# SMOTE: Credit Card Fraud Detection Using Supervised Machine Learning Methods

[1]R.P.Shanthi Rani, [2] Allimalli Durgabhavani, [3] R.Reeja Igneshia Malar, [4]Amit Kumar Singh

[1,4]*Assistant Professor, Dept of CSE (AI&ML),CMR Engineering College, Hyderabad*
[2,3]*Assistant Professor, Dept of CSE (Cyber Security),CMR Engineering College, Hyderabad*

**Abstract-**Credit card transaction frauds are normal today as a large portion of us are utilizing the charge card installment strategies all the more much of the time. This is because of the development of technology and the rise in online transactions, which has led to frauds that cause significant financial losses. As a result, effective means of lowering the loss are required. Moreover, fraudsters track down ways of taking the Mastercard data of the client by sending counterfeit SMS and calls, likewise through disguising assault, phishing assault, etc. The goal of this paper is to use support vector machine (SVM), k-nearest neighbor (KNN), and artificial neural network (ANN), Synthetic Minority over-sampling Technique (SMOTE) algorithms to predict when fraud will occur. we have achieved accuracy of 95.81% using logistic regression and 93.62% using naive Bayes and 93.88% using decision tree and we step into deep learning, we used ANN achieved better accuracy then all other algorithms of 97.67%.

***Keywords:*** *Synthetic Minority over-sampling Technique, SVM, KNN, ANN*

# 1. Introduction

There are growing number of new companies all around the world. All of that companies are trying to provide best service quality for their customers. In order to succeed in that, companies are processing a lot of data on a daily basis. These data come from vast number of resources and are in different formats. Moreover, this data contains some of the key parts of the company's future business. Because of that, companies need to store that data, to process it and what is really important, to keep it safe. Without securing data, a lot of it can be used by other companies or even worse, it can be stolen. In most cases, financial information is stolen, which can harm whole company or individual.

The rapid growth in credit card transactions has resulted in a significant rise in instances of fraud. Fraud can be detected using a variety of statistical and data mining techniques. Pattern matching, or artificial intelligence, is used in many fraud detection methods. It is very important to use effective and safe methods to find fraud. Due to fraud, credit card fraud is on the rise. Financial losses are also on the rise. Nowadays, the Internet—also known as online transactions—is expanding as new technologies emerge on a daily basis. The credit card holds the largest share in these transactions. Losses from credit card fraud in London were estimated at 844.8 million US dollars in 2018. It is necessary to prevent or detect fraud in order to lessen these losses. The rapid development of technology has resulted in a variety of frauds. Therefore, numerous machine algorithms, including hybrid algorithms and artificial neural networks, are utilized to detect fraud due to their superior performance.

## Supervised Machine Learning:

Supervised learning is a machine learning method in which models are trained using labeled data. In supervised learning, models need to find the mapping function to map the input variable (X) with the output variable (Y).

$$Y = f(X)$$

Supervised learning needs supervision to train the model, which is similar to as a student learns things in the presence of a teacher. Supervised learning can be used for two types of problems: **Classification** and **Regression**.

Supervised machine learning requires labeled input and output data during the training phase of the machine learning model lifecycle. This training data is often labelled by a data scientist in the preparation phase, before being used to train and test the model. Once the model has learned the relationship between the input and output data, it can be used to classify new and unseen datasets and predict outcomes.The reason it is called supervised machine learning is because at least part of this approach requires human oversight. The vast majority of available data is unlabelled, raw data. Human interaction is generally required to accurately label data ready for supervised learning. Naturally, this can be a resource intensive process, as large arrays of accurately labelled training data is needed.

Supervised machine learning is used to classify unseen data into established categories and forecast trends and future change as a predictive model. A model developed through supervised machine learning will learn to recognise objects and the features that classify them. Predictive models are also often trained with supervised machine learning techniques. By learning patterns between input and output data, supervised machine learning models can predict outcomes from new and unseen data. This could be in forecasting changes in house prices or customer purchase trends.

