National College of Ireland

Classification of Credit Card Fraudlent transactions using Neural Network and Oversampling Technique.

MSc Research Project
Msc in Data Analytics

Ketan R.Chalwadi
Student ID: 19222840

School of Computing
National College of Ireland

Supervisor: Prof.Hicham Rifai

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Ketan R.Chalwadi
Student ID:	19222840
Programme:	Msc in Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Prof.Hicham Rifai
Submission Due Date:	20/04/2021
Project Title:	Classification of Credit Card Fraudlent transactions using Neural Network and Oversampling Technique.
Word Count:	5286
Page Count:	14

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Ketan R.Chalwadi
Date:	23rd September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Credit Card Fraudlent transactions using Neural Network and Oversampling Technique.

Ketan R.Chalwadi
19222840

Abstract

Credit card fraud is a financial type of fraud that involves the use of credit card details to purchase products and withdraw specific amounts without the permission of the person who holds the credit card.Since the advent of the online payment method in the banking sector, there has always been someone or a group of individuals who have discovered new techniques or approaches to obtaining finance/funds through unlawful means.It is noted that the card owner is not aware of illegal transactions that have performed with his/her credit card until any kind of purchase is made, as physical credit cards are not used in online purchases. In recent years, many credit card companies have deployed an automated system with machine learning technique commonly known as Fraud Detection system so as to analyse fraudulent transaction.Every new fraudulent activity raises the demand for software systems that detect fraudulent credit card transactions. Based on the logs of transactions performed, many researchers have built credit card fraud detection systems that use various data mining and deep learning techniques, machine learning algorithms to determine whether the transaction performed is fraudulent.However, the intricacy of fraudulent transactions is created in such a way that it resembles the genuine ones every time.. For the identification of frauds, the suggested method employs unbalanced severely skewed transactional data and a convolutional network. The dataset utilised here is the highly skewed machine learning kaggle dataset for credit card fraud detection. The characteristics that have been assessed are 1 for the fraud class and 0 for the non-fraud class.The present research uses credit card fraud dataset,where the dataset is preprocessed with the help of Principal Component Analysis , Adaptive Synthetic Sample technique and then Neural Network Classifiers are applied with different number of hidden layers and performance of these classifiers has been evaluated on the basis of accuracy,precision and recall rate.

Keywords— KDD,Principal Component Analysis(PCA),Adaptive Synthetic Sample(ADASYN),Random Forest,Decision Tree, Neural Network(NN)

1 Introduction

With the growing reliance on electronic payment methods, especially the global increase in the use of credit cards by customers and businesses , most of the stakeholders, including e-card companies, payment service sectors and individuals with cardholders, continue to look for various measures to combat risks of credit card fraud.Prevention of Credit card frauds is beneficial for a number of reasons. The most obvious advantage of implementing a good fraud prevention system is that it minimises and regulates potential financial losses that occured because of fraudulent activity.Many of credit card users and issuers suffer significant financial losses each year as a result of fraudulent credit card transactions, and large amount of money might be saveed if successful and adequate fraud prevention techniques will be implemented.

Financial fraud is rising on a daily basis, resulting in significant losses for the banking industry, business groups, and the government sectors. Credit card fraud refers to the situation in which hoaxer utilises a credit card for their own purposes while the owner of that credit card is unaware.There are two forms of credit card fraud: fist is stealing the credit card physically and other is by refering the card's sensitive information such as card number, CVV, expiration year, and name without the card holder's permission. This credit card information in the hands of a fraudster, can be used to begin transaction and withdraw significant sums of money or purchases before the cardholder is aware.

