TOWARD IMPROVEMENT OF CREDIT CARD FRAUD DETECTION BASED ON MACHINE LEARNING TECHNIQUES

Daya Shankar Srivastav[1], Dr. Shweta Rai[2],[1] *Assistant Professor*,[2] *Associate Professor*[1]*United College of Engineering and Research, Greater*,[2] *GNIOT MBA Institute, Greater Noida*

Corresponding Author Email: shwetarai@gniotmba.net

Received:14.02.24/ Published:01.09.24

Abstract:

Financial organizations are now expanding their financial facilities with the use of innovativeservices, such as credit card, automated distributors (ATMs), internet and mobile banking. Furthermore, using credit cards along with the rapid progress of e-commerce are becoming a convenience and an important component of financial life. The Credit Card is a payment method issued by the customer's credit card. The use of credit cards has many advantages, like:

Easy to buy: Credit cards will improve your life. We allow customers to buy credit arbitrary without carrying cash in an arbitrary way without the need to provide an easy method of payment for online purchases, either through the telephone, via ATM machines, etc.

Maintain a history of company credit: Regular customers with a good credit history are sometimes important to understand. Perhaps not credit cards, but also other financial services such as loans leases and even some employment are significant towards this background. Customer credit score and history can be checked at the appropriate times by credit card companies, banks, credit card companies, retail and utility companies in order to see the timely, accountable payment of their debts by customers.

Purchase insurance: Credit cards can also provide consumers with extra security when you buy items they're lost, hurt or robbed. The statement of the purchaser's credit card and the company can confirm that the client purchased it if it lost or stolen the original receipt. However, several credit card companies offer great purchase insurance.

1. Introduction

Credit card plays a very important role in today's economy. It is an inevitable part of households, companies and worldwide business. That use credit cards provides huge advantages when used thoughtfully and sensibly, considerable financial damage and credit can be caused by fraud. Credit is a way to sell services or goods without cash in hand from the buyer. Even an automatic means of giving a consumer credit is a credit card. Each credit card today carries an identifier which speeds up shopping.Credit card plays a very important role in

today's economy. It is an inevitable part of households, companies and worldwide business. That use credit cards provides huge advantages when used thoughtfully and sensibly, considerable financial damage and credit can be caused by fraud. Credit is a way to sell services or goods without cash in hand from the buyer. Even an automatic means of giving a consumer credit is a credit card. Each credit card today carries an identifier which speeds up shopping.Each credit card now carries an identifier that speeds up shopping. "The use of credit cards started in the United States as during 1920s, when large companies, such as big corporations and hotel chains, began issuing them to consumers," according to Urban Dictionary. Credit cards, on the other hand, have been listed in Europe since 1890. Fraud occurs in the credit card industry when a lender is duped by a borrower who offers him/her transactions, assuming that the borrower's credit card issuer will cover the cost. In an ideal world, there will be no bill. The credit card company will reclaim the money if the payment is made. Today, as e-commerce expands, half of all fraud in credit cards is performed on the internet. Fraudsters frequently have ties to the company in question. It may be an internal faction, but it's more likely to be an external party in the credit card industry. As a prospective customer or a prospective/existing seller, fraud is committed as an external party. External fraudsters can be classified into three categories: average perpetrator, violent offender, and organised crime offender. Initial credit cards included direct sales between the retailer issuing the credit and the consumer of that merchant [1].

Fraud is a major issue for e-banking services, particularly when it comes to credit card transactions. Fraud is an intentional failure with a view to receiving or indirectly or specifically trickled financial gain. Fraud represents a violation of public law, which gives the fraudster an illegal benefit or unauthorized damage. The calculation of the fraud-related damages demonstrates that perhaps the cost of fraud is high. Digital technology has increased substantially with credit card fraud resulting in the loss of trillions of dollars globally each year. Internet Crime Complaint Centre statistics indicate that the amount of fraud recorded has increased dramatically in the past decade. Today, the need of credit and debit cards has increased due to contribution of new payment solutions, for example, individual to-individual payments and mobility card-based contactless vicinity payments. Basically, credit card fraud is an unauthorized activity committed by any individual from another individual’s account for money theft. And the operation which involves taking actions to prevent activities like these and indulge risk management practices to ensure against comparable activities later on is called fraud detection. Fraud requires the fraudulent use details about credit or debit card for transactions [2].

1.1 Types of Fraud

Financial fraud can be described as a deliberate use of illegal methods or practices to make a financial gain [4, 5]. It is a big issue for individuals, organizations, governments, and other sectors. The rise of the internet, cloud computing, automation, and different e-payment channels is fuelling this issue more even more. The most common financial fraud can be categorized as shown in Figure 1.

.