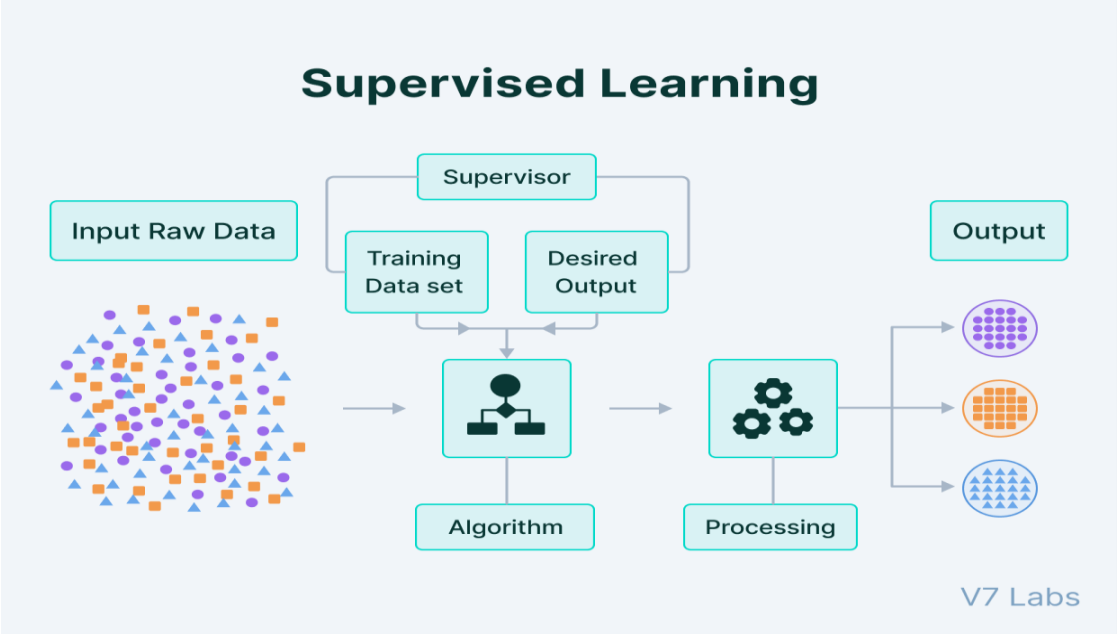


Fig.1. Supervised Learning

The objective this project is to predict the diabetes using various machine learning algorithms and comparing the result based on the prediction and accuracy of the different algorithms of machine learning with the specified dataset of diabetes mellitus.

Check Fraud occurs when person forges a check or pays for something with check knowing that there is not enough money. Internet sales are fraud where fraudster sale fake items or counterfeit items, or taking payment without delivering the item. There are a couple more, such as charities fraud, identity theft, credit card fraud, debt elimination, Insurance fraud and others. Due to increasing popularity of cashless transactions, one of the most common frauds are credit card frauds. Credit card fraud refers to the situation where fraudster uses credit card for their needs while owner of that credit card is not aware of that.

# 2. Related Background Survey

Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques , Author: Ankit Mishra ,Year: 2018 , Nowadays, as internet speed has increased and the prices of mobile have decreased very much in past few years. Also, the data prices too are very much affordable to most of the people. This has resulted into the digitization of most of the institutes as it is easy and convenient for the people and also for the authority to maintain the records. So, it resulted in most of the banks and other institutes receiving and transferring money through credit cards. But with the hackers and other cyber criminals around there is always chances of the frauds in the transactions. The possibility of the fraud transaction is very less but it is not negligible and even having one fraud transaction is unacceptable because it is crime and we can't neglect it even if it is very less as it harms both the customer and credibility of the institute. So this paper aims at analyzing various classification techniques using various metrics for judging various classifiers. This model aims at improving fraud detection rather than misclassifying a genuine transaction as fraud.

