Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6, June 2021: 664-674

An Augmentation of Credit Card Fraud Detection using Random Undersampling

Sandeep Kumar Nayak[a*], Vipin Khattri[b]

[a*] Department of Computer Application, Integral University, Dasauli, Bas-ha Kursi Road, Lucknow – 226026, India.

***Corresponding author:** vipinkhattri@gmail.com; nayak.kr.sandeep@gmail.com

Abstract

The two most prominent necessities which triggered the advent and growth of Digital Transactions are ease of doing payments and security of the transactions. However, even after a lot of research into the field, financial cyber crime is still marred with lots of digital frauds and corruption; credit card fraud is one of them. A plethora of patents and research papers have tried to solve this issue, but the secure transaction is still a distant dream. As per a survey, $24.26 Billion was lost worldwide due to payment card fraud only in 2018. This study aims to augment the performance of credit card fraud detection using an undersampling based data analysis technique. The study used an ensemble method to detect credit card fraud. The data used for simulation was highly imbalanced. Hence, a random undersampling technique was applied to datasets to make it a balanced dataset. The validation of performance augmentation was done based on the predefined performance measure metrics such as accuracy, precision, recall, f-score, geometric-mean, and the area under curve score with the receiver operating curve. The main focus was to check the ensemble method's performance on imbalanced data and improve it by providing it a balanced data. The study results showed that the augmentation of the detection technique of credit card fraud was improved with a random under-sampling method on a credit card transaction imbalanced dataset. The study performed a comparative analysis of the augmentation of detection technique of credit card fraud models before and after incorporating random under-sampling techniques on credit card fraud imbalanced datasets.

Keywords: credit card fraud detection; classification method; random undersampling; random forest classifier; imbalanced dataset

1. Introduction

Credit card fraud is a severe offense against an authorized person with an objective of unauthorized purchase or money transfer or takes cash [1-2]. The most straightforward meaning of credit card fraud is "committing fraud using a payment card" [3]. Credit card fraud's primary purpose is to purchase the product or obtain services or deliver money fraudulently [4]. Fraud related to credit card takes place when hackers steals all information related to credit card [5] by using illegal methods [6-7] like identity theft or counterfeit card or card stolen or card skimming or account takeover or data breach or phone fraud or dark web or fake website or fake application [8]. Fraud related to credit card may happen in both cases; either card is present (card-present fraud) or not preset (card-not-present fraud). In card-not-present fraud, the thief steals data related to credit cards and makes online payments on portals where the credit card is not required, only data related to credit card requires authentication of payment. Credit card data can be obtained with the help of fake applications, identity theft, data breach, account takeover, phone fraud, dark web, or fake websites. In card-present fraud, the thief steals a

credit card physically and makes payment where the purchase card is required. The physical credit card can be obtained with the help of activities like card lost or stolen, counterfeit card, card skimming, point of sale fraud, or ATM fraud [9]. In all the case, the authorized user only realizes the fraud when he receives the payment statement. Activities related to the stolen card or counterfeit card and skimming card had increased. The fraud pattern is changing [10], and presently, account takeover fraud, application fraud, and social engineering fraud are the most common types of fraud in this industry. Many customers face the consequences of credit card fraud that not only results in loss of customer amount but also affects the country's economy [5, 11-12].

The United Kingdom faced a loss of around £ 844.8 million in 2018 due to the payment card, cheques, and remote banking [13]. In 2016, a $12.7 million amount was breached by stealing data of the African credit cards in Tokyo, and thieves took only 3 hours for completing all activities. As per Australian payment card fraud report [14], a total $574 million amount was lost using payment card fraud in 2018. The primary type of card fraud here was card-not-present (figure 1).

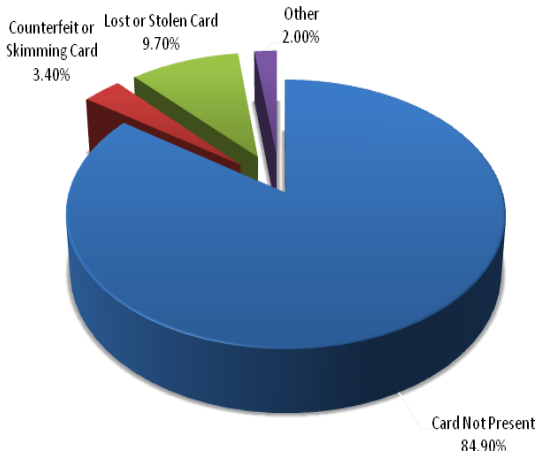


Figure 1. Percentage of Payment Card Fraud Categories [14].

