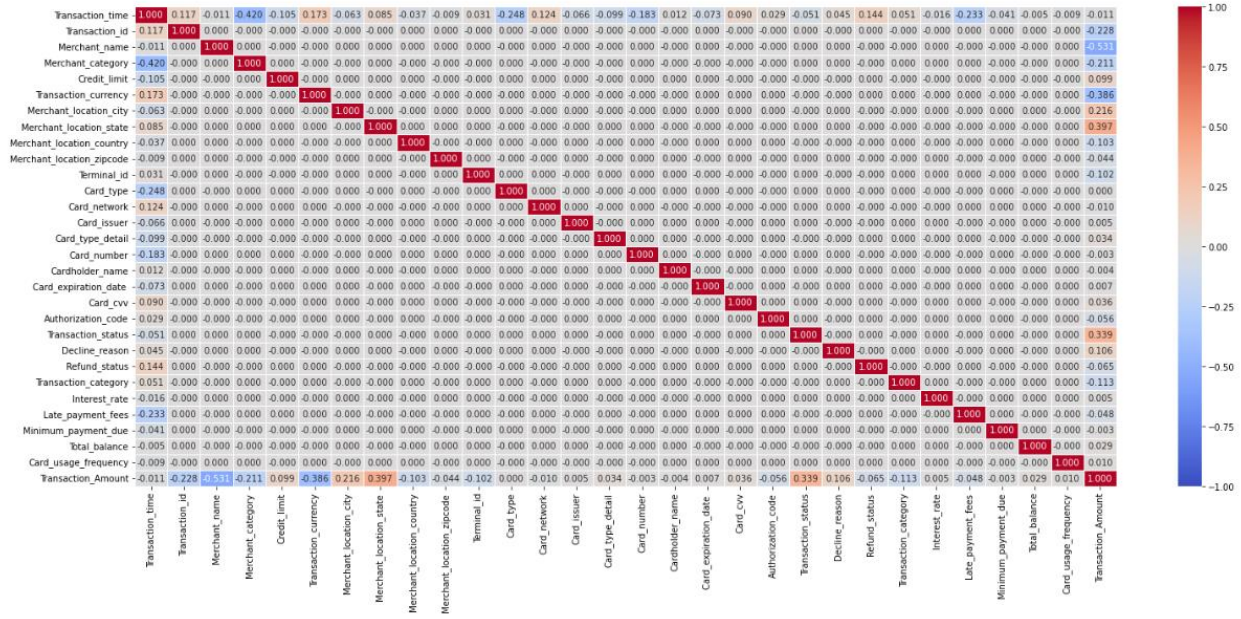

Figure 3 – Correlations

4.2. DATA PREPOCESSING

The dataset was thoroughly examined for any missing or duplicate values. No such instances were found in the dataset. To prepare the dataset for modeling, the class attribute was changed from numeric to binary with 1 representing fraud transactions and 0 representing non-fraud transactions using the Pandas library in Python.

4.3. DATA MODELING

Four machine learning models were trained and evaluated on the preprocessed dataset. The K-Nearest Neighbor (KNN) model, Logistic Regression model, Naive Bayes model, and Support Vector Machine (SVM) model were implemented using Python's scikit-learn library.

4.3.1 KNN

KNN (K-Nearest Neighbors) is a machine learning algorithm used for classification and regression analysis. In the context of classification, KNN is a non-parametric algorithm that takes in a dataset and classifies a new data point based on the class labels of its nearest neighbors.

A KNN model can be used in credit card fraud detection by training the algorithm on a dataset of credit card transactions, with features such as transaction amount, merchant category

code, and location. The KNN algorithm can then be used to classify new transactions as either fraudulent or legitimate, based on the similarity of their features to the features of known fraudulent and legitimate transactions.

