Volume- 5
Issue- I
June-2020

Rajarata University Journal

ISSN: 2362-0080

www.ruj.ac.lk/journals/

Category: Research Article

Fraud Detection Mechanism for Credit Card Transactions of Banks Using Machine Learning

*[1]Lojenaa N & [2]Nawarathna RD

*[1]*Department of Information and Communication Technology, Vavuniya Campus of the University of Jaffna, Vavuniya, Sri Lanka*

[2]*Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka*

ARTICLE DETAILS

Article History
Published Online: 30 June 2020

Keywords
Digital banking, Classification techniques, Statistical features, Fraud transaction, Credit card transaction

***Corresponding Author**
Email:jenaalogi@gmail.com

ABSTRACT

Banks provide their services to make money transactions and other money-based banking activities. The basic concept of the bank is the transaction of money; it can be either manual or digital through multiple channels. Currently, electronic-based money transactions are more popular among customers. Due to the advancement in technologies, perpetrators also get a chance to steal money in unauthorized ways. This study considers fraud transactions among credit card transaction history as a digital forensic footprint to analyze and detect fraud transactions using machine learning techniques. Therefore, a dataset from an American Bank, USA, was used to build a model for fraud detection. The dataset consists of genuine and fraud transactions having 31 attributes each. The dataset was preprocessed, and the statistical features were extracted by applying different sampling techniques. Further, some combination of feature sets is used with different classifiers to create a better model. Classification methods were analyzed through confusion matrix and execution time. Finally, it originates from a suitable model to recognize fraud detection with credit card transaction data. The predicted model produces 99.999% of accuracy in detecting fraudulent transactions using a random forest classifier.

1. Introduction

Currently, credit card transactions, online payment, point of sale payment, ATM, CDM, Kiosk transactions, and many more various types of money transactions are available in banks and other financial institutions to provide their services for customers effectively [1]. The banking industry produces a large amount of sensitive and confidential data in various money-related transactions due to the digitalization of banking activities. In the meantime, the number of fraud transactions also increased due to the development of advanced technologies and tools.

Fraudulent transactions can be analyzed and assessed by several shreds of evidence from banks and other financial organizations such as customer transaction history, credit card transaction history, ATM transaction summaries, bank statements, etc. [2]. A fraud transaction detection mechanism will help to detect fraudulent activities in the banking sector to provide better and secure services to the customers. Credit card transaction history is one of the valuable digital footprints to analyze the patterns between fraud and genuine transactions.

Banking scams are increasing rapidly in the current digital era. The banks are introducing innovative and new banking activities with the latest technologies and smart ideas to attract young beneficiaries. Therefore, not only young customers but also the general public can transform from the traditional time-consuming banking procedures to digital banking and are able to adopt new features efficiently. Meanwhile, perpetrators also get advantages with centralized autonomous activities of the banks [3].

Banking activities have improved on opening of accounts, online transactions, credit, and debit card transactions and other relevant modernized activities. The fraud transactions were made through various attempts such as stolen debit and credit cards, misuse of other cards and online accounts and, the hacked username and password of others [4]. Therefore, detecting fraud transactions is an

essential aspect of identifying and rectifying the losses.

The objectives given below can be achieved with the study proposed:

- To build an efficient model for auditing of credit card transactions to eliminate fraudulent access.
- To build and ensure the policies, rules, and law enforcement.

Data mining techniques and Machine Learning methods are used to detect abnormal behaviours of transactions [5]. The proposed research is mainly focused on detecting fraud transactions of the credit cards in the banking domain using classification techniques with the selected statistical features of the dataset.

A dataset from an American Bank, USA, was used to build a model for fraud detection. The dataset consists of genuine and fraud transactions having 31 attributes each [6]. The dataset was pre-processed, and the statistical features were extracted by applying different sampling techniques. Further, some combinations of feature sets are used with different classifiers to create a better model. Classification methods are analyzed through confusion matrix and execution time. Finally, it originates a suitable model to recognize fraud detection with credit card transaction dataset.