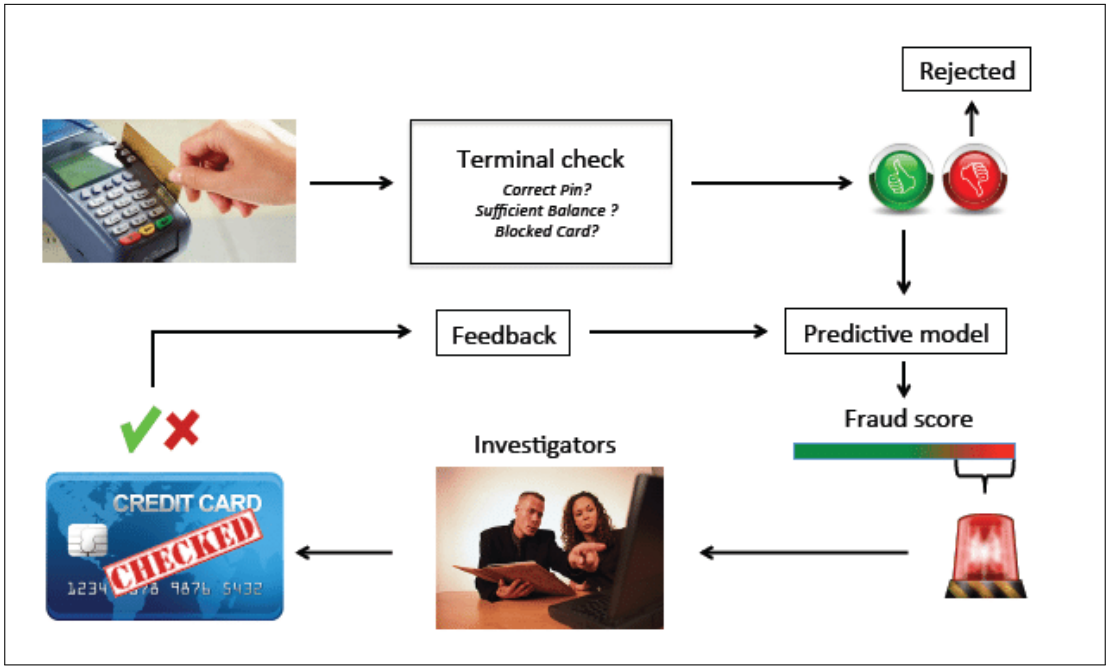


Figure 1: Credit Card Fraud Mechanism

While businesses continue to evolve and migrate to the internet, and money is increasingly exchanged electronically in a paperless banking environment, effective fraud detection remains a top priority for modern technology banking systems. This is not just about decreasing immediate costs caused by fraudulent transactions, but also about ensuring that honest consumers are not harmed by automated and manual fraud detection.Tanouz et al. (2021) According to a Statista poll, 42% of worldwide internet users bought items by online in 2013. In 2011, there were 792.6 million digital purchasers globally. The figure climbed to 903.6 million a year later. In 2016, 43.2% of all worldwide internet users made an online purchase.This proportion is anticipated to rise to 46.4% in 2017. Roy et al. (2021) According to a BBC News poll, losses from online banking fraud increased by 48% in 2014 compared to 2013, as customers increas-

ingly performed their financial transactions online. As the number of cashless transactions and online purchases grows, so does the number of fraudulent transactions.

In the research study provided byWang et al. (2018) several machine algorithms such as Naive Bayes, Logistic Regression, J48, and Adaboost were employed. The Naive Bayes Algorithm is a classification kind of algorithm that uses Bayes' theorem to determine the likelihood of an event occurring. Logistic regression is frequently used for classification problems.The J48 method, which creates a decision tree, is used to solve the classification issue. J48 is an Iterative Dichotomieser ID3 extension. J48 is regarded as one of the most frequent and well-studied topics of Machine Learning, and its methods mostly deal with categorical and constant variables.The method is primarily used to increase the accuracy of the decision tree. The J48 method was employed in this study to categorise transactions as fraudulent or non-fraudulent. According to their findings, the J48 algorithm and Logistic Regression both produce the highest percentage of accuracy. Because they are both correct, the sensitivity factor was utilised to select the best algorithm.Every year, more credit card fraud incidents occur as a result of the widespread use of the internet and other mobile technologies.It has been estimated that the credit card industry loses around $2 billion each year due to fraudulent transactions.At the same time, cardholders are unsure how to defend themselves. Both credit card issuer company and consumers want to know how to avoid or identify fraud as quickly as possible.

In the research study Benson Edwin Raj and Annie Portia (2011a) The Federal Republic of Germany and the United Kingdom have the most web users, while credit cards are the most popular form of payment 59%. Barclaycard, the largest credit card firm in the United Kingdom, allegedly handled 350 million transactions per year around the turn of the century.Retailers such as Wal-Mart often handle a much broader range of credit card transactions, and fraud occurs.Benchaji et al. (2018) The total credit card fraud in the United States is estimated to be 2.7 billion in 2005 and 3.0 billion in 2006, with 1.6 billion and 1.7 billion being estimates of online frauds in the near future.

The major purpose of this proposed research project is to develop a classification-based model to detect fraudulent credit-card transactions. In addition, the proposed research will attempt to meet the following objectives.Identifying the various forms of purchase where the fraudulent transaction happens. Identifying the pattern of the fraudulent transaction so that preventive actions may be taken to minimise monetary losses.

For a variety of reasons, including the fact that data distribution varies over time owing to new fraudulent tactics and seasonality, fraud detection systems are regarded as a key problem for Machine learning.This research study's contribution is to uncover the factors that influence in classifying credit card fraudulent transactions.The proposed research work is divided into different sections.Section 2 describes various research work carried out by researchers in detecting credit card fraudulent transactions.Section 3 discusses research methodology which explains Dataset Selection,Data preprocessing,Data Cleaning and transformation.Section 4 and 5 describes the Research project implementation and Design specification of the credit card fraud detection model.Key findings of the research work has been discussed in section 6 and Final section describes about Conclusion and Future Work.

2 Related Work

Machine learning techniques,Statistical approach,Deep learning techniques are essential in a wide range of successful data analysis fields, including the identification of credit card fraud. The detection of credit card fraud is dependent on analysing card behaviour during transactions. This part delves into a variety of significant research. A brief explanation of all the studies conducted by various researchers for credit card fraud detection is provided below.

2.1 Stastical Approach

Research study proposed byNadim et al. (2019) detecting credit card fraudlent transactions using a customised or aggregated approach. As a result of credit card frauds, the financial industry has suffered losses in the millions of dollars. To identify fraud, enormous effort, time, and money have been expended. To detect fraud, a customised model is developed for each credit card holder. In addition, the efficacy of the customised model is compared to that of the aggregated model.For this aim, real transaction and other data are collected via an online question section, and different models are constructed using random forest and naive bayes. Model performed good, however the accuracy of the customised model was typically lower than that of the aggregated model. This application has issues such as missing values, incomplete values, and false data provided by the individual.As a result, adopting a credit rating model can help to enhance this.