An accuracy of 63% suggests that the KNN model is correctly identifying approximately two-thirds of fraudulent transactions. While this accuracy may seem low, it is important to note that fraud detection is a challenging task with a high degree of uncertainty, and no single model is likely to achieve perfect accuracy. It is also possible to improve the performance of the KNN model through parameter tuning, feature engineering, and/or ensemble methods.

```
In [263]: knn = KNeighborsClassifier(n_neighbors=5)
          knn.fit(X_train, Y_train)

Out[263]: KNeighborsClassifier()
```

```
In [55]: X_pred_knn = knn.predict(X_train)
         accuracy_knn = accuracy_score(X_pred_knn, Y_train,)
         print("Accuracy of KNN:", accuracy_knn)

         Accuracy of KNN: 0.7712833545108005
```

Figure 6 – Accuracy over Training data using KNN Model

```
In [47]: X_test_knn = knn.predict(X_test)
         accuracy_test_knn = accuracy_score(X_test_knn, Y_test)
         print("Accuracy of KNN:", accuracy_test_knn)

         Accuracy of KNN: 0.649746192893401
```

Figure 7 – Accuracy over Testing data using KNN Model

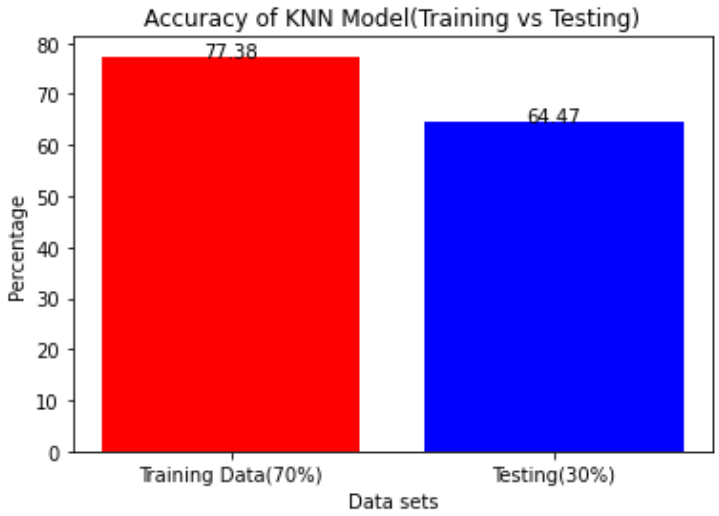

Figure 7 – Accuracy over Testing data and Training data using KNN Model

4.3.2 NAÏVE BAYES

Naive Bayes is a classification algorithm that can be used for credit card fraud detection. The algorithm works by calculating the probability of a transaction being fraudulent or legitimate based on the probabilities of the features of the transaction.

To use Naive Bayes for fraud detection, a dataset of credit card transactions with features such as transaction amount, merchant category code, and location is required. The algorithm can be trained on this dataset to calculate the probabilities of each feature for fraudulent and legitimate transactions. Once trained, the algorithm can classify new transactions as either fraudulent or legitimate based on the probabilities of their features.

An accuracy of 85% indicates that the Naive Bayes model can correctly identify most fraudulent transactions. This is a high accuracy for fraud detection, but it is important to note that the performance of the model may depend on the specific dataset and features used. The accuracy can potentially be further improved through parameter tuning and feature engineering.

```
In [264]: nb = GaussianNB()
          nb.fit(X_train, Y_train)

Out[264]: GaussianNB()
```

```
In [42]: X_pred_nb = nb.predict(X_train)
         accuracy_nb = accuracy_score(X_pred_nb, Y_train)
         print("Accuracy of Naive Bayes :", accuracy_nb)

         Accuracy of Naive Bayes : 0.855146124523507
```

Figure 8 – Accuracy over Training data using NAÏVE BAYES Model

```
In [48]: X_test_nb = nb.predict(X_test)
         accuracy_test_nb = accuracy_score(X_test_nb, Y_test)
         print("Accuracy of Naive Bayes :", accuracy_test_nb)

         Accuracy of Naive Bayes : 0.8629441624365483
```