2. Background

Banks are the leading financial institutions of Sri Lanka and have changed their manual works into digital two decades ago. The problems faced by banks and customers due to the digitization of banks were investigated in this study. Banks are the most famous and vital institutions to deliver financial services to the general public, either traditional or online. Banks are doing several promotional activities to serve and attract customers. The banks were improved with a lot of technological facilities step by step. Very recently, banks promoted their key banking activities with new terms such as green banking, paperless banking, self-banking, etc. [1]. Customers can make their transactions by themselves without interacting with employees of the bank in person.

The modern digital methods and technological developments provide more opportunities for both perpetrators and investigators of fraud. This kind of fraudulent activity decreases the hope for adopting modern digitized banking methods. Several banking fraudulent activities can be identified in certain banking activities such as Credit Cards, Deposits, Internet Banking, Loans, Cheque / Demand Drafts, Cash Transactions, Advances, ATM / Debit Cards, etc. [7].

The probability of making a fraud transaction, hacking and unauthorized access was increased due to this digitization of banking activities. Many issues may arise due to unauthorized access. Fraud has evolved from being committed by casual fraudsters to an organized crime [8]. Fraud rings use sophisticated methods to take over control of accounts for committing frauds.

The proposed study contains a digital forensics analysis on the credit card transaction of the banking sector to detect and report fraudulent and unauthorized activities. After any un-authorized access happened, a fraud investigator or audit department has to analyze the digital shreds of evidence [9]. Therefore, the dataset is collected from online and mobile banking platforms as digital information. Finally, a dataset with fraud transactions is gathered based on credit card transactions found in the Kaggle Website [6]. Further, the methods of the research applied to a particular dataset.

This fraud detection mechanism may be beneficial to the banking sector to build customer trust towards online, green banking mechanisms [10]. Further, this mechanism increases the number of customers to the banks and removes the barriers in the application of advanced technologies to improve the banking activities in the long run.

3. Literature Review

Rajdeepa and Nandhitha conducted a study on the banking sector in 2015 [7]. This study includes how data mining techniques influence banking activities such as marketing, risk management, and fraud detection. Such data mining techniques used here are classification, clustering, prediction, association rules, and neural network to analyze the banking data. Banks have examined the following for different needs: Customer's transaction history using data mining techniques to promote their products and services based on the interest of customers. Customer transaction behaviours to issue new credit cards, loans and credit limits to avoid risks. This study provides data mining techniques to analyze customer transaction patterns after any complaints to the banks. Further, this technique was used to analyze the customer's banking interest to promote the latest banking activity, loan facility, credit card facility to low-risk customers and reduce real-time banking frauds [7].

John at el proposed a method in 2016 [11]. The method investigates bank fraud using data mining techniques; association, clustering, forecasting, and classification to analyze the behaviours and patterns

of customers in banking transactions to find the bank fraud in credit card, insurance, accounting, loan, etc. Customers' transaction history and data of loan beneficiaries have been assessed for fraud detection. The analysis was made to predict some common patterns of customers for security enhancement regarding banking activities. For example, adding a threshold limit as his maximum withdrawal amount by examining the maximum withdrawal of a particular customer and associating the occurrences of the maximum withdrawal amount. It asks security questions related to the customer when the customer or others try to withdraw more than the limit set in either counter or the ATM. Analyzing marital status and credit outstanding of the customer from loan records and predict who is eligible to sanction a loan. That means, sanctioning a loan for singles is advisable based on the analysis. This study proposed that the techniques are suitable for knowing a customer very well, giving awareness to the public regarding bank fraud, and ability to save the losses of money for customers and banks. It discourages the fraudsters from continuing new techniques regarding bank fraud [11].