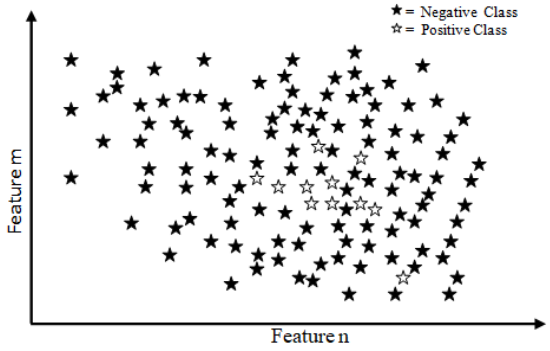


Figure 2. Imbalanced Data

The detection of credit card fraud can be contained using one of the two different paths. The first way is by using an authentication process, and the second is to use a wholesome credit card fraud detection model (CCFDM). The CCFDM [15] is an automatic process to categorize the payment between genuine and fraud using previous transaction patterns [16] using a classifier related to a machine learning algorithm. CCFDM is developed concerning training using a credit card fraud dataset of previous transactions [17]. In real life, most fraudulent transactions are significantly less than genuine transactions [18-20]. The interpretation is that the dataset of a genuine and fraudulent transaction is imbalanced [21]. Classifiers, which are an integrated part of CCFDM, were unable to produce correct results if trained using an imbalanced dataset [22]. Hence, most of the CCDM struggles with accuracy.

The imbalanced data is an imbalanced class problem (figure 2) [23], which means that one class's instances are less than the second class's instance. The imbalanced dataset can be found in many areas of real-life [20], such as medical (cancer), information technology (network intrusion), and security (data breach). As per the argument given in last paragraph, training of CCFDM should be on a balanced credit card fraud dataset. The

imbalanced dataset can be change to a balanced one using the undersampling technique. This study's objective is to improve the CCFDM's performance using a random undersampling technique. This study produces a comparative analysis to measure the CCFDM's performance with an imbalanced and balanced credit card fraud dataset.

To achieve the objective of the proposed study, the complete research paper has separated into six sections. The second section reviews the various methodologies used in augment the performance of CCFDM. The third fragment defines the method utilized in the study to achieve the objective. The fourth part elaborates on the complete experimental setup (procedure dataset and performance measures) to achieve the study's objective. The fifth section provides comparative results and analyzes the result using data, facts, and figures. The sixth section presents the study's objective and results and defines the study's future scope

2.Related Work

This section analyzed the previous works related to improvement in credit card fraud detection and imbalanced dataset. The different research papers were studied to find the different techniques of credit fraud detection and to handle imbalanced dataset. The related work and gave a concrete direction to move forward.

The study of Li [24] focused on improving the performance of the CCFDM. The improvement was achieved by developing a new loss function named as a full center loss. The full center loss function was designed to find the difference between genuine and fraudulent transactions based on feature differences by considering distance and angle. The study performed an experiment using two different datasets of the credit card transaction. The result of the study showed that the proposed function performed superior as compared with other models.

Zhu [25] studied the detection of credit card fraud and handled the imbalanced dataset using a weighted extreme learning machine algorithm. Zhu found two different features that impacted on the performance of the system. The algorithm proposed by the proposed work was compared with the numerous optimization algorithms such as genetic algorithm, bat algorithm, swarm algorithm, and self learning and concluded that the proposed algorithm worked perfectly with the dandelion algorithm. The proposed algorithm showed the achievement of high performance concerning the credit card fraud detection.

The study of Rtayli [26] focused on developing an improved CCFDM with an imbalanced dataset. The proposed model was developed based on a hybrid approach, including recursive feature elimination used in selecting significant features, GridSearchCV used for optimization, and synthetic minority oversampling method utilized to grip imbalanced datasets. The author tested the proposed CCFDM on a real dataset and found good results to support the study.

Study by Mittal and Tyagi [27], reviewed various detection methods of credit card fraud. The study discussed the various safety issues in using the credit card and explained various solutions of credit card fraud. The authors also discussed some significant problems such as imbalanced dataset, classifier performance, and concept drift for credit card fraud detection. The authors also proposed new challenges and directions for detection of credit card fraud.

The study of Makki [28] discussed the current issue with detecting credit card fraud using an imbalanced dataset. The authors elaborated on various solutions to CCFDM based on machine learning. The authors conducted an experimental study to demonstrate the degradation in detecting credit card fraud results with an imbalanced dataset.

The study of Randhawa [29] discussed the unavailability of the real credit card dataset. Firstly, the author developed a CCFDM based on the machine learning algorithm. After that, the authors developed another CCFDM based on a hybridization of AdaBoost and the majority voting algorithm. The authors tested the proposed hybrid CCFDM using a credit card fraud dataset. The results of the experiment showed the better performance to detect credit card fraud.

The study of Roy [30] applied deep learning approach to developing a credit card fraud detection system based on credit card fraud transactions' historical data. The authors used long short term memory with different parameters using pre-labeled credit card transaction. The authors also handled the problem of credit card fraud imbalanced dataset and presented a framework for deep learning to tune the CCFDM's parameters.

