National College of Ireland

Financial Analysis

MSc in Fintech

Continuous Assessment 1

Group Members:

[OLADIPO OLADOTUN – X20230532]

[SOBAYO HIKMOT-X20241089]

[NWACHUKWU IKECHUKWU EUGENE-X20257929]

The Use of Machine Learning Models For Credit Card Fraud Detection

Abstract— The introduction of credit cards has afforded customers the option to make purchases for goods and services by applying for loans from banks and making repayments with little additional interest. The rate of fraudulent transactions has increased over the years which has been difficult for both consumers and firms. The financial industry continues to change because of the introduction of new technologies in the sector. Machine learning models have been able to provide solutions for businesses for the purpose of large data circulating around the world. The main aim of this research proposal is to make use of machine learning models to correctly predict transactions that are fraudulent and genuine. This research work will make use of machine learning models such as logistic regression, Decision tree, and neural network. The use of Random sampling is been adopted for the purpose of balancing the dataset.

Keywords— Machine learning, Logistic regression, Decision tree

I. INTRODUCTION

With the ever-growing reach and influence of information technology and its influence on the banking sector, many financial transactions have been modernized and innovative ways have been introduced in which financial transactions can be established, maintained, and monitored. One such invention is the introduction of credit cards which is an incentive that affords cardholders the option to apply for loans to make purchases by paying instantly and additional interest accruals is applied if the loans are not made back after the grace period [3]. With the invention of credit cards, customers can take loans to make purchases instantly and payment for these loans can be made periodically [3]. Credit cardholders enjoy numerous benefits like one-time bonuses, minimal fees, and interests with high rewards rates, reward points, a sufficient grace period for loan repayment, and insurance [4].

Financial institutions can offer financial benefits through the issuance of credit cards to foster and strengthen relationships and financial engagements with clients and customers. Financial institutions risk losing their customers to other competitors with the existence of a robust credit card program that offers a platform for servicing individuals, small businesses, and commercial credit card programs [1]. With the establishment of mobile banking applications, cardholders can have better control of their credit card spending, view the extent of grace periods, and the availability of other incentives like consumer protection [4]. In Europe, Purchase transactions among credit cardholders amounted to €135billion in 2021 and are estimated to rise to €180billion by 2026 [7].

With most financial transactions conducted on the internet and the use of mobile banking applications for payments and other monetary engagements, banks have at their disposal, a large variety of customers' confidential information and data which makes them susceptible to cybercrime and credit card fraud [3]. With the advent of digitization and automation of financial systems, cybercrimes have become more electronically sophisticated and impersonal. One of these malicious cyber-attacks plaguing the banking industry is credit card fraud which compromises the personal information of cardholders and gives unauthorized access to confidential data and makes payments without due authorization from the cardholder or the issuing bank [1]. In 2020, financial fraud in the UK amounted to £783.3million with losses incurred via payment cards, remote banking, and cheques [6].

With the growth and expansion of credit card usage and the increase in the volume of purchases and payments using credit cards, credit card fraud detection is crucial for financial institutions to avoid financial losses and reduction in customer engagement [5]. Financial organizations must detect fraudulent activities to eliminate exposure to such activities, reduce costs associated with credit card fraud and mitigate the impact of such attacks on the organization to ensure improvement in results and build and maintain the trust of both customers and shareholders [3]. Credit card fraud detection analysis can be carried out to ascertain the causes of the fraud and the possible effect it might have on the cardholder and the financial institution. Data analytics has been adopted by many financial organizations to help identify traces of fraudulent transactions and through visualization, banks can identify future effects of such malpractice on financial operations [4]. Data and fraud analytics enable financial institutions to hide patterns and trends, structure uneven or unbalanced data for easy analysis and possible solutions and integrate various data sources and records to build an effective model for identifying fraudulent activities to increase performance and continuous operation [3]. Due to fraudsters and criminals adapting and inventing new ways to hijack financial operations and get access to customers' confidential details, financial institutions have incorporated the use of Artificial Intelligence and Machine Learning Algorithms that will detect fraudulent transactions effectively and provide consistent results [2].