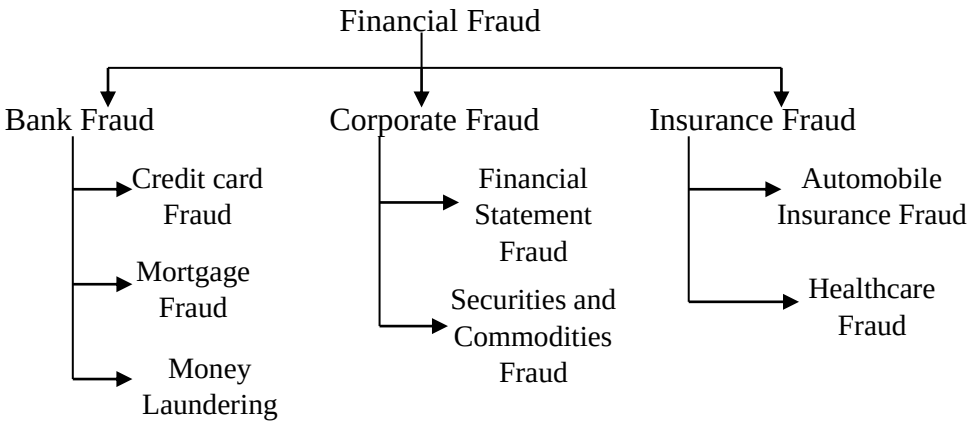


Figure 1: Types of financial fraud

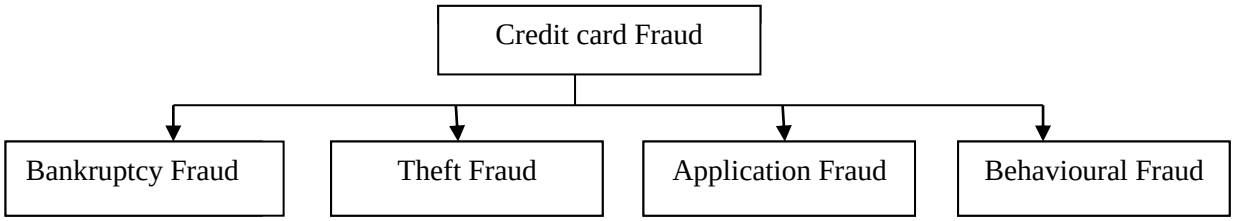


Figure 2: Credit card Fraud

Consequently, the threat of credit card fraud is also on the rise. Fraudulent detection is difficult and the issuing authorities shall avoid losses. This is why the system proposed uses fraud detection models for an anonymized dataset to assess the performance. The main terms of this proposed system is to contribute a variety of machine learning techniques which predict the real world fraud detection in datasets. The fraudulent used a publicly available dataset. The data set is used to extract from the real credit card transactions for the last three months. This system machine learning technique is used for financial applications and detects fraudulent activity.

1.2. Credit card fraud

Credit card fraud shall be referred to as illegally using credit card or its information without the knowledge of the owner. The two main application and behavioural fraudes are different credit card fraud tricks. Application fraud occurs when, by using falsified or other information, fraudsters apply new banking or issuing cards. A user with a set of user details (so-called duplication fraud) or other users with identical detail may submit multiple applications (called identity fraud). Conductive fraud, on the other hand, consists of four main types: card robbery/loss, mail robbery, counterfeit card, and "not present card holders." Stolen/lost card fraud happens when a card is stolen or accessed in the hands of the fraudsters. Mail theft fraud happens if a fraudster receives a credit card from the bank or personal data before they reach the real one. Credit card details are acquired and

without information of the card holders in both counterfeit and „card holders not recognise fraud." The former can be done by remote mail, telephone or the Internet with card details. Both are made using card information to produce counterfeit cards. According to statistics from 2012, credit card fraud threats to high-risk countries. With a stunning 19 percent, Ukraine has the highest rate of fraud.With its staggering 19%, Ukraine has the highest fraud rate, largely dominated by Indonesia at 18.3%. Yugoslavia is the riskiest country after these two countries at a rate of17.8%. Malaysia (5.9%), Turkey (9%) and eventually the United States are the next highest rate of fraud. The main differences between analysis of user behaviour and fraud analysis approaches are to be highlighted. With a low false positive rate, the fraud analysis method can detect known fraud tricks. These features represent the document and framework of fraud tricks described in the dataset of oracles and can then easily find out what frauds are currently suffering in the system. If the test data is not included. Where there are no fraud signatures in the test data, there will be no alarm. This allows a huge reduction in the false positive rate. However, as it is based on limited, specific fraud records that the learning of a fraud analysis system cannot detect new frauds. Thus, depending on how naively the fraudsters are, the rate of false negatives may be extremely high. The analysis of user behaviour, however, highly solves the problem of novel fraud detection. These approaches do not look for specific patterns of fraud but instead compare incoming activities to the built-up model of legitimate user behaviour [4].

2. Literature review

Xinwei Zhang et al. (2019) suggested that card transaction fraud costs card issuers billions of dollars per year. The effective identification of identity theft program is seen as necessary to reduce the losses of fraud with a cutting-edge fraud detection model. The main contribution to their research is to create a fraud prevention system using a profound learning architecture and progressed HOBA-based feature design. First, a collection of characteristics worth researching to prevent fraud must be defined in their proposed system. Four forms of consumer behavior analysis will then be done for each of these characteristics [5].

M.Suresh Kumar et al. (2019) focused mainly on the detection of fraud in real-life tickets. The detection of fraud by credit card is here based on transactions fraudulent. In general, online and offline, credit card fraud may occur. Yet online fraud operations are rising every day in today's world. Therefore, various methods have been used in the existing system to detect online fraud transactions. We use the Random Forestry Algorithm (RFA) in the proposed system for the determination of fraudulent transactions and their accuracy. This methodology is based on a supervised learning technique that is used to identify a database by decision trees. A confusion matrix is acquired after categorization of the data set. Random Forest Algorithm's performance is assessed on the basis of the confusion matrix. The results of information processing are approximately 90 per cent accurate [6].

