Graphical user interface, text

Description automatically generated

**Master of Science Project**

Credit Card Fraud Detection via Machine Learning Techniques

THIS RESEARCH PROJECT IS SUBMITTED IN PARTIAL FULFILMENT OF THE DEGREE OF MASTER OF SCIENCE IN DATA ANALYTICS AT THE TECHNOLOGICAL UNIVERSITY OF THE SHANNON: MIDLANDS MIDWEST UNDER THE DEPARTMENT OF BUSINESS & FINANCIAL SERVICES

**AUTHOR**

Vahdet Sufyan Taskin

**SUPERVISOR**

Dr. David Leonard

**Table of Content**

[*Abstract* 1](#_Toc113102666)

[Acknowledgements 3](#_Toc113102667)

[List of Tables 4](#_Toc113102668)

[List of Figures 5](#_Toc113102669)

[List of Abbreviatons 7](#_Toc113102670)

[1 Introduction 9](#_Toc113102671)

[1.1 Fraud Detection 10](#_Toc113102672)

[1.2 Literature Review 11](#_Toc113102673)

[1.3 Machine Learning Techniques 15](#_Toc113102674)

[1.3.1 Supervised Learning (SL) 16](#_Toc113102675)

[1.3.2 Unsupervised Learning (UL) 17](#_Toc113102676)

[1.3.3 Reinforcement Learning (RL) 18](#_Toc113102677)

[1.4 Taxonomy of Fraud Detection Techniques 19](#_Toc113102678)

[1.4.1 Evolutionary Algorithm (EA) 20](#_Toc113102679)

[1.4.1.1 Genetic Algorithm (GA) 20](#_Toc113102680)

[1.4.2 Data Mining (DM) 21](#_Toc113102681)

[1.4.2.1 Classification 22](#_Toc113102682)

[1.4.2.1.1 Random Forest (RF) 22](#_Toc113102683)

[1.4.2.1.2 Decision Tree (DT) 23](#_Toc113102684)

[1.4.2.1.3 Support Vector Machines (SVM) 24](#_Toc113102685)

[1.4.2.1.4 Naïve Bayes (NB) 25](#_Toc113102686)

[1.4.2.1.5 Bayesian Belief Network (BBN) 25](#_Toc113102687)

[1.4.2.1.6 K-Nearest Neighbors (KNN) 27](#_Toc113102688)

[1.4.3 Prediction 27](#_Toc113102689)

[1.4.3.1 Deep Learning (DL) 27](#_Toc113102690)

[1.4.3.1.1 Recurrent Neural Networks (RNN) 29](#_Toc113102691)

[1.4.3.1.2 Convolutional Neural Network (CNN) 29](#_Toc113102692)

[1.4.3.1.3 Artificial Neural Network (ANN) 30](#_Toc113102693)

[1.4.4 Clustering 31](#_Toc113102694)

[1.4.4.1 Fuzzy Clustering (FC) 32](#_Toc113102695)

[1.4.5 Regression 32](#_Toc113102696)

[1.4.6 Logistic Regression (LR) 32](#_Toc113102697)

[1.4.6.1 Linear Regression 33](#_Toc113102698)

[1.4.6.1.1 Gradient Boosting 33](#_Toc113102699)

[1.4.6.1.2 Oversampling 36](#_Toc113102700)

[1.4.6.1.3 Synthetic Minority Oversampling TEchnique (SMOTE) 36](#_Toc113102701)

[1.4.6.1.4 Policy of SMOTE 37](#_Toc113102702)

[2 Research Objectives and Approach 40](#_Toc113102703)

[2.1 Research Methodology 40](#_Toc113102704)

[2.2 Measures of Performance 42](#_Toc113102705)

[2.2.1 Confusion matrix 42](#_Toc113102706)

[2.2.2 Accuracy 44](#_Toc113102707)

[2.2.3 Precision 44](#_Toc113102708)

[2.2.4 Recall 44](#_Toc113102709)

[2.2.5 F-Measure 45](#_Toc113102710)

[2.2.6 ROC (Receiver Operating Characteristic Curve) area 46](#_Toc113102711)

[2.3 AUC (Area under the ROC Curve). 46](#_Toc113102712)

[2.4 Tools 46](#_Toc113102713)

[3 Current Work and Preliminary Results 47](#_Toc113102714)

[3.1 Information about the dataset 47](#_Toc113102715)

[3.2 Exploratory Data Analysis 47](#_Toc113102716)

[3.2.1 Robust scaling technique 49](#_Toc113102717)

[3.2.2 Principal Component Analysis (PCA) 51](#_Toc113102718)

[3.2.3 Time Feature 51](#_Toc113102719)

[3.2.4 Amount Feature 53](#_Toc113102720)

[3.2.5 Class Feature 53](#_Toc113102721)

[3.2.6 V-features 55](#_Toc113102722)

[3.2.7 Correlation Matrices 58](#_Toc113102723)

[3.3 Comparison of models with scale (No Smote) in OrIginal DataSET 59](#_Toc113102724)

[3.3.1 Threshold optimization approach 64](#_Toc113102725)

[3.4 SMOTE optimization approach 69](#_Toc113102726)

[4 Conclusions 77](#_Toc113102727)

[REFERENCES 79](#_Toc113102728)

# *Abstract*

Credit cards are one of the products that have entered our daily life rapidly with the advancement of technology. These cards have become a form of payment that is frequently used by people in a very short time with the help of its strong infrastructure.

Nowadays, people do not prefer to carry cash. Furthermore; the increasing demand for e-commerce and the demand of people to make their payments quickly and easily over the internet increases usage of credit cards. Increasing use of credit cards attracts forgers' attention and that causes an increase in fraud cases. As a result of this situation, banks and people suffer serious loss.

Detection for fraud of credit card is a serious issue for companies and financial corporations. These institutions carry out various research for detecting and prevent credit card fraud.

Studies are conducted with the deep neural network method to detect fraudulent transactions in the literature. These operations cost much; however, they can work well when used on larger datasets. Actually, better results can be achieved by using fewer resources instead of Deep Learning (DL) methods.

The goal of this study is to reveal performance of Decision Tree (DT) along with Random Forest (RF) algorithms, which are Machine Learning (ML) algorithms, by using appropriate pre-processing techniques. Performances of algorithms; will be compared for accuracy, recall, and precision. While applications with under sampling are common in the literature, this study was conducted using oversampling as a different approach.

**Keywords**

Fraud Detection, Machine Learning, Python, Random Forest, Decision Tree.

# Acknowledgements

I would first like to thank my thesis advisor David Leonard

I would also like to thank Matthew Horrigan, Eric McNamara, Paul OKeeffe

Finally, I would like to thank my family for always being there for me

This accomplishment would not have been possible without them.

# List of Tables

[Table 1‑1 - Table of Literature Review 18](#_Toc112239276)

[Table 3‑1: Description of Dataset and Attributes 54](#_Toc112239277)

[Table 3‑2: Credit Card Fraud Dataset 57](#_Toc112239278)

[Table 3‑3: Amount Details of Fraudulent Transaction 68](#_Toc112239279)

[Table 3‑4: Prediction Results of RF and DT on Testing Data for Original Data 74](#_Toc112239280)

[Table 3‑5: Prediction Results of RF and DT on Testing Data for Threshold 81](#_Toc112239281)

[Table 3‑6: Prediction Results of RF and DT on Testing Data for SMOTE 88](#_Toc112239282)

[Table 3‑7: Prediction Results of RF and DT 89](#_Toc112239283)

# List of Figures

[Figure 1‑1: Types of machine learning 21](#_Toc112239295)

[Figure 1‑2: Taxonomy of fraud detection techniques 24](#_Toc112239296)

[Figure 1‑3: Diagram of a decision tree 29](#_Toc112239297)

[Figure 1‑4: Propagation in a tree network 31](#_Toc112239298)

[Figure 1‑5: The relationship between AI, ML and DL 33](#_Toc112239299)

[Figure 1‑6: ANN’s workflow diagram algorithm 36](#_Toc112239300)

[Figure 1‑7: SMOTE 43](#_Toc112239301)

[Figure 3‑1: Confusion matrix 49](#_Toc112239302)

[Figure 3‑2: Amount Distribution for Fraud and Genuine transactions 57](#_Toc112239303)

[Figure 3‑3: Variation with transaction amount. 58](#_Toc112239304)

[Figure 3‑4: Distribution of Time Feature for Fraud Transaction 60](#_Toc112239305)

[Figure 3‑5: Distribution of Time Feature for non-Fraud Transaction 60](#_Toc112239306)

[Figure 3‑6: Distribution of Fraud & Normal Transaction 61](#_Toc112239307)

[Figure 3‑7: Number of Non-Fraudulent and Fraudulent Transactions 62](#_Toc112239308)

[Figure 3‑8: Frequency of per Feature in the dataset 64](#_Toc112239309)

[Figure 3‑9: Features Density plot 65](#_Toc112239310)

[Figure 3‑10: Correlation Matrices 67](#_Toc112239311)

[Figure 3‑12: Confusion Matrix of RF – Original Dataset 69](#_Toc112239312)

[Figure 3‑13: Confusion Matrix of DT – Original Data 69](#_Toc112239313)

[Figure 3‑14: ROC curve and AUC value for RF 70](#_Toc112239314)

[Figure 3‑15: ROC curve and AUC value for DT 71](#_Toc112239315)

[Figure 3‑16: Precision-Recall Curve For RF 72](#_Toc112239316)

[Figure 3‑17: Precision-Recall Curve For DT 72](#_Toc112239317)

[Figure 3‑11: Features That Affect Fraud 74](#_Toc112239318)

[Figure 3‑18: Confusion Matrix of RF – Threshold Data 75](#_Toc112239319)

[Figure 3‑19: Confusion Matrix of DT – Threshold Data 76](#_Toc112239320)

[Figure 3‑20: ROC curve and AUC value for RF (ROC AUC Score:0.9997) 76](#_Toc112239321)

[Figure 3‑21: ROC curve and AUC value for DT 77](#_Toc112239322)

[Figure 3‑22: Precision-Recall Curve For RF 78](#_Toc112239323)

[Figure 3‑23:Precision-Recall Curve For DT 79](#_Toc112239324)

[Figure 3‑24: After Apply Over-Sampling Approach 81](#_Toc112239325)

[Figure 3‑25: Confusion Matrix of RF – With SMOTE 82](#_Toc112239326)

[Figure 3‑26: Confusion Matrix of DT – With SMOTE 82](#_Toc112239327)

[Figure 3‑27: ROC curve and AUC value for RF 83](#_Toc112239328)

[Figure 3‑28: ROC curve and AUC value for DT 84](#_Toc112239329)

[Figure 3‑29: Precision-Recall Curve For RF 85](#_Toc112239330)

[Figure 3‑30: Precision-Recall Curve For DT 86](#_Toc112239331)

[Figure 3‑31: Prediction results of RF and DT on testing data 89](#_Toc112239332)

# List of Abbreviatons

|  |  |
| --- | --- |
| ACC | Accuracy |
| AI | Artificial Intelligence |
| ANN | Artificial Neural Network |
| BBN | Bayesian Belief Network |
| CNN | Convolutional Neural Network |
| DM | Data Mining |
| DNN | Deep Neural Network |
| DT | Decision Tree |
| DL | Deep Learning |
| EA | Evolutionary Algorithm |
| FC | Fuzzy Clustering |
| FN | False Negatives |
| FP | False Positives |
| GA | Genetic Algorithm |
| KNN | K-Nearest Neighbourhood |
| LR | Logistic Regression |
| NB | Naïve Bayes |
| NBTree | Naive Bayes Tree |
| OLightGBM | Light Gradient Boosting Machine |
| RF | Random Forest |
| RNN | Recurrent Neural Networks |
| RF | Reinforcement Learning (RL) |
| SMOTE | Synthetic Minority Oversampling TEchnique |
| SL | Supervised Learning |
| SVM | Support Vector Machines |
| TN | True Negative |
| TP | True Positive |
| UL | Unsupervised Learning |

# Introduction

Financial institutions, companies and people experience huge loss due to fraudulent credit cards (Adewumi and Akinyelu, 2017). In order to overcome this problem, most companies gather their employees to detect credit card fraud and work to eliminate this problem. One of the most effective methods of preventing such frauds is to take action for the transaction that is thought to be fraudulent. Thanks to the artificial intelligence software developed today, if it is considered that the person spends on unusual (contrary to her past) shopping items, it gives a suspicious transaction warning to the system.

A great deal of models and approaches are studied in order to prevent and detection of credit card fraud. Financial institutions and organizations are constantly working to improve the security of their systems continuously, effectively and automatically. The effective use of these systems is one of the uttermost considerable factors in detecting credit card fraud. Therefore, development of fraud detection has become a necessity in securing legal transactions. Studies carried out within this framework are also aimed at detecting illegal movements in the financial sector (Özmen and Özcan, 2019).

## Fraud Detection

Even though the considerable contribution of academics who have done similar studies in discovering the best methods in order to decrease frauds through utilizing Artificial Intelligence (AI) or DM techniques. Incidents of fraud is also increasing in order to steal money along with boosting use of online payment via card (Hazım, 2018).

Card owners are allowed for credit card activities by using the parameters like numbers on credit card, signatures, expiry date, name of card holder, etc.

The term “Credit Card Fraud” refers to illicit usage of card info without card holders' permission; therefore, a criminal act of deception. Detection of credit card fraud is reasonably non-public. The most common fraud detection ways are rule-induction techniques, Logistic Regression (LR), DT, Support Vector Machines (SVM), Artificial Neural Network (ANN) and metaheuristics methods such as k-Means Clustering, GA and K-Nearest Neighbour (KNN) algorithms (Raj and Portia, 2011). Fraud is a type of human attitude which is associated with stealing, misrepresenting, misunderstanding, cheating, etc. Companies occasionally cope with millions of exterior parties, it is cost-prohibitive to control the most of the exterior parties' act and identification manually. Undoubtedly, investigation of each suspicious transaction cost a direct overhead for companies. The investigation isn't worthy although that is sceptical if the value of transaction is more little than overhead cost (Chaudhary, Yadav and Mallick 2012).