Chaudhary and Mallick [12] had analyzed the data mining techniques for credit card fraud detection. This analysis studied how data mining techniques involve in credit card fraud detection. Authors have chosen the clustering data mining technique in credit card fraud detection to analyze the patterns of customer behaviour from past transactions. Here, the authors analyzed sudden changes in regular patterns using various techniques. Bayesian Classification involves propagating the evidence in the network. Fuzzy logic rules classify the credit card transaction into suspicious and non-suspicious. Credit card fraud detection can be classified into supervised and unsupervised methods. Supervised methods are used to a known dataset and the past events have been studied to detect the fraud. Unsupervised learning has been applied to unknown datasets. However, in this study the Fuzzy Darwinian fraud detection systems prove 100% of the fraud detection rate. The neural network shows good accuracy in fraud detection and gives high processing speed, but it is limited to one-network per customer. Between the number of fraud detection algorithms, neural networks approach is a trend among banks and other institutions [12].

Sunil and Lobo proposed a system in 2012 [13]. This paper proposes a model using Hidden Markov Models (HMM) with a statistical approach to detect and prevent fraud in internet banking. HMM is suitable for this analysis because it can detect the different states of transactions effectively. Authors have proposed an architecture with a valid client who has the authorized internet banking account with the bank, a fraudulent client who doesn't have any internet banking account in a bank, bank server which is dedicated to retrieving the history of the transaction and able to detect the behaviour of customer during transactions. The database is used to store the transactions of the customer. Authors have developed a dummy database with bank account details and designed an online banking facility with TomCat Apache web server to execute the client-server model. Then they have tried some transactions as trial. Based on the transactions, the patterns of the customer during the previous transactions were detected and evaluated with the current transaction using the HMM algorithm to check whether the current user id is original or fake. The customer login to the online banking framework using the proper username and password and then it asks further information regarding the transaction and allows the current transaction with the HMM method. If the method is satisfied, it allows further procedure; otherwise, it confirms the customer with other security questions. Here the transaction amount is observed to detect the behaviour of customers using the HMM algorithm. It helps to detect and prevent unauthorized transactions during online banking activity [13].

Roy at el. [14] analyzed fraud detection mechanisms using deep learning techniques in 2018. This study analyses four different learning techniques to detect credit card fraudulent activity and reduce losses by preventing fraudulent activity. The data was collected from a financial institution engaged in retail banking. The dataset consists of nearly 80 million transaction details for the eight-month duration of credit cards. The dataset contains both transaction history and account details. Here '1' is labeled for the fraud transaction, and '0' is labeled for the non-fraud transaction. Around 0.14% are fraudulent; nearly 99.86% are legitimate among dataset. In the methodology part, they have analyzed four deep learning topologies that include six hidden layers, each with 150 nodes. Artificial Neural Network (ANN) produces the worst result in accuracy score as 0.899 and best performance for learning rate as 0.05. The best performance of the accuracy score of Recurrent Neural Network (RNN) is 0.90433, and the learning rate is 0.5. The Long Short-term Memory (LSTM) shows the best performance inaccuracy score as 0.912, and the learning rate is 0.5. Finally, Gated Recurrent Units (GRUs) make each recurrent unit adaptively capture dependencies of different time scales. It contains the accuracy rate as 0.916, and the learning rate is 0.05. Therefore, GRU conclude a better performance than other topologies [14].

The researchers have analyzed mainly data mining techniques and machine learning models with clustering algorithms. Further, there were several analyses based on recognizing the transaction pattern over time and spending behaviours with possible pieces of evidence. No research identifies the maximum accuracy of fraud detection by up to 99%. Moreover, nobody discussed the time factor to identify fraud detection. Therefore, this literature survey intends to achieve the objectives of this research by around 100% accuracy with the least execution time.

4. Methodology

The selected dataset is greatly uneven and the original meaning of the dataset was hidden, and the real features had changed to numerical format using 'Principle Components' [15] except three attributes, Time, Amount and Class. Column class is labeled either Genuine or Fraud. The 'time' attribute defines the time is taken between two transactions. The 'Amount' is referring to the transaction amount. Other attribute names have hidden their original schema because of security reasons and replaced with V1, V2, V3, up to V28. Altogether 284,807 transactions were there; among them, 492 transactions were labeled as fraud transactions. Therefore, this transaction dataset is pointed out as an unbalanced dataset because the percentage of fraud and genuine transactions are not equal. Altogether the percentage of fraudulent transactions is 0.173%.