In contrast to the previous studies, Benson Edwin Raj and Annie Portia (2011b)and Agrawal et al. (2015a) have also worked with an industry partner to fraud detection issues that realistically describes the working framework of fraud detection model through the analysis of enormous number of credit card transactions.They employed the isolation forest in combination with local outlier factor to find anomalies.It works best on data that are unlabelled.The technique allows for the avoidance of detection subtasks and the completion of the behaviour profile of logical graphs, which are total ordering based methods for presenting the logical relationship of characteristics of transaction records. It is based on the likelihood of a path transition from one characteristic to another. They described an entropy-based data diversity coefficient to reflect the variety of transactions.

Mareeswari and Gunasekaran (2016) proposed privacy protection of sensitive data using distributed k-means clustering method. An economical, privacy protective using k-means clustering algorithmic program during a social network setting is introduced.The objective was to prevent the service provider from knowing the user's sensitive private information while without affecting the efficiency of the k-means clustering algorithm technique.The feasibility of privacy-protecting using k-means clustering technique takes half hour to cluster one lakh individuals. The result obtained was good, however it maximises the total of pairwise dissimilarity since it utilises a sum of squared Euclidean distance, thus a mix of methods may be used to address this problem.Hence Genetic algorithm, clusters of previous transaction records and Support vector machine-SVM based approach were used. Furthermore, by filtering the collected data, fraud is identified in order to achieve a better result.Also, For this issue, an intelligent fraud detection approach was used.They worked on an improved light gradient boosting machine LightGBM and showed promising results when compared to other current methods.

2.2 Machnine learning technique

Yu et al. (2020) and Guo and Li (2008) conducted research using matching algorithms to differentiate the pattern of a newly performed credit card transaction with the previous immediate patterns of each customer, and on the basis of this technique, they created patterns of fraudulent and genuine transactions for each customer, and whenever a new transaction takes place, they check if it is same as fraudulent transaction pattern or genuine transaction pattern.To distinguish between the two patterns, i.e. fraudulent and genuine transactions patterns, firstly they differentiate each customer's transactions, then separate the fraudulent transactions from the genuine transactions for each customer, and finally apply Apriori algorithm to the list of genuine and fraudulent transactions.Because there is a genuine and fraudulent transaction for every consumer, any new transaction from that customer makes would be compared to existing patterns, making it straightforward to identify whether the transaction is fraudulent or genuine transaction.In addition,Shen et al. (2007) suggested machine learning algorithms such as Bayesian belief networks (BNNs), Apriori Algorithm, and decision trees (DTs) were utilised to

identify fraud in financial transactions. In this study, financial transactions data from 78 Greek industrial firms were collected.In the data gathering, assessors discovered that 38% of financial transactions were fraudulent. Following the execution of the fore mentioned algorithms, it was discovered that Bayesian belief networks (BNNs) produced the highest percentage of accuracy, i.e. 90.34%, and a recall value of 82 %.

Zhang et al. (2009) conducted research that created a Credit Card Fraud Detection System based on Observation Probabilistic in Hidden Markov model, The k-means clustering technique was employed in this model. In this case, observation probabilistic in an HMM-based model initially examines cardholder amount spending on the basis of transaction. Using data aggregation of parameter areas, a clustering model is utilised to classify legitimate and fraudulent transactions.The Markov process displays the starting state and the transaction state immediately, whereas the Hidden Markov Model gives the observation state of the initial state and the transaction success and the K-means algorithm is used for training the model.The cardholder's spending habits are evaluated based on price ranges between high and low.In addition, a research study conducted by Rai and Dwivedi (2020) and Zhang and Huang (2020) suggested the Multiple Classifier System (MCS) to address difficulties based on their analysis utilising single classifiers. They demonstrated through their experimental results that the MCS model outperformed prior works. They have also completed the experimental work with a rigorous experimental investigation and the outcomes to confront the imbalanance issue.

In the reserach performed by Kazemi and Zarrabi (2017) Standardized models were initially employed as a machine learning-based approach for identifying credit card based fraudulent transaction, but hybrid models that incorporated XGBoost and support vector machine-SVM was developed .A publicly available dataset was utilised to evaluate the model's performance, while the feature variables in the dataset from are used to examine fraudulent transactions. Gradient Boosting (GB), AdaBoost and support vector machine (SVM), Decision Tree (DT), Logistic Regression Random Forest, and a mix of various categorization machine learning algorithms are used which resulted in recall of 89%.Futhermore It has been discovered that high value percentage of recall was obtained by dataset balancing via undersampling of the given dataset.In this study, a comparison of four classification algorithms, namely Logistic Regression and Random Forest,, AdaBoost and support vector machine(SVM), was performed and among these models, AdaBoost and support vector machine(SVM) proved to be one of the best models with an accuracy value of 94.5%.