## Literature Review

Woyemi et al., (2017) researched the performance of Naïve Bayes (NB), K-Nearest Neighbourhood (KNN) and Logistic Regression (LR) on extremely unbalanced data of credit card fraud in this study. Dataset of credit card operations is obtained from card owners in Europe including 284,807 operations. A hybrid technique for oversampling and under-sampling is applied on the unbalanced data. Researchers carried out three technics on raw and pre-processed data. Accomplishment of these technics is assessed based upon accuracy, sensitivity, specificity, precision, Matthews correlation coefficient as well balanced classification ratio. Comparative results indicate KNN implements better than NB and LR technics.

Zeager et al. (2017) aimed to incorporate data on fraud motivations into an adaptive fraud detection system. In the study, the most successful strategies of fraudsters were modeled and fraudulent systems were determined. They used this system, along with a game-theoretic enemy learning approach, as a pre-warning system for better classification for future fraudulent transactions.

In this study, HAZIM (2018) analysed actual credit card transactions among European cardholders; furthermore, based them upon four data mining techniques; NB, SVM, KNN and RF. The outcome of research indicates the highest correctness for the NB, SVM, KNN as well RF classifiers in order of 97.46%, 95.04%, 97.55% and 97.7%. When outcomes are compared, the RF algorithm achieves better than NB, SVM and KNN.

Özmen and Özcan (2019) conducted a classification study to identify illegal behaviors (fraud activities) in the financial sector. They applied the CART algorithm in their studies and then obtained GA-CART by optimizing the CART algorithm with the GA method. As a result, 64.28% accuracy was obtained with the CART algorithm, while 87.95% accuracy was obtained with the GA-CART algorithm.

Safa and Ganga (2019) analysed the performance on credit card fraud data usage Logistic Regression, Pure Bayesian Algorithm Together With K-Nearest Neighbor Algorithms. These techniques’ performances were evaluated according to the ratio of balanced classification accuracy and time duration. When the results were analyzed comparatively, they determined that logistic regression showed higher performance than pure Naïve Bayesian and KNN methods. These results show that optimal accuracies for the KNN, Pure Bayesian, and Logistic Regression (LR) classifiers are 54.86%, 83.0%, and 97.69%, respectively.

They used data mining to develop a model that could discover hidden patterns and predict the probability of fraud based on these discovered patterns. Kamusweke, et al., (2019). They advanced their work in two stages; First, they identified hidden patterns and trends in the data, and second, they built a model that predicted fraud based on the patterns discovered.

Taha and Malebary (2020) offer a smart way for detecting fraud in credit card actions put into service improved Light Gradient Boosting Machine (OLightGBM) at their article. The Bayesian-based hyperparameter method is cleverly jointed to adjust parameters of the LightGBM in the suggested way. Tests are applied by utilizing two of actual credit card operation dataset legitimate operations and deceitful operations in order to showed the success of the suggested OLightGBM in detection of fraud in credit card operations. The suggested method surpassed other ways; furthermore, they found that the approach reached top performance in respect of accuracy (98.40%), F1-score (56.95%), Precision (97.34%) as well Area Under receiver operating Characteristic curve (AUC) (92.88%) based upon a crosscheck with other ways using the two data sets.

Kirelli et al., (2020) analyzed the order data of an online commerce site and identified possible suspicious transactions. In this study, whole order data were analysed, filtered and top variable selection algorithms defined on account of the classification process. Afterwards, they applied their classification algorithms and successfully identified 92% of suspicious orders. RBF Network, KNN, Naive Bayes Tree (NBTree), NB and J48 DT were used as data mining methods in the study.

Keskenler et al., (2021) developed a heuristic algorithm called Majority Vote Decision Making System (MOCS) in their study. Thanks to the algorithm they have developed, it has been observed that the performances accomplished in previous works was surpassed This new method, which aims at financial security, showed an accuracy rate of 99.93% with an accuracy rate of 95.60% in the tests they carried out. Moreover, it showed that it was also possible to categorize process in the dataset kind of "fake" either " lawful" with a ROC AUC value of 80.0%.

Asha and Kumar (2021) have suggested a new way for detection of credit card fraud actions which is based upon DL in this research. Firstly, it is compared with ML algorithms like KNN, SVM, etc.. Eventually, the neural network is used as well as the model is trained by force that would fit well to the model in order to detect credit card fraud operations. ANN supplies accuracy of the model is almost to 100% that is the finest suited for detection of credit card fraud at this model. ANN, allows more accuracy than unsupervised learning algorithms. Normalization, data pre-processing and under-sampling conducted to handle troubles encountered through applying an unbalanced dataset in this research.

Alharbi et al., (2022), they proposed an approach which is another way for credit card fraud detection. They used a dataset of Kaggle to improve a DL-based way in order to resolve text data issue in this study. A new text2IMG changing technic is suggested that creates little images. Those are provided for a CNN structure along with class weights handling reverse frequency approach in order to solve the class unbalance problem. ML and DL are ways that were applied to confirm validity as well robustness of recommended model. The accuracy of 99.87% was accomplished by Coarse-KNN utilizing deep features of recommended CNN. This new method that uses text2IMG conversion was suggested herein, and it accomplished promising detection of credit card fraud outcomes.

Table 1‑1 - Table of Literature Review

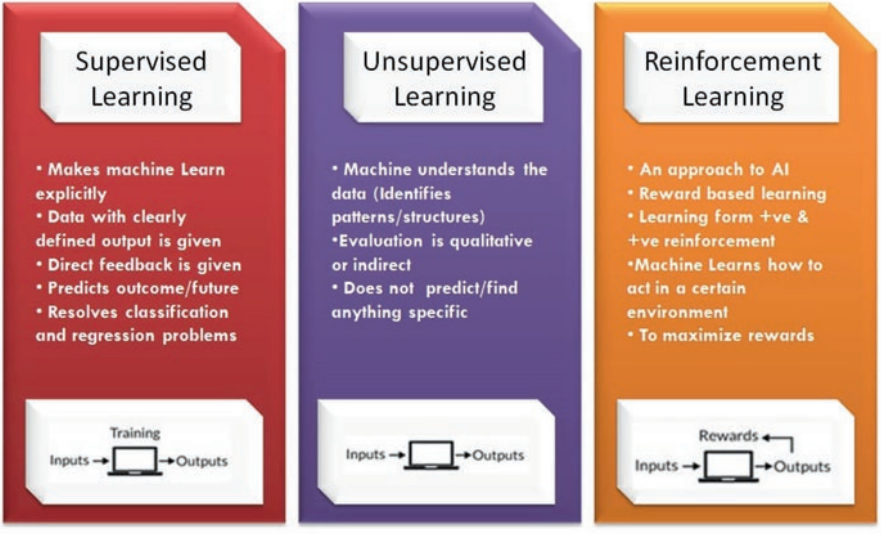
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Research | Fraud Detected | Year | Method Investigated | Accuracy |
| Woyemi et al. | Credit card fraud detection using machine learning techniques: A comparative analysis | 2017 | NB  KNN  LR | 97.92%,  97.69%  54.86% |
| Zeager et al. | Adversarial Learning in Credit Card Fraud Detection | 2017 | AUC - Adversarial Learning  AUC - Classifier Stays The Same | 81.26%  78.69% |
| Hazım | Four classification methods Naïve Bayesian, support vector machine, K-nearest neighbors and random forest are tested for credit card fraud detection | 2018 | NB  SVM  KNN  RF | 97.46%  95.04%  97.55%  97.70% |
| Özmen and Özcan | Hybrid Classification and Regression Tree Application On Fraud Detection In Finance Sector | 2019 | CART  GA-CART | 64.28%  87.95% |
| Safa and Ganga | Credit Card Fraud Detection Using Machine Learning | 2019 | NB  LR  KNN | 83.00%,  97.69%  54.86% |
| Kamusweke et al. | A Data Mining Model for Predicting and Forecasting Fraud in Banks | 2019 | DM-Arima | 86.10% |
| Kirelli et al. | Detection of Credit Card Fraud in E-Commerce Using Data Mining | 2020 | NB  RBF  KNN  NBTree  J48 | 92.76%  93.40%  94.66%  94.60%  87.70% |
| Taha and Malebary | An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine | 2020 | RF  ANN  LightGBM | 95.50%  92.86%  98.40% |
| Keskenler et al. | Detection of Credit Card Fraud Using Artificial Intelligence Supported ÇOKS Method | 2021 | DT  KNN  NB  ÇOKS | 99.92%  99.83%  99.30%  99.930% |
| Asha and Suresh Kumar | Credit card fraud detection using artificial neural network | 2021 | SVM  KNN  ANN | 93.49%  99.82%  99.92% |
| Alharbi et al. | A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach | 2022 | Fine-KNN  Medium-KNN  Coarse-KNN  LP-Boost Ensemble  Bagged-Boost Ensemble  Subspace Ensemble | 99.81%  99.87%  99.87%  90.75%  99.86%  99.77% |

It is noticed that most of researches carried out in the area of detecting credit card fraud benefit from ML algorithms. When the studies are examined, the common point in the methods applied for these studies is that statistical, mathematical and logical methods were used. Furthermore, as that is observed at Table 1-1, more accurate results are obtained in detection of credit card fraud in recent years.

## Machine Learning Techniques

A number of pioneers in the industry have operated to direct us in the right way for the last sixty years. Several different algorithms might be used by anyone in ML. Along with the acquired output is decided which to use. ML is broadly used in computer programme in order to provide an advanced experience for user. Robots can gain skills or learn to comply with the environment along with method of ML in which they are working. They can gain skills like locomotion skills, object positioning and figuring out objects via either automated learning or learning human intervention (Bell 2022).

Figure 1‑1: Types of machine learning



It is significant to figure out each learning method for choosing right way, to fulfil the learning needs and to solve the problem, like demonstrated at Figure 1-1 These three algorithms will be explained down.

### Supervised Learning (SL)

Instances are used by tagging to train the algorithm in supervised learning method. (Silva 2017). The meaning of tagging the instances is that the outcomes are contained within dataset. The aim of this approach is to unveil the relationship network within the data by regarding outcomes of the model. As an instances of the method, if the data set includes info about the typical that can cause the available customers to go away from the bank and whether they go away from the bank for an analysis of the possibility of departing the bank of recently acquired customers of a bank branch, the present customers are tagged. The model considers whether the present customers depart the bank and generates results in accordance with in this method (Şenocak, 2021).

### Unsupervised Learning (UL)

Unsupervised learning looks like pretty harder: the aim of method is letting the ML how to carry out anything which we do not teach it how to act! The unsupervised learning has two main approaches. The initial approach is training the agent through utilizing any kind of award procedure to point out achievement instead of giving obvious categorizations. In general, the aim is to not generate a classification, take decisions will maximize awards. The way fairly might be generalized to actual world in which agents could be awarded for performing particular acts as well penalized for doing another. Frequently, a form of RL could be utilized for UL, in which agent depends its acts on the former punishments and awards without inevitably learning whichever info concerning certain methods which its activities effect world. In some way, whole info is redundant since via learning a function for an award, the agent merely recognizes what to perform without an operating due to the fact unsupervised learning realizes the certain award that waits for accomplishing every action which it could take. The situation might be highly useful when it computed whole probability is plenty time consuming. Contrarily, that might be plenty time-intensive to learn, due to error and trial. However, that type of learning might be strong since that accepts none of pre-discovered classification of samples. For instance; our classifications might not be the best probable in some cases. An interesting instance is that the typical mind about the game of backgammon was changed while a set of software (TD-gammon and neuro-gammon) which learned via UL became better than top chess participants through playing with themself over and over. This software found several rules which surprised backgammon professionals as well it achieved better than backgammon programmes trained on pre-classified samples. The second kind of UL is named “Clustering”. This approach’s aim is discovering similarities within training data in this kind of learning instead of lessening a utility function. The general thought is that clusters found will pair quite fine with an intuitional classification. Moreover, clustering people depends upon demographics may ensue in grouping rich people in a group and poor ones in other group. Furthermore, the method can produce names in order to set these clusters then this may utilize those clusters to set new samples to others. This approach is a data-driven way which can run well while data is sufficient (Ayodele, 2010).

### Reinforcement Learning (RL)

The user might have to enter a great amount of data for several problems. Initially, data is tagged as (X), later others are tagged as (Y) in Reinforcement Machine Learning. This way is between SL and UL. Photo archive is a great example in which items of just some photographs are tagged (e.g., bird, house, person, dog; however, a museum is untagged) (Kamath et al., 2016).

A huge amount of society subjects learning associated issues are directed to this particular group since they could be time consuming or costly. Access might have to be obtained via server administrators in order to label data. Furthermore; untagged data is inexpensive relatively simple to collect and find. UL ways might be operated to achieve ideal matters. They might assist to learn and find out the various reasonable components which happen in the kind of variable (Dhanaraj, Rajkumar and Hariharan, 2020).

## Taxonomy of Fraud Detection Techniques

There are many methods used in detection of fraud techniques. These methods are basically divided into two. The first of these is evolutionary methods. Optimization and search methods are Evolutionary Algorithms which find their inspiration and origin within natural world. Data mining is the second and most common area of these methods. Given the current global economic context, increasing endeavour are being made to both fraud prevent and detect. Because of eternally rising levels of data needs to be analysed, DM techniques and methods are being used frequently (Sabau, 2012).

A taxonomy of fraud detection techniques is given in Figure 1-2. (Tiwari et al. 2021)

Figure 1‑2: Taxonomy of fraud detection techniques

### Evolutionary Algorithm (EA)

Optimization and search methods are Evolutionary Algorithms which find their inspiration and origin within natural world. Evolution theorem of Darwin, underlines the survival of the befitting in a lively habitat, as known as and generally admitted, at least on the ground of evidence collected so far on Earth. EA is a common expression surrounds a few involved methodologies which are based upon natural evolution (Tomassini, 1999).