The study of Pumsirirat and Yan [31] analyzed and discussed fraud transactions that were not detected earlier using the CCFDM. The authors developed a model based on a restricted Boltzmann machine and Auto-encoder to find the regular transactions' anomalies. The authors also applied a back propagation algorithm to the model. The authors evaluated the model's performance with standard performance measures such as the area under the curve, root mean squared error, and mean squared error.

3.Methodology

The proposed study's main objective is to strengthen the performance of the CCFDM. It is observed that the percentage of fraud of credit card transactions is less than the rate of genuine credit card transactions. Therefore, the fraud of credit card dataset is an imbalance, and the credit card dataset's imbalanced nature degrades classifier performance. A random undersampling technique was used to handle the dataset's imbalanced nature, and it was implemented as CCFDM with a random forest classifier.

3.1.Random Undersampling Technique

This mechanism's scope is to balance the dataset due to class imbalance issues by reducing the degree of bias. The mechanism of random undersampling eliminates the observations from the majority class of the dataset [22]. Observations of the majority class from the dataset are selected randomly and eliminate from the dataset. Random under-sampling aims to reduce the majority class's observation level to the minority class level, and the following figure 3 shows an algorithm of random undersampling technique.

$$|D_{mod}| = |A_{min}| + |B_{maj}| - |C_{maj}|$$

$$|D_{mod}| < |D_{org}|$$

D_{org}= Original dataset before applying random undersampling

D_{mod} = Modified dataset after applying random undersampling

A_{min} = Total number of observations of the minority class

B_{maj} = Total number of samples of the majority class

C_{maj} = Total number of samples removed from the majority class.

Figure 3. Algorithm of Random undersampling [22] (He and Garcia, 2009)

Reducing the observation from the majority class is repeated until the total number of observations of the majority class becomes equal to the total number of observations of the minority class (figure 4). This mechanism is straightforward and suitable for those datasets in which sufficient observations of minority class is available for training the classifier. This method has a disadvantage also. When observations are eliminated from the majority class randomly, it is possible to lose boundary class information from the majority class and degrade the classifier's performance. While removing the samples, it is possible to eliminate observation, which has rich information for classification. There is no mechanism in random under-sampling to differentiate between significant and non significant observations.

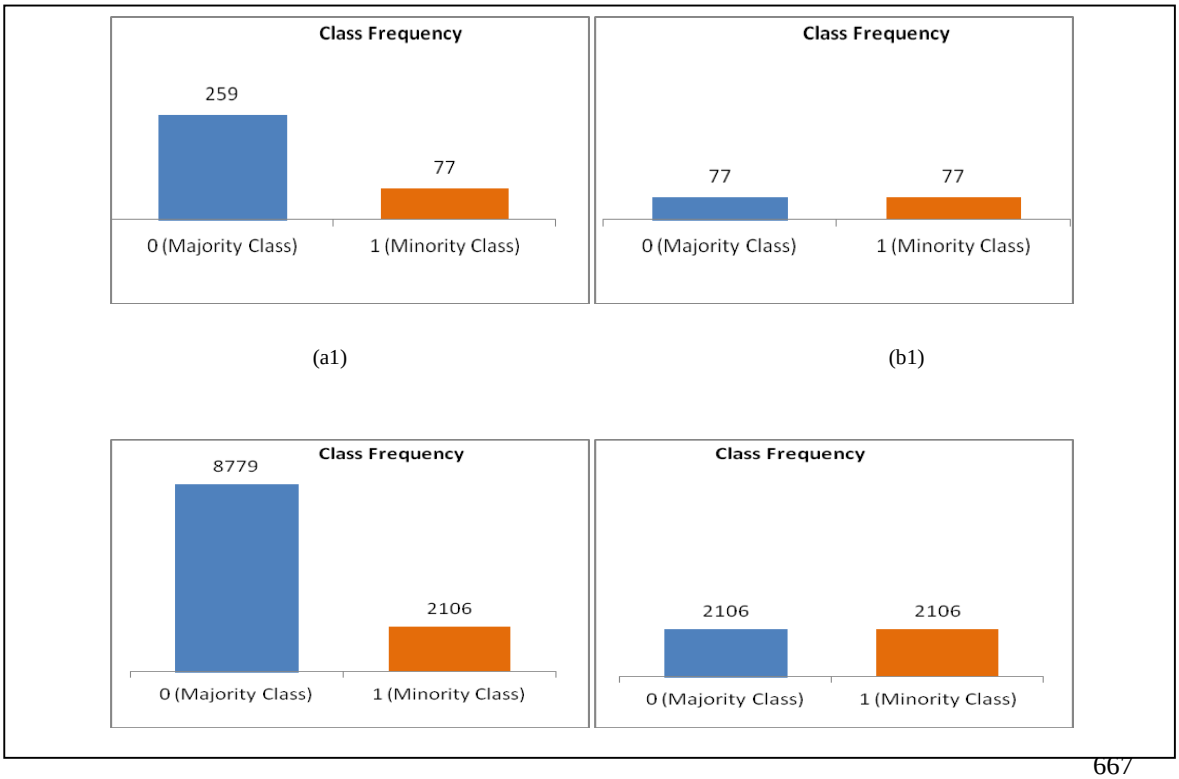


Figure 4 (a1) and (a2) Imbalanced dataset and (b1) and (b2) Balanced dataset after random undersampling