The credit card transaction dataset is highly unbalanced; therefore, the resampling technique is adopted to handle the unbalanced dataset. The oversampling technique is used to increase the number of minority class objects to make a balance with the majority class [16]. In the credit card dataset, the genuine class is the majority, and fraud is a minority; therefore, the fraud transactions were increased from 492 to 284,315. Altogether 568,630 (twice of original dataset) transactions are available to proceed further to detect fraud transactions.

Initially, the CSV format of the dataset has been imported into the program. The dataset contains 31 different attributes as columns and 284,807 transactions as rows. Time, Amount, and Class are eliminated from the original dataset. The rest of the twenty-eight attributes (V1 – V28) that are already composed of PCA values analyzed for fraud transaction detection.

The selected attributes (V1 – V28) are applied with different feature measurements in different patterns. The newly composed dataset and the class attributes are labeled as different matrices and those two matrices are split into train and test datasets. Figure 1 illustrates the flow of procedures through this research.

There are several compositions of statistical measures such as Skewness, Kurtosis, Median, Variance, Standard Deviation, and IQR applied to the credit card transaction dataset. Around 57 different compositions were composed to apply for training to detect the fraud transactions. It helps to increase the identification rate of either fraud or genuine transactions in an efficient manner.

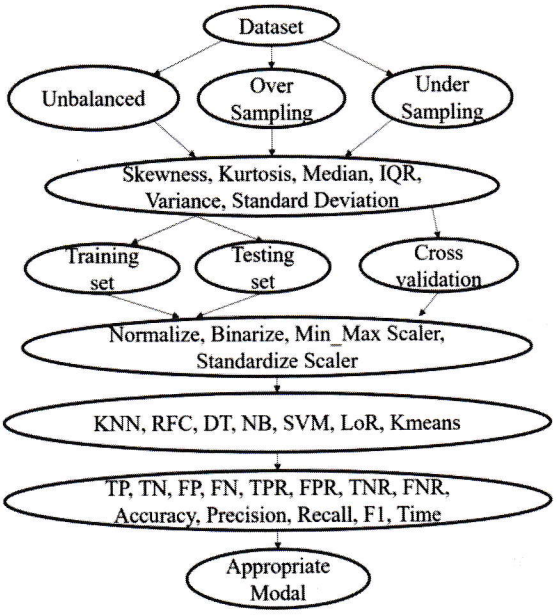

Figure 1. Process of the research

Each composed dataset with statistical features is separated into two parts called training and testing dataset [17].The two types of dataset have been separated 60% of training and 40% of testing due to the low number of fraud transactions (total 284315 genuine transactions and 492 fraud transactions). Altogether around 170589 genuine transactions and 295 fraud transactions were separated in the training part, and around 113726 genuine transactions and 197 fraud transactions expected in the testing part from the total amount of transactions. Figure 2 illustrates the distribution of the proposed credit card transaction dataset.

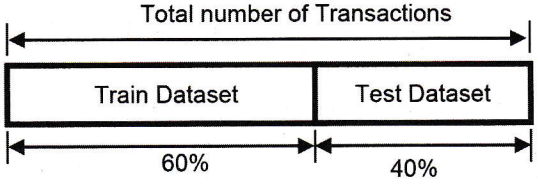

Figure 2. Splitting the total number of transactions

The dataset partition not only focused on the train-test partition but also the dataset was again partitioned into five portions for cross-validation [17] named as CV1, CV2, CV3, CV4, and CV5. The accuracy value is evaluated in each partition of the dataset. Further, among the five different accuracies, the mean value of the accuracies is calculated to identify the average value of accuracy. Figure 3. shows the partition of the dataset into five equal portions for the cross-validation.

Preprocessing techniques such as rescale data, standardize data, normalize data, and binarize data have been applied on the dataset after the split of the dataset into training and testing [18]. The preprocessing technique can remove noisy data before applying a classification algorithm.