Sangeeta Mittal et al. (2019) explained that today, credit card transactions have become commonplace and frauds have also become commonplace. One of the most common methods of fraud is to illegally obtain card details and to use them to shop online. It is impossible to detect such fraudulent transactions between thousands

of normal transactions for creditcard enterprises and merchants. In order to resolve this problem, master-learning algorithms can be applied if sufficient data are collected and available. In this study, common monitored and unmonitored master-learned algorithms were implemented in an extremely imbalanced dataset to detect card fraud.Unsupervised machine learning algorithms are found to handle the skewness and to give best results of classification [7].

Hamzah Ali Shukur et al. (2019) mentioned that rapid participation in mainly online transactional activities is causing fallacious cases and massive personal and financial losses everywhere. Although various criminal activities are taking place in commercial activities, online consumers are mainly affected by fraudulent e-card activities. In order to verify Suspect and fraudulent payment trends and parameters and normal information was used for processing data. Data processing techniques on the opposite, machine learning software (ML) was used to automatically mechanically predict suspicious and foolish transactions by victim's classifications. The author discusses the identification based on the supervision. All classifiers achieve more than 95.0% of the accuracy opposed to the outcomes they accomplished before the set of data was processed when the data set is pre-processedwith normalization and main component analysis [8].

Tamanna Chouhan, Ravi Kant Sahu, (2018) has presented data mining concept, "classification technique",used to prevent fraud by credit or debit card. The mining is a term, to predict some that means, but the data mining is a technique, is to find out the hidden predictive information. The data mining introduce and describe the technique is called "Prediction analysis". The SVM is a classifier and it is proposed into the detecting the credit card fraud. In the concept is to take the input data and it is divided into test and training sets, and it is predicted the precision and recall [9].

MaliniN. et.al (2017) has been identified for bank card fraud detection. This paper compares various methods for the identification of fraudulent credit card transactions such as machine learning, genetic coding, fugitive logic and sequence alignment. KNN algorithms and outlier approaches are used together with these strategies to find the best solution to the problem of fraud detection. The false alarm rates and the fraud detection rates were reduced to a minimum [10].

Maira Anis, Mohsin Ali, (2015) have presented the Credit card fraud identification decision tree algorithms for group unbalancing training, which is reduces the financial problems. This concept is to apply comparative method in decision tree techniques. The paper concept introduce the term, "Resembling", it is related to the imbalanced data. In the paper concept, aim is to find out the best classifier based on distribution. In this concept is applied in to the RUS and feature selection for the family of classifier that is "Decision tree classifier". In this concept finally, provides the result is denotes the improved performance for the decision tree classifiers are already known, so for this system is very efficient to detect the fraud [11].

Wee-Yong Lim, Amit Sachan, (2014) kindly presented the concept namely, "conditional weighted transaction aggregation for credit card fraud detection", which reduce the problem of substantial losses for credit card companies and consumers. In order to identify this problem, the conditionally weighted transaction aggregation software uses supervised learning methods, so fraudulent transactions can be detected [12].

Vijayalakshmi Mahanra Rao, (2013) have presented the evaluation of the Decision Tree for the identification of fraud through ensemble training. The credit card fraud problem is mostly affected the banking industries. The rise of web services give the advantages (banking) at the same time raises the banking frauds. The banking systems every second have robust, safe and secured one. It is order to detects and prevents the fraudulent activities of physical and virtual (any kind of) transactions. But the machine learning technique is to minimize the frauds. In this paper aim is to reduce the banking frauds. This system is used to (i) genetic algorithm and (ii) ensemble tree learning techniques these two techniques are to indicate the ensemble of decision trees, so databases for the credit card transaction are identified and to prevent the bank fraud [13].

3. Problem Statement

While Credit card fraud identification has been a major issue in literature, researchers are still confronted with certain problems (Substantial open issues) and were not properly addressed. This analysis will focus on future studies to improve the efficiency and trustworthiness of fraud prevention structures. These are the following questions:

No standard and detailed benchmark or data set for credit cards

Inherently, credit cards are private properties, so it is very difficult to establish a correct benchmark for this purpose. Failed data sets can partially lead to the learning of fraud tricks or normal behaviour. The lack of a standard data set, however, makes it problematic or impossible to compare different techniques. Most researchers used only authors ' datasets which are not publishable for reasons of confidentiality.

Nonexistence of standard algorithm

Among credit card fraud literature there is no efficient algorithm proven to be above all others. As demonstrated in previous chapters, each technique has its own advantages and disadvantages. It would be extremely important to combine these algorithms in request each other to help and covering another's faults.

Nonexistence of suitable metrics

To evaluate the performance the lack of decent indicators in fraud prevention structures is still an open issue. Unable to compare different methods or to define targets for more successful fraud detectio n systems, these indicators are not applicable to researchers and practitioners.

Lack of adaptive credit card fraud detection systems

Although various works in the area of card fraud has been conducted, there are no or few predictive strategies that can learn about the flow of transaction information. This framework can over time, without needing to be studied offline, upgrade its internal model and mechanisms. It can then immediately add new fraud (or usual conduct) to a template of learning tricks and later detect them.