Machine Learning for Credit Card Fraud Detection System ,Author: Lakshmi S V S S Year: 2018 , The rapid growth in E-Commerce industry has lead to an exponential increase in the use of credit cards for online purchases and consequently they has been surge in the fraud related to it .In recent years, For banks has become very difficult for detecting the fraud in credit card system. Machine learning plays a vital role for detecting the credit card fraud in the transactions. For predicting these transactions banks make use of various machine learning methodologies, past data has been collected and new features are been used for enhancing the predictive power. The performance of fraud detecting in credit card transactions is greatly affected by the sampling approach on data-set, selection of variables and detection techniques used.

This paper investigates the performance of logistic regression, decision tree and random forest for credit card fraud detection. Dataset of credit card transactions is collected from kaggle and it contains a total of 2,84,808 credit card transactions of a European bank data set. It considers fraud transactions as the “positive class” and genuine ones as the “negative class”. The data set is highly imbalanced, it has about 0.172% of fraud transactions and the rest are genuine transactions. The author has been done oversampling to balance the data set, which resulted in 60% of fraud transactions and 40% genuine ones. The three techniques are applied for the dataset and work is implemented in R language. The performance of the techniques is evaluated for different variables based on sensitivity, specificity, accuracy and error rate. The result shows of accuracy for logistic regression, Decision tree and random forest classifier are 90.0, 94.3, 95.5

respectively. The comparative results show that the Random forest performs better than the logistic regression and decision tree techniques.

Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection , Author: N. Malini ,Year: 2017 , Popular payment mode accepted both offline and online is credit card that provides cashless transaction. It is easy, convenient and trendy to make payments and other transactions. Credit card fraud is also growing along with the development in technology. It can also be said that economic fraud is drastically increasing in the global communication improvement. It is being recorded every year that the loss due to these fraudulent acts is billions of dollars. These activities are carried out so elegantly so it is similar to genuine transactions. Hence simple pattern related techniques and other less complex methods are really not going to work. Having an efficient method of fraud detection has become a need for all banks in order to minimize chaos and bring order in place.

There are several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc are used for detecting credit card fraudulent transactions. Along with these techniques, KNN algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate. Any of these methods can be implemented on bank credit card fraud detection system, to detect and prevent the fraudulent transaction.

Credit Card Nearest Neighbor Based Outlier Detection Techniques ,Author: Mrs.C.Navamani ,Year: 2018 , Popular payment mode accepted both offline and online is credit card that provides cashless transaction. It is easy, convenient and trendy to make payments and other transactions. Credit card fraud is also growing along with the development in technology. It can also be said that economic fraud is drastically increasing in the global communication improvement. It is being recorded every year that the loss due to these fraudulent acts is billions of dollars. These activities are carried out so elegantly so it is similar to genuine transactions. Hence simple pattern related techniques and other less complex methods are really not going to work. Having an efficient method of fraud detection has become a need for all banks in order to minimize chaos and bring order in place. There are several techniques like Machine learning, Genetic Programming, fuzzy logic, sequence alignment, etc are used for detecting credit card fraudulent transactions. Along with these techniques, KNN algorithm and outlier detection methods are implemented to optimize the best solution for the fraud detection problem. These approaches are proved to minimize the false alarm rates and increase the fraud detection rate. Any of these methods can be implemented on bank credit card fraud detection system, to detect and prevent the fraudulent transaction.

Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis Author: John O. Awoyemi ,Year: 2017, Machine Learning is considered as one of the most successful technique used for creating a fraud detection algorithm for fraud identification. In rule-based approach, algorithm cannot recognize the hidden patterns as they are strictly rule based. We use Machine Learning as it makes machine to learn by itself using classification and regression approach for recognizing fraud in credit card transaction. Due to its fast computing power it has become one of the efficient ways of detecting fraud. The machine learning algorithms are divided into two types, supervised [4][8] and unsupervised [6] learning algorithm. Many supervised and unsupervised machine learning techniques have been presented for fraud detection in credit card transaction which includes logistic regression [3], decision tree [4][5], neural networks [10][9][1][2], Naive Bayes [6], K-Nearest Neighbors [6] ,Support Vector Machines [5] and Random Forest [1][2]. This

paper proposes a FDS using Random Forest which can identify transaction fraud in credit card. Random Forest is the advance version of Decision Tree and has better efficiency and accuracy than any other Machine Learning algorithm. The system also uses learning to rank approach to rank the alert generated by the model so that alert with highest rank will only be notified thereby reducing the number of alert detected by rule-based approach FDS.