Most credit card fraud scenarios happen in seconds and companies have a large volume of data and information, machine learning models need to be employed to track fraudulent transactions within seconds and identify

irregularities in large chunks of data [2]. Machine learning models (ML) provide greater accuracy, reduce operational cost, detect patterns, provide insights, and can easily adapt to the innovative tactics of fraudsters [2]. ML algorithms are generally more efficient than human participants as a large quantity of data can be processed and analyzed in seconds. ML techniques become better at predicting consumers' purchase behavior and deciphering patterns and traces of fraudulent activities [5]. For instance, ML algorithms like Logistics regression (LR) and Random Forest (RF) can predict the number of fraudulent transactions and genuine transactions [2]. Algorithms like Artificial Neural networks (ANN) can trace consumers' daily activities on the internet before making an order or purchasing a product on a website or online store [4]. With the constant increase in the size of data available to most financial institutions, it's important to adopt ML algorithms because the larger the data, the more efficient the model is at predicting and detecting credit card fraud [1].

The purpose of this research is two-fold. Firstly, this paper proposes the use of three Machine Learning Algorithms (Logistics Regression, Decision Trees, and Artificial Neural networks) to ascertain which model best identifies fraudulent transactions. Secondly, due to the highly imbalanced nature of the adopted datasets, two resampling techniques (the Random Over-Sampling Technique (ROSE) and the Synthetic Minority Oversampling Technique (SMOTE)) have been used to balance out the dataset.

Section II discusses various studies on various ML algorithms and how the accuracy levels of these classifiers in detecting credit card fraud. Section III describes the proposed methodology, techniques adopted, and the results derived from the models. Section IV contains the conclusion of the paper and future work and recommendations.

II. Related works

Many studies have deployed various machine learning algorithms to create various detection systems and models to determine the frequency of fraudulent transactions in credit card datasets. Based on the high imbalance of most credit card fraud datasets where non-fraudulent transactions exceed fraudulent transactions, resampling techniques have been employed to balance the datasets such as the Random Over-Sampling Technique (ROSE), the Synthetic Minority Oversampling Technique (SMOTE), and other resampling techniques. Key findings from various studies will be discussed and criticized below and insights on the success of various Machine learning algorithms will be evaluated.

[8] implemented a machine learning model based on detecting credit card fraud using the genetic algorithm(GA), a heuristic algorithm used to provide high-quality solutions to solve optimization problems and is efficient for large and complex datasets, to select certain features that predict the occurrence of credit card fraud accurately and also adopted Machine learning(ML) classifiers such as Decision Tree(DT), Artificial Neutral Networks(ANN), Logistic Regression(LR), Random Forest(RF) and Naïve Bayes(NB) to analyze and categorize the data and gather valuable insights. The outcome showed that the combination of the GA feature selection together with any of the classifiers had higher accuracy in predicting fraudulent transactions than implementing the Classifier techniques only.

[9] also adopted the GA technique for feature selection of the most important and fittest features while adopting ML classifiers like NB, RF, and SVM for data analysis. The result showed that RF – GA method was most accurate in predicting features for credit card fraud with an accuracy of 96.4 compared to SVM and NB which had an accuracy of 96.3 and 94.3 respectively. In Research [10] feature selection was used on the European cardholders' dataset to balance out the data and various ML algorithms including RF, ANN, LR, and NB were used with each having accuracy of 99.96%, 99.93%, 97.46%, and 99.23% respectively. Due to the dataset being imbalanced, the Synthetic Minority Oversampling Technique (SMOTE) was used for random sampling [11].

[12] used the European cardholders' dataset to evaluate the performance of six ML classifiers including RF, LR, DT, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and K-nearest neighbors (KNN). The results showed that RF, XGBoost, and LR with an accuracy of 0.986%, 0.984%, and 0.977% were highly effective and provided the highest business value. [13] also implemented the SMOTE technique on the insurance and credit card datasets extracted from Kaggle and adopted machine learning models with the SMOTE technique. Results proved that SMOTE adapted with machine learning algorithms performed better than the Original SMOTE technique with K-means SMOTE and Borderline SMOTE having better accuracy of 99.93% and 98.26% compared to pure SMOTE with an accuracy of 97.76% indicating that oversampling with machine learning better predicts fraudulent activities than just using oversampling methods.