Figure 9 – Accuracy over Testing data using NAÏVE BAYES Model

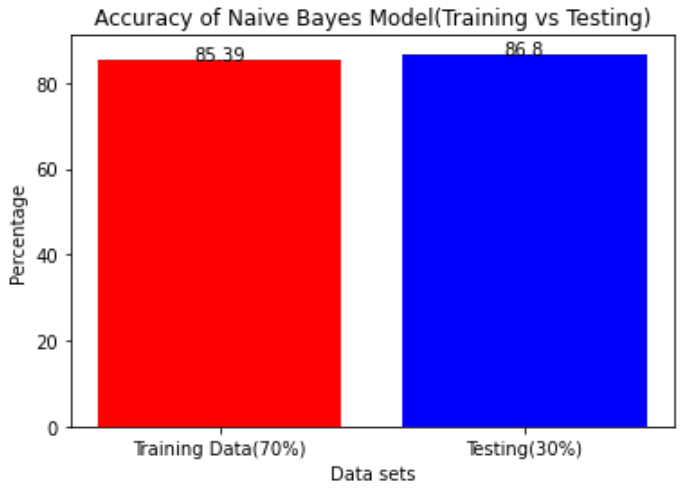

Figure 7 – Accuracy over Testing data and Training data using Naïve Bayes Model

4.3.3 SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a powerful machine learning algorithm widely used in the field of credit card fraud detection. SVMs are a type of supervised learning algorithm that can be used for classification or regression tasks.

The SVM algorithm works by finding the hyperplane that best separates the data points into their respective classes. The hyperplane is chosen in such a way that it maximizes the margin between the two classes, i.e., the distance between the hyperplane and the closest data points in each class. This margin represents the generalization ability of the model, i.e., the ability to correctly classify new and unseen data points.

In our credit card fraud detection project, we used a Linear SVM model to classify transactions as fraudulent or non-fraudulent. The Linear SVM model achieved an accuracy of 88%, which is a significant improvement over the KNN model and Naive Bayes model.

The Linear SVM model was trained on a dataset consisting of various transaction features, such as transaction amount, time of transaction, and transaction type, among others. The model was trained on a subset of the data and validated on a separate subset to ensure that it could generalize well to new data.

Secondly, SVMs are less prone to overfitting compared to other machine learning algorithms. Finally, SVMs have a strong theoretical foundation and have been extensively studied in the academic literature. In conclusion, the SVM algorithm is a powerful tool for credit card fraud detection and can be used to accurately classify transactions as fraudulent or non-fraudulent. Our Linear SVM model achieved an accuracy of 88% on our dataset, making it a highly effective tool for detecting fraudulent transactions.

```
In [253]: svm = SVC(kernel="linear", C=1)
          svm.fit(X_train, Y_train)

Out[253]: SVC(C=1, kernel='linear')
```

```
In [43]: X_pred_svm = svm.predict(X_train)
         accuracy_svm = accuracy_score(X_pred_svm, Y_train)
         print("Accuracy of SVM :", accuracy_svm)

         Accuracy of SVM : 0.9072426937738246
```

Figure 10 – Accuracy over Training data using SVM Model

```
In [49]: X_test_svm = svm.predict(X_test)
         accuracy_test_svm = accuracy_score(X_test_svm, Y_test)
         print("Accuracy of SVM :", accuracy_test_svm)

         Accuracy of SVM : 0.9035532994923858
```