2.3 Deep learning technique

Deep learning is a popular and efficient approach for identifying credit card fraud. This network has a complex data transmission system that is tough to comprehend.In order to implement Deep learning in Fraud detection system reseracher Agrawal et al. (2015b) and Thennakoon et al. (2019) developed a deep autoencder and was employed at several phases to recover the best features in the dataset and for classification purposes. High percentage of accuracy and low percentage of variance are two most significant performance matrices that have been identified using these Neural network.It also provides a good starting point for the sensitivity analysis of the suggested feature parameters in terms of fraud detection effectiveness. This approach has been tested on European credit card datasets from Neural Networks. This approach is based on an artificial neural network that has time and memory aspects built in, such as long and short-term memory, along with several other feature variables.They presented a novel paradigm for detecting credit card fraud that integrates artificial neural networks and spectraal graphical analytics. The Whale algorithm approach was utilised to optimise the back propagation method of neural networks employed in the research study. Furthermore, the researchers proposed a technique for parameter tuning using Deep Learning Network topologies for detecting fraudulent transactions.The financial organisation is able to cut costs by reducing fraudulent credit

card practises. The neural network that was implemented included 2 input layers, 10 hidden layers, and 2 output layers. On 820 research samples, they achieved remarkable findings. They also built and tested two neural networks for fraud detection using a deep auto encoder and an artificial neural network.Their studies demonstrate that their recommended approach for identifying credit card fraud is more effective. The Optimization of Whale method yielded an accuracy rate of 95.40 % and a recall percentage of 96.73%.

Modi and Dayma (2017) Conducted a study to define customer credit rating using an Hidden Markov Model and Group Method of Data Handling GMDH hybrid model.The aim is to define a score a customer's credit score on the basis of the transactions. By self-organizing, GMDH may automatically choose input variables from dataset and obtain the best output model. As a result, a hybrid model integrating HMM and GMDH was suggested.The hybrid model takes into account customer information and employs customer attributes as input variables in the modelling process, and the findings suggest that it can increase the effect of credit scoring. The study was decent, however the resultant model's accuracy may be enhanced by including secondary arguments and neural networks based models with active neurons.Secondary features variables can be calculated usin subsamples since they take less computing time.

3 Research Method and Specification

As shown in Figure 2 Knowledge Discovery from Database research methodology has been used for the present reseach study explains the overall procedure starting from data collection,pre processing of data , feature-engineering process,data modelling,results evaluation and prediction of model visualisation. The inclusion of feature-engineering process to the present research approach contributes in determining the most important elements in predicting fraudlent transactions that has happened using credit card.

The present research study makes use of a credit card frauds dataset containing both types of transactions that is genuine and fraudlent transactions experienced by European Cardholders . This data set on fraudulent transactions has been taken from Kaggle dataset repository. There are about 284807 transactions details along with 30 feature variables and one target variable present in the dataset.The second stage is pre-processing of data, which includes identifying and then imputing missing values. Furthermore, data transformation is used to normalise and standardise the data. Inaccuracy in the results may occur beacuse of the missing values. As a result, missing values must either be removed or replaced with the mean or NA values.

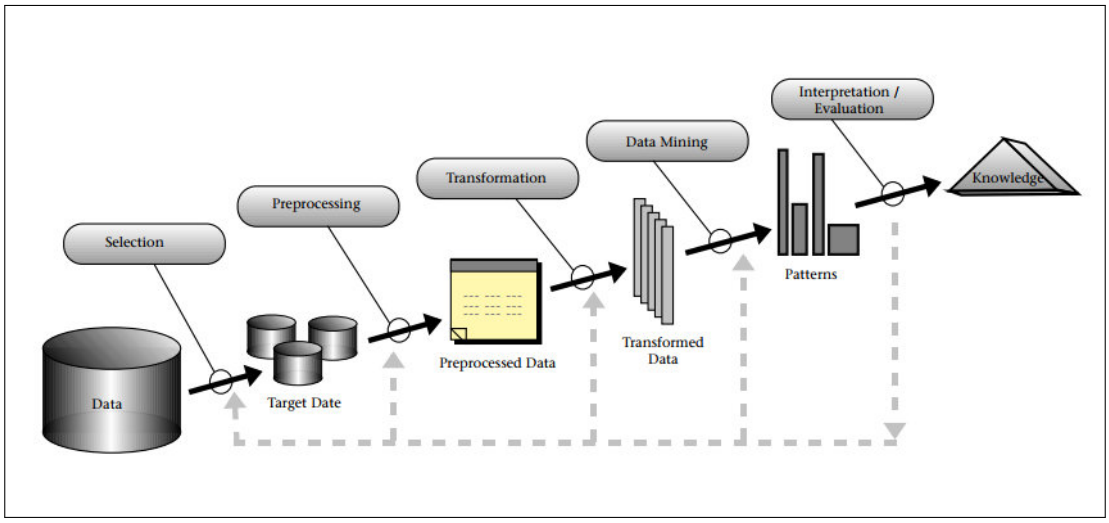


Figure 2: KDD Approach

One of the feature variables of numerical type in the credit card frauds dataset in this present research had missing values. The fillna calculation approach is used to replace missing values with the mean of the feature varaibles that has values. Furthermore, using ordinal and label encoder which is data transformation methods were used to correctly assign values to categorical feature variables.The 'Standard scaler' standardisation based tool is used to standardise the nominal type of variables.

The third stage is feature engineering process and is done in order to identify the most important feature variable influencing credit card fraud detection, which improves the performance of Neural Netwok multi layer model with resampling technique using Adaptive Synthetic Sample(ADASYN) and dimension reductionality technique(PCA) .This present research study helps to determine the optimal feature engineering approach based on selection of important feature varibles which direcly contributes to performance of the model.

The fourth stage entails the use of machine learning methodologies. In order to predict credit card based frauduent transactions, data modelling entails the use of dimension reduction techniques namely PCA and resampling technique namely ADASYN along with Neural Network based Machine learning technique.