# 3. Proposed Implementation

There are two types of credit card frauds. One is theft of physical card, and other one is stealing sensitive information from the card, such as card number, cvv code, type of card and other. By stealing credit card information, a fraudster can broach a large amount of money or make a large amount of purchase before cardholder finds out. Because of that, companies use various machine learning methods to recognize which transactions are fraudulent and which are not.

Experiment included back propagation neural network that was optimized with Whale algorithm. Neural network consisted of 2 input layers, 20 hidden and 2 output layers. Due to optimization algorithm, they achieved exceptional results on 500 test samples: 96.40% accuracy and 97.83% recall. Authors of paper and used neural networks, in order to demonstrate improvement in results when ensemble techniques are used. In paper three datasets were used for comparison between Auto-encoder and Restricted Boltzmann Machine algorithms, which led to the conclusion that algorithms like MLP can be suitable for credit card fraud detection.

**Data Gathering:**
Data collection is the process of gathering and measuring information on targeted variables in an established system, which then enables one to answer relevant questions and evaluate outcomes. The goal for all data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed. Here we need to gather the data which used for detecting the credit card frauds.

**Data Exploration:**
After gathering the data from web, we will explore the data that is contained in the credit card data frame. We will proceed by displaying the credit card data using the head function as well as the tail function. We will then proceed to explore the other components of this data frame.

**Data Manipulation:**
We will apply this to the amount component of our credit card data amount. Scaling is also known as feature standardization. With the help of scaling, the data is structured according to a specified range. Therefore, there are no extreme values in our dataset that might interfere with the functioning of our model.

**Data Modeling:**
Data modeling is the process of creating a data model for the data to be stored in a dataset. This data model is a conceptual representation of Data objects, the associations between different data objects and the rules. Data modeling helps in the visual representation of data and enforces business rules, regulatory compliances, and government policies on the data. We will split our dataset into training set as well as test set with a split ratio.

**Fitting:**
We will implement a decision tree algorithm. Decision Trees to plot the outcomes of a decision. These outcomes are basically a consequence through which we can conclude as to what class the object belongs to. We will now implement our decision tree model.