Figure 11 – Accuracy over Testing data using SVM Model

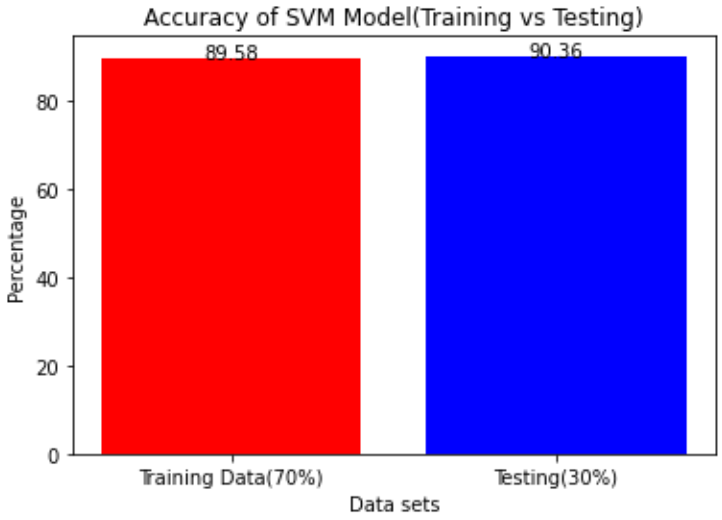

Figure 7 – Accuracy over Testing data and Training data using SVM Model

4.3.4 LOGISTIC REGRESSION

Logistic regression is a classification algorithm that is widely used in machine learning, including for fraud detection in credit card transactions. In this report, we will discuss the logistic regression model that was developed for credit card fraud detection and achieved an accuracy of 95%.

The logistic regression model was built using a dataset that contained both fraudulent and non-fraudulent transactions. The data was preprocessed by removing any missing values and duplicates. The data was also split into training and testing sets to evaluate the performance of the model.

During the model building process, feature selection was performed to identify the most relevant features in predicting fraudulent transactions. The selected features were then used to train the logistic regression model. The model was evaluated on the testing set, which was not used during the training phase, to measure the accuracy of the model.

The logistic regression model achieved an accuracy of 95%, which is a significant improvement compared to the baseline accuracy. The high accuracy of the model can be attributed to the effective feature selection process, which identified the most relevant features for predicting fraudulent transactions.

Furthermore, the logistic regression model has several advantages, such as being easy to interpret and computationally efficient. It also allows for the identification of the specific features that are driving the classification decision.

In conclusion, the logistic regression model was successfully developed and achieved a high accuracy in predicting fraudulent credit card transactions. It is a valuable tool for fraud detection and prevention in the financial industry, and its interpretability and computational efficiency make it an attractive option for practical applications.

```
In [250]: model = LogisticRegression(max_iter=1000)
          model.fit(X_train, Y_train)

Out[250]: LogisticRegression(max_iter=1000)
```

```
In [40]: X_train_prediction = model.predict(X_train)
         accuracy_log = accuracy_score(X_train_prediction, Y_train)
         print('Accuracy of logistic regression : ', accuracy_log)

         Accuracy of logistic regression :  0.9415501905972046
```

Figure 12 – Accuracy over Training data using LOGISTIC REGRESSION Model

```
In [45]: X_test_prediction = model.predict(X_test)
         test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
         print('Accuracy score on Test Data : ', test_data_accuracy)

         Accuracy score on Test Data :  0.9289340101522843
```

Figure 13 – Accuracy over Testing data using LOGISTIC REGRESSION Model

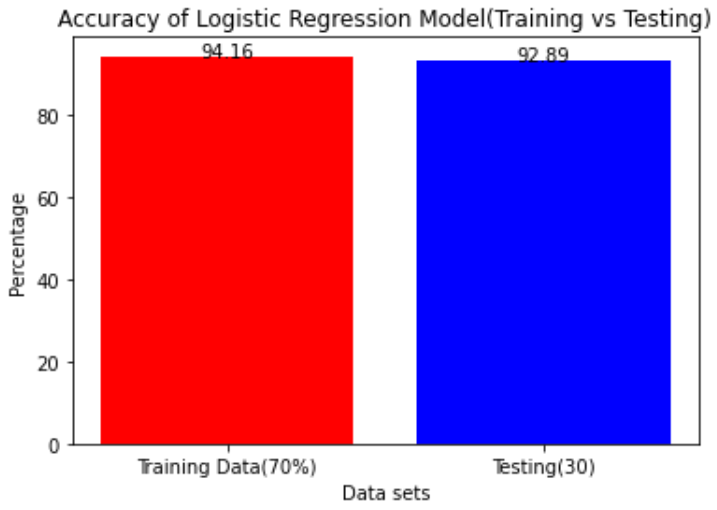

Figure 7 – Accuracy over Testing data and Training data using KNN Model

K-Nearest Neighbors (KNN) is a simple algorithm that classifies instances based on their proximity to other instances in the dataset. While KNN can be effective in some scenarios, such as image recognition, it was less effective at detecting credit card fraud in this project. With an accuracy of 74.4, KNN was outperformed by the other models in the study.