3.2.Random Forest

Fraud detection of credit card is a binary classification job. Therefore random forest classification model was used to perform classification tasks for detection of credit card fraud. A random forest (RF) is used in classification and regression problems [11]. RF is a combination of multiple decision trees [32-33]. RF's basic working is that it creates many decision trees with different attributes and samples [34]. These multiple decision trees are randomly selected to combine and create a robust classifier known as an RF. Individual decision tree classifier is weak, but these weak decision trees combine to make a robust classifier (figure 5).

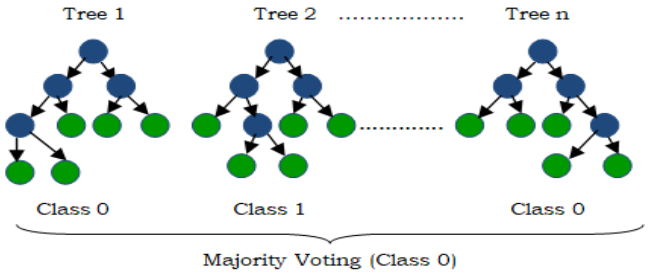


Figure 5. Random Forest Classifier

The final result of an RF classifier is the mean of all the randomly selected decision trees' predictions. RF performs accurately without scaling of features of numerical features and encoding of categorical features. The randomness of the decision tree is of two types. The first type of randomness is chosen based on samples, and the other randomness is determined based on different subsets of features. The fundamental objective behind combine the decision tree is to lower their variance

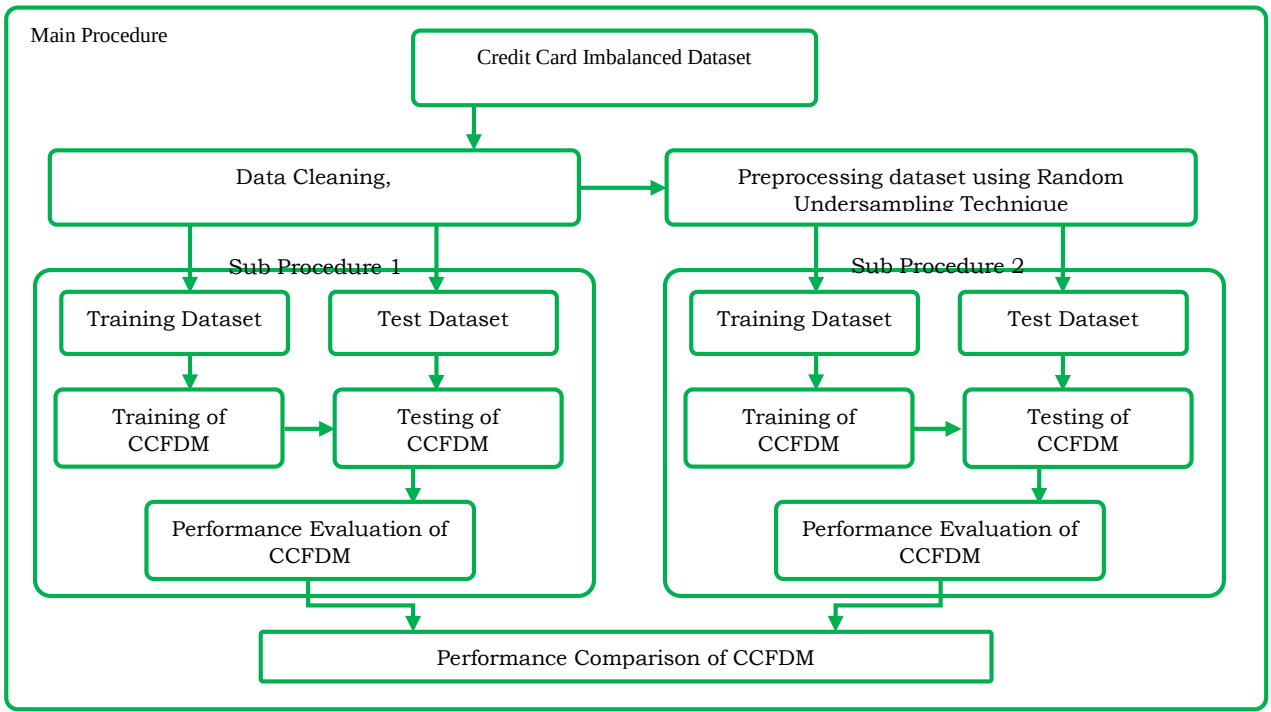
3.3.Experimental Setup

The study's experimental setup included execution procedure, credit card fraud dataset, performance metrics, and methodology (random undersampling and random forest defined in section 3.1 and 3.2)