4. Objectives

Fraud is a false or fraudulent disappointment that has the intention of gaining financial or personal gain. Two methods can be used to stop damage from fraud, namely avoidance of fraud and detection of fraud. Fraud avoidance is a constructive way of preventing fraud first and foremost. However, when a fraud transaction is attempted by a fraudster, fraud detection is necessary. The main objectives of this research are as follows:

- Analyse and predict all important features or parameters of credit card fraud.
- Performance investigations are carried out on the machine learning algorithms which are used to detect the fraudulent activities.
- The specific objective of the performance investigations is to determine the algorithms efficiency in terms of precision, recall and F-measure rate.
- Evaluation and comparison of predicted results with other available techniques.

5. Proposed Methodology

Some popular machine learning algorithms in supervised and unsupervised categories were chosen to be evaluated for the above-mentioned problem. The data collection is analysed and transactions identified as fraud. For our proposed Kaggle dataset, we have used three separate algorithms for fraud using python in the credit card framework. Which are briefly explained below and compared their performance. Comparison are made for these algorithms. In order to determine which machine learning, provide better outcomes and can be developed to classify fraud by credit dealers.

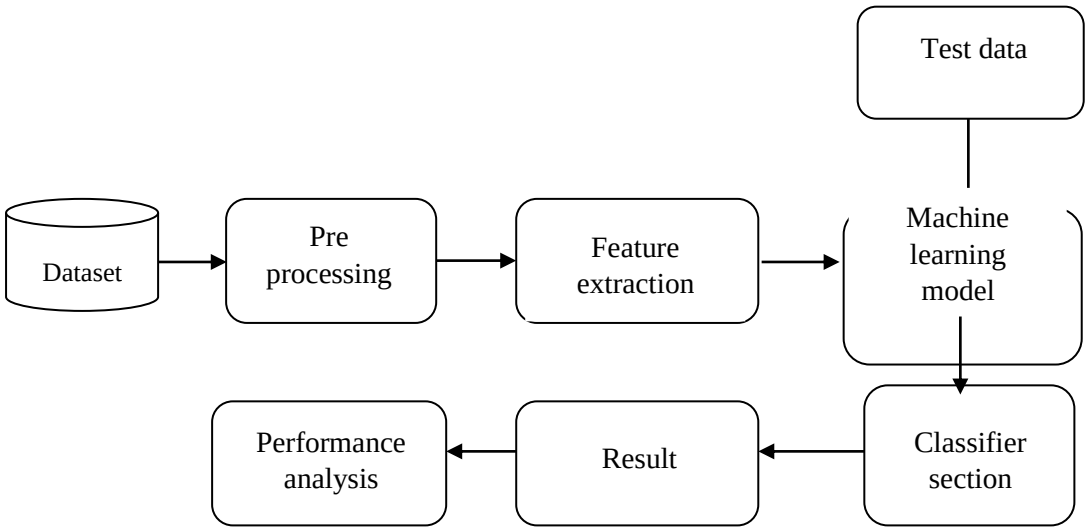


Figure 1. Proposed methodology

Algorithm steps:

Phase 1: Please read the dataset.

Stage 2: The data set is sampled randomly to balance it.

Phase 3: Split the data into two sections such as test and train.

Stage 4: The proposed models will be applied to the selection of features.

Phase 5: Accuracy and efficiency metrics for various algorithms were measured to know the efficiency.

Phase 6: Retrieve the best algorithms for the data set based on efficiency.

A. Local Outlier Factor

In this year 2000 M. In order to find the anomalous data points by measuring local deviations of a given database in regard to its neighbours, Breunige, Hans-peter Kriegel, Anthony T. Ng and Jörg Sander introduced the local Outlier Factor (LOF) dataset. Outliers based on the local density are detected using this algorithm. Locality is given by nearest neighbours and density is calculated by their distance.By equating the local volume of an instrument with the local concentrations of its neighbouring countries, regions of the same density and points which are significantly lower than their surroundings can be identified. The data point is considered as an outlier if it has very small density as compared to its neighbours [15].

Outer trends can be broken down into two types: national and global outliers. The object which is significantly having a large distance with respect to its k-th neighbour is called Global outlier while as object while a local outlier has a fairly large length from his k-th neighbour to its closest neighbours in the median wage [16].

B. Random forest

Random Forest is an approximation focused on ensemble training Random Forest. Ensemble learning is a predicted algorithm, wherein different models or related models are assembled or bundled many times. The random algorithm of the forest operates in the same way and uses many different algorithms called "the random forest." The random forest algorithm can also be used for estimation and recognition functions. We have obtained the exact fraud percentage from the dataset with this Local Outlier Factor algorithm and Random Forest algorithm by analysing their output [17].