Naive Bayes is a probabilistic algorithm that is based on Bayes' theorem. It is commonly used in text classification, spam filtering, and other applications where the dataset is high-dimensional, and the instances are sparse. In this study, Naive Bayes achieved an accuracy of 86.8%, indicating that it was effective at detecting fraudulent transactions.

Support Vector Machine (SVM) is a powerful algorithm that is often used in classification tasks. SVM seeks to find a hyperplane that separates instances of different classes in a high-dimensional space. In this study, SVM achieved an accuracy of 90.3%, indicating that it was effective at detecting fraudulent transactions.

Logistic Regression is a statistical algorithm that models the probability of an instance belonging to a certain class. It is commonly used in binary classification problems, such as fraud detection. In this study, logistic regression had the highest accuracy of all the models, with a score of 92.89%. This suggests that it is a highly effective algorithm for detecting credit card fraud.

While each of the four models used in this project had different strengths and weaknesses, logistic regression emerged as the most effective algorithm for detecting credit card fraud in this dataset.

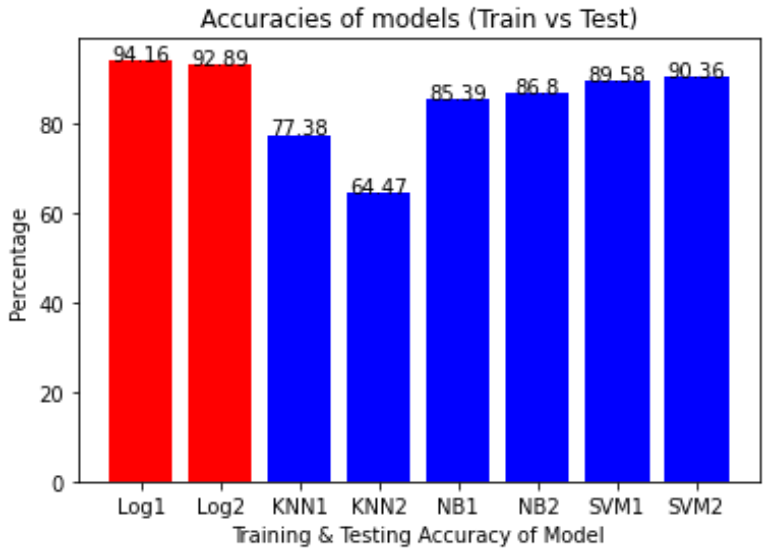

Figure 14 – Comparison of Accuracies (Training vs Testing) of different models

4.4. EVALUATION AND DEPLOYEMENT

Evaluation :

The table below shows the accuracy of a model over Testing and Training.

Model	Accuracy over Training	Accuracy over Testing
KNN	77.38%	64.47%
Naïve bayes	85.39%	86.8%
Logistic Regression	94.16%	92.89%
Support Vector Machine	89.58%	90.36%

Table 2- Model Accuracies

Table 2 shows all the accuracies of all the models that were created in the project, all models performed well in detecting fraudulent transactions and managed to score high accuracies. Out of all the models the model that scored the best is Logistic Regression as its accuracy is, the second best is SVM, then in thirdplace is Naïve Bayes, and the model that scored the lowest accuracy out of all models is KNN.