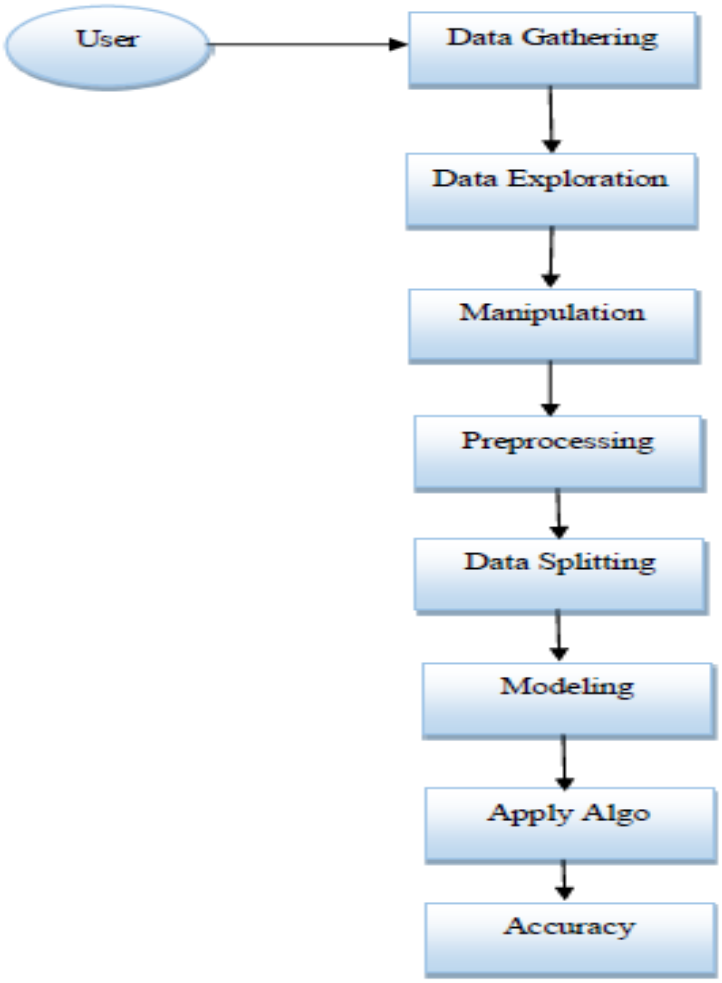


Fig.2 Proposed Implementation Flow

**Multilayer Perception**
The purpose of this paper is to analyze various machine learning algorithms, such as Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and Multilayer Perceptron (MLP) in order to determine which algorithm is most suitable for credit card fraud detection.

Our multi-modal event tracking and evolution framework is suitable for multimedia documents from various social media platforms, which can not only effectively capture their multi-modal topics, but also obtain the evolutionary trends of social events and generate effective event summary details over time. Our proposed mmETM model can exploit the multi-modal property of social event, which can effectively model social media documents including long text with related images and learn the correlations between textual and visual modalities to separate the visual-representative topics and non-visual-representative topics