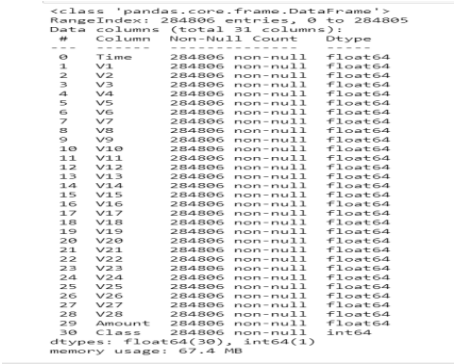
C. Support Vector Machine

Support vector machines do a class division by looking for a divisional point for categorization in a hyper plane in a high-dimensional space. The SVM model shows the instances as spaces, mapped so as to separate the examples from the various categories into a simple, as large, distance. The hyperplane or set of Hyperplanes in large or infinite spaces is installed in the support-vector machine that can be used for identification, correlation and other functions like detecting outliers. This algorithm can detect fraud very effectively. Hyperplanes are characterized as a sequence of points, the dot product of which is constant with a vector in this space, where a set of vectors is an orthogonal set of vectors which defines hyper-planes [18].

5.1. Dataset and Features

The data set and experimental context for calculate the relationships most explicitly between Time, class and amount. More precisely, and provided that each modality used is correlated with a credit card data would be based personalized learning. Finally, here analyze each experiment's findings and present a qualitative results review. Here we use credit card data for personalized learning. In the data variables available are Time, V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V20, V21, V22, V23, V24, V25, V26, V27, V28, class and amount are around 284806as shown in Figure.2. Since we must introduce a character-level model, all the lines of our dataset will be divided into character lists. The lower the frequency value, the higher the char (among the data set) the more frequently. Here need to remove null values, to get the best output. Main aim is to predict total fraud transaction based on features such as time, class and amount.

In the credit card dataset, need to remove null values and then identify the amount, time and class. To train and test our models, we used a publicly available Kaggle for credit card dataset. The data set is for credit card dataset and consists of well over 284806 examples with 28 features categorized as follows:



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284806 entries, 0 to 284805
Data columns (total 31 columns):
 #   Column  Non-Null Count   Dtype
---  ------  --------------   -----
 0   Time    284806 non-null  float64
 1   V1      284806 non-null  float64
 2   V2      284806 non-null  float64
 3   V3      284806 non-null  float64
 4   V4      284806 non-null  float64
 5   V5      284806 non-null  float64
 6   V6      284806 non-null  float64
 7   V7      284806 non-null  float64
 8   V8      284806 non-null  float64
 9   V9      284806 non-null  float64
 10  V10     284806 non-null  float64
 11  V11     284806 non-null  float64
 12  V12     284806 non-null  float64
 13  V13     284806 non-null  float64
 14  V14     284806 non-null  float64
 15  V15     284806 non-null  float64
 16  V16     284806 non-null  float64
 17  V17     284806 non-null  float64
 18  V18     284806 non-null  float64
 19  V19     284806 non-null  float64
 20  V20     284806 non-null  float64
 21  V21     284806 non-null  float64
 22  V22     284806 non-null  float64
 23  V23     284806 non-null  float64
 24  V24     284806 non-null  float64
 25  V25     284806 non-null  float64
 26  V26     284806 non-null  float64
 27  V27     284806 non-null  float64
 28  V28     284806 non-null  float64
 29  Amount  284806 non-null  float64
 30  Class   284806 non-null  int64
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

Figure.2. Data Features

The first move is to test the dataset. While most of the datasets were complete, some data were missing. Features such as time,class and amount were measured in case of scheduled and actual check-out and check-out through monthly, hourly and weakly. The missing values were difficult to quantify for features such as Time, class and amount and thus examples are omitted from our data collection for missing values.

5.2. Data Pre-processing

The program will read in a single text file that contains all the features such as time, class and amount. We need to fill the values that are blank with 0. Main aim is to predict time, class and amount. An outlier is easily visible for total interactions, at around 284806. The next step is to pre-process the dataset.

After getting all necessary parameters from the credit card, we have done some pre-processing operations on the data to make them compatible for the sentiment analysis. They were also used for pre-processing to a certain degree in most of the research papers. A small number of scientists used hopping, deletion of words and spelling. A growing number of the uses of stemming, elimination of words, indexing of text, reduced dimensionality and weighting of expressions.For pre-processing sentiment analysis alone, these approaches cannot be used. For example, the accuracy of the sentiment analysis will decrease if we stop deleting words. Few scientists used tokenization. To make a sentiment study, we cannot tokenize the sentences. What we should do, then, is to isolate the sentences in one paragraph as much as possible.

6. Results and Analysis

In this evaluation, we divide the entire feature set by credit card fraud detectionactivity type into different categories. Every experiment was performed on the same computer with 1.6 GHz processor and 8 G memory configurations. A 10-fold cross validation was conducted on the same dataset as previously used. Evaluation was performed for all study classifiers and three algorithms were employed for training and testing or model assessment. Various studies have been performed in the fields of credit card fraud detectiondata. Excessive research conducted on predicting and analysing credit card fraud detection. The accuracy and instability matrix to select the algorithm is thoroughly examined and the execution on the Jupyter notebook is carried out. Is implemented on the Jupyter notebook platform with python and machine learning language. The parameters in the data set are as follows. The accuracy of the machine-learning model is the method used for determining which framework is best classified. The accuracy of the machine learning model is the metric used to assess the model best for classifying data-based trends and relationships within a data set between variables. Precision is one metric used to evaluate classification techniques. It thoroughly examines the accuracy of its models and provides insight into the likelihood of factors like target loss that could impair model accuracy and thus adversely affect the process of selection. Informally, the model gets right is the predicted component. Companies use machine learning and artificial intelligence realistic business choices and have more precise results of model models that help to make better choices. Many methods and techniques for forecasting credit card fraud detectionwere developed. Because of the noise and stochastic characteristic of the data should be clean, not null values for accuracy calculations. The more elevated the It will err if the data set contains a number of null values. Because there are many various methods and advanced features for predictive analysis, these are discussed in this chapter separately, to help select the best techniques to effectively predict credit card fraud detectionfor intelligent households. The most successful forecasting of large datasets is based on a set of algorithms and tools collected under an umbrella called 'Machine Learning. Cross validation is common to be used for model selection.