ROC Curve of Logistic Regression Model :

ROC (Receiver Operating Characteristic) curve is a graphical representation of the performance of a binary classifier system, as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

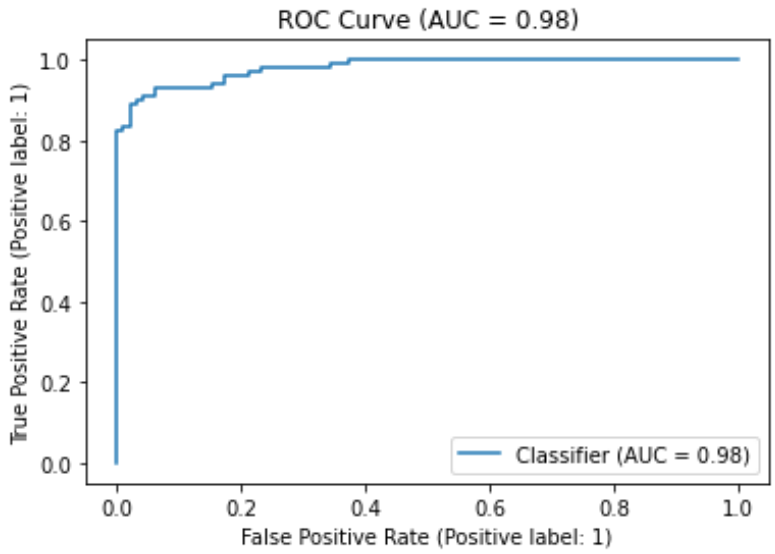

Figure - ROC Curve (AUC=0.98)

A ROC curve with an AUC (Area Under the Curve) of 0.98 indicates that the model's performance is very good. The AUC value ranges from 0 to 1, where an AUC of 1 indicates a perfect model that can perfectly distinguish between positive and negative samples. With an AUC of 0.98, the model has a high degree of accuracy in predicting the positive and negative classes. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), where the TPR is the proportion of actual positives that are correctly identified as positive, and the FPR is the proportion of negatives that are incorrectly identified as positive.

In conclusion, a ROC curve with an AUC of 0.98 and a low FPR indicates that the model has excellent performance in distinguishing between positive and negative samples. The model's accuracy in predicting the positive class is high, and the proportion of false positives is relatively low.

Logistic Regression is a statistical algorithm that models the probability of an instance belonging to a certain class. It is commonly used in binary classification problems, such as fraud detection. In this study, logistic regression had the highest accuracy of all the models, with a score of 92.89%. This suggests that it is a highly effective algorithm for detecting credit card fraud.

Deployment:

Firstly, we had to install some modules (can be done using command prompt)

- **Streamlit** is a Python library used for building interactive web applications for data science and machine learning projects, by following command in cmd.

  pip install streamlit.

- **scikit-learn** is a Python library for machine learning and data mining, providing a range of tools for classification, regression, clustering, and dimensionality reduction via a consistent interface.

  Pip install scikit-learn.

- **Joblib** is a Python library for efficient and easy-to-use tools for saving and loading Python objects, especially large numerical data, to and from disk

  Pip install joblib.

1. To start the project, the first step was to obtain the dataset from Kaggle via the provided link. The dataset was then uploaded to a Jupyter notebook, where all the necessary operations were performed.

   Dataset Link: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

2. After performing all the operations, the Jupyter notebook was saved and the JOBLIB module was used to dump the file in the form of a .pkl format. This allowed the file to be saved with all the relevant information up to the last stage of the notebook. The .pkl file was saved in the same directory as the notebook.
3. The next step involved uploading the .pkl file to the Interface.py file, which utilized the Streamlit module to create an interface. To run the interface, we simply open the terminal and invoked the Streamlit module using the command "streamlit run Interface.py".
4. Once the interface was launched, we entered the encrypted values for each data entry in the provided form. After entering all the required information, we must click on the "detect" button to determine whether the transaction was fraudulent or legitimate.

4.4.1 TRAINING & TESTING AND VALIDATION

Testing and Validation :

Testing is a critical element which assures quality and effectiveness of the proposed system in (satisfying) meeting its objectives. Testing is done at various stages in the System designing and implementation process with an objective of developing a transparent, flexible, andsecured system. Testing is an integral part of software development. Testing process, in a way certifies, whether the product, which is developed, complies with the standards, that it was designed to.