3.4.Procedure of Experimen

The main procedure is presented in figure 6, which is needed to carry out the complete experiment. The core procedure was further subdivided into two sub-procedures. The first sub-procedure was defined to find the performance of the CCFDM using a random forest classifier based on a credit card fraud imbalanced dataset. The second sub-procedure is designated to find the CCFDM's performance using a random forest classifier based on the credit card fraud balanced dataset, after applying a random undersampling technique on the credit card fraud imbalanced dataset. Data cleaning such as missing values, duplicate rows, normalization, and scaling was done on the credit card fraud dataset before executing the sub-procedures.

Figure 6. The main procedure of execution of the experimental setup



Performances were calculated after completing the sub-procedure1 and sub-procedure2. Finally, the performance of both the sub-procedures was compared to identify that which credit card fraud detection model performed better before and after applying the random undersampling technique on the credit card imbalanced dataset.

3.6 Dataset

Dataset plays an essential role in building an accurate credit card fraud detection model by training along with testing the model. Using the dataset, the study finds the accurateness of the model. The proposed comparative study used a credit card dataset for achieving an objective of the study. The dataset contained the transaction of credit cardholders of Europe [1] [35]. The dataset possessed the following characteristics (Table 1).

Table 1. Characteristics of Credit Card Fraud Dataset

S.No.	Characteristics	Value
1	Total Number of Transactions	2,84,807
2	Total Number of Transactions (Negative Class – Genuine)	284,315
3	Percentage of Genuine Transaction	99.83%
4	Total Number of Transactions (Positive Class- Fraudulent)	492
5	Percentage of fraudulent Transaction	0.17%
6	Total Variables in Number	31
7	Data Type of Variables	Numeric
8	Target Value	0 = Genuine 1 = Fraudulent

The credit card fraud dataset was separated into two sections. The first division was the training dataset (70%) utilized for training the fraud detection model of credit card, and second category of a dataset (30%) was the testing dataset used to test the performance and accuracy of the CCFDM.

3.7.Performance Assessment

Fraud detection of credit card is a binary classification task that finds a fraudulent transaction among genuine and imitation transactions. The evaluation of the credit card fraud detection classification task was achieved by using confusion metrics [11-12, 18, 36] and performance metrics [11, 23, 36]. The confusion matrix classifies the transaction into one of the four categories (figure 7). The first category is a true positive (TP) that classifies how many genuine transactions are classified as genuine transactions. The second category is a true negative (TN) that classifies how many fraudulent transactions are classified as fraudulent transactions. The third category is a false negative (FN) that classifies how many genuine transactions are categorized as fraudulent transactions. The fourth category is a false positive (FP) that classifies how many fraudulent transactions are classified as genuine transactions. A confusion matrix was created after testing the credit card fraud detection model's performance using a testing dataset of credit card fraud. All transactions were classified as per the confusion matrix, and on the arrangement categories of the confusion matrix, performance of the credit card fraud detection model was measured using the following five performance metrics (figure 7), the area under curve score (AUC) score and receiver operating curve (ROC).

Confusion Matrix

Actual	Prediction	
	True	False
True	TP (True Positive)	FN (False Negative)
False	FP (False Positive)	TN (True Negative)

Performance Metrics

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

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

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

$$\text{F} - \text{score} = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

$$G - mean = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{TN + FP}}$$

Figure 7. Confusion Matrix and Performance Metrics

The credit card fraud dataset is highly imbalanced; therefore, the accuracy metric alone is not sufficient to find the CCFDM's performance. The accuracy of the CCFDM produces the output inclined towards the negative majority class. This inclination can be handled using AUC score, F-Score, recall, precision performance metrics. The precision performance metric is utilized to find the exactness of the model. The F-score demonstrates the steadiness between sensitivity and precision. The geometric mean is a performance measure metric that finds the balance between the fraudulent and genuine transactions based on the classification performances.

4.Results and Discussion

The study was executed under a set of procedures using a random forest classifier as a CCFDM, random undersampling to tune imbalanced dataset, and the credit card dataset to test the CCFDM's performance. The dataset utilized in the credit card fraud detection model was imbalanced under the degree of bias category with the positive class (0.17%) and negative class (99.83%). The credit card fraud dataset's distribution of data patterns can be seen in figures 8 and 9 as an imbalance between positive and negative class before and after implementing a random undersampling technique. The results of the study were produced and compared based on various performance measures as shown in Table 2.