5.1. Summary of findings

We have used two credit card fraud detectiondatasets with machine learning and deep learning algorithm. Machine learning detection techniquesis a part of a supervised machine learning task aimed at the amount ofcredit card fraud detectionappliance based on parameters such as climate, humidity and pressure. In particular, certain features such as Time, class and amount. For example, we have found lots of difference of Time, class and amount. These findings suggest that credit card fraud detectionissues build because of appliances and lights. It is important to note that the class and amountis different. The application of machine learning techniques produced good results with 0.58 to 100 % accuracy.

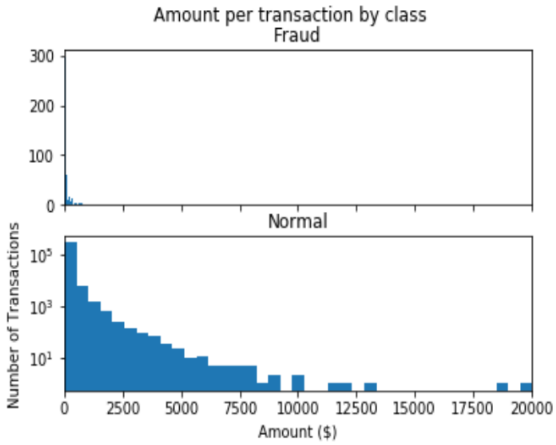


Figure.3.Amount per transaction by class fraud

Fig.3 shows that amount per transaction by class fraud have clearly the greatest effect on the number of transactions in comparison of normal transaction. This result is in line with findings that the most fraud amount vs normal amount. In addition, even though a similar conclusion has been reached for fraud transaction and normal transaction to another conclusion that they receive more normal fraud amount or normal amount. "Fraud amount transaction" is the second most important function to be considerably less significant than "normal amount" even though it has still an impact of 17 percent. This input feature refers to the fraud amount that the normal amount transaction at the time the transaction.

Figure.4.Transaction class distribution

However, Observation of Fig.4 showed that the transaction class distribution frequency is high with respect to normal. Transaction class distribution frequency fraud is very less in comparison of normal. It was decided to have more importance to the transaction class distribution. These findings indicate that transaction class distribution contributes little to the maximum, and after an hour it is the most significant factor in fraud transaction.

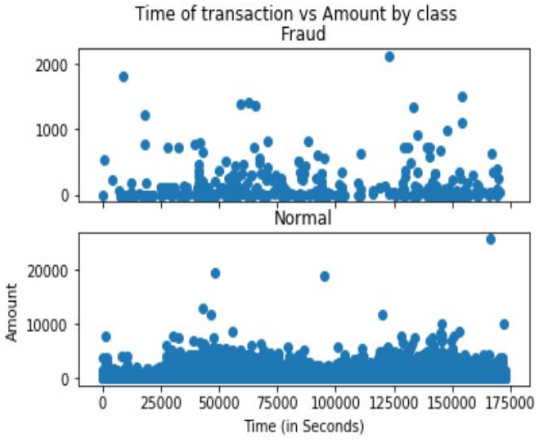


Figure.5.Time of transaction vs amount by class fraud

As the number of fraud transaction occur duringtransaction on the amount. A statement from Fig.5 shows, however, that after time of transaction vs amount by class fraud to decrease, that is to say that amount in time is reluctant to the normal amount. This issue may reveal some transaction fraud with respect to time because normal transaction fraud chances are less in comparison of higher amount transaction.

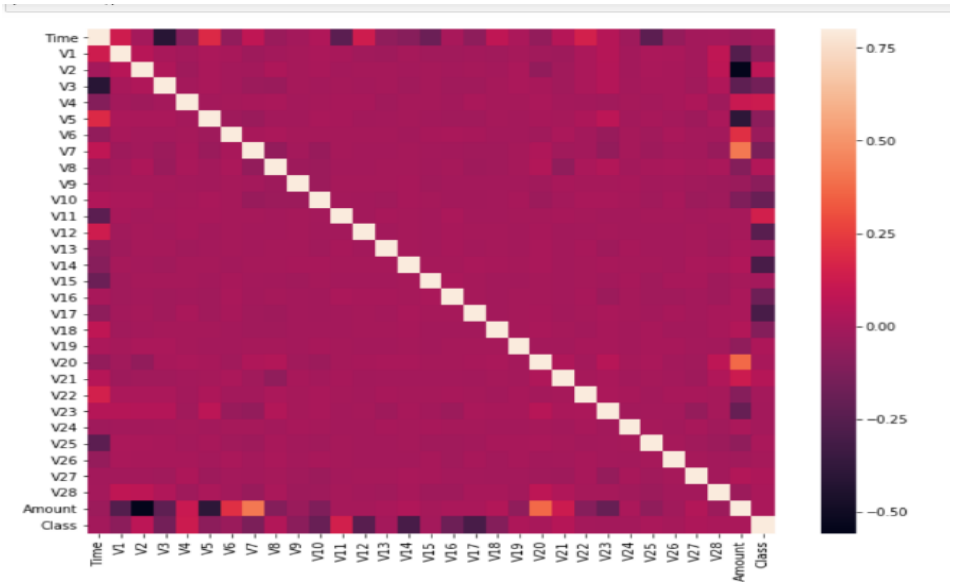


Figure.6. Corelation Matrix

Fig.6 shows the correlation matrix of the V1 to V28 features that have any correlation with each other moreover if we classify Class feature which has some negative and positive form correlations with the V features but have not any correlation with Amount and Time.