# 4. Results Analysis

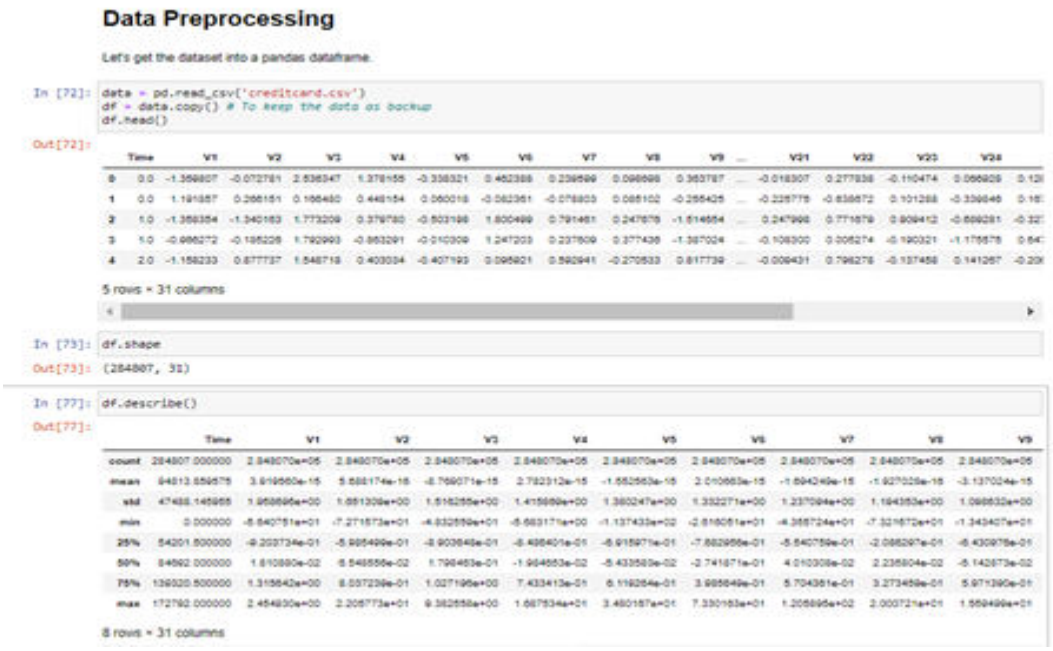


Fig.3 Data Preprocessing

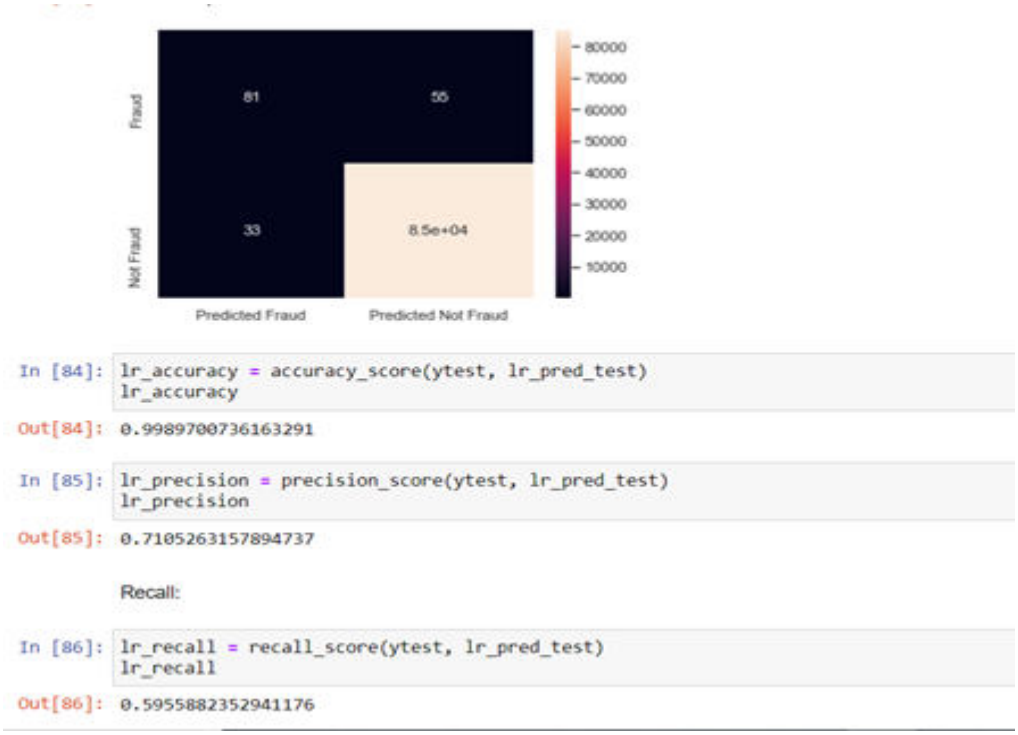


Fig.4 Logistic Regression Algorithm

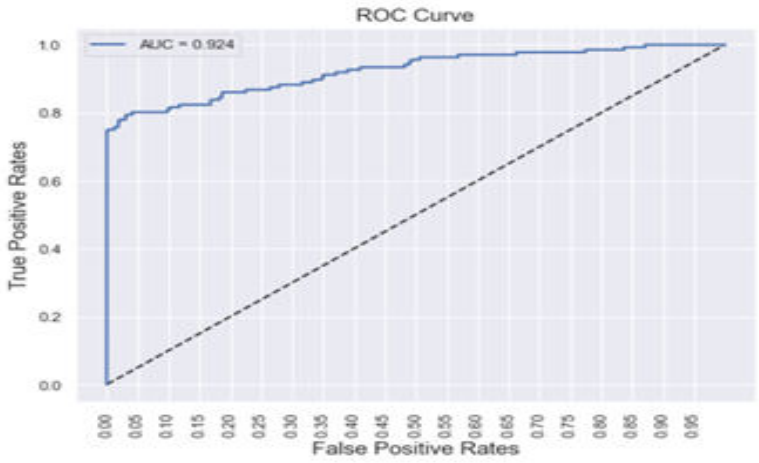


Fig.5 Support Vector Machine Algorithm

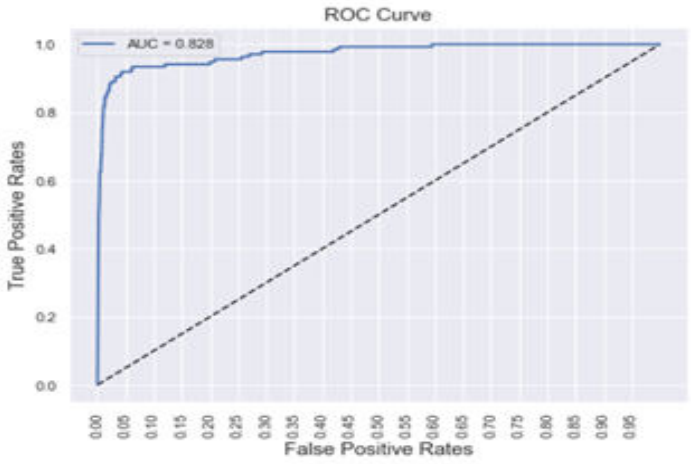


Fig.6 Naive Bayes Algorithm

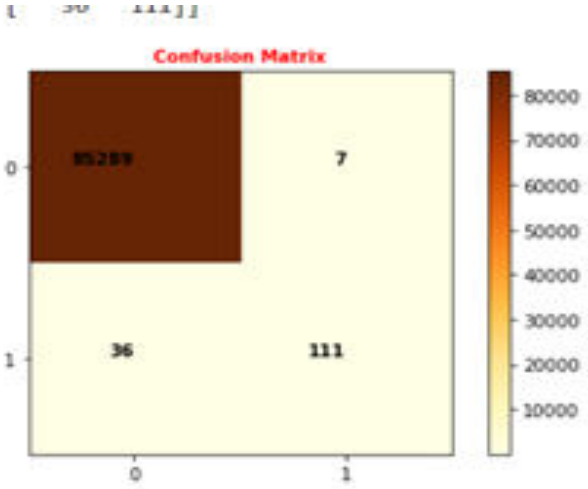


Fig.7 Random Forest Algorithm

# 5. Conclusion

Credit card frauds signify a very severe business problem. These frauds can result in significant personal and business losses. As a result, businesses are spending more and more money on developing novel strategies and concepts that will assist in the detection and prevention of fraud. Comparing various machine learning algorithms for detecting fraudulent transactions was the primary objective of this paper. Thus, correlation was made and it was laid out that Irregular Woods calculation gives the best outcomes for example best arranges regardless of whether exchanges are extortion. This was laid out utilizing various measurements, like review, exactness and accuracy. It's critical to have high-value recall for this kind of issue. The selection of features and the balancing of the dataset have been shown to be crucial for obtaining significant results.

## References:

[1] Global Facts (2019). Topic: Startups worldwide. [online] Available at: https://www.statista.com/topics/4733/startups-worldwide/ [Accessed 10 Jan. 2019].