[14] combined the SMOTE technique with the edited nearest-neighbor (ENN) to provide a balance of oversampling and undersampling methods used on the datasets. The SMOTE-ENN technique was evaluated based on three metrics: sensitivity, specificity, and area under the curve (AUC). The result showed that the SMOTE-ENN technique performed better based on those metrics compared to if the technique wasn't implemented. The authors of the paper [15] combined the SMOTE resampling method with the RT classifier to predict the occurrences of fraudulent transactions, with results showing that the combination of RT with the SMOTE technique provided the best outcomes with a ROC-AUC score of 0.97.

In this paper [16] DT was the preferred model to ascertain credit card fraud as it had a better sensitivity and precision in comparison to other ML methods such as KNN, LR, NB, and RF. Even with its high precision, the author advised that DT should be used with unsupervised machine learning methods to get the best results. DT and RF were also identified as having better accuracy in predicting credit card fraud with an accuracy of 98.4% and 98.5% respectively in comparison to other ML methods such as LR and KNN which had an accuracy of 98.2% and 94% respectively [17].

[18] adopted the Extreme Gradient Boosting (XGBoost) with the Long-Short Term Memory (LSTM) to evaluate the effectiveness of the XGBoost classifier with or without LSTM in predicting credit card defaulters. Based on performance metrics like accuracy, recall, and precision, the XGBoost-LSTM model outperformed the XGBoost model with 0.936% for accuracy, 0.736% for recall, and 0.928 for precision compared to the XGBoost model which had

0.870% for accuracy, 0.507% for recall and 0.768 for precision.

[18] adopted the Adaboost technique, a technique used for boosting or strengthening the performance of ML algorithms to solve problems of classifications and improve the accuracy, of certain ML classifiers namely: Support vector machine (SVM), Logistic regression(LR), Decision Tree(DT), Random Forest(RF) and Extreme gradient boosting (XGBoost). The results showed that the ML classifiers performed better when adopted with the AdaBoost technique having high accuracy, precision, and recall than when the ML methods were not applied with the AdaBoost technique. The RF with the AdaBoost technique had a higher accuracy of 99.97%, recall of 99.77%, and a precision of 99.91% in comparison to an accuracy of 99.95%, recall of 79.83%, and precision of 97.19% when the AdaBoost technique wasn't used. [22] identified the AdaBoost technique as an efficient enabler to improve predicting power of ML classifiers in identifying fraudulent transactions since most credit card fraud datasets are highly skewed due to most variables or features being mainly non-fraudulent engagements. With ML classifiers like NB, DT, and RT displayed an increase in accuracy from 7.4% to 94.1% in predicting fraudulent transactions once AdaBoost was adopted.

[19] used various ML algorithms on a credit card datasets extracted from Kaggle with findings indicating that based on the performances metrics used such as accuracy, sensitivity, specificity, and precision, the Support vector machine (SVM) proved to be the best predictor of both fraudulent and non-fraudulent transactions with accuracy and precision of 94.13%. Even though Naïve Bayes (NB) was a better predictor of fraudulent credit card activities, SVM was better at predicting non-fraudulent activities as most credit card transactions are non-fraudulent [19]. [24] proposed a dual approach model for detecting credit card fraud occurrences by resampling the data to increase the frequency of fraudulent activities and then applying ML Classifiers to ascertain the best model to predict fraudulent transactions. Results showed that of all the models used, RF had the best performance metrics attaining a true positive score of 0.96 compared to LR, SVM, DT, and ANN with true positive scores of 0.86, 0.85, 0.88, and 0.83 respectively.