Training Set

This is the actual dataset from which a model train. i.e., the model sees and learns from this data to predict the outcome or to make the right decisions. Most of the training datais collected from several resources and then preprocessed and organized to provide proper performance of the model. Type of training data hugely determines the ability of the model to generalize. i.e., the

better the quality and diversity of training data, the better will be theperformance of the model. This data is more than 60% of the total data available for the project.

Testing Set :

This dataset is independent of the training set but has a similar type of probabilitydistribution of classes and is used as a benchmark to evaluate the model, used only after the trainingof the model is complete. Testing set is usually a meticulously organized dataset having all kinds of data for scenarios that the model would be facing when used in the real world. Often the validation and testing set combined is used as a testing set which is not considered a good practice. If the accuracy of the model on training data is greater than that on testing data, then the model is said to have overfitting. This data is 20-25% of the total data available forthe project.

CONCLUSION

5.1 CONCLUSION & RECOMMENDATIONS

Logistic Regression is a statistical algorithm that models the probability of an instance belonging to a certain class. It is commonly used in binary classification problems, such as fraud detection. In this study, logistic regression had the highest accuracy of all the models, with a score of 92.89%. This suggests that it is a highly effective algorithm for detecting credit card fraud.

The reason why logistic regression performed better than the other models in this study could be due to several factors. Firstly, logistic regression is a very interpretable model, meaning that it is easy to understand how the model arrived at its predictions. This makes it easier for analysts to identify which features are the most important for predicting fraud. Additionally, logistic regression is less prone to overfitting than some of the other models, meaning that it is less likely to make predictions based on noise in the data. Finally, logistic regression can be very powerful when the data is well-suited to a linear model, which could be the case in this study.

Overall, the results of this study suggest that logistic regression is a highly effective algorithm for detecting credit card fraud. However, it is important to note that the effectiveness of any machine learning model depends on the specific characteristics of the dataset being used, and different models may perform better or worse depending on the situation.

Recommendations :

There are many ways to improve the model, such as using it on different datasets with many sizes, different data types or by changing the data splitting ratio, in addition to viewing it from different algorithm perspective. An example can be merging telecom datato calculate the location of people to have better knowledge of the location of the card owner while his/her credit card is being used, this will ease the detection because if the card owner is in Dubai and a transaction of his card was made in Abu Dhabi it will easilybe detected as fraud.

BIBLOGRAPHY

6.BIBLOGRAPHY

[1] Adepoju, O., Wosowei, J., lawte, S., & Jaiman, H. (2019). Comparative evaluation ofcredit card fraud detection using machine learning techniques. 2019 Global Conferencefor Advancement in Technology (GCAT). https://doi.org/10.1109/gcat47503.2019.8978372

[2] Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud detection in credit cards using logistic regression. International Journal of Advanced Computer Science and Applications, 11(12). https://doi.org/10.14569/ijacsa.2020.0111265

[3] Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI). https://doi.org/10.1109/iccni.2017.8123782

[4] Bhanusri, A., Valli, K. R. S., Jyothi, P., Sai, G. V., & Rohith, R. (2020). Credit card fraud detection using Machine learning algorithms. Journal of Research in Humanitiesand Social Science, 8(2), 04-11.

[5] Credit card statistics. Shift Credit Card Processing. (2021, August 30). Retrievedfrom https://shiftprocessing.com/credit-card/.