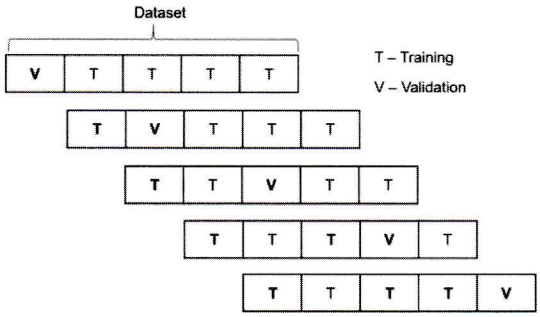

Figure 3. Cross-validation of the dataset

In this approach, several classification algorithms were used to find the best technique with the composed statistical feature of the dataset to suspect the fraud case with high accuracy. Igor Anohhin is suspected around 0.99 of accuracy while detecting fraud using three different data mining techniques: Logistic Regression, Decision Tree, and Self Organizing Map [19].

In this study, seven different data mining techniques, especially, the clustering algorithms such as K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest Classifier (RFC), Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), and K-means (KM) were studied to perform fraud detection [20].

After applying the classification algorithms on the composed dataset, several matrices are identified to predict and calculate as a model to find the fraud transaction. The evaluation of the clustering technique learns a model using the training dataset and predicts the result using a testing dataset. The predicted output retrieves a matrix format of true and false values called the confusion matrix. The illustration of the confusion matrix is shown in Figure 4.

Accuracy, Precision, Recall, F1 measurement, False positive rate (FPR), True positive rate (TPR), False negative Rate (FNR), True negative rate (TNR) are important measurements from confusion matrix to analyze the parameter of fraud detection. Time measures the execution time of the standard classification algorithms [21]. Here, t1 is the beginning time and t2 is the ending time of the classification algorithm.

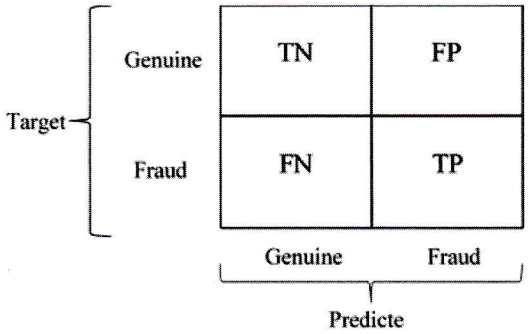

Figure 4. Construction matrix

The measurements are defined as follows [22]:

$$TPR = \frac{TP}{TP + FN}$$

$$TNR = \frac{TN}{FP + TN}$$

$$FPR = \frac{FP}{FP + TN}$$

$$FNR = \frac{FN}{TP + FN}$$

$$Precision = \frac{TP}{FP + TP}$$

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN}$$

$$Recall = \frac{TP}{FN + TP}$$

$$F1\ measurement = 2 * \frac{Precision * Recall}{Precision + Recall}$$

$$Time = t2 - t1$$

ROC curve is defined as Receiver Operating Characteristic Curve and it is used to denote the graphical representation of the difference of true positive rate (TPR) and false positive rate (FPR) thresholds where TPR on y-axis and FPR is on the x-axis.

5. Results and Discussion

Collection of the 57 different datasets were derived from a combination of statistical features and randomly tested with three different ways: the unbalanced dataset, under sampled dataset, and oversampled dataset.

Each tested dataset was divided as 60% for the training and 40% for testing. Table 1 illustrates the number of fraud and genuine transactions available in the unbalanced, under sampled, and oversampled dataset after the separation of the train-test dataset.