6.2. Train Test Split

Training can be divided into two parts, the trained sentence context model and the trained generative model. Both training processes may be considered without manual marks as unregulated training. We divide three parameters, namely sharing and comment, for the training of the random forest model, lasso and support vector machine algorithm. This project is mainly aimed at predicting that committed users prefer time, class and amount on a creditcard data. In the data variables available are Time, V1, V2, V3, V4, V5, V6, V7, V8, V9, V10, V11, V12, V13, V14, V15, V16, V17, V18, V19, V20, V21, V22, V23, V24, V25, V26, V27, V28, class and amount are around 284806as shown in Figure.2.

Data cleaning is an initial phase for the final analysis assessment of the dataset. Databases are susceptible to noisy, incomplete, and incoherent data because of the large amount of data available. This project's credit card data are derived from kaggles that have 15 variables of various kinds and may not be consistent with the format in which Python allows the data to be used. Data Cleaning helps to erase noisy details and to eliminate incoherence. A missing value, sequentially entered, given a constant value or a mean value, may be ignored in the data cleaning process.In this case, the whole frame of the data is structured and arranged to preserve and delete the appropriate attributes. This is done to make the usage simpler and more feasible. The fill factor tells

us how much room can be added on each page. The created filling factor value can generally be defined from 1 to 100 as a percentage.

```
Local Outlier Factor: 93
Accuracy Score :
0.9967346652154068
Classification Report :
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28434
           1       0.02      0.02      0.02        47

   micro avg       1.00      1.00      1.00     28481
   macro avg       0.51      0.51      0.51     28481
weighted avg       1.00      1.00      1.00     28481
```

Figure.7. Local Outlier Factor Algorithm Accuracy

Table.7. clarifies the accuracy of local outlier factor algorithm. This algorithm is used to detect fraud transaction. It is predicting 99% accuracy. The Precision score is 100%, recall score is 100%, F1-score is 100% and support values are 18434. It is predicting higher accuracy in comparison with support vector machine.

```
Random Forest Classifier: 69
Accuracy Score :
0.9975773322565921
Classification Report :
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     28434
           1       0.27      0.28      0.27        47

    accuracy                           1.00     28481
   macro avg       0.63      0.64      0.64     28481
weighted avg       1.00      1.00      1.00     28481
```

Figure.8. Random Forest Classifier Algorithm Accuracy

Fig.8. iterating through a random forest using the most important variables led to some improvement, and the robust scoring metrics suggest that Time and class can be modelled for to a reasonable degree. A simple (3 feature) Random Forest performed better than a Lasso Regression model to a large degree, with a difference in test r^2 values of .05. However, both models demonstrated no signs of over fitting, with the test r^2 values all being close to or higher than the train r^2 value. This model can provide value by giving posters an accurate estimate of how many users they engaged by post, particularly useful for optimizing for post scheduling, and measuring post performance. Best results came from RF parameters are 500 estimators, 15 min sample split and

Train/test split of 0.3. Here had solid performance in the test set, with 772 r^2 and 919 Spearman Correlation. But the model showed some signs of over fitting when exposed to the test sets are reduction of .15 in the r^2 and reduction of .10 in test Spearman Correlation. One reason that there could be over fitting is the large number of features in the model. We can take the feature importance to get the top 15 features, then iterate through the Random Forest again, and see if over fitting persists. The error increases as the value of engaged users increases. Thus, this violates the one of the main assumptions of a support vector machine model.

```
Support Vector Machine: 8411
Accuracy Score :
0.7046803131912504
Classification Report :
              precision    recall  f1-score   support

           0       1.00      0.71      0.83     28434
           1       0.00      0.34      0.00        47

   micro avg       0.70      0.70      0.70     28481
   macro avg       0.50      0.52      0.42     28481
weighted avg       1.00      0.70      0.83     28481
```

Figure.9. Support Vector Machine Algorithm Accuracy

Fig.9 illustrates the model performed no differently when taking the top 3 features by importance from the first model. The support vector machine model performed solid overall. No over fitting: the R2 value rose .045 points from train to test. Moderate predictive power: .64 R2 train, .68 in test. For the accuracy calculation the R2 score is stated in a formula. In addition, time to complete the classification is also reported. For tests of various SVM assemblies, the test data set and the accuracy values are presented in Fig.9. As shown in Fig.9, the SVM labelling ensemble carries out the best 100% value accuracy measurement. This is consistent with the R2 score and spearman and Pearson shown in Fig.9, where most of the approaches have higher scored as compared to other two algorithms. The simpler model (with the top 4 features only) had higher performance than the model with the k-best 20 features. So, the final features are: Total Interactions, Status, Page total likes and Photo.