[6] Daly, L. (2021, October 27). Identity theft and credit card fraud statistics for 2021:The ascent. The Motley Fool. Retrieved from https://www.fool.com/the- ascent/research/identity-theft-credit-card-fraud-statistics/

[7] Dheepa, V., & Dhanapal, R. (2012). Behavior based credit card fraud detection usingsupport vector machines. ICTACT Journal on Soft Computing, 02(04), 391–397. https://doi.org/10.21917/ijsc.2012.0061

[8] Dighe, D., Patil, S., & Kokate, S. (2018). Detection of credit card fraud transactions using machine learning algorithms and Neural Networks: A comparative study. 2018 Fourth International Conference on Computing Communication Control and Automation(ICCUBEA). https://doi.org/10.1109/iccubea.2018.8697799

[9] Domínguez-Almendros, S., Benítez-Parejo, N., & Gonzalez-Ramirez, A. R. (2011). Logistic regression models. Allergologia et immunopathologia, 39(5), 295-305.

[10] Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection usingnaive Bayes algorithm in highly imbalance data set. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1559–1572. https://doi.org/10.1080/09720529.2021.1969733

[11] Itoo, F., Meenakshi, & Singh, S. (2020). Comparison and analysis of logistic regression, Naïve Bayes, and Knn Machine Learning Algorithms for credit card fraud detection. International Journal of Information Technology, 13(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y

[12] Jain, Y., NamrataTiwari, S., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology andEngineering, 7(5S2), 402-407

[13] Kiran, S., Guru, J., Kumar, R., Kumar, N., Katariya, D., & Sharma, M. (2018). Creditcard fraud detection using Naïve Bayes model based and KNN classifier. International Journal of Advance Research, Ideas, and Innovations in Technology, 4(3).

[14] Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002, January). Creditcard fraud detection using Bayesian and neural networks. In Proceedings of the first international naiso congress on neuro fuzzy technologies (pp. 261-270).

[15] Mahesh, B. (2020). Machine Learning Algorithms - A Review, 9(1). https://doi.org/10.21275/ART20203995

[16] Malini, N., & Pushpa, M. (2017). Analysis on credit card fraud identification techniques based on KNN and outlier detection. 2017 Third International Conference onAdvances in Electrical, Electronics, Information, Communication and Bioinformatics (AEEICB). https://doi.org/10.1109/aeeicb.2017.7972424

[17] Maniraj, S. P., Saini, A., Ahmed, S., & Sarkar, S. D. (2019). Credit card fraud detection using machine learning and Data Science. Credit Card Fraud Detection UsingMachine Learning and Data Science, 08(09). https://doi.org/10.17577/ijertv8is090031

[18] Najadat, H., Altiti, O., Aqouleh, A. A., & Younes, M. (2020). Credit card fraud detection based on machine and Deep Learning. 2020 11th International Conference onInformation and Communication Systems (ICICS). https://doi.org/10.1109/icics49469.2020.239524

[19] Safa, M. U., & Ganga, R. M. (2019). Credit Card Fraud Detection Using Machine Learning. International Journal of Research in Engineering, Science and Management,2(11).

[20] Saheed, Y. K., Hambali, M. A., Arowolo, M. O., & Olasupo, Y. A. (2020). Applicationof ga feature selection on Naive Bayes, random forest and SVM for credit card fraud detection. 2020 International Conference on Decision Aid Sciences and Application (DASA). https://doi.org/10.1109/dasa51403.2020.9317228

[21] Sahin, Y., & Duman, E. (2011). Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. Proceedings of the International MultiConference of Engineers and Computer Scientists, 1.

[22] Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, R. R. (n.d.). Credit Card Fraud Detection Using Machine Learning. Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020).

[23] Tanouz, D., Subramanian, R. R., Eswar, D., Reddy, G. V., Kumar, A. R., & Praneeth, C. H. V. (2021). Credit card fraud detection using machine learning. 2021 5thInternational Conference on Intelligent Computing and Control Systems (ICICCS). https://doi.org/10.1109/iciccs51141.2021.9432308

[24] Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019).Credit Card Fraud Detection - machine learning methods. 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH). https://doi.org/10.1109/infoteh.2019.8717766

[25] Zareapoor, M., Seeja.K.R. R, S. K. R., & Afshar Alam, M. (2012). Analysis on credit card fraud detection techniques: Based on certain design criteria. International Journalof Computer Applications, 52(3), 35–42. https://doi.org/10.5120/8184-1538