Table 1. The number of transactions in train and test dataset

Dataset	Train / Test	Total number of Transactions	Genuine Transactions	Fraud Transaction
Un balanced dataset	Train (60%)	170,884	170,589	295
	Test (40%)	113,923	113,726	197
Under sampled Dataset	Train (60%)	590	295	295
	Test (40%)	394	197	197
Over sampled Dataset	Train (60%)	341,178	170,589	170,589
	Test (40%)	227,452	113,726	113,726

The balanced and unbalanced dataset was created over the associated statistical featured dataset and four different preprocessed techniques applied on it. Moreover, seven different classification algorithms were applied to the 57 types of statistical featured dataset using cross-validation and train-test to evaluate machine learning parameters to identify the accuracy and execution time. Therefore, initially, a confusion matrix was calculated with seven different classification algorithms on the dataset. After that, necessary calculations such as TPR, FPR, TNR, FNR, Accuracy, Precision, Recall, F1 measurement, and Time are calculated according to the confusion matrix.

Finally, oversampling techniques almost provides good performance in various classifiers with a different combination of statistical features. DT and RFC classifiers show almost 100 % of performance from the ROC curve in the normalized preprocessing technique as shown in Figure 5.

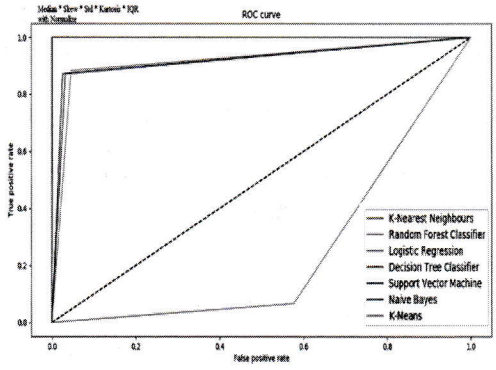

Figure 5. ROC curve on Median, Skew, Kurtosis, IQR and Standard Deviation with Normalized preprocessing scaler

Decision tree, Random forest classifier, and Support vector machines show high performance in binarized preprocessing technique. Again both DT and RFC show high accuracy in minimum and maximum scaler. KNN shows high performance in Standardized preprocessing scalers.

Table 2 shows the accuracy and the other relevant measurement on oversampled datasets with various classifiers and preprocessing techniques. The accuracy shows 0.9999 percentage on normalized preprocessing techniques and 0.9994 percentage on binarized preprocessing technique with RFC from the measurements. Further, RFC with a normalized preprocessed technique shows different results for two different sets of the combined dataset with statistical features. One set of dataset shows 0.9999 on precision and F1 measure and 100% on recall. The other set of dataset shows 0.9998 on precision, 0.9999 on F1 measure, and 100% on recall. The first combinations of RFC normalized feature took12.3739 seconds, and another combination of the same algorithm took12.7020 seconds.

Table 2. Result of the combined statistical features with various preprocessing techniques on an oversampled dataset

Features	Median * Skew * Std * Variance	Median * Skew * Std * Kurtosis * IQR	Kurtosis* Std * IQR * Variance
Classifiers	RFC(Normalize)	RFC(Normalize)	RFC(Binarize)
TN	113686	113645	113547
FP	18	13	129
TP	0	0	0
FN	113748	113794	113776
TPR	1.0000	1.0000	1.0000
FPR	0.0002	0.0001	0.0011
TNR	0.9998	0.9999	0.9989
FNR	0.0000	0.0000	0.0000
Accuracy	0.9999	0.9999	0.9994
Precision	0.9998	0.9999	09989
Recall	1.0000	1.0000	1.0000
F1	0.9999	0.9999	0.9994
Time	12.3739	12.7020	3.4149

Further, Table 3 shows the accuracy value of each partition of the oversampled dataset using cross-validation technique. It shows 100% in a different cross-validation partition with a different combination of statistical values. That is Median, Skewness, Standard deviation, and Variance Combined dataset gives 100% on CV4 with normalized preprocessing technique. Except for binarized preprocessing technique with Kurtosis, Standard deviation, IQR, Variance combined dataset, other datasets give nearly 0.9999 accuracy value. Further, almost every combination of datasets gives better results using the Random Forest Classifier.