#### Genetic Algorithm (GA)

GA, introduces natural evolution opinion, in which more vital or fitter members of the population have a greater chance of survival than weaker members of the inhabitants. Robust and healthy members are selected to breed, thereby improving all inhabitants' average fitness. More compatible associates of a particular generation are selected as next generation's parents, and fewer fit members are discarded. The number of parents and children rest on many parameters like fitness and selection functions, some termination criteria and reproduction (Kumar et al., 2010).

### Data Mining (DM)

The operation of detecting significant shapes and patterns out of big data is DM (Linoff and Berry, 2011). The process of DM is useful to take out significant and precious data out of great data heaps, allows foreseeing the risks and opportunities that may occur and enabling decision mechanisms to work according to these risks and opportunities (Şenocak, 2021). Though math and statistics is the basis of data mining, today, progress of data mining activities by using techniques like Computer Vision, Natural Language Processing and ML is becoming common in different sectors.

Data mining techniques are used for many analyses such as measuring customer satisfaction, evaluating the profitability and performance of the company, making future planning of the company by creating predictive models for the future. There are five basic stages in the processing of the data mining technique. This method is mentioned below by Restivo (1999).

* Identification of the Problem
* Arrangement of Data
* Forming and assessment of the Model
* Using the Model
* Model Tracing

#### Classification

The process of partitioning objects according to their similar properties is called classification. For ML, it is necessary to classify the data we have in a method. In particular, classification processes are often preferred in order to detect fraudulent operations in Credit Card Fraud Detection.

##### Random Forest (RF)

RF is an utmost favored ML approaches that can be used to develop prediction methods. It was first introduced in 2001 by Breiman (2001). RF uses binary splits on forecast variables to determine forecast results. The models it uses are a collection of regression and classification trees.

The RF is tree estimators' combination of in which per tree counts on the independently exampled random vector values and has the identical distribution for whole trees in the forest. Computational performance of the RF is relatively good as each tree is built independently of the other. RF is also a classification tree and/or a set of regression which obtain variance among their trees. Therefore, it is easy to use because of just two sources of parameters or randomness that generate trees using the trained data. There is also a subset of data properties that randomly creates per tree as specified (Bhattacharyya, 2011).

Common aspects of RF techniques are creating a random vector Θk on account of Kth tree, unattached of the former random vectors Θ1,...,Θk-1 ;however, beside equal distribution; furthermore, a tree is grown preferring training data set and Θk, ensuing at a classifier h(x, Θk) in which x is an input vector. For example, the random vector k is created as regards in N boxes ensuing from N darts that put down into boxes at random in which N is number of samples in training dataset Θ contains a range of free random numbers from 1 to K in the random divide choice in bagging. Dimensionality along with nature of RF rely on use of Θ in tree creation.

Afterwards a lot of trees are created and they select the utmost favored class. We call selection made with these steps random forests (Breiman 2001).

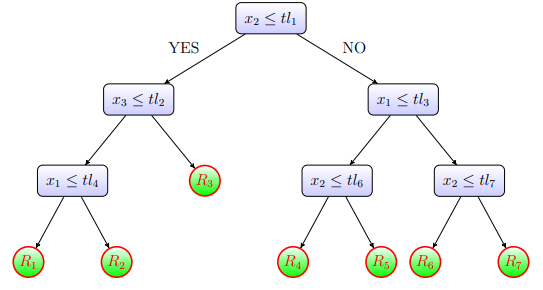
##### Decision Tree (DT)

DT algorithm is one of common tools utilized in machine learning. This method is frequently utilized in DM due to its easy use and successful results.

A DT is a formalism in order to demonstrate such like mappings. One of trees is either a structure complies with a test node bonded to more subtrees or a leaf node tagged with a class. A test node calculates several results which are depends upon properties value of a sample; in which each probable result is related with a subtree. A sample of this way is classified with the beginning of a node of tree root. Result of sample is defined and operation keeps to prefer the suitable subtree if this node is a test. Its tag returns forecasted class of samples when a leaf is encountered at the end (Oliver, 1992).

A test node computes any result based upon properties value of an example, where each likely outcome is associated with a subtree A sample is classified beginning from root node of structure (Quinlan, 1996). Figure 1-3 illustrates a sample of a DT with three explaining variables: X = [x1, x2, x3]. Division of example is accomplished at threshold degree tl1 to tl7, that outcome in areas R1 to R7. At this area that spread of branches, the decision tree establishes a relationship among explanatory variables, X and the dependent variable, Y.

Figure 1‑3: Diagram of a decision tree



A DT might be built out of a group of samples through a divide and conquer tactics. If each the samples pertain the very same class, the tree is a leaf with that class as label. Else, a test is selected that has unlike results for at least two of the samples, that are partitioned with respect to this result. A node at the root of the tree points to the test, and correlation with the subtree for each result in turn is obtained by applying any procedure to the subset of samples with that result (Quinlan, 1996).

##### Support Vector Machines (SVM)

SVM are high-dimensional linear classifiers. Because at higher dimensions, a nonlinear task in the input becomes linear. This, therefore, makes support vector machines very useful for detecting fraud. Furthermore, support vector machines are capable of very high generalisation due to the two most important features of the input data point projection, a core function representing the classification function in the dot product and for trying to find a hyperplane to maximize the division among classes while minimizing overfitting of the training data (Bhattacharyya, 2011).

##### Naïve Bayes (NB)

NB uses Bayes' Theorem in order to calculate possibility of a theorem and define if it is true or false. Furthermore, a classifier prefers to compute depending on possibilities for every likely class as well NB add them to the class with the top conditional possibility value on account of a given X value. Schematically, it is represented in the shape of an inherently directed graph. Graph nodes represent instances and their dependencies are mirrored from the leaded edges. The two variables are free whether no link is edge among them (Suraj, Varsha and Kumar, 2018).

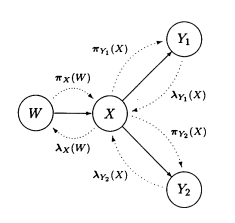
##### Bayesian Belief Network (BBN)

Belief Network, Bayesian Network or Bayesian Belief Network(BBN) are Probabilistic Graphical Model (PGM) which symbolizes conditional reliances among random variables via a Directed Acyclic Graph (DAG) (De Iuliis, 2021).

A BBN is PGM model. BBM occur a set of interconnected nodes, where per node symbolizes a variable in the dependency model and the link arcs symbolize the causal relations among these variables. Any node can get a number of probable values or states. Belief or certainty in per of those situations is defined as follows: It is specified from the belief of every probable state of every node straight linked to it and its relation to all of them. A node's belief in per status is updated when the belief in per status of any straight connected node modifies.

At present, there is a case of various child nodes, like showed at Figure 1-4 (Krieg, 2011) Process made here are integrating the ʎs while they are propagated up the tree along with separating the πs while they spread out down the tree.

Figure 1‑4: Propagation in a tree network



These networks are well suitable for the target recognition issue where the type, class and identity of a target trace is to be determined. These three properties can be modelled by a hypothesis node, that per valid case symbolize a dissimilar hypothesis. Every one of these three track features might be modelled by a hypothesis node, that every one of state symbolize a different hypothesis (Krieg, 2011).

##### K-Nearest Neighbors (KNN)

KNN technique is a SL technic which supplies steady quite upper success that compared to other detection of fraud practices of supervised computerized pattern recognition. Three factors effect the achievement of this technique and these factors are listed below:

* Distance to determine the farthest neighbours
* Some rules for extracting a classification from the k-nearest neighbour
* Number of neighbours to tag new instance

KNN method classifies operations through calculating closest point to that certain operation. At the same time, if this furthest neighbour is classified as rogue, the new action is called rogue. The Euclidean distance is a pretty good selection in order to calculate distances in this procedure. Euclidean distance is a fine option for calculating distances in this method. The advantage of this technique is that it is fast, but it causes error warnings. However, the achievement of KNN can be enhanced by distance metric optimization (Sudha, 2017).

### Prediction

In this section, an introductory note into the base notions of DL is provided.

#### Deep Learning (DL)

Modern DL ensures a strong framework for SLA deep network can symbolize functions of boosting complication by joining more units and more layers inside of a layer. Most duties which are composed of mapping an input vector to an output vector, and which are simple for an individual to carry out quickly, can be achieved through DL, given enough big models and adequately great datasets of tagged training samples (Goodfellow, Bengio and Courville, 2016).

The aim of AI is to provide human intelligence for machines. ML is an approach in order to apply AI by utilizing algorithms to analyse data, learn from data, and take decisions and estimations concerning actual world cases. DL is a innovation in order to recognize ML that lets ML to figure out plenty of apps and enlarges range of AI. Figure 1-5 shows the relation among DL, ML and AI (Ji, Alfarraj and Tolba, 2020).

Figure 1‑5: The relationship between AI, ML and DL

**Gradient Descent.**

Regarding these problems, it was stated that the gradient might increase or decrease exponentially while the input sequence increases' lengthiness. Issues like these might be come through by eliminating the dependence of the series state vectors over the weight matrix. In addition, it is possible to overcome the problem by using techniques such as the introduction of memory cells named as LSTM (Long Short-Term Memory) Network.

##### Recurrent Neural Networks (RNN)

RNNs are networks with loops capable of holding information. They can be opinion of as a multilayer perceptron, per transmitting data to its inheritor. Recurrent neural networks have interior states that are updated after per unit of time. Although iterative neural networks are a powerful and simple model, it is difficult to train them with gradient descent appropriately.

##### Convolutional Neural Network (CNN)

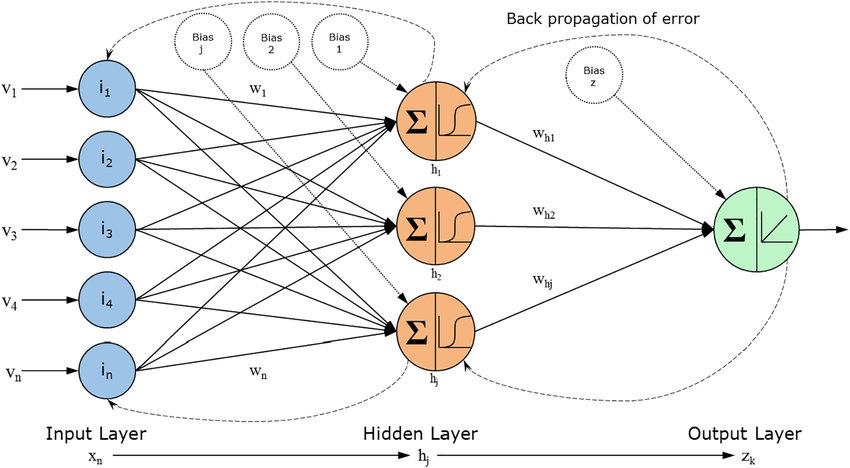
CNN contain neurons might be able to learn values, biases, and weights. Per neuron gets various input vectors and then computes the total weighted on them, crosses valuation through an activation function, and gives a result. Initially, properties choice algorithm was implemented to the original data. The following step includes the property conversion, together with properties were converted to a matrix specifies as input. These networks utilized the soft-max activation function.

CNN contain neurons learnable values, biases and weights. Per neuron gets various input vectors and then computes total weighted on them, passes the value through an activation function as well as gives a result. Initially, the property choosing algorithm is applied to original data. Then, SMOTE (Synthetic Minor Oversampling Technique) helps generate fraudulent synthetic transactions via balancing the dataset in order to overcome the unbalance distribution in data. Next phase is transformation characteristic, and properties are transformed to a matrix given as input. This approach utilizes the soft-max activation function. (Suraj, Varsha and Kumar, 2018)

##### Artificial Neural Network (ANN)

ANN is a kind of neurons’ network that consist of a perception, a linear binary classifier that provides classify input data. Every sensor has four parts:

Figure 1‑6: ANN’s workflow diagram algorithm



(Elbeltagi et al., 2009)

ANN is consisted of three kind of layers:

1. Input layer is primary data for NN,
2. Hidden layers are a secondary layer among independent input layer and dependent output layer in which whole the computation accomplished well,
3. Output layer generates outcomes for given inputs.

Figure 1-6 indicates structure of ANN method that is a class of ML approaches which utilize multiple layers in order to take-out higher-level characteristics out of raw input. “5 circles” are shown as input layers which are marked as vector i. Three circles symbolize hidden layers as well as they symbolize “activation” nodes and generally they are noted as weight (W). Latest circle symbolizes the output layer which demonstrates estimated value of result. (Elbeltagi et al., 2009).

Data which will be trained goes from input to medium latent and afterwards to the output layer called Forward Propagation. Cardholder's normal behaviour initially trains the network. Processes which appear to be fraudulent might be Back-propagated over the network are classified like non-fraudulent and fraudulent works. That technic is pretty effective since a NN does not require to be reprogrammed, and the process speed of process is high (Suraj, Varsha and Kumar, 2018).

### Clustering

Clustering is a usual way for statistical data analysis that is utilized in plenty areas which includes ML, DM, bioinformatics, pattern recognition and image analysis. This is the operation of gathering alike items into various groups or exactly, splitting a data set into subsets; therefore, it splits data in every subset in accordance with some determined distance measure. Clustering section is concerning clustering algorithms (Madhulatha, 2012).

#### Fuzzy Clustering (FC)

A great deal of algorithms for FC count on first estimates of cluster prototypes as well as on suppositions assumed for the amount of subgroups existing in data. Evaluation of cluster validity is depending upon achievement calculates utilizing density criteria as well as hyper volume. (Gath and Geva, 1989).