[23] proposed a model that used several unsupervised techniques such as Neural networks (NN), Local Outlier Factor(LOF), Isolation Forest (IF), and K-Means clustering (KMC) to evaluate the occurrence of credit card fraud. NN technique proved to be the best with an accuracy of 99.87% compared to IF and LOF which both had an accuracy of 98% and KMC with an accuracy of 99.75%. [25] analyzed the efficiency of ML algorithms in predicting mobile money fraud detection with Support Vector Machines (SVM), Decision Trees(DT), and Naïve Bayes(NB) algorithms used for predicting and identifying mobile money fraudulent transactions. The performance metrics used to evaluate the performance of the classifiers were accuracy, precision, recall, and F1-score. The findings displayed Decision trees as the best classification algorithm for predicting fraudulent transactions in an imbalanced dataset with an accuracy of 99.90%, precision of 99.99%, recall of 100%, and F1-Score of 99.95%

[26] developed an enhanced model for predicting credit card fraud by selecting three machine learning algorithms out of nine machine learning algorithms and 19 resampling techniques used alongside each classifier. After the first stage, three ML algorithms were selected which were CatBoost, Extreme Gradient Boosting (XGBoost), and Random Forest(RF), then the 19 resampling techniques were used for each classifier to identify the best predictor for fraudulent transactions. The All K-Nearest Neighbours resampling technique and CatBoost classifier proved to be the best with AUC, Recall, and F1-Score of 97.94%, 95.91%, and 87.40%. [25] adopted the resampling technique on a credit card fraud dataset by using the SMOTE AND Adaptive Synthetic (ADASYN) to resample the imbalanced dataset and then adopt seven ML algorithms are used on the newly balanced dataset and then the Deep Reinforcement Learning (DRL) is used to evaluate the reliability of the dataset. Results showed that ML classifiers performed better with the SMOTE resampling technique in comparison to the ADASYN where results were low. DRL was also discovered to be ineffective in evaluating the reliability of the imbalanced dataset [23]. Two ML classifiers (XGBoost and RF) gave the best results when dealing with both resampling techniques for the balanced dataset [27].

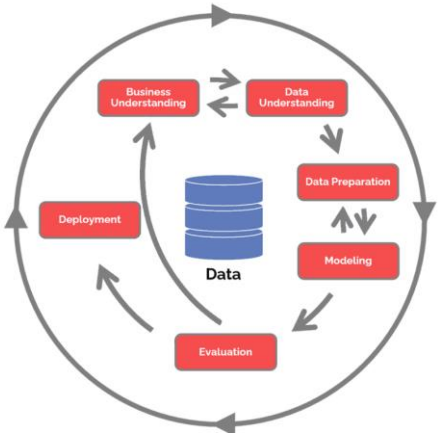
[23] developed a method for credit card fraud detection by adopting the ML algorithms such as Fuzzy-Rough nearest neighbor (FRNN), Sequential Minimal Optimization (SMO), and Logistic Regression (LR). SMO showed the highest detection rate with 76.40% better than LR at 76.40% and FRNN at 68.50%. Being used on two separate datasets, SMO and LR had high approval ratings on both datasets with SMO having 0.8525 and 0.6810 on the first and second datasets respectively, and the approval rating of LR amounting to 0.8550 and 0.6810 on both the first and second dataset [24].

[24] applied two resampling techniques, the Random Oversampling technique (ROSE) and the SMOTE technique to balance the dataset adopted by the paper due to the imbalanced nature of the dataset. SMOTE proved to be the most accurate of the two due to its ability to create new synthetic samples to balance the dataset. After balancing the dataset, five ML algorithms were applied to the dataset with results showing that LR had the best accuracy and precision of 97.04% and 99.99% respectively. [24] also adopted the ROSE and SMOTE sampling techniques for balancing the dataset and applied several ML algorithms to evaluate which was the best indicator of fraudulent transactions. The SMOTE produced the best results as it was able to ascertain accurately the fraudulent and non-fraudulent transactions. Random forest was the best predictor of credit card fraud based on two performance metrics, accuracy, and F1-Score, with RF having a score of 0.984 and 0.992 respectively.