Table 3. Accuracy and the mean value of the accuracies of five-fold cross-validation on oversampling technique

Features	Median * Skew * Std * Variance	Median * Skew * Std * Kurtosis * IQR	Kurtosis* Std * IQR * Variance
Classifiers	RFC (Normalize)	RFC (Normalize)	RFC (Binarize)
CV1	0.9999	0.9999	0.9995
CV2	0.9999	0.9999	0.9994
CV3	0.9999	0.9999	0.9992
CV4	1.0000	0.9999	0.9995
CV5	0.9999	0.9999	0.9995
Mean of the Accuracies	0.9999	0.9999	0.9994

The Normalized preprocessing technique with oversampling dataset gives equal or above 0.9999 on Accuracy, Precision, F1 measure, and Recall with a specified combination of statistical features Median, Skewness, Kurtosis, IQR, and Standard deviation. The time taken for above mentioned featured dataset is 12.7020 seconds, which is the adequate time compared to other normalized preprocessing techniques with RFC. Further, cross-validation analysis also gives 0.9999 accuracies in each partition. Compared to other features, the combination of the dataset mentioned above is more suitable as it gives 0.9999 accuracy using the training-testing dataset and the mean value of the cross-validation also gives 0.9999 of accuracy.

6. Conclusion

Fraudulent banking activities are the biggest headache in the current world due to technological advancement. Fraud transactions make customers lose their hope in the banking industry. Therefore, the fraud detection mechanism with higher accuracy and quick predicting technique is essential in the banking industry to attract more customers, promise their security and increase their status among the public and other financial institutions.

The credit card transaction dataset is used to carry out this research to predict a suitable model for fraud detection. The main part of this research is the statistical feature extraction on the dataset with Median, Skewness, Kurtosis, IQR, Variation, and Standard deviation. The formation of a different combination of the statistical features also used to apply in the dataset. Therefore, around 57 datasets used to test fraud detection with higher accuracy.

Under-sampling and oversampling techniques are used to resample the dataset to balance genuine and fraud transactions. Further preprocessing techniques such as normalization, binarization, standardization scale and, minimum and maximum scale are used to clean the raw data.

The datasets are divided into training and testing for learning the behaviour and identifying the pattern of fraud and genuine transactions and then testing for predicting the new transaction as fraud or genuine. Therefore, 60% of the dataset are considered for training and 40% of the dataset for testing and thereafter the classification algorithms such as KNN, RFC, DT, NB, SVM, LR, and KM clustering are used to detect the fraud transactions. Further, cross-validation is also applied by separating the dataset into five portions and five times accuracy has been calculated and the accuracy value is finalized with the mean of accuracies.

The confusion matrix from each classifier is formulated with various types of datasets. Accuracy, Precision, Recall, F1measure, FPR, TPR, FNR, TNR and the execution time have been calculated. The ROC curve is used to analyze and compare the performance between classifiers.

The normalizer shows better performance compared to other preprocessing techniques on the various statistical featured dataset. The Oversampled dataset gives more accuracy than an unbalanced, and under-sampled dataset. A combined statistical featured with a normalized and oversampled dataset gives 0.9999 in accuracy, precision, and F1 measurement and 100% on recall with proper execution time. ROC curve shows better performance on RFC and DT rather than other classifiers.

Finally, a proper model for fraud detection is chosen from the analysis with the random forest classifier and the different combinations of statistical features such as Median, Skewness, Kurtosis, IQR and Standard deviation with the oversampled dataset and normalized preprocessing technique gives 0.9999 of accuracy, precision, F1 measure and 100% of recall and also with adequate time 12.7020 seconds. Therefore, it has been chosen as the finalized best model from this research.

Moreover, Oversampled dataset with normalizer, minimum and maximum scaler, and standardized scaler give an average of 0.9999 of accuracy using cross-validation technique. Several partitions of the dataset of cross-validation give 100% accuracy. Thus, above mentioned statistical feature based oversampling dataset gives better result in all type of measurement. Therefore, it shall be concluded that this model is suitable to detect the fraud on credit card transactions more accurately and much faster.