### Regression

Regression is a statistical instrument for the analysis of relationships among variables. Generally, the researcher looks for figuring out the causativeness of a variable on other—the effect of a cost rise on request otherwise the effect of changes in the cash supply on the inflation ratio. The researcher collects data about main variables of concerned and uses regression to guess the quantitative impact of the causal variables on the variable which they effect in order to explore such issues. Researcher usually determines the “statistical significance” of the estimated relations which is the rank of reliance, true relation is near to the estimated relation (Sykes, 1993).

### Logistic Regression (LR)

It is a suitable technique preferred in predictive in study while depending on variable is binary. This technique can be used because the classification of transactions as fraudulent is a two-pronged variable. This classification shape depends upon possibilities uncovers fraud utilizing the logistic curve. Because logistic curve’s valuation ranges out of 0-1, that might be utilized to explain class membership possibilities.

Data set that feeds the model classified to train the model as well as to test it. After training the model, this model has been tested for any minimal threshold interrupt value to estimate. Essential variables are then choses along with this model has adjusted accordingly. Logistic regression can split the plane preferring a single line and split the data-set points into two parts based on some threshold probabilities. Therefore, outliers are not dealt with effectively. This model uses the natural logarithmic function in order to compute possibility along with demonstration which outcomes are transferred into the specific class.

#### Linear Regression

ML is mostly used in various fields to resolve hard problems which cannot be easily resolved in based upon computer approaches. Linear Regression is one of the basic and furthest common ML algorithms. This is a mathematical way used to achieve predictive analysis. Liner Regression lets mathematical or continuous/real variables projections. Sir Francis Galton earliest proposed the concept of Liner Regression in 1894. This is a mathematical test used for assessing and quantifying the relationship among regarded variables (Maulud, Abdulazeez, 2020).

##### Gradient Boosting

Gradient Boosting is a particular strong approach in order to create prognostic methods. This is a useful gradient algorithm which chooses a function that leads in the way of a low hypothesis or negative gradient again and again so that it can lessen a loss function. Generally, it merges some low learning methods for generating a powerful estimation model.

Various ML apps has a typical duty which is to generate classification model or a nonparametric regression through data While mapping out the model in domain-specific domains, a kind of strategies can be building the model through theory and calibrate parameters of model with respect to the monitored data. Almost all this kind of models do not exist in real life, with regret.

In many cases, even the first expert-guided estimates of possible relation among input variables are unavailable to the searchers. A model can be deficient whether non-parametric ML ways like SVM, NN or some other algorithm at another choice are applied to construct the model straight through data. Those models can be controlled that express data beside expected and aimed variables must be present prior to.

The utmost common way for data-driven modelling can be to construct a predictive-only solid model. An alternative way would be to create a collection of models or a bucket for a specific learning task. We might take into account to construct collection of "strong" models, such as NN that can be more integrated in order to generate a stronger estimation.

Nevertheless, the community approach relies on combining numerous relatively weak basic models to acquire a more robust community estimate. The most significant samples of such ML ensemble technics are neural network ensembles and random forests, which have found many successful applications.

Standard ensemble techniques, such as RF count on plain averaging of the models within community. Empowerment methods’ group is stand on the distinctive, effective community-building tactics. The keynote of Boost is adding fresh models for community in order. The fundamental and a fresh user model can be trained on the whole community's error that learned up to a point for every particular iteration.

The first significant enhancement technics were merely algorithm-driven that formed a comprehensive analysis of their performance challenges as well properties. It has led to some speculation about why these algorithms outperform all other methods or cannot be implemented due to overfitting. Therefore, a gradient descent-based formulation of the augmentation ways is reproduced to link alongside statistical framework. Related models and this formularization of boosting methods are named gradient boosting machines.

Moreover, the framework supplied the essential criteria of formed the methodological basis and model hyperparameters for other gradient boosting model development. This reinforcement learning method repeatedly adjusts current models to supply further accurate prediction for output variable. Fundamental idea at the back of the algorithm is to construct beginner learners; therefore, they are ultimately match up with the loss function’s negative gradient incorporated beside entire community.

This loss functions that put in might be optional; however, in order to receive better perception, whether error function can be classical square-error loss, training process will outcome within sequential error adjusting. Generally, loss function’s selection can be according to the searcher both with the wide range of loss functions acquired until now and the probability of applying another task-specific loss. The technique’s high adaptability provides Gradient Boosting Models extremely personalized for each of data-driven duty. This also gives a great deal of freedom for approach planning. Therefore, choosing optimal loss function is an issue for test as well error. Nonetheless, the amplification algorithms can be allow experimentation with different model designs and relatively simple. In addition, this models have shown notable performance not only in practical applications but also in different data mining challenges and ML (Bhattacharyya, 2011).

##### Oversampling

Training from class-skewed data proceeds in order to be a usual as well as tough issue in SL like ordinary classification methods are planned to hold balanced class distributions. When various plannings present in order to handle the issue, ways that create synthetic data to succeed a balanced class distribution are further multiskilled than changes for classification method. These technics which are named over samplers change training data allows each classifier to be utilized with class-skewed datasets. A great deal of algorithms are suggested for this task; however, most of them are complicated and prone to compose redundant noise. (Last, Douzas and Bacao, 2017).

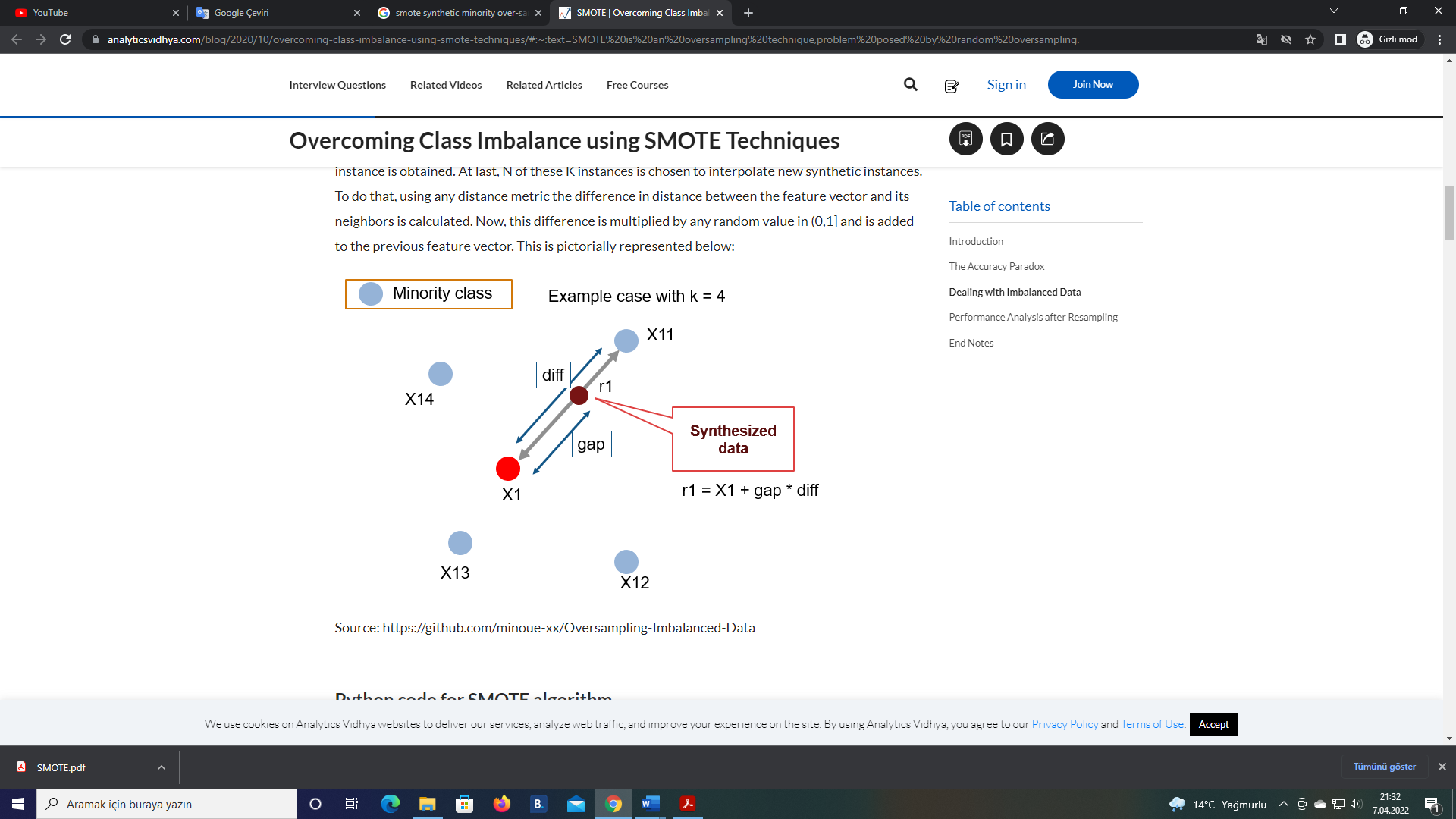
##### Synthetic Minority Oversampling TEchnique (SMOTE)

Initially, SMOTE is used as an oversampling method which is utilised when synthetic instances are obtained on account of minority class. The algorithm that overcomes the overfitting issues due to random oversampling. This is mainly focused on properties area to create new samples beside aid of interpolating the positive samples which stand together (Chawla et al., 2002).

##### Policy of SMOTE

Initially, N is initiated as a total number of observations. Generally, it is chosen as binary class distribution 1:1. However, it could be tuned down according to needs. After this, the iteration will be started by selecting the positive class sample randomly. Afterwards, KNN is obtained for this sample. Lastly, N of those K samples is chosen for interpolating new synthetic examples. For achieving this, to some degree distance metric differentiation in distance among the feature vector as well neighbors of it are found the metric difference is multiplied with a random value (0,1), as well as it is put on to the property mentioned above vector. The procedure is shown at the following Figure 1-7,

Figure 1‑7: SMOTE



Steps of POLICY OF SMOTE may be summarized as below:

Step-1: The k Near Neighbors of each observation existing to the minority class are researched,

Step-2: The distinction between the observation existing in the minority class as well as the observation with k Near Neighbors is catches,

Step-3: At the moment that a random numeral (α) is notified (0,1), this numeral is multiplied by the one found in Step 2,

Step-4: A new synthetical observation is acquired by using Equation 1.

𝑥𝑦𝑒𝑛𝑖 = 𝑥𝑖 + (𝑥𝑗 − 𝑥𝑖) \* α

Step-5: Reiterate steps 1-4 to generate expected replicable observables.

Fraudulent activities lead to significant losses. This motivates researchers to solve the fraud detection problem. In literature, different techniques are suggested and tested previously. In this chapter, some of these methods are revised.

There are usual algorithms like GB, SVM, DT as well RF. According to the literature review, Random Forest has the best performance (Dejan 2019). According to (Malini and Pushpa, 2017) and (Navamani, Phil, and Krishnan, 2018), KNN and other detection approaches that are effective in detection of card fraud. These are helpful for minimising incorrect alert values and rising detection for fraud. In this study (Kazemi and Zarrabi, 2017), a cross check is conducted among several usual algorithms and DL technics. These technics result in an correctness of nearly 80%.

KNN, RF, LR, GB, DT, MLP, SVM, XGBoost (XGB), NB along with stacking classifier (various machine learning classifiers' combination) are used, and the European dataset is used. (Dhankhad, Far, Mohammed, 2018). According to pre-processing results, whole algorithms had achieved with 90% correctness. These stacking classifiers have the top accomplishment with the accuracy of 95% and recall of 95%.

There are many studies conducted for the detection of credit card fraud with neural networks within literature. A NN was tested upon a European dataset. These experiments include of BNN, which is optimized by Whale algorithm (Wang et al., 2018). This NN is made up two input layers, twenty hidden and two output layers.  In accordance with optimization algorithm, they provided extraordinary outputs on 500 test instances. The correctness is 96.40%, and recall is 97.83%.

Neural networks improve results if ensemble techniques are used (Kalaiselvi, Rajalakshmi, Padmavathi, 2018) (Ghobadi and Rohani 2016). In another study, three datasets had used. Finally, the comparison was made among Auto-encoder as well Restricted Boltzmann Machine Learning algorithms (Pumsirirat, Yan, 2018).

# Research Objectives and Approach

The primary aim of the study is to implement some successful ML algorithms for fraud detection. RF and DT algorithms are used in order to accomplish the purpose of the project and estimation outcomes are crosschecked. A Common dataset on credit card actions is utilized at the study. Properties of dataset are changed via PCA techniques because of personal data security matters.

The fraud of credit card means losing of sensitive credit card info and/or lost credit card. A lot of studies are conducted to detect fraudulent operations with DNN in the literature. Nevertheless, DNNs cost much computationally; in addition, they can act preferable on bigger datasets. DNN may have good outcomes. However, the same or yet preferable results might be reached with fewer sources. ML is used in prediction, classification, detection and optimization applications. There are many types of ML algorithms that might be utilized for detection.

## Research Methodology

The aims of scrutinizing the effectiveness of conventional algorithms of DM in coping with administration of data requires of credit card fraud are assessed via utilizing two AI methods. The confirmation of doubtful actions of the cardholder is a significant component of fraud inquiry and cannot be bypassed. Consequently, every approach which clarifies the inquiry preference operation via decreasing the number of redundant calls in such situations is embraced from banks of world.

First, train - test split technique has been used. Train - test split technique is a process of splitting dataset to two subsets. The initial subset entitled “training” is utilized for fitting the approach and latter subset entitled “test” is utilized to evaluate fitted model. In this project, 30% of the dataset allocated for test, and 70% of the scaled data allocated for train of the model. The main idea of using 30% of the dataset for the test of the model is to compare different pre-processing techniques, SMOTE and under-sampling on the imbalanced data.

The study aims to reveal the performance of two algorithms when appropriate pre-processing techniques are used. Under-sampling is common in literature, but oversampling is used as a different approach in this study.