III. METHODOLOGY

In relation to the data mining process, the use of the Cross-Industry Standard is being adopted for this project work. It can be defined as a guide that can be used for a data mining process that comprises different phases for analysis. The Cross industry-standard refers to a model that can be used as a foundation for the data science process. It comprises six important phases namely Business understanding, Data

understanding, Data preparation, Modelling, Evaluation, and Deployment. Below is a diagram that can be used to describe the process of data mining.



The process goes through a clockwise format from business understanding to the final phase which is deployment.

A. *Business Understanding*

The project focuses on credit card fraud detection that is used to help financial firms gain control against fraudsters on the internet. The dataset is based on private credit card information that was used in the past with different time stamps. Credit card fraud has been suggested to be an easy target for fraudsters against e-commerce businesses. The e-commerce businesses focus on primarily receiving payments online from customers through debit or credit cards. The use of cards as a means of payment is assigned to a specific individual for the main purpose of conducting transactions within a specified credit limit in advance.

B. *Data Understanding*

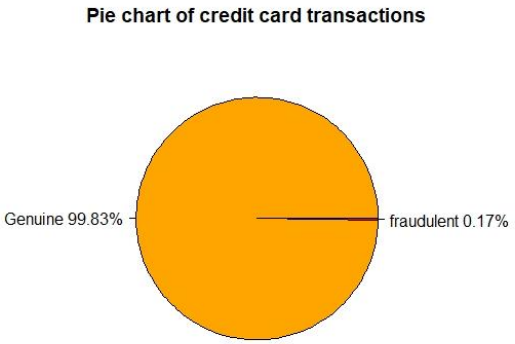
It is important to provide a better understanding of the dataset, it contains transactions that were made by European cardholders in the year 2013. The dataset is able to provide transactions that are fraudulent or genuine. It was discovered that 492 credit card purchases were fraudulent out of the total transaction of 284,807 in the dataset. It is also important to point out that the dataset is highly imbalanced because of the huge difference between fraudulent and genuine transactions. The dataset contains majorly numerical values which were used as a result of principal component analysis. PCA is a statistical procedure that can be used to summarise a huge sum of data through the process of smaller pieces of information. The reason this technique was adopted simply because of issues in relation to the confidentiality of users. It is important that essential information such as background information for customers is not published to the public.

After the use of PCA

No	Feature	Description
1	Time	The time frame between the first transaction and the current transaction
2	Amount	Transaction amount
3	Class	0– Genuine, 1-fraudulent

C. *Data Preparation*

A credit card fraud detection dataset is been gathered through the Kaggle website. This dataset consists of numerous transactions conducted by individuals which are have been classified into fraudulent and genuine transactions. The dataset is evaluated using logistic regression as the primary model to solve the classification problem. It is important to highlight any missing values in the data set and also use pie charts for the purpose of providing visualization. The dataset comprises 28 predictor variables with a response variable in form of class. In the process of cleaning the data, it was discovered there were 0 missing values. There was no form of outliers and the dataset is an imbalanced dataset that produces a high accuracy. The purpose of logistic regression was to correctly predict fraud transactions for a financial firm. Below is a diagram of a pie chart that shows the percentage of genuine transactions and fraudulent ones.



D. *MODELING*

1) *Logistic Regression*

The concept of logistic regression is adopted for this data set. Logistic regression is a machine learning model that focuses on a target variable in form of a category that can be used to solve a classification problem. The above model problem is used to solve a problem to determine whether a transaction is fraudulent or genuine. The values that are adopted for this classification problem are strictly between 0 and 1 . This

project is based on binary logistic regression that has two possible outcomes.