References

1. ATM operations support services expression of interest (EOI), State Bank of India ATM operations department, Global IT Centre, CBD Belapur, Navi Mumbai, 400614.
2. Adam PK, Imam R, Ahmad L. Mobile Forensics Development of Mobile Banking Application using Static Forensic. International Journal of Computer Applications 2017; 160(1):5-10.
3. Madan LB, The Role of Technology in Combatting Bank Frauds: Perspectives and Prospectives. ECOFORUM 2016; 5(2): 200-212.
4. Sayali KR. Study of data mining on banking database in fraud detection techniques. International Research Journal of Engineering and Technology 2016; 3(5):1831-1835.
5. Safia A. Deposit subscribe Prediction using Data MiningTechniques based Real Marketing Dataset. International Journal of Computer Applications (0975 – 8887) 2015; 110(3).
6. Credit card Fraud Detection [Internet]. Available from: https://www.kaggle.com/mlg-ulb/creditcardfraud. (Accessed on: 28/01/2019).
7. Rajdeepa B, Nandhitha D. Fraud Detection in Banking Sector using Data mining. International Journal of Science and Research 2015; 4(7):1822-1825.
8. Roohollah FN. The Fraud Detection in the Bank Payments and its Methods. Asian Journal of Information Technology, Medwell Journals. 2015; 14(6): 239-245.
9. Natasa S, Gojko G, Nenad R. Forensic Accounting in the Fraud Auditing Case. The European Journal of Applied Economics. 2016; 15(2): 45-56.
10. Rodrigo C, Michael G, Sadie C.Applying Semantic Technologies to Fight OnlineBanking Fraud. 2015 European Intelligence and Security Informatics Conference, IEEE Computer Society. 2015; 61-68. DOI 10.1109/EISIC.2015.42
11. John SN, Anele C, Okokpujie KO, Olajide F, Chinyere GK. Real-time fraud detection in the banking sector using data mining techniques. International Conference on Computational Science and Computational Intelligence (IEEE). 2016;1186-1191.
12. Choudhary K. Mallick B. Exploration of Data mining techniques in Fraud Detection: Credit Card. International Journal of Electronics and Computer Science Engineering 1(3):1765-1771.
13. Sunil SM, Lobo LMRJ. Internet Banking Fraud Detection Using HMM. ICCCNT'12 (IEEE-201S0). 2012;
14. Roy A, Jingyi S, Robert M, Loreto A, Stephen A, et al. Deep Learning Detecting Fraud in Credit Card Transactions. Systems and Information Engineering Design Symposium (SIEDS). 2018;129-134.DOI: 10.1109/SIEDS.2018.8374722
15. Balasupramanian N, Ben GE, Imad SAB. User Pattern Based Online Fraud Detection and Prevention using Big Data Analytics and Self Organizing Maps. 2017 International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) (IEEE). 2017; 691-694.
16. Resampling strategies for imbalanced datasets [Internet]. Available from: https://www.kaggle.com/rafjaa/resampling-strategies-for-imbalanced-datasets [Internet]. (Accessed on: 26/01/2019).
17. Train/Test Split and Cross-Validation in Python [Internet]. Available from: https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6. (Accessed on: 22/01/2019).
18. Data Preprocessing for Machine learning in Python [Internet]. Available from: https://www.geeksforgeeks.org/data-preprocessing-machine-learning-python/. (Accessed on: 22/01/2019).
19. Igor A. Data mining and machine learning for fraud detection. Faculty of Information Technology, Tallinn University of Technology. 2017;
20. Sagar BB, Pratibha S, Mallika S. Online Transaction Fraud Detection Techniques: A Review of Data Mining Approach. 2016 International Conference on Computing for Sustainable Global Development (INDIACom) (IEEE). 2016; 3756-3761.
21. Kesavaraj G, Sukumaran S. A Study On Classification Techniques in Data Mining. 4th ICCCNT – 2013 (IEEE). 2013;
22. Krishna M, Reshma D. Review On Fraud Detection Methods in Credit Card Transactions. 2017 International Conference on Intelligent Computing and Control (I2C2'17). 2017