7. Conclusion

In this research work on theoretical machine learning model has been proposed for credit card detection. Credit card frauds are increasing massively with the increase in usage of credit cards for transactions. Fraud on credit cards is certainly an act of fraudulent deception. The most popular techniques of fraud and methods for detection were described in this article and recent studies were discussed in this area. This paper described in detail how apply machine learning to detect credit card fraud along with the algorithm. The aim of the study is to classify and detect credit card fraud with best machine learning algorithms with high percentage of accuracy between the number of valid and number of genuine transactions.We have used three algorithm such as Random Forest Algorithm, Local outlier Factor Algorithm and Support Vector Machine. These are the best algorithms in

comparison with other two works. We are getting 99% percent accuracy with Local outlier algorithm. It is predicting best results. These algorithms are able to handle large and overfit data. These algorithms are able to predict higher accuracy in comparison of other two works. Each comparison works defines the user interest on social media with the use of machine learning. But their results are not effective in comparison of our results. We have used the best algorithms and technique to predict the user interest and accuracy.Based on the analysis in this research work it was concluded as there is getting higher accuracy from Local outlier factor algorithm in comparison of other two algorithms. However, both models demonstrated no signs of over fitting, with the test r^2 values all being close to or higher than the train r^2 value.

Hence, we have acquired the algorithms in methodology section to accurate value of Detection of credit card fraud with a random method of forest with new enhancements. This proposed module is implemented for the larger dataset compared to current modules and offers more precise findings from other related studies. The Local Outlier Factor algorithm, Random forest algorithms and Support vector machine are the best algorithms for fraud detection. These algorithms are the best and innovative algorithms to detect fraud. It will provide better performance with testing and training results.

References

1. KhyatiChaudhary, JyotiYadav, BhawnaMallick, " A review of Fraud Detection Techniques: Credit Card", International Journal of Computer Applications Volume 45– No.1 2012.
2. Michael Edward Edge, Pedro R, Falcone Sampaio, "A survey of signature based methods for financial fraud detection", journal of computers and security, Vol. 28, pp 3 8 1 – 3 9 4, 2009.
3. Linda Delamaire, Hussein Abdou, John Pointon, "Credit card fraud and detection techniques: a review", Banks and Bank Systems, Volume 4, Issue 2, 2009.
4. L., Delamaire, & Abdou, H. (2009), Credit card fraud and detection techniques: a review. Banks and Bank Systems, 4(2).
5. Adnan M. Al-Khatib, "Electronic Payment Fraud Detection Techniques", World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 4, 137-141,2012.
6. Altman, E. (2001). Bankruptcy, credit risk and high yield junk bonds, Part 1 Predicting financial distress of companies: revisit
7. Zhou W and Kapoor G (2011) Detecting evolutionary financial statement fraud. Decision Support Systems 50, 570-5.
8. Zhang, X., Han, Y., Xu, W. and Wang, Q., 2019. HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. Information Sciences
9. Kumar, M.S., Soundarya, V., Kavitha, S., Keerthika, E.S. and Aswini, E., 2019, February. Credit Card Fraud Detection Using Random Forest Algorithm. In 2019 3rd International Conference on Computing and Communications Technologies (ICCCT) (pp. 149-153). IEEE

10. Mittal, S. and Tyagi, S., 2019, January. Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 320-324). IEEE
11. Shukur, H.A. and Kurnaz, S., Credit Card Fraud Detection using Machine Learning Methodology, International Journal of Computer Science and Mobile Computing, Vol.8 Issue.3, March- 2019, pg. 257-260
12. Tamanna Chouhan, Ravi Kant Sahu, "Classification technique for the credit card fraud detection," International Journal of Latest Trends in Engineering and Technology Vol.(10) Issue(2), pp.283-286, 2018.
13. Malini, N. & Pushpa, M., " Analysis on credit card fraud identification techniques based on KNN and outlier detection," Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB),255-258, 2017.
14. Maira Anis, Mohsin Ali, Amit Yadav. "A Proportional Study Of Decision Tree Algorithms For Class Imbalanced Learning In Credit Card Fraud Detection". International Journal Of Economics, Commerce And Management. Vol. III, Issue 12, December 2015.
15. Wee-Yong Lim, Amit Sachan, Vrizlynn Thing. "Conditional Weighted Transaction Aggregation For The Credit Card Fraud Detection". IFIP International Conference On Digital Forensics. Springer, Berlin, Heidelberg, 2014.
16. Vijayalakshmi Mahanra Rao, Yashwant Prasad Singh. "Decision Tree Induction For Financial Fraud Detection Using En Masse Learning Techniques". Proceeding Of The International Conference On Artificial Intelligence In Computer Science And ICT (AICS 2013). 2013.
17. Parthasarathy, G., Ramanathan, L., JustinDhas, Y., Saravanakumar, J. and Darwin, J., 2019. Comparative Case Study of Machine Learning Classification Techniques Using Imbalanced Credit Card Fraud Datasets. Available at SSRN 3351584.
18. Karthikeyan, K., Raj, K.S., Ramaganesh, S., Parthasarathi, P. and Suguna, N., 2019. Credit Card Fraud Detection Using Machine Learning.