The outputs of the Neural Netwok multi layer classification model with ADASYN and PCA technique are then evaluated using performance metrics like recall, precision, F1-score, and AUC-ROC. Finally, the performance indicators from different algorithms are displayed and compared in order to choose the optimal credit card fraud prediction model.

4 Design Specification

The major goal of this research study is to demonstrate that Neural networks along with PCA as dimension reduction and Adaptive Synthetic Sample resampling technique models can effectively and more accurately predict credit card based fraudulent transaction that has taken place by means of various payment methods . Figure 2 depicts the process of designing and implementing the credit card fraud detection model.

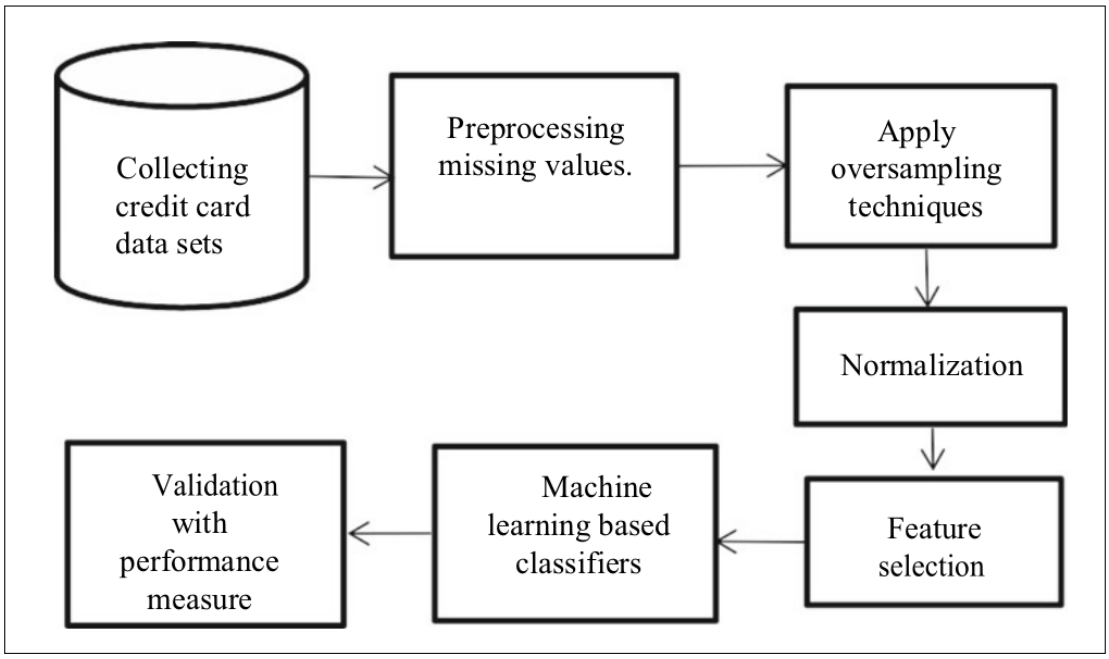


Figure 3: System Architecture

This reserach work presents also highlights on comparison of several classification based

models. To make predictions, Neural Network with PCA models employs a well defined classifier technique that is arbitrarily associated with the real classification and is constructed by using Adaptive Synthetic Sample(ADASYN) resampling in order to deal with imbalanced dataset.This can also be termed as sequential ensemble based approach.

4.1 PCA

Principal Component Analysis is a dimension reduction based techniques which is used for reducing the dimensions of the dataset by preserving important feature variables without loss of information. In the present research PCA is applied on credit card fraudlent dataset and this PCA will create set of new feature variables from original set of feature variables. These newly obtained feature variables are known as Principal Components(PCs). Principal component analysis is considered to be one of the well known technique for Anamoly detection,PCA looks for correlations between feature variables in the instance of credit card fraudlent transactions, these may be location,time of the transaction ,amount of transaction done — and identifies which combination of values leads to the variability in the result. These combined feature variables enable the construction of a more constrained feature space known as Principal components(PCs).

4.2 ADASYN

In the present research Adaptive Synthetic Sample(ADASYN) is used as resampling technique for imbalanced dataset.The basic idea behind ADASYN is to utilise weightd distribution for various minority class related examples on the basis of their level of learning difficulty. Excess amount of synthetic data is being created for minority class related examples which are considered more easy to learn.

The ADASYN method enhances learning about data distributions in two different ways that is It reduces the biasness caused due to class imbalance.It moves the categorization decision boundary toward the harder cases in an adaptive manner.

4.3 Neural Networks

Topologies or architectures of Neural networks are built by arranging nodes into section of layers and then connecting these layers with changeable weighted based interconnections.Researchers in Neural networks have also included statistical and numerical analytic tools into their networks.Since Neural networks are a nonlinear based mapping connection from input to output region, they may learn from previous cases and conclude the principle of data even when the prospective data principles are unknown. It may also adjust its own behaviour to the new environment as a consequence of the creation of a general capability of evolution from the current situation to newly environment.The Multilayer Perceptron (MLP) is a supervised method that consists of single input layer, one or multiple hidden layers and single output layer, each of which contains neurons and is a feed forward neural network.The architecture of a neural network is determined by number of neurons,hidden layers and interconnections between nodes.In the present research application for detecting credit card fraudulent transactions, the nonlinear neural networks based technique with more than one hidden layers outperforms the statistical techniques in terms of performance.