Classifiers must indicate whether the relevant transaction is a fraud in this context. In this thesis, two ML algorithms will be used. These are RF and DT. Their performances are compared regarding precision, recall and accuracy. This dataset is particularly unbalanced. Therefore, SMOTE technics will be utilized in order to overcome oversampling. Selection of properties will be made. The dataset will be divided to two pieces. These are test data and training data. Proposed models will be tested for detection of fraud detection regarding F-Measure, Precision, Recall, Accuracy, ROC Area and AUC**.**

Scaling ought to be applied in order to obtain entire features in a alike setting in numerous ML algorithms. The Robust Scaler method is preferred because it gives better results in data with outliers. The method is sensitive for datasets with outliers, due to the fact it uses IQR (Quarters Aperture) for scaling. Outliers are excluded in the Robust Scaler. The median value is eliminated for later use and the values are fitted to the 1st and 3rd quartile range.

Information about the dataset and tools will be provided in this chapter. Afterwards, the steps of this study will be given.

## Measures of Performance

The measure of performance metrics used in this study are listed below.

* Confusion Matrix
* Accuracy
* Precision
* Recall
* F-measure
* ROC Area
* AUC

### Confusion matrix

The Error Matrix which is also called as "Confusion Matrix" includes "two columns" and “two rows" that symbolizes True Negatives (TN), True Positives (TP), False Negatives (FN) and False Positives (FP) (Maimon and Rokach, 2010).

Initially, each of values obtained in the TP cell is estimated results pairing the present values within dataset. That indicates an estimation which classifies as original while actual value is also authentic within data operations.

Second, the TN is precisely vice versa, an estimated consecutively a real value is both tagged like fraudulent. Third, the FP are accepted like (type two errors) which means that results are classified as fraudulent; actually, they are true. Lastly, estimated value is negative; however, the existing value is positive, i.e., the approach falsely estimated the positive class tags to be negative (Hazım, 2018).

Figure 2‑1: Confusion matrix

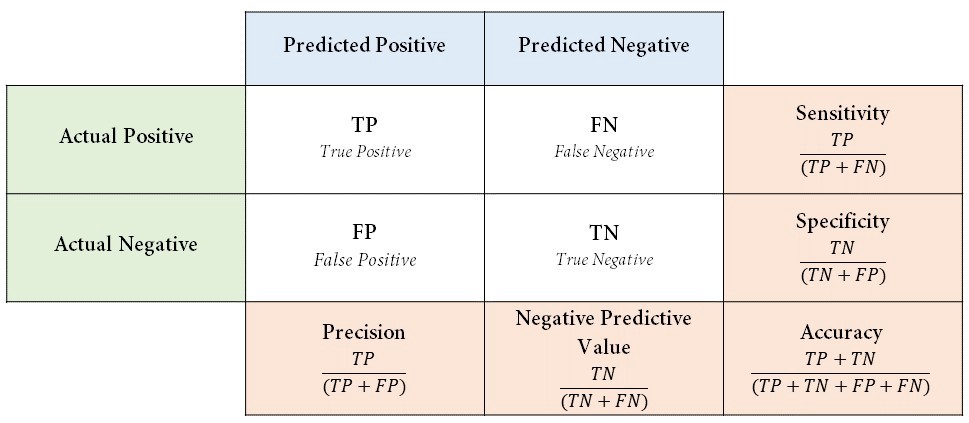


Figure 2-1 provides a visual description of the evaluation accomplishment measurements with confusion matrix. Initially, confusion matrix utilized with under sampled dataset then confusion matrix utilized with entire dataset and it used for comparison of them.

That could be a diagonal matrix in which valuations of primary diagonal could be zero, specifying false estimations, if we had excellently estimated each item. It demonstrates that we have comparatively a small amount of FPs, comparatively a small amount of legitimate transactions were incorrectly flagged in this situation. Although, we could wish to have less FNs in spite of cost for boosting amount of FPs. The trade-off might be suitable since FNs could let fraudulent actions to pick over, whilst FPs can give rise to send a SMS or an e-mail for a customer in order to ask them verifying their card actions.

### Accuracy

Classification DM approaches are suggested as a fitness function to assess per subset which is generated in this measurement.

Formulation of the accuracy measurement is shown below:

### Precision

Precision is a qualified measurement in order to specify, while costs of FP is extreme. For example, e-mail spam detection. The FP signifies that an e-mail which is anti-spam (Actual Negative-AN) is recognized as junk e-mail in email spam detection. The person who has an email account may lose critical e-mails if the precision is low by spam detection system.

Formulation of the precision measurement is shown below:

### Recall

Recall computes the amount of TP of our model catch via tagging it as Positive (Actual Positive-AP). While a high cost related with FN, Recall will be the evaluation metric which is used in order to select best way. For instance, this method is used in fraud detection or patient detection.

The result can be extremely problematic for the bank whether an AP (fraudulent action) is estimated as Predicted Negative-PN (non-fraudulent). Likewise, for patient detection. The cost associated with FN will be extremely high if the illness is infectious if a patient (AP) goes via the test and predicted as not ill (PN). The charge related with FN will be high-cost if the illness is infectious.

Formulation of the recall measurement is shown below:

### F-Measure

F-measure is required when you wish to search for a stability between Recall and Precision. F-measure is a preferable measure if a balance is sought between Precision & Recall and there is an uneven class distribution.

Formulation of the F-Measure is shown below:

### ROC (Receiver Operating Characteristic Curve) area

The ROC is the chart demonstrates accomplishment of the classification approach on overall classification thresholds. The ROC curve plots have two parameters:

True Positive Rate (TPR) is an equivalent for recall as well it is determined below:

FPR is defined like follows:

A ROC curve plots FPR vs TPR for various classification thresholds. Decreasing with classification threshold categorizes more objects as positive, consequently, values of both FP and TP increase.

## AUC (Area under the ROC Curve).

AUC supplies a total for performance’s measure over entire probable classification thresholds. Another approach of explaining AUC is as possibility which approach rates a random positive instance extremely than a random negative instance.

## Tools

* Python - 3.x
* Pandas - 1.4.2
* Numpy - 1.21.5
* Seaborn - 0.11.2
* Matplotlib - 3.5.2
* Scikit-learn - 0.20
* Imbalanced-learn – 0.9.1

# Current Work and Preliminary Results

## Information about the dataset

The credit card fraud dataset contains 284,807 transactions, and 492 of them are fraud. The dataset contains 31 parameters. It is highly imbalanced. It is obtained from Kaggle.. It consists of actions done by European cardholders in September 2013. Link for this dataset: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

## Exploratory Data Analysis

The property "Amount" is represented as transaction amount and that property might be utilized for dependant cost-sensitive learning in this dataset. Properties like V1, V2, … V28 are main parts acquired with PCA. Furthermore; just "Amount" and "Time" are not converted with PCA. The property "Time" consists of seconds which are passed among first action and per action in dataset. The property of "Class" is the reaction variable as well as "Class" receives value "1" in the event of fraud; otherwise, it is "0".

Table 3‑1: Description of Dataset and Attributes

|  |  |  |
| --- | --- | --- |
| Attributes | Variable Type | Variable Description |
| Time | int | Time between each transaction |
| V1 | Num | Attribute variable converted with PCA |
| V2 | Num | Attribute variable converted with PCA |
| . | . | ------- |
| . | . | -------- |
| V28 | Num | Attribute variable converted with PCA |
| Amount | Num | Total transactions |
| Class | İnt | Response feature (0= non-Fraudulent and 1 = Fraudulent) |

Table 3.1 shows that all variables are numerical. No need for changing data types. Dataset has no missing data.

Table 3‑2: Credit Card Fraud Dataset

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Time** | **V1** | **V2** | **V3** | **…** | **V27** | **V28** | **Amount** | **Class** |
| **0** | 0.0 | -1.359.807 | -0.072781 | 2.536.347 | … | 0.133558 | -0.021053 | 149.62 | 0 |
| **1** | 0.0 | 1.191.857 | 0.266151 | 0.166480 | … | -0.008983 | 0.014724 | 252.35 | 0 |
| **2** | 1.0 | -1.358.354 | -1.340.163 | 1.773.209 | … | -0.055353 | -0.059752 | 378.66 | 0 |
| **3** | 1.0 | -0.966272 | -0.185226 | 1.792.993 | … | 0.062723 | 0.061458 | 123.50 | 0 |
| **4** | 2.0 | -1.158.233 | 0.877737 | 1.548.718 | … | 0.219422 | 0.215153 | 69.99 | 0 |
| . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . |
| **284802** | 172786.0 | -11.881.118 | 10.071.785 | -9.834.783 | … | 0.943651 | 0.823731 | 0.77 | 0 |
| **284803** | 172787.0 | -0.732789 | -0.055080 | 2.035.030 | … | 0.068472 | -0.053527 | 24.79 | 0 |
| **284804** | 172788.0 | 1.919.565 | -0.301254 | -3.249.640 | … | 0.004455 | -0.026561 | 67.88 | 0 |
| **284805** | 172788.0 | -0.240440 | 0.530483 | 0.702510 | … | 0.108821 | 0.104533 | 10.00 | 0 |
| **284806** | 172792.0 | -0.533413 | -0.189733 | 0.703337 | … | -0.002415 | 0.013649 | 217.00 | 0 |

### Robust scaling technique

Features such as “Time” and “Amount” are extremely skewed as it can be observed at Figure 3-2 and Figure 3-3. These characteristics requires to be scaled with respect to values in other columns. Certain statistical techniques help us in order to come through skewness. Robust scaling technique is preferred in order to come through skewness in the study.

The Robust Scaler method is preferred because it gives better results in data with outliers. The method is sensitive for datasets with outliers, due to the fact it uses IQR (Quarters Aperture) for scaling. Outliers are excluded in the Robust Scaler. The median value is eliminated for later use and the values are fitted to the 1st and 3rd quartile range.

The interquartile range approach detected 11.2% (31904) outliers. Owing to the losing of a great deal of data on account of the ML approach, taking out outliers from the dataset might not be appropriate.

Figure 3‑1: Amount Distribution for Fraud and Genuine transactions

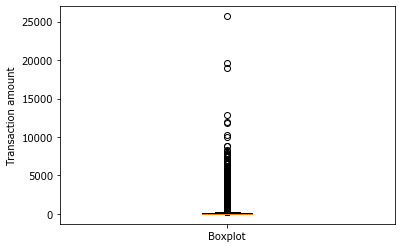
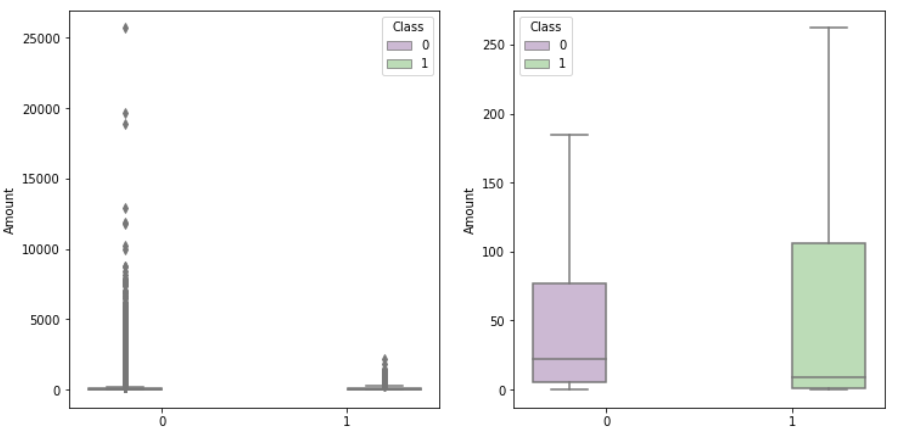


Figure 3‑2: Variation with transaction amount.



### Principal Component Analysis (PCA)

Dataset includes merely numeric input variables that are outcome of a PCA conversion. Moreover, the dataset does not give original properties as well as more background info about the data because of confidentiality issues.

PCA method is multivariate which analyzes data table where observings are defined through some inter-correlated quantitative related variables. Aim of this method is taking out significant data from table, in order to symbolize PCA as a group of new orthogonal variables which is named as principal components, as well to show the similitude pattern of observings and variables like points in maps. The success of PCA model might be assessed by utilizing cross-validation methods like Jackknife and Bootstrap (Abdi and Williams, 2010).

This study usually focuses on determining infrequent data in classification problems in which data distributions are skewed. The success of ML approach should be usually evaluated according to results that obtained with prediction of minority class. Eventually, it is analysed that how to deal with the imbalanced class distribution and observe results which are obtained by various scaling methods. The Robust Scaler method is preferred in this study because it gives better results in data with outliers.

### Time Feature

Time indicates seconds which are passed onward the initial transaction in the dataset. This may be confirmed database actions which are made during “2” days when values are demonstrated graphically (look at Table 3-2). "Time" indicates bimodal acts while remarkable drop is seen in the amount of transactions after a period of almost 24 hours as it is demonstrated at Figure 3-3 and Figure 3-4. Eventually, this variable is removed due to the fact that it is not proper in ML, because data is extremely near to others till the recent transaction.