- **Result**

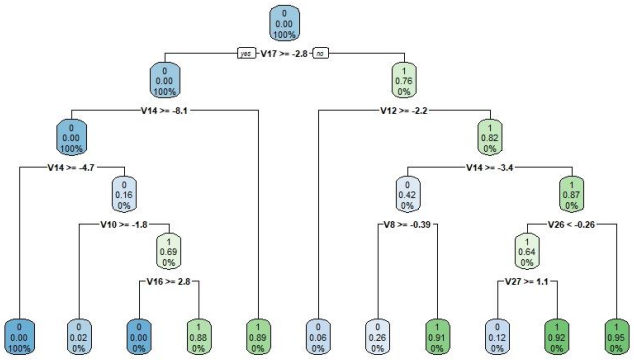
The logistic regression model makes use of 28 predictor variables to correctly predict a response variable. Below is a diagram of logistic regression describing the process.

```
Call:
glm(formula = Class ~ ., family = binomial(), data = test_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0932  -0.0258  -0.0169  -0.0113   4.1464

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -9.225344   0.415251 -22.216  < 2e-16 ***
V1          -0.019437   0.115409  -0.168  0.86625
V2           0.171945   0.248592   0.692  0.48914
V3          -0.029636   0.131294  -0.226  0.82142
V4           0.334886   0.160529   2.086  0.03697 *
V5           0.168722   0.205370   0.822  0.41133
V6          -0.152158   0.194218  -0.783  0.43337
V7          -0.063784   0.227020  -0.281  0.77874
V8          -0.141855   0.096201  -1.475  0.14033
V9          -0.635029   0.260651  -2.436  0.01484 *
V10         -0.392055   0.234060  -1.675  0.09393 .
V11         -0.173557   0.200231  -0.867  0.38606
V12          0.806153   0.310982   2.592  0.00953 **
```

Pr (>|z|): The use of P-value can be used to determine how significant the use of logistic regression is for prediction. It is important to point out that the smaller the p-value, the more significant the estimate is.

2) Decision tree

The model decision can be used to solve both regression and classification problems. It belongs to a supervised learning algorithm. The classification process of the machine follows two steps process from the learning and prediction step. In the learning step, a model can be developed based on producing training data. In the prediction step, such a model can be used to predict the response of a given data. It is considered to be one of the easiest and most popular classification algorithms. There are important terminologies such as root node, splitting, decision node, leaf/terminal node, pruning, branch, and parent node. Below decision that comprises key important predictor variables that can be. Below is a diagram of the decision tree applied to the dataset.



3) Neural Network

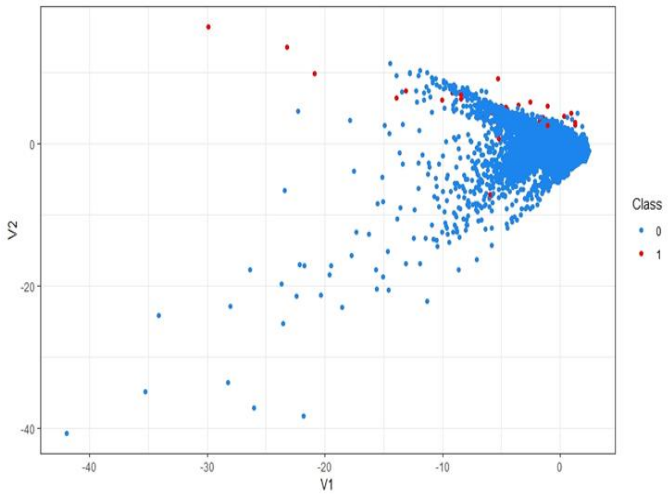
The use of the model can adapt to show an underlying relationship in a set of data through the process of mimicking how the human brain operates. It operates in the same format as that of the system of neurons either in an organic or artificial nature. It consists of three layers which are input, hidden, and output. This model can be referred to as a general circuit of neurons that work on any number of inputs and can be suitable for dealing with nonlinear datasets. It can be used to solve both classification and regression problems. The model made use of multiple predictor variables and a single response variable.

4) Random Resampling of an Imbalanced Dataset

The main purpose of using this technique helps to provide an individual with the ability to be able to create a new transformed version of the training dataset in which the chosen examples have different class distributions. It can also be referred to as a simple strategy that can be used to solve classification problems. The main idea behind this strategy would be on a particular dataset randomly through the process of random sampling. There are two approaches that can be adopted for random sampling which are Random Oversampling and Random under-sampling. These techniques were adopted on the selected dataset

5) Smote

The term SMOTE represents a Synthetic minority oversampling technique that can be used to create a replica of an existing minority class. It is used to solve a dataset that is imbalanced in nature. The concept uses a technique that allows for the oversampling of the minority class in a synthetic way. it creates synthetic data that is based on similarities within an existing minority instance. A synthetic instance is created based on finding the K – nearest neighbors of each minority instance. The main purpose of this technique is to avoid the issue of overfitting in data analysis. It can be adopted for different applications such as data analysis, statistics, and others.