5 Data Collection and Preprocessing

A dataset is essentially a grouping of linked data. We use a publicly accessible unbalanced dataset in this article. A dataset that is unbalanced has disparities in the dependent variables.

Imbalanced indicates that the distribution of classes is uneven. The dataset that we are using is likewise unbalanced.This dataset provides a history of transactions done by European cardholders. It includes records of 284,807 transactions and the dataset is taken from kaggle dataset repository.The dataset presented in this research work contains feature variables ranging from V1 to V28, which are the major components produced by PCA. The only feature variables that are not changed by PCA are 'Time' and 'Amount.' The seconds between any transaction and the first transaction in the dataset are included in the 'Time' feature variable. The 'Amount' feature variable represents the transaction amount.Class is the target variable which has the value=1 for fradulent transactions and value=0 for Genuine transactions.

5.1 Data Pre-Processing

It involves three common steps which are Data Formatting,Data Cleaning and Data Sampling.

5.1.1 Data Formatting

The dataset that has been selected cannot be in format that allows to work on it. There are chances that data will be in a relational database system and want it in a csv file format.In order to deal with this issue Data formatting step should be carried out before uploading the data for futher processing.

5.1.2 Data Cleaning

The removal of missing data or filling up the field value having NA with mean value is referred to as data cleaning. There might be some data instances that will have incomplete values and do not include the information that is required to solve the problem. These instances have been deleted. Furthermore, various attributes that carry sensitive information or private information has been excluded from the dataset.

5.1.3 Data Sampling

There might be considerably more chosen data available than required. More the number of data longer will be the algorithm running times and higher computational as well as memory needs. Before examining the entire dataset, it has taken smaller number of the selected data, which may be considerably faster for exploring and developing ideas.

5.1.4 Outlier detection

The outlier identification approach, like clustering, examines the distance between each data point, but is also discover particular data and rules that are isolated from the entire data. Outliers are values that are not in the flow of lineaer graph.In the present research decrease outliers in order to have a better trained model. Python's numpy package is used for this.

5.2 Feature Engineering

In the detection of credit-card fraud, a fraud is said to be a property of both the transactions which are fraudulent and the context in which it occurred.Transaction may raise suspicion if the amount of money spent is particularly large and it occurs at a specific sort of merchant at a specific time of day, such fraud detection approach examines transactions in different way than it was previously occured. The customer's previous purchases, as well as subsequent transactions with the merchant with whom he interacted, are not taken into account.

The majority of feature engineering work in credit card based fraud detection use transaction based aggregation techniques similar to those outlined in previous research. The goal is to

describe a spending of amount pattern from account, known as an activty record, by gathering transaction related data over a span of time.Using exploratory data analysis, the researchers discovered various features of transactions that are significant to credit card based fraud detection, like no.of transactions recently performed and overall amount of the transactions.

Financial related systems handle a massive amount of data regarding accounts, clients, cards and transactions performed by the client. In order to find the correct target variable it is necessary to have previous domain knowledge, which substantially contributes to the development of prediction models for detecting credit card based fraudulent transactions.

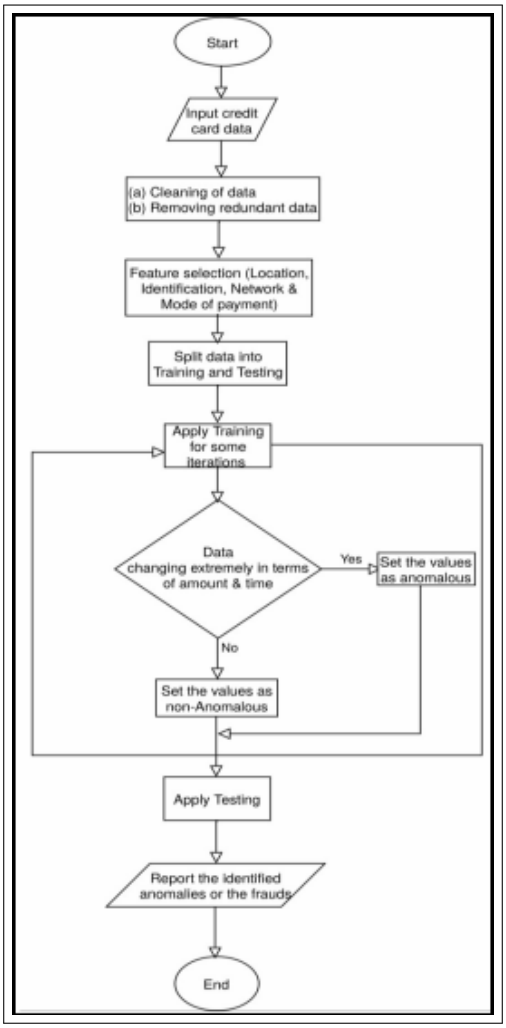


Figure 4: Flow Chart for Credit Crad Fraud detection

Most research appear to concur that the Recency - Frequency - Monetary Value (RFM) paradigm offers the required information for detecting fraud. Recency is the period since the last transaction performed, frequency is the number of transactions as time interval, and monetary value is the total amount of money spent during the particular interval.In the present research study several important feature variables are have been identified which are Inter transaction time gap that is time between previous and current one,Transactions performed per day,week,month,transaction frequency and time of transaction.