[2] Legal Dictionary (2019). Fraud - Definition, Meaning, Types, Examples of fraudulent activity. [online] Available at: https://legaldictionary.net/fraud/ [Accessed 15 Jan. 2019].
[3] European Central Bank (2018). Fifth report on card fraud, September 2018. [online]. Available at: https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport201809.en.html#toc1 [Accessed 21 Jan. 2019].
[4] En.wikipedia.org. (2019). Credit card fraud. [online] Available at: https://en.wikipedia.org/wiki/Credit_card_fraud [Accessed 24 Jan. 2019].
[5] A. Mishra, C. Ghorpade, “Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques” 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) pp. 1-5. IEEE.
[6] S. V. S. S. Lakshmi, S. D. Kavilla “Machine Learning For Credit Card Fraud Detection System”, unpublished
[7] N. Malini, Dr. M. Pushpa, “Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection“, Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on pp. 255- 258. IEEE.
[8] Mrs. C. Navamani, M. Phil, S. Krishnan, “Credit Card Nearest Neighbor Based Outlier Detection Techniques”
[9] J. O. Awoyemi, A. O. Adentumbi, S. A. Oluwadare, “Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis”, Computing Networking and Informatics (ICCNI), 2017 International Conference on pp. 1-9. IEEE.
[10] Z. Kazemi, H. Zarrabi, “Using deep networks for fraud detection in the credit card transactions”, Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on pp. 630-633. IEEE.