Figure 3‑3: Time Distribution Feature in Fraud Process

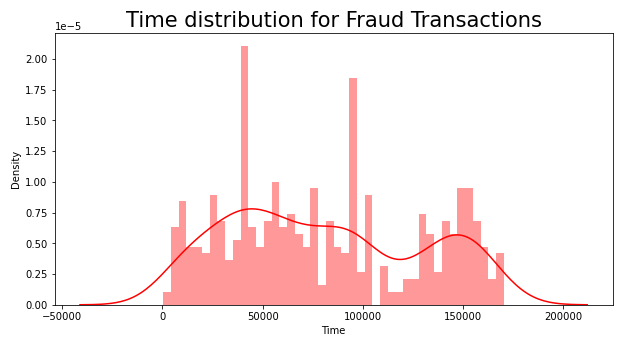
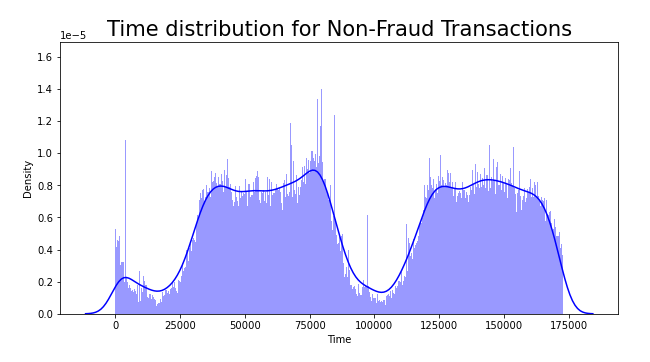


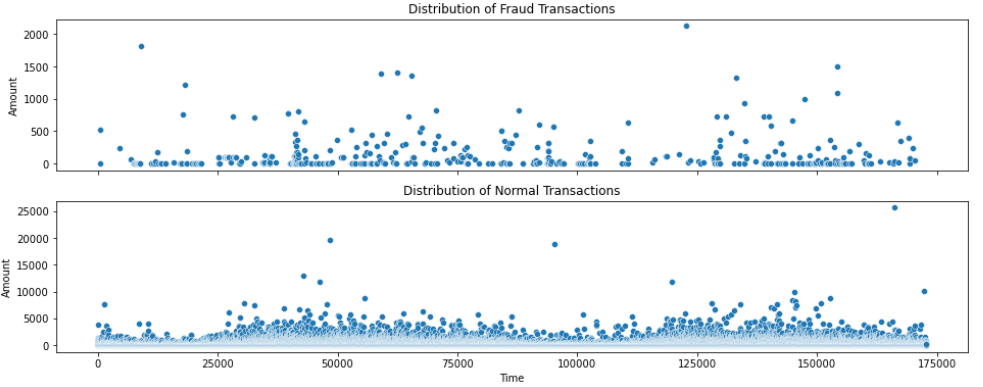
Figure 3‑4: Time Distribution Feature in non - Fraud Process



### Amount Feature

"Amount" property indicates the quantity of money in per transaction. It is observed that the highest transactions are $25,691.16 when mean of actions is $88.35 in this dataset. Figure 3-6 demonstrated which data is frequently intensified at minor values near to "0" while just a few transactions come close to the maximal value found. At the same time, Figure 3-6 also demonstrates extraordinary actions which differ from others. Those are named as outliers as well as they are actions where a great quantity of capital is sent. Reasonably, those values take notice for being probable frauds, but that situation is which fraudsters wish to stay away absolutely. Fraudsters usually prefer to transfer tiny money in order to maintain scamming by an undetectable way.

Figure 3‑5: Distribution of Fraud & Normal Transaction



### Class Feature

The class feature shows whether transactions are fraudulent or not and the variable gets value in the event of nonsuspicious as well if it is fraud then value will be Figure 3-7 shows that fraudulent actions represent 0.17% of whole data when non-fraudulent actions are equivalent to 99.83%. ML needs sustainable methods in order to split data as well as provide the training of the procedure effectively due to data is extremely unbalanced.

Figure 3‑6: Number of Non-Fraudulent and Fraudulent Transactions

.

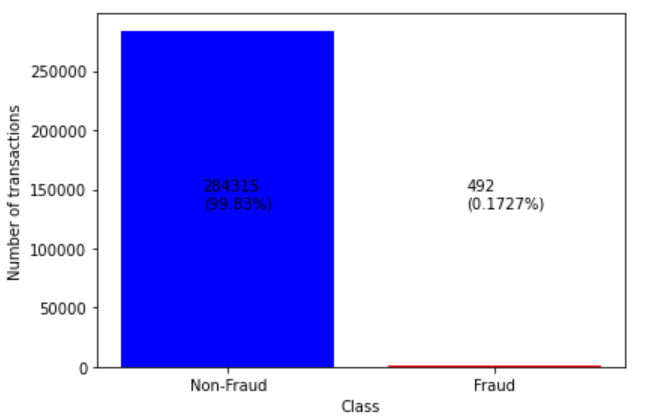


Figure 3-6 shows that the classes are extremely unbalanced with 99.83% of monitoring relating to non-fraudulent transactions as well as just 0.17% of monitoring labelled as fraudulent. Needs to handle the imbalance. ML needs sustainable methods in order to split data as well as provide the training of the procedure effectively due to data is extremely unbalanced.

### V-features

Histogram representation of v-Features is shown in Figure 3-7. These graphs provide a simple concept of data distribution. This is vital to control whether whichever significant relation is exist among features, particularly regarding "Class" property.

Figure 3-7 demonstrates a matrix of correlations among each of properties. These graphs highlight some features which are connected to the Class though many features are in the data which have comparatively few important correlations. Thus, it can be interpreted that properties are PCA that are outcome of preceding arrangement with the PCA method. As a result, “Time” and “Amount” properties does not have correlation with the Class property; hence, these features are not related with ML operation. Consequently, these variables are removed from the dataset due to that reason.

Figure 3‑7: Frequency of per Feature in the dataset

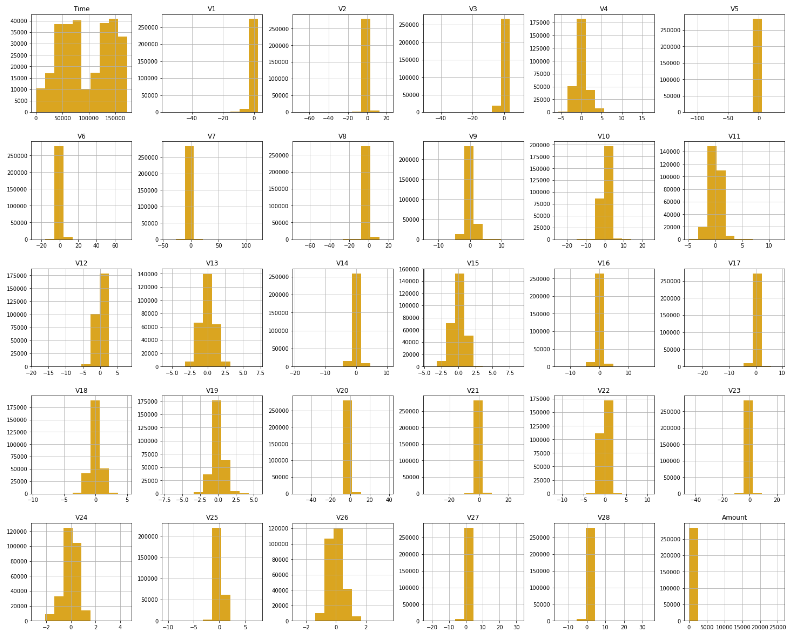
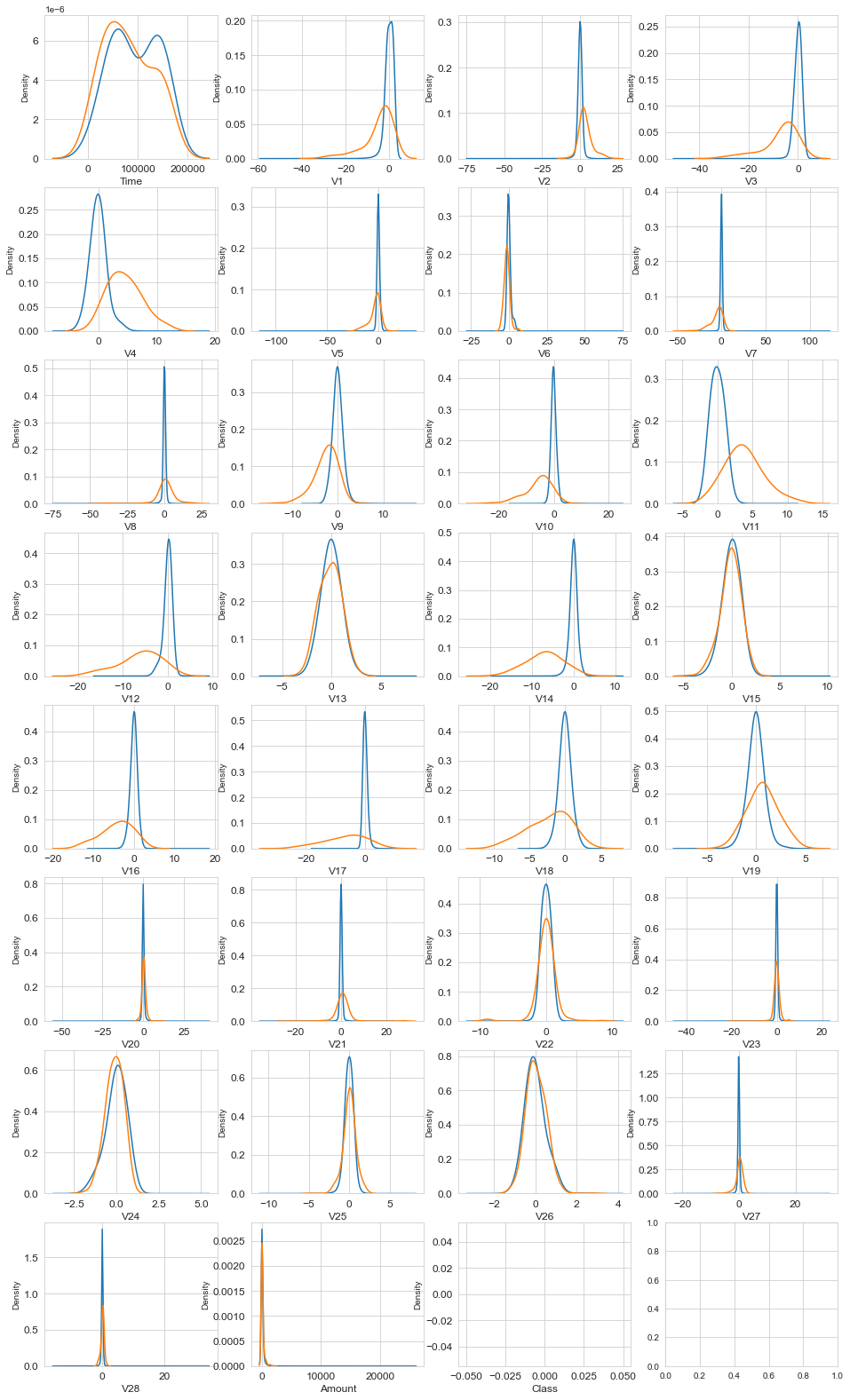


Figure 3-8 Plotting the Form of a Distribution of per Feature in “Fraudulent” and “Non-Fraudulent”. It can be noticed that a fine selectivity regarding distribution for two values of Class for several features:

Figure 3-8 demonstrates these:

* V4 and V11 have obviously split distributions for Class values "0" and "1".
* V12, V14 and V18 are partly apart from each other,
* V1, V2, V3 and V10 have a rather various feature,
* V25, V26 and V28 have parallel properties for Class value.

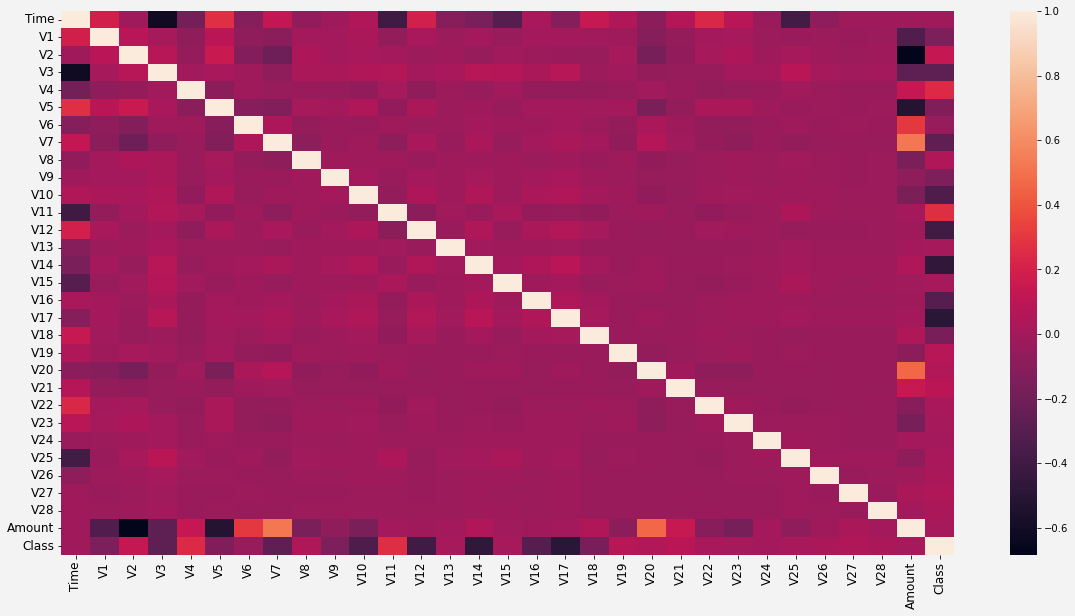
Figure 3‑8: Features Density plot



### Correlation Matrices

Actually, Correlation Matrices are basis for figuring out data. This helps to determine whether a particular transaction is fraudulent and which features have an important impact on fraud.

Figure 3‑9: Correlation Matrices



Several insights on the visualization in Figure 3-9 reveal the following:

* V7 and V20 are positively correlated with the amount.
* V2 and V5 are negatively correlated with the amount.

The statistical analysis which is shown at Table 3-3 plainly emphasizes the mean money action for the fraudulent ones that is perceived more compared to non-fraudulent ones.