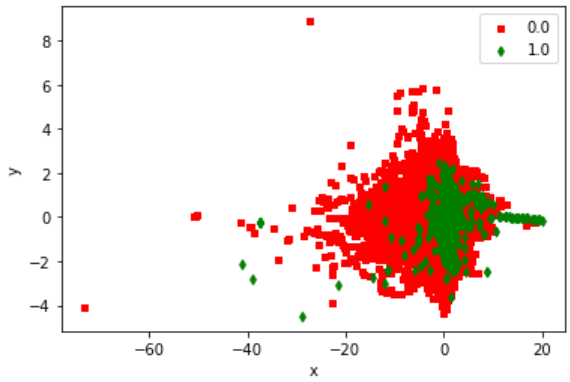


Figure 8. Pattern of data distribution of credit card fraud imbalanced dataset before implementing random undersampling technique

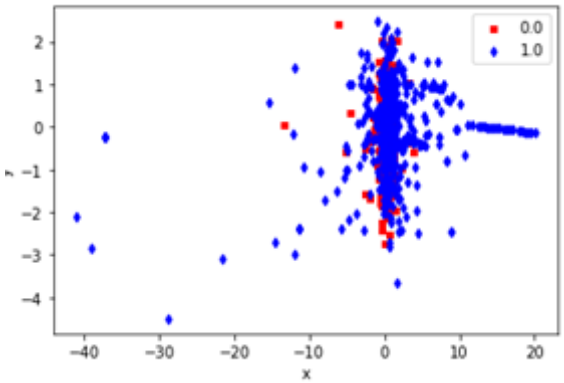


Figure 9. Pattern of data distribution of credit card fraud imbalanced dataset after implementing random undersampling technique

Table 2. Performance of CCFDM using Random Forest Classifier based on performance metrics

S.No.	Performance Measures	Implementing a random undersampling technique on credit card fraud imbalanced dataset	
		Before	**After**
1.	Accuracy	0.99905	0.92905
2.	Precision	0.85870	1.00000

3.	Recall	0.53741	0.85906
4.	F Score	0.66109	0.92419
5.	Geometric mean	0.73303	0.92686
6.	AUC Score	0.76863	0.92953

Performance comparison of the CCFDM before and after implementing a random undersampling technique on the credit card fraud dataset is shown in figure 10, based on the results generated during the execution of an experiment. The comparison graph is showing performance improvement under all the performance measures except accuracy. Performance measure (accuracy) of the credit card fraud detection model decreased after implementing a random undersampling technique, and it was done because accuracy is biased towards the negative (majority) class and accuracy of the model shown by considering the negative (majority) class. Therefore, an accurate picture cannot be seen by only the model's accuracy metrics, and the result was not balanced with respect to the positive and negative class. Therefore, the proposed comparative study used various standard performance measures that showed the credit card fraud detection model's performance improvement by making the balance between positive (minority) class and negative (majority) class. The most significant area under curve score metrics and operating receiver curve (figure 11 and 12) showed remarkable results.

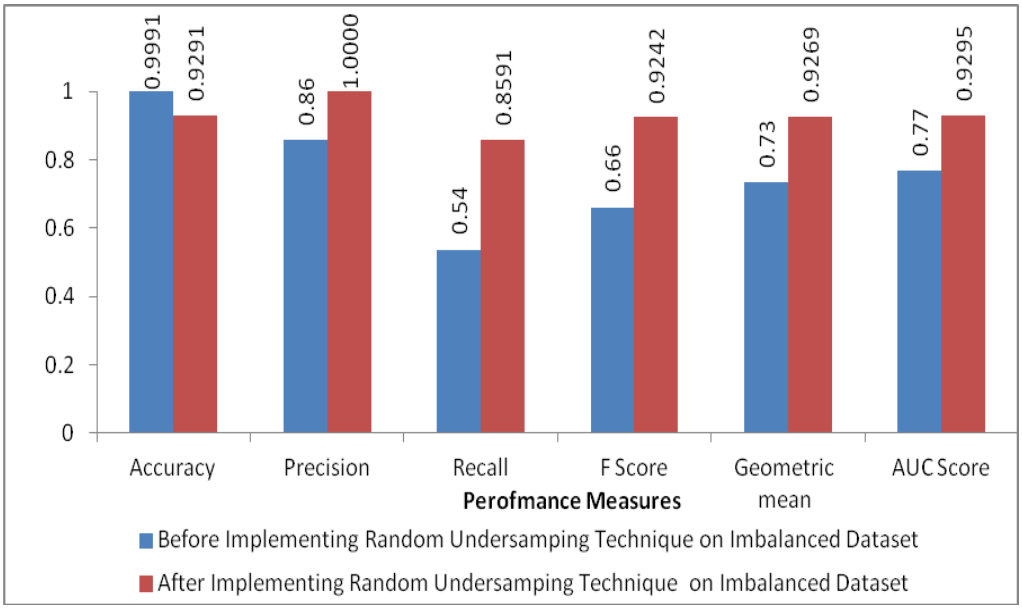


Figure 10. Performance of CCFDM using a Random Forest Classifier based on performance metrics