E. EVALUATION

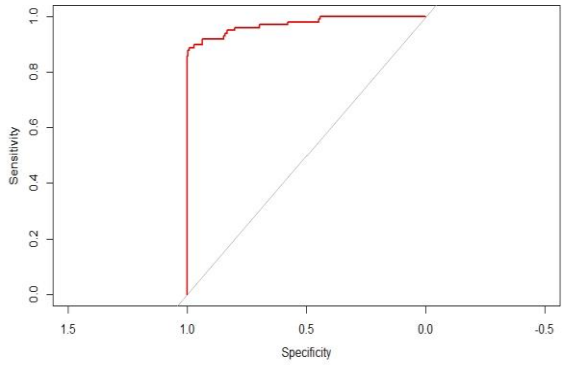
Under this concept, an evaluation technique was adopted for this research work. The credit card dataset would use techniques such as Logistic regression, Decision tree, and Neural network to solve this classification problem. The dataset is partitioned into training data and test data on the basics of 80% and 20% respectively. The use of ROC CURVE is considered to be known as the receiver operating characteristic curve. It is referred to as a graph that can be used to show the performance level of a classification model. The graph comprises of two parameters in the name of True positive and False positive rates.

F. DEPLOYMENT

The final stage of the Crisp industry methodology can be used to help financial firms fight against fraudsters on the internet. The main goal of this machine learning process is to determine how machine learning models can solve classification problems. The use of logistic regression is able to provide an accurate answer to the classification problem with the use of multiple predictor variables.

IV. EVALUATION/RESULT

The use of ROC curve is used to justify how good the binary classification is in classifying based on true positive and false-positive rates. ROC curve was used to justify how effective logistic regression was able to solve the classification problem with the dataset. The range below this curve falls within the range 0 to 1. The ROC curve is able to justify based on the logistic regression used for prediction, it was to accurately make predictions of true positive and false-positive rates.



V. CONCLUSIONS AND FUTURE WORK

The concept of credit card fraud is considered to be a major challenge for e-commerce businesses globally. It is important to point out that firms have looked to develop various types of models to help reduce fraud. In the past, most financial firms used traditional approaches that allow an analyst to use fixed code to curb fraudulent transactions. This approach was not successful because of the large amount of data. This project is based on machine learning models such as logistic regression, decision tree, and neural network to help tackle fraud. The use of logistic regression was applied to the dataset that produced a high accuracy of 90%. It was obvious that this dataset was imbalanced, there was a need to make use of random sampling techniques.

In relation to future works, the amount of data on the internet would always increase and fraudsters will always look for alternative ways to disrupt machine learning models to hide. It would be advised that advanced models should be introduced to help against these challenges.

REFERENCES

[1] I. Emmanuel, S. Yanxia and W. Zenghui, “A machine learning-based credit card fraud detection using the GA algorithm for feature selection,” *JOURNAL OF BIG DATA*, vol. 9, pp. 17, 2022.

[2] M. T. Chung and D. H. Phan, “Comparing ML Algorithms on Financial Fraud Detection,” 2019.

[3] S. Aanchal, G. Harshvardhan and K. G. Mahendra, “A Dual Approach for Credit Card Fraud Detection using Neural Network and Data Mining Techniques,” *7th India Council International Conference* , 2020.

[4] M. Jincy and S. Prathiksha, “An Analysis on Fraud Detection in Credit Card Transactions using Machine Learning Techniques,” *Proceedings of the Second International Conference on Artificial Intelligence and Smart Energy* , 2022.

[5] H. N. Abrar, M. S. Ibrahim and S. C. Mohammad, “Analysis of Machine Learning Techniques for Credit Card Fraud Detection,” *International Conference on Machine Learning and Data Engineering* , 2019.