Table 3‑3: Amount Details of Fraudulent Transaction

|  |  |
| --- | --- |
| Amount details of fraudulent transaction | Amount details of non-fraudulent transaction |
| count 492.000000 | count 284315.000000 |
| mean 122.211321 | mean 88.291022 |
| std 256.683288 | std 250.105092 |
| min 0.000000 | min 0.000000 |
| 25% 1.000000 | 25% 5.650000 |
| 50% 9.250000 | 50% 22.000000 |
| 75% 105.890000 | 75% 77.050000 |
| max 2125.870000 | max 25691.160000 |

## Comparison of models with scale (No Smote) in OrIginal DataSET

It is observed that FPs are more than FNs according to Confusion Matrix which is indicated below (Figure 3-10 and Figure 3-11). That may be an unintelligent approach due to ignoring possibility of an action being fraud while essentially it is very unsafe. Therefore, it is required to reduce the possibility of threshold for boosting TP rate.

Recall evaluation feature is more significant than precision in the event of detection for credit card fraud. Due to the issue is admissible for a degree which a non-fraud action is tagged as fraud. In such situation, a fraud action will proceed with no control so that is risky to tag a fraud as non-fraud.

Figure 3-10 and Figure 3-11 demonstrates confusion matrix showing results obtained by verification RF and DT methods.

Figure 3‑10: Confusion Matrix of RF – Original Dataset

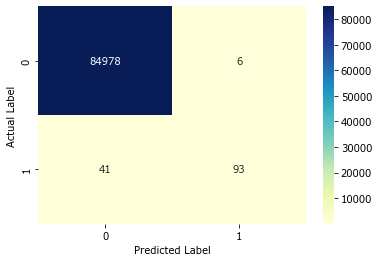
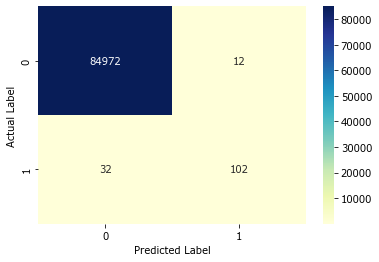


Figure 3‑11: Confusion Matrix of DT – Original Data



When the ROC Curve graph of the RF and DT algorithm given in Figure 3-12 and Figure 3-13 is analysed, the space under the blue line corresponds to almost 97% of the entire area for RF and 86% for DT. These rates for detection of credit card fraud in banking transactions is deficient in spite of the fact that this means a very successful classification process.

Figure 3‑12: ROC curve and AUC value for RF

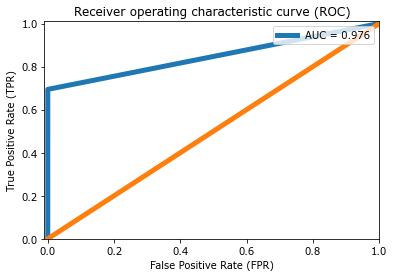
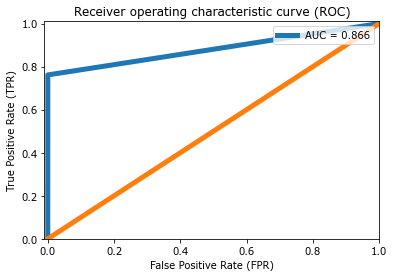
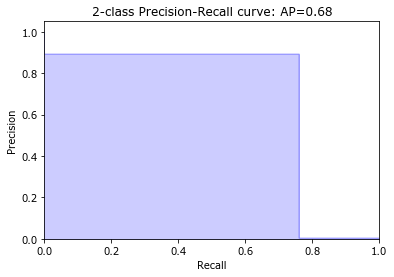


Figure 3‑13: ROC curve and AUC value for DT



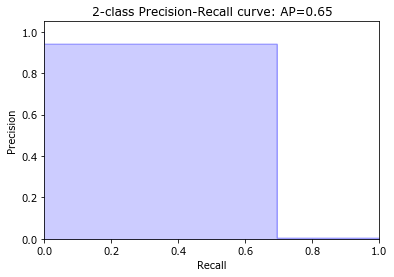
A precision-recall curve is an important metric in order to demonstrate trade-offs among recall and precision for skewed datasets. While number of positive instances increases (high recall), the accuracy of classifying per instance correctly reduces (low precision). This curve allows to determine the point in which both the recall and precision are up. Calculating precision call back pairs for different probability thresholds was done with the "precision\_recall\_curve()" command. The best point of threshold is 0.68 as it is shown in Figure 3.14.

Figure 3‑14: Precision-Recall Curve For RF



The best point of threshold is 0.65 as it is shown in Figure 3-15.

Figure 3‑15: Precision-Recall Curve For DT



The results obtained from the RF model are shown in Table 3-4.

Table 3‑4: Prediction Results of RF and DT on Testing Data for Original Data

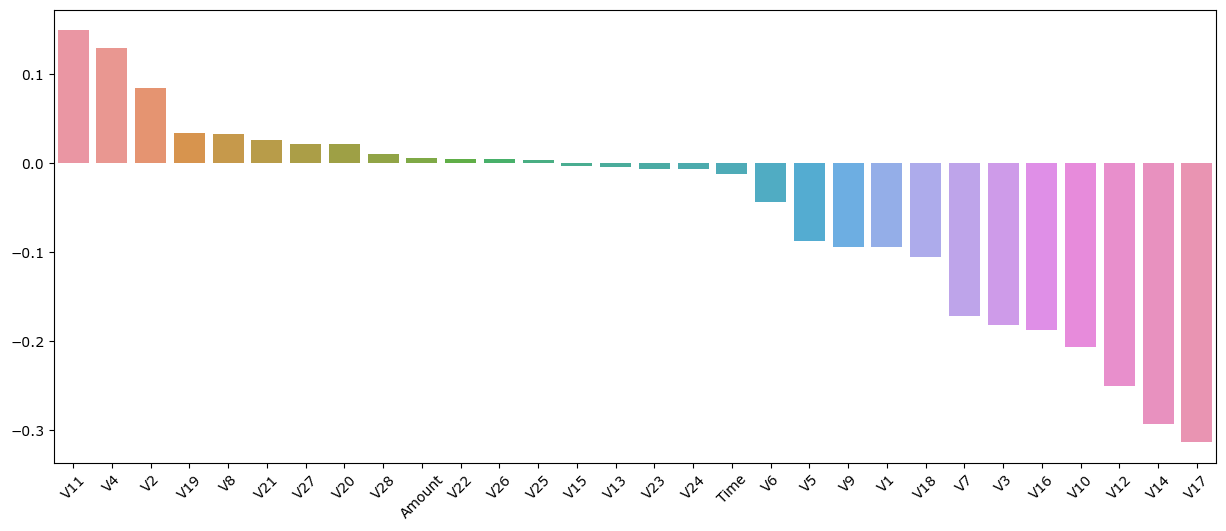
|  | **Model** | **Accuracy** | **Recall** | **F1 Score** | **Precision** |
| --- | --- | --- | --- | --- | --- |
|  | RF\_OrijnalData | 0.999448 | 0.694030 | 0.798283 | 0.939394 |
|  | DT\_OrjinalData | 0.999483 | 0.761194 | 0.822581 | 0.894737 |

### Threshold optimization approach

Normally, each of ML algorithm utilizes a possibility threshold of 0.5 to classify among negative and positive classes. The recall for fraudulent actions would be increased if the possibility of threshold could be adjusted to any other values that boosts TPR.

Coping with class imbalance in different circumstances is to adjust the possibility threshold which is utilized in order to make estimations (standard threshold is 0.5). Selection for threshold provides the higher accuracy for making future estimations.

Figure 3‑16: Features That Affect Fraud

Figure 3-16 shows the features that affect fraud. It is seen that the columns "V17", "V14", "V12" and "V10" have a high impact on fraud. Setting threshold values is necessary for these features in order to avoid overfitting.

Many attempts have been made to set the threshold value for these four features that are highly fraudulent. Prediction results are obtained by performing machine learning with the determined thresholds. As the results improved, threshold values are tested for the next feature. When the following threshold values are determined RF and DT can produce better results is observed for these features as a consequence of these improvements.

* "V17" < 0.4
* "V14" < 0.5
* "V12" < 0.1
* "V10" <=0.2

Better predictions could be obtained by modifying some options in the model. The ROC curve plots TPR versus FPR, over various threshold values. That is actually expected that blue line would be as near as to upper left corner.

Figure 3-17 and Figure 3-18 shows the confusion matrix showing the results obtained by the verification RF and DT methods.

Figure 3‑17: Confusion Matrix of RF – Threshold Data

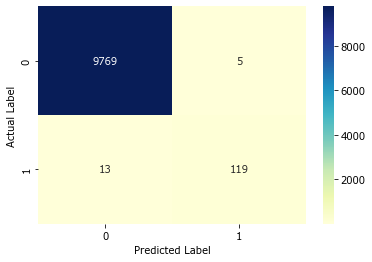
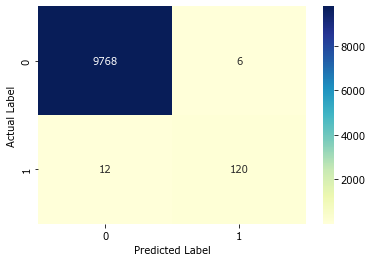
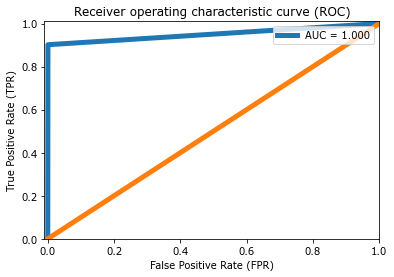


Figure 3‑18: Confusion Matrix of DT – Threshold Data



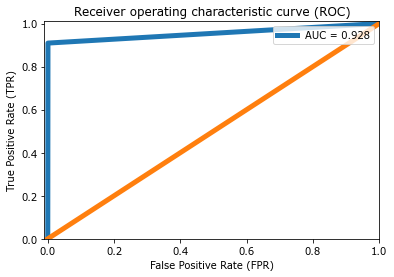
When the ROC Curve graph of the RF algorithm given in Figure 3-19 is analysed, the space under the blue line corresponds to almost 99.97% of the entire area for RF. When AUC is high, TN and TP distributions do not intersect. That can be interpreted as the model performs well with the help of correct threshold value which model can distinguish two classes from each other without error. These rates for detection of credit card fraud in banking transactions is this means a very successful classification process.

Figure 3‑19: ROC curve and AUC value for RF (ROC AUC Score:0.9997)



When the ROC Curve graph of the DT algorithm given in Figure 3-20 is analysed, the space under the blue line corresponds to almost 92% of the entire area for RF. These rates for detection of credit card fraud in banking transactions is deficient in spite of the fact that this means a very successful classification process. The model needs to be improved.

Figure 3‑20: ROC curve and AUC value for DT



The best point of threshold is 0.67 as it is shown in Figure 3-21.

Figure 3‑21: Precision-Recall Curve For RF



The best point of threshold is 0.87 as it is shown in Figure 3-22.

Figure 3‑22:Precision-Recall Curve For DT



The results of the optimized model with defined threshold values are shown in Table 3-5. Nevertheless, classification of credit card fraud requires to be finer. The model stands in need of to be improved farther.

Table 3‑5: Prediction Results of RF and DT on Testing Data for Threshold

|  | **Model** | **Accuracy** | **Recall** | **F1 Score** | **Precision** |
| --- | --- | --- | --- | --- | --- |
|  | **RF\_Threshold** | 0.998183 | 0.901515 | 0.929688 | 0.959677 |
|  | **DT\_Threshold** | 0.998183 | 0.909091 | 0.930233 | 0.952381 |

## SMOTE optimization approach

As it can be seen from Figure 3-7, this dataset contains 284,807 transactions and only 492 of that are suspicious. In other words, the number of regular transactions is larger than that of the fraudulent ones. As most actions are non-fraud, the dataset is highly imbalanced. In other words, this data would lead to inaccurate overfitting and correlations. More specifically, the accuracy score may have a quite high accuracy rate since the algorithm can smoothly predict fraud ones by focusing on non-fraud transactions. There are many techniques used to deal with such an imbalance. SMOTE is used in this project. SMOTE creates new minority class samples among available ones. These new virtual instances created by KNN of examples in the minority class. Under-sampling randomly balances the distribution through eliminating majority class randomly until a balanced distribution is reached (Sudha, 2017) Because of their success in dealing with class imbalance, SMOTE and random under-sampling feature scaling methods are used.

Figure 3‑23: After Apply Over-Sampling Approach



Distributions of the dataset has the high imbalance as shown in Figure 3-7. The subject is how to process the skewed dataset in this part of the study. Figure 3-23 shows the status of the dataset following the application of SMOTE method.

Figure 3-24 and Figure 3-25 demonstrates confusion matrix showing results obtained by verification RF and DT methods.

Figure 3‑24: Confusion Matrix of RF – With SMOTE

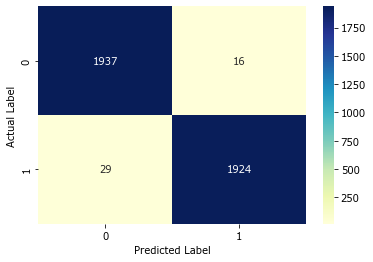
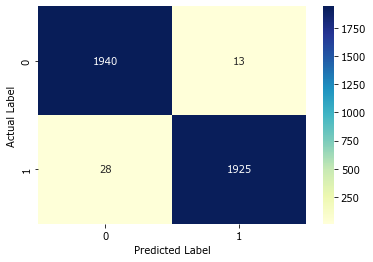


Figure 3‑25: Confusion Matrix of DT – With SMOTE



When the ROC Curve graph of the RF and DT algorithm given in Figure 3-26 and 3-27 is analysed, the space under the blue line corresponds to almost 99.9% of the entire area for RF and 98.9% for DT. It means that the model developed for the detection of credit card fraud in banking transactions make a very successful classification.