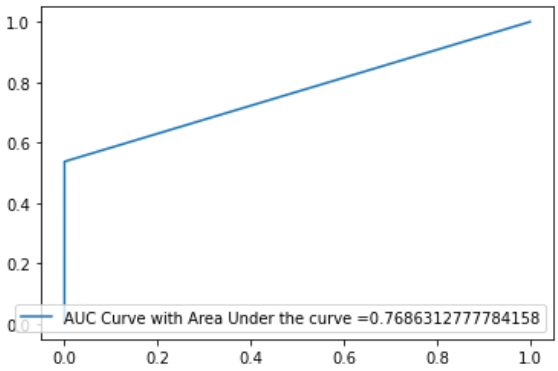


Figure 11. AUC score and ROC curve of CCFDM using credit card fraud imbalanced dataset before implementing a random undersampling technique

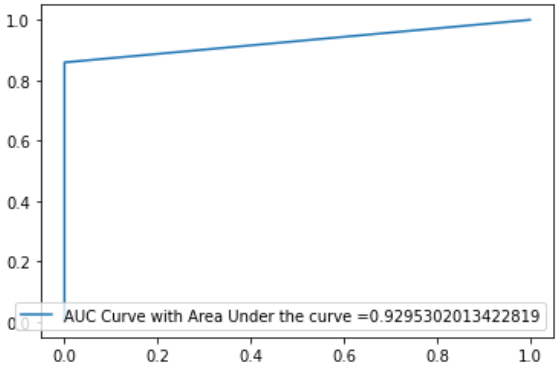


Figure 12. AUC score and ROC curve of CCFDM using credit card fraud imbalanced dataset after implementing a random undersampling technique

5.Conclusion and Future Scope

The study analyzed the credit card fraud and problems associated with an imbalanced dataset that degraded the credit card fraud detection system's performance. The study reviewed literature based on handling the imbalanced dataset using the undersampling technique and improved classification task performance. The study performed a comparative analysis by comparing the CCFDM's performance using random forest before and after implementing the random undersampling technique on credit card fraud imbalanced dataset. This study ensured the CCFDM's performance improvement using standard performance metrics such as the area under curve, receiver operating curve, geometric mean, f-score, recall, and precision. The purpose behind taking various performance metrics was to ensure the performance in all aspects. The result of the study showed remarkable improvement. The CCFDM's performance improvement was found to about a 20% increase (AUC score=0.76863 before implementing random undersampling on an imbalanced dataset; AUC-score=0.92853 after implementing random undersampling on an imbalanced dataset). The result was remarkable and supported the study. In the future, the study will work further on performance augmentation of the CCFDM by handling the outlier and feature engineering of an imbalanced dataset.

Acknowledgements

We give our thank and gratitude towards the Integral University for supporting research work and providing Manuscript Communication Number-IU/R&D/2020-MCN001000

References

[1] A. Dal Pozzolo, O. Caelen, Y. A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert systems with applications*, Vol. 41, No. 10, pp. 4915-4928, 2014

[2] X. Zhang, Y. Han, W. Xu, and Q. Wang, "HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture," *Information Sciences*, Vol. 557, 2019

[3] V. Khattri and D. K. Singh, "Parameters of automated fraud detection techniques during online transactions," *Journal of Financial Crime*, Vol. 25, No. 3, pp. 702-720, 2018

[4] S. B. E. Raj and A. A Portia, "Analysis on credit card fraud detection methods," *IEEE, International Conference on Computer, Communication and Electrical Technology*, pp. 152-156, 2011

[5] N. S. Halvaiee and M. K. Akbari, "A novel model for credit card fraud detection using Artificial Immune Systems," *Applied soft computing*, Vol. 24, pp.40-49, 2014

[6] M. Abomhara and G. M. Køien, "Cyber security and the internet of things: vulnerabilities, threats, intruders and attacks," *Journal of Cyber Security and Mobility*, Vol. 4, No. 1, pp. 65-88, 2015

[7] D. A. Williams, "Credit card fraud in Trinidad and Tobago," *Journal of financial crime*, Vol. 14, No. 3, pp. 340-359, 2007

[8] V. Khattri, S. K. Nayak, and D. K. Singh, "Plastic card circumvention an infirmity of authenticity and authorization," *Journal of Financial Crime*, Vol. 27, No. 3, pp. 959-975, 2020

[9] V. Khattri and D. K. Singh, "Implementation of an additional factor for secure authentication in online transactions," *Journal of Organizational Computing and Electronic Commerce*, Vol. 29, No. 4, pp.258-273, 2019