In the next stage of implementation process,the selected feature variables are applied to Machine learning models.Class Imbalance is handled with ADASYN approach as the target variable has large number of genuine transactions.

5.3 Neural Network Multi layer classification model with resampling and dimension reductionality technique

In the present research work Neural networks Multi-layer perceptron with different number of hidden layers has been implemented using ADASYN as resampling technique and PCA as dimension reductionality technique.All these project work has been implemented using python libraries in Jupyter notebook.

5.4 Evaluation and Result

In order to evaluate the performance of the credit card fraudulent transaction based classification model,performance metrics namely accuracy,precision,recall and F1-score has been taken into consideration.As this research study deals with imbalanced dataset resampling technique will be applied for the same.Finally,AUC-ROC ranking metrics will be used in order to evaluate the efficitiveness of the model in classifying fraudulent and genuine transactions.

Table 1: Classification Models and Hyperparameters

Classification Models	**Hyper-parameters**
AdaBoost	n_estimators:300,random state:789
Random Forest	max_depth: 5, max_features: 20, min_samples_leaf: 100, min_samples_split: 100, n_estimators: 300
Decision Tree	max_depth=5, min_samples_leaf=100, min_samples_split=100, random_state=100
XGBoost	n_estimators=500,max_depth=18

6 Evaluation

A complete series of experiments has been carried out in present research study so as to evaluate the proposed strategy.The main goal of these experiments was to compare different classification models and their performance with PCA based Neural networks Multi-layer perceptron model having different number of hidden layers on credit card fraud dataset. The performance Neural Network based model is evaluated using Accuracy,Precision,Recall, F1 Score.

6.1 Experiment 1

Machado et al. (2019) proposed Standard Classification models like Random Forest,XGBoost and the Adaboost methodologies to predict credit card fraudulent transactions that occur in financial sector using credit card transactions of all clients of the Spanish bank BBVA. This research replicated the implementation stages from state of the art paper and extended the research to Credit card fraudulent transaction dataset of European Card holders. Replicating the state-of-the-art paper has been very helpful in optimizing the Imbalanced data for Neural networks Multi-layer perceptron with multiple hidden layers based PCA model algorithms.

6.2 Experiment 2

Neural networks with multiple hidden layers based PCA model has been implemented as the part of present research work to predict credit card fraudulent transacation using European credit cardholder dataset.The hyperparameters that has been obtained from Experiment 1 has

been optimised and analysed so as to generate model with much better accuracy rate.For all Neural networks with multiple hidden layers based PCA model, balacing of data has been done using ADASYN technique with lower error value and highest fraud prediction rate are determined.Furthermore, for Neural Network with multiple hidden layers, fixed number of iterations have been provided, and after execution behavior of the model is evaluated.However, choosing between 40 and 100 iterations is appropriate because it reduces and stabilises the error rate.

6.3 Discussion

Table 2: Comparison of Standard Classification Model

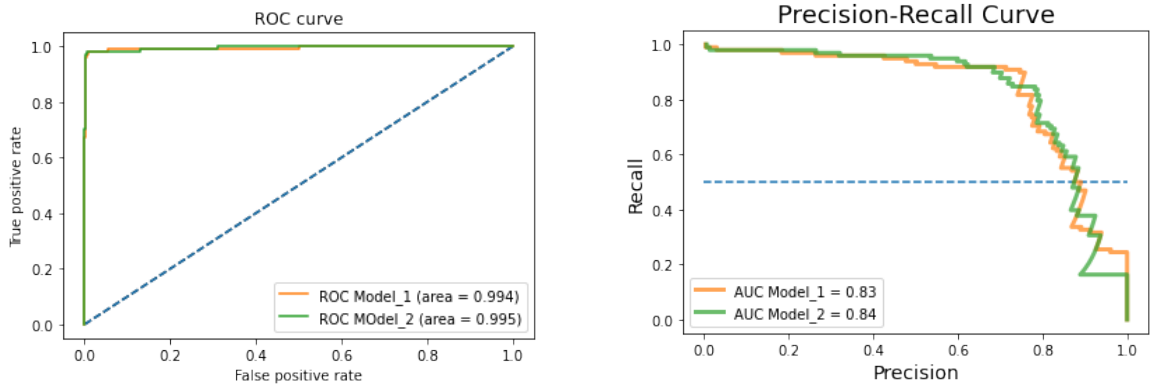
Machine Learning Classification Models	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Logistic Regression	98.93%	87%	63.52%	76%
Naive Bayes	96.23%	84.2%	65.48%	72.56%
XGBoost	97.45%	90.28%	79.31%	84.68%
ADABoost	94.58%	81.30%	57.28%	69.43%

Table 3: Comparison of Neural Network MLP based and Standard Classification Model

Machine Learning Classification Models	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Neural Network MLP with one hidden layer	99.82%	92.85%	50.27%	65.23%
Neural Network MLP with two hidden layer	99.83%	95.91%	50.81%	66.43%
Random Forest	99.91%	68.36%	99.96%	78.23%
Decision Tree	99.89%	61.22%	99.95%	66.29%

It is evident from the Table 3.that the accuracy obtained by using accuracy obtained by using Neural Network (NN) Multi-Layer Perceptron (MLP) with one hidden layer is 99.82% and by using Neural Network(NN) Multi-Layer Perceptron (MLP)with two hidden layers is 99.83%.Also,from the Table 3 it can also be seen that the Precision rate obtained by using NN MLP with one hidden layer is 92.85% and with two hidden layers it is 95.61%,which is the highest precision rate on the credit card fraud detection model as compared to other standard Machine learning classification algorithm.The recall,F1-score percentage obtained on NN MLP with one hidden layer using one hidden layer is 50.27%,65.23% and using 2 hidden layers is 50.81%,66.43%.