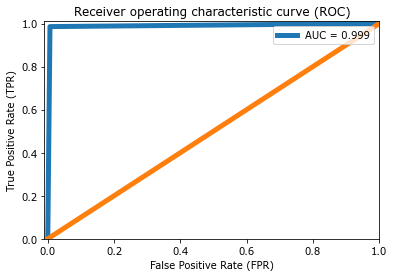
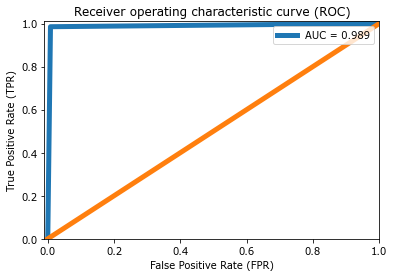
Figure 3‑26: ROC curve and AUC value for RF

Figure 3‑27: ROC curve and AUC value for DT



When the ROC Curve graph of the DT algorithm given in Figure 3-28 is analysed, the space under the blue line corresponds to almost 99% of the entire area for RF.

Figure 3‑28: Precision-Recall Curve For RF



When the ROC Curve graph of the DT algorithm given in Figure 3-29 is analysed, the space under the blue line corresponds to almost 98% of the entire area for DT.

Figure 3‑29: Precision-Recall Curve For DT



Entire measure of performance metrics in the improved model are above 98% while Table 3-6 is examined. The model is developed in order to categorize a prosperous classification for the detection of credit card fraud in banking transactions.

Table 3‑6: Prediction Results of RF and DT on Testing Data for SMOTE

|  | **Model** | **Accuracy** | **Recall** | **F1- Score** | **Precision** |
| --- | --- | --- | --- | --- | --- |
|  | RF\_Smote | 0.989503 | 0.985663 | 0.989463 | 0.993292 |
|  | DT\_Smote | 0.988479 | 0.985151 | 0.988441 | 0.991753 |

A continuous increase is taking notice in the performance metrics of the model when Table 3-7 and Figure 3-30 are examined; however, a slight decrease is noticed in the Accuracy metric. The decrease which occurs with SMOTE reminds "overfitting". As a result of the improvements which are applied to the data set, the expected results were obtained. The model is developed for the detection of credit card fraud in banking transactions in order to categorize classification prosperously. The SMOTE Oversampling Threshold Model is the final model that is improved in this study.

Table 3‑7: Prediction Results of RF and DT

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Accuracy | Recall | F1 Score | Precision |
| 1 | RF\_OriginalData | 0.999448 | 0.694030 | 0.798283 | 0.939394 |
| 2 | DT\_OriginalData | 0.999483 | 0.761194 | 0.822581 | 0.894737 |
| 3 | RF\_Threshold | 0.998183 | 0.901515 | 0.929688 | 0.959677 |
| 4 | DT\_Threshold | 0.998183 | 0.909091 | 0.930233 | 0.952381 |
| 5 | RF\_Smote | 0.989503 | 0.985663 | 0.989463 | 0.993292 |
| 6 | DT\_Smote | 0.988479 | 0.985151 | 0.988441 | 0.991753 |

| Figure 3‑30: Prediction results of RF and DT on testing data |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |

# Conclusions

Credit card fraud means actual losing of credit card either sensitive info of a credit card. A great deal of studies are conducted to detect fraudulent transactions with DNN in the literature. But those techniques are calculational high-cost; furthermore, they can only carry out finer on big data. The method might have good outcomes. However, the same or even better results might be accomplished with fewer sources. This is the starting point for this study. There are many types of ML algorithms that might be used for detection. This study aims to reveal the performance of two algorithms when appropriate pre-processing techniques are used.

Classifiers must indicate whether the relevant transaction is a fraud in this context. Goal of the research is to reveal the most effective ML ways for detecting fraudulent transactions for credit cards. In this study, two ML models which categorize classification of credit card transactions are utilized on the data set. These are RF and DT. Their performances will be compared in terms of F-Measure, Precision, Recall, Accuracy, ROC Area and AUC**.** The dataset is highly imbalanced. SMOTE method is utilized in order to cope with like a skewed data set since quantity of fraud transactions is lesser than non-fraud transactions. Therefore, SMOTE technique will be used to overcome oversampling. Feature selection will be performed. The dataset will be divided to two pieces which are test data and training data. Proposed approaches will be tested for detection of fraud detection in terms of F-Measure, Precision, Recall, Accuracy, ROC Area and AUC**.**

Recall evaluation feature is more significant than precision in the event of detection for credit card fraud. Cause of the issue is admissible to a degree which a non-fraud action is tagged as fraud. In such situation, a fraud action will proceed with no control so that is risky to tag a fraud as non-fraud. A great deal of experimental tests and solution methods are tried and the best solution model was presented with the improvements made along within study. RF and DT classifiers are acquired 98.56% and 98.51% Recall results, respectively with the proposed model.

Furthermore, columns "V17", "V14", "V12" and "V10" were found to have a high impact on fraud.

Considering the results obtained, two algorithms bring out excellent outcomes. Results of both algorithms are so close to each other. Despite the size of the data set used, the availability and reliability of the obtained accuracy rates are at a suitable level for the detection of credit card fraud.

**Future Work**

The results of this study can be tested via other algorithms used in DM in order to improve this study. Thus, rates of this research can be compared.

Evaluation criteria can be increased by including other unbalanced data processing techniques for classification problems.

Experiments can be made by setting different threshold values for features that affect fraud less.

The variation of the results can be looked at using other normalization techniques.

# REFERENCES

*Abdi, H. & Williams, L.J. (2010) Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), pp. 433-459.*

*Adewumi, A. O. & Akinyelu, A. A. (2017) A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(2), pp. 937-953.*

*Alharbi, A., Alshammari, M., Okon, O. D., Alabrah, A., Rauf, H. T., Alyami, H., & Meraj, T. (2022) A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach. Electronics, 11(5), pp. 756-774.*

*Asha, R. B., & Suresh Kumar K.R. (2021). Credit card fraud detection using artificial neural network. Global Transitions Proceedings, 2(1), pp. 35-41.*

*Ayodele, T. O. (2010). Types of machine learning algorithms. New advances in machine learning, 3, pp.19-48.*

*Bell, J. (2022). What Is Machine Learning?. Machine Learning and the City: Applications in Architecture and Urban Design, pp. 207-216.*

*Bhattacharyya, S. Jha, K. Tharakunnel, J. Westland. (2011) Data mining for credit card fraud: A comparative study, Decision Support Systems, 50, pp. 602-613.*

*Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.*

*Chaudhary, K., Yadav, J., & Mallick, B. (2012) A review of fraud detection techniques: Credit card. International Journal of Computer Applications, 45(1), pp. 39-44.*

*Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002) SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, pp. 321–357.*

*De Iuliis, M., Kammouh, O., Cimellaro, G. P., and Tesfamariam, S. (under review). "Quantifying restoration time of power 743 and telecommunication lifelines after earthquakes using Bayesian belief network model." Journal of Management 744 in Engineering. Pp. 1-15*

*Dhanaraj, R. K., Rajkumar, K., & Hariharan, U. (2020) Enterprise IoT modelling: supervised, unsupervised, and reinforcement learning. In Business Intelligence for Enterprise Internet of Things (pp. 55-79). Springer, Cham.*

*Dhankhad, S., Far, B., & Mohammed, E.A. (2018) Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study, IEEE International Conference on Information Reuse and Integration (IRI), pp. 122-125.*

*F. Ghobadi, & M. Rohani, (2016) Cost Sensitive Modelling of Credit Card Fraud using Neural Network strategy", Signal Processing and Intelligent Systems (ICSPIS), International Conference of IEEE pp. 1-5.*

*Gath, I., & Geva, A. B. (1989) Unsupervised optimal fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence, 11(7), pp. 773-780.*

*Goodfellow, I., Bengio, Y., & Courville, A. (2016) Deep learning. MIT press.*

*Hazım, L. R. (2018) Four classification methods Naïve Bayesian, support vector machine, K-nearest neighbors and random forest are tested for credit card fraud detection Unpublished Master's thesis, Altınbas University, Turkey.*

*Jain, Y., Tiwari, N., Dubey, S., & Jain, S. (2019) A comparative analysis of various credit card fraud detection techniques. Int J Recent Technol Eng, 7(5S2), pp. 402-407.*

*Ji, H., Alfarraj, O., & Tolba, A. (2020) Artificial intelligence-empowered edge of vehicles: architecture, enabling technologies, and applications. IEEE Access, 8, pp. 61020-61034.*

*Kamath, G., Agnihotri, P1., Valero, M., Sarker, K., & Song, W. -Z. (2016) Pushing analytics to the edge. In GLOBECOM, IEEE, pp 1–6.*

*Kamusweke, K., Nyirenda, M., & onde Kabemba, M. (2019) A Data Mining Model for Predicting and Forecasting Fraud in Banks. Khera, SN and Divya (2018)." Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques." Vision, 23(1), pp. 12-21.*

*Kalaiselvi, N., Rajalakshmi, S., Padmavathi, J. (2018) Credit card fraud detection using learning to rank approach", Internat2018 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC) National conference on computation of power, energy, Information and Communication (ICCPEIC) pp. 191-196.*

*Kazemi, Z., Zarrabi, H., (2017) Using deep networks for fraud detection in the credit card transactions, Knowledge-Based Engineering and Innovation (KBEI), IEEE 4th International Conference, pp. 630-633.*

*Keskenler, M.F., Deniz, D.A.L., & Aydın, T. (2021) "Yapay zeka destekli ÇOKS yöntemi ile kredi kartı sahtekarlığının tespiti." El-Cezeri 8(2), pp. 1007-1023.*

*Kirelli, Y., Arslankaya, S. A., & Zeren, M. T. (2020) Detection of credit card fraud in e-commerce using data mining. Avrupa Bilim ve Teknoloji Dergisi, (20), pp. 522-529.*

*Krieg, M. L. (2001). A tutorial on Bayesian belief networks.*

*Kumar, M., Husain, M., Upreti, N., & Gupta, D. (2010) Genetic algorithm: Review and application. Available at SSRN 3529843.*

*Last, F., Douzas, G., & Bacao, F. (2017) Oversampling for imbalanced learning based on k-means and smote. arXiv preprint arXiv:1711.00837.*

*Linoff, G.S., & Berry, M.J.A., (2011) Data Mining Techniques for Marketing, Sales and Customer Relationship Management, Wiley,3rd.ed., Canada.*

*Madhulatha, T. S. (2012). An overview on clustering methods. IOSR Journal of Engineering, 2(4) pp. 719-725.*

*Maimon O., & Rokach, L. (2010) Data Mining and Knowledge Discovery Handbook, Database Management & Information Retrieval, pp. 875-886.*

*Malini, N., & Pushpa, M., (2017) Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection, Advances in Electrical, Electronics, Information, Communication and Bio- Informatics (AEEICB), Third International Conference on pp. 255- 258.*

*Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), 140-147.*

*Navamani, C., & Krishnan, S. (2018) Credit card nearest neighbor based outlier detection techniques. Int. J. Comput. Tech, 5(2), pp. 56-60.*

*Oliver, J. (1992). Decision graphs: an extension of decision trees, Monash University, Department of Computer Science, pp. 343-350.*

*Özmen E. P., & Özcan T. (2019) Finans Sektöründe Dolandırcılık Tespiti Üzerine Melez Sınıflandırma ve Regresyon Ağacı Uygulaması, 5(2), pp. 12-20.*

*Pumsirirat, A. & Yan, L. (2018) Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine, International journal of advanced computer science and applications, 9(1), pp. 18-25.*

*Quinlan, J. R. (1996). Learning decision tree classifiers. ACM Computing Surveys (CSUR), 28(1), pp. 71-72.*

*Raj, S. B. E., & Portia, A. A. (2011) Analysis on credit card fraud detection methods. In 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET), pp. 152-156.*

*Restivo, K. (1999). The drill on data mining. Computer Dealer News, 15(14), pp. 29-30.*

*Sabau, A.S. (2012) Survey of clustering based financial fraud detection research. Informatica Economica, 16(1), pp. 110-122.*

*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), pp. 372-374.*

*Shetty, N. P., Shetty, J., Narula, R., & Tandona, K. (2020) Comparison study of machine learning classifiers to detect anomalies. International Journal of Electrical and Computer Engineering, 10(5), pp. 5445- 5452.*

*Sudha, T.R.C. (2017) Credit Card Fraud Detection in Internet using K Nearest Neighbour Algorithm," IPASJ international journal of computer science, 5(11), pp. 22-30.*

*Suraj, P., Varsha, N., & Kumar, S.P. (2018) Predictive modelling for credit card fraud detection using data analytics, Procedia Computer Science, 132, pp. 385-395.*

*Sykes, A. O. (1993). An introduction to regression analysis. Coase-Sandor Institute for Law & Economics Working Paper No. 20.*

*Şenocak, B. (2021) Data mining and machine learning applications in bank audit / Banka denetiminde veri madenciliği ve makine öğrenimi uygulamaları, Unpublished Master's Thesis, Bahcesehir University, Turkey.*

*Taha, A. A., & Malebary, S. J. (2020) An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access, 8, pp. 25579-25587.*

*Tomassini, M. (1996) Evolutionary algorithms. In Towards Evolvable Hardware. Springer, Berlin, Heidelberg, pp. 19-47.*

*Tiwari, P., Mehta, S., Sakhuja, N., Kumar, J., & Singh, A. K. (2021).Credit Card Fraud Detection using Machine Learning: A Study. arXiv preprint arXiv:2108.10005.*

*Wang, C. Wang, Y., Ye, Z., Yan, l., Cai, W., Pan, S. (2018) Credit card fraud detection based on whale algorithm optimised BP neural network, 13th International Conference on Computer Science & Education (ICCSE), pp. 1-4.*

*Zeager, M. F., Sridhar, A., Fogal, N., Adams, S., Brown, D. E., & Beling, P. A. (2017) Adversarial learning in credit card fraud detection. In 2017 Systems and Information Engineering Design Symposium (SIEDS), pp. 112-116*