[10] S. Jha, M. Guillen, and J. C. Westland, "Employing transaction aggregation strategy to detect credit card fraud," *Expert systems with applications*, Vol. 39, No. 16, pp.12650-12657, 2012

[11] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C.Westland, "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, Vol. 50, Vo. 3, pp.602-613, 2011

[12] E. Duman and M.H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," *Expert Systems with Applications*, Vol. 38, No. 10, pp.13057-13063, 2011

[13] K. Worobec, "FRAUD THE FACTS" (https://www.ukfinance.org.uk/system/files/Fraud%20The%20Facts%202019%20-%20FINAL%20ONLINE.pdf, accessed December 2020)

[14] Australian Payments Network Limited, "AUSTRALIAN PAYMENT CARD FRAUD 2019" (https://www.auspaynet.com.au/sites/default/files/2019-08/AustralianPaymentCardFraud2019_0.pdf, accessed December 2020).

[15] Y. Sahin and E. Duman, "Detecting credit card fraud by ANN and logistic regression," *IEEE International Symposium on Innovations in Intelligent Systems and Applications*, pp. 315-319, 2011

[16] J. R. Dorronsoro, F. Ginel, C. Sgnchez, and C. S. Cruz, "Neural fraud detection in credit card operations," *IEEE transactions on neural networks*, Vol. 8, No. 4, pp.827-834, 1997

[17] P. K.Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, "Distributed data mining in credit card fraud detection," *IEEE Intelligent Systems and Their Applications*, vol. 14, no. 6, pp.67-74, 1999

[18] A. Kundu, S. Panigrahi, S. Sural, and A. K. Majumdar, "Blast-ssaha hybridization for credit card fraud detection," *IEEE transactions on dependable and Secure Computing*, Vol. 6, No. 4, pp.309-315, 2009

[19] H. Patel, D.S. Rajput, G. T. Reddy, C. Iwendi, A. K. Bashir, and O. Jo, "A review on classification of imbalanced data for wireless sensor networks," *International Journal of Distributed Sensor Networks*, Vol. 16, No. 4, pp.1550147720916404, 2020.

[20] Y. Sun, A. K. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *International journal of pattern recognition and artificial intelligence*, Vol. 23, No. 04, pp.687-719, 2009

[21] B. Liu and G. Tsoumakas, "Dealing with class imbalance in classifier chains via random undersampling," *Knowledge-Based Systems*, Vol. 192, p.105292, 2020

[22] H. He and E.A. Garcia, "Learning from imbalanced data," *IEEE Transactions on knowledge and data engineering*, Vol. 21, No. 9, pp.1263-1284, 2009

[23] H. Zhang, H. Zhang, S. Pirbhulal, W. Wu, and V. H. C. D. Albuquerque, "Active Balancing Mechanism for Imbalanced Medical Data in Deep Learning–Based Classification Models," *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol. 16, No. 1s, pp.1-15, 2020

[24] Z. Li, G. Liu, and C. Jiang, "Deep representation learning with full center loss for credit card fraud detection," *IEEE Transactions on Computational Social Systems*, Vol. 7, No. 2, pp.569-579, 2020

[25] H. Zhu, G. Liu, M. Zhou, and Y. Xie, A. Abusorrah and Q. Kang, "Optimizing Weighted Extreme Learning Machines for Imbalanced Classification and Application to Credit Card Fraud Detection," *Neurocomputing*, Vol. 407, pp. 50-62, 2020

[26] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," *Journal of Information Security and Applications*, Vol. 55, pp.102596, 2020

[27] S. Mittal and S. Tyagi, "Computational Techniques for Real-Time Credit Card Fraud Detection," Handbook of Computer Networks and Cyber Security Springer, Cham. pp. 653-681, 2020

[28] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. S. Hacid, and H. Zeineddine, “An experimental study with imbalanced classification approaches for credit card fraud detection,” *IEEE Access*, Vol. 7, pp.93010-93022, 2019.

[29] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K.Nandi, "Credit card fraud detection using AdaBoost and majority voting,” *IEEE access*, Vol. 6, pp.14277-14284, 2018

[30] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, "Deep learning detecting fraud in credit card transactions," *IEEE Systems and Information Engineering Design Symposium*, pp. 129-134, 2018

[31] A. Pumsirirat and L. Yan, "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine,” *International Journal of advanced computer science and applications*, Vol. 9, No. 1, pp.18-25, 2018

[32] M. Bader-El-Den, E.Teitei, and T. Perry, "Biased random forest for dealing with the class imbalance problem," *IEEE transactions on neural networks and learning systems*, Vol. 30, No.7, pp. 2163-2172, 2018

[33] D. Veganzones and E. Séverin, "An investigation of bankruptcy prediction