[6] K. Samidha, A. Aishwarya and P. A. Arun, “Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison,” *10th International Conference on Cloud Computing, Data Science & Engineering.*

[7] I. EMMANUEL, S. YANXIA and W. ZENGHUI, “Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost,” *IEEE ACCESS*, 2021.

[8] S. Tanmay and H. S. Gururaja, “Explainable Machine Learning in Identifying Credit Card Defaulters,” *Global Transitions Proceedings*, 2022.

[9] A. Saurabh, B. Sushant, S. Survesh and K. N. Vinay, “Prediction of credit card defaults through data analysis and machine,” *Materials Today: Proceedings*, vol. 51, pp. 110-117, 2022.

[10] E. EBENEZER, D. M. IBOMOIYE, A. KEHINDE and O. GEORGE, “A Neural Network Ensemble With Feature Engineer For Improved Credit Card Fraud Selection,” *IEEE Access*, 2022.

[11] M. A. TALHA, S. KAMRAN and S. SHAKIR, “An Investigation of Credit Card Default Prediction,” *IEEE Access*, 2020.

[12] S. Nikolaos, P. Stelios and G. Alexandros, “Default avoidance on credit card portfolios using accounting,

demographical and exploratory factors: decision making based on machine learning (ML) techniques," *Annals of Operations Research,* pp. 715–739, 2020.

[13] D. Kamil, "PREDICTING CREDIT CARD CUSTOMER CHURN USING SUPPORT VECTOR MACHINE BASED ON BAYESIAN OPTIMIZATION," vol. 70, pp. 827-836 , 2021.

[14] T. Huei-Wen and L. Michael, "Estimation Procedures of Using Five Alternative Machine Learning Methods for Predicting Credit card default," *Review of Pacific Basin Financial Markets and Policies,* vol. 22, pp. 27, 2019.

[15] G. Jing, S. Wenjun and S. Xin, "Research on Default Prediction for Credit Card Users Based on XGBoost-LSTM Model," *Discrete Dynamics in Nature and Society,* vol. 2021, pp. 13, 2021.

[16] Z. Yiming, D. Haoyun, L. Chenglong and D. Ning, "Comparative study on credit card fraud detection based on different support vector machine," *Intelligent Data Analysis,* vol. 25, pp. 105–119, 2021.

[17] M. M. Nhlakanipho and N. Nalindren, "Solving Misclassification of the Credit Card Imbalance Problem using near miss," *Mathematical Problems in Engineering,* vol. 2021, pp. 16, 2021.

[18] M. Elena-Adriana and M. Gabriela, "Methods of handling unbalanced data sets in credit card detection," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience,* vol. 11 , pp. 131-143, 2020.

[19] S. R. Routhu and R. P. Alwyn, "Detection of phishing websites using an efficient feature-based machine learning framework," *Neural Computing and Applications,* pp. 3851–3873, 2019.

[20] M. Kazi, A. Rokibul, S. Nazmul and A. Hojjat, "A dynamic ensemble learning algorithm for neural networks," *Neural Computing and Applications (,* vol. 3, pp. 8675–8690, 2020.

[21] Y. Meltem, u. N. Ozge and Y. Seda, "Using machine learning techniques to develop prediction models for detecting unpaid credit card customers," *Journal of Intelligent & Fuzzy Systems,* vol. 39, pp. 6073–6087, 2020.

[22] K. A. FAWAZ, M. IQRA and U. K. HIKMAT, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithm," *IEEE ACCESS,* 2022.

[23] A. E. D. EMAD, E. ESSAMEDEAN and M. FAHIMA, "Improve Profiling Bank Customer's Behavior Using machine learning," *IEEE ACCESS,* 2019.

[24] I. EMMANUEL, S. YANXIA and W. ZENGHUI, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," *IEEE ACCESS,* 2021.

[25] I. Emmanuel, S. Yanxia and W. Zenghui, "A machine learning-based credit card fraud detection using the GA algorithm for feature selection," *Journal of Big Data,* 2022.