Fig 4. shows graphical representation of AUC and ROC curve for both the Models that is the Neural Network MLP with one hidden layer-Model 1 and Neural Network MLP with two hidden layers-Model 2.From the Fig 4 it can be seen that ROC and AUC values for Model 2 is greater than that of Model 1.Hence,Also from the finding mentioned above it can be said that Neural Network MLP with two hidden layers better than Neural Network MLP with one hidden layer.



(a) ROC Curve

(b) AUC Curve

Figure 5: ROC and AUC Curve for NN MLP with 2 hidden layers

7 Conclusion and Future Work

Credit cards are the most popular form of payment. And the number of scammers is growing by the day. A good fraud detection system should be able to correctly and rapidly identify fraudulent activity.Present research study proposed a Neural Network Multi layer classification model with resampling and dimensionality reduction technique approach to accurately detect and classify credit card transactions as Fraudulent or genuine. In this research study,Neural Network MLP classifier model with different hidden layers has been implemented and compared with different classification model.In this research accurate model has been implemented successfully that identifies and classifies the fraud transaction with the genuine one.As a part of the Future Work in order to detect and prevent credit card fraudulent activities combination of various classification models can be taken into consideration which in turn reduces the financial loss.

References

Agrawal, A., Kumar, S. and Mishra, A. K. (2015a). Implementation of novel approach for credit card fraud detection, *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1–4.

Agrawal, A., Kumar, S. and Mishra, A. K. (2015b). Implementation of novel approach for credit card fraud detection, *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1–4.

Benchaji, I., Douzi, S. and ElOuahidi, B. (2018). Using genetic algorithm to improve classification of imbalanced datasets for credit card fraud detection, *2018 2nd Cyber Security in Networking Conference (CSNet)*, pp. 1–5.

Benson Edwin Raj, S. and Annie Portia, A. (2011a). Analysis on credit card fraud detection methods, *2011 International Conference on Computer, Communication and Electrical Technology (ICCCET)*, pp. 152–156.

Benson Edwin Raj, S. and Annie Portia, A. (2011b). Analysis on credit card fraud detection methods, *2011 International Conference on Computer, Communication and Electrical Technology (ICCCET)*, pp. 152–156.

Guo, T. and Li, G.-Y. (2008). Neural data mining for credit card fraud detection, *2008 International Conference on Machine Learning and Cybernetics*, Vol. 7, pp. 3630–3634.

Kazemi, Z. and Zarrabi, H. (2017). Using deep networks for fraud detection in the credit card transactions, *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pp. 0630–0633.

Mareeswari, V. and Gunasekaran, G. (2016). Prevention of credit card fraud detection based on hsvm, *2016 International Conference on Information Communication and Embedded Systems (ICICES)*, pp. 1–4.

Modi, K. and Dayma, R. (2017). Review on fraud detection methods in credit card transactions, *2017 International Conference on Intelligent Computing and Control (I2C2)*, pp. 1–5.

Nadim, A. H., Sayem, I. M., Mutsuddy, A. and Chowdhury, M. S. (2019). Analysis of machine learning techniques for credit card fraud detection, *2019 International Conference on Machine Learning and Data Engineering (iCMLDE)*, pp. 42–47.

Rai, A. K. and Dwivedi, R. K. (2020). Fraud detection in credit card data using unsupervised machine learning based scheme, *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 421–426.

Roy, P., Rao, P., Gajre, J., Katake, K., Jagtap, A. and Gajmal, Y. (2021). Comprehensive analysis for fraud detection of credit card through machine learning, *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 765–769.

Shen, A., Tong, R. and Deng, Y. (2007). Application of classification models on credit card fraud detection, *2007 International Conference on Service Systems and Service Management*, pp. 1–4.

Tanouz, D., Subramanian, R. R., Eswar, D., Reddy, G. V. P., Kumar, A. R. and Praneeth, C. V. N. M. (2021). Credit card fraud detection using machine learning, *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 967–972.

Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S. and Kuruwitaarachchi, N. (2019). Real-time credit card fraud detection using machine learning, *2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pp. 488–493.

Wang, C., Wang, Y., Ye, Z., Yan, L., Cai, W. and Pan, S. (2018). Credit card fraud detection based on whale algorithm optimized bp neural network, *2018 13th International Conference on Computer Science Education (ICCSE)*, pp. 1–4.

Yu, X., Li, X., Dong, Y. and Zheng, R. (2020). A deep neural network algorithm for detecting credit card fraud, *2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pp. 181–183.

Zhang, Y., You, F. and Liu, H. (2009). Behavior-based credit card fraud detecting model, *2009 Fifth International Joint Conference on INC, IMS and IDC*, pp. 855–858.

Zhang, Z. and Huang, S. (2020). Credit card fraud detection via deep learning method using data balance tools, *2020 International Conference on Computer Science and Management Technology (ICCSMT)*, pp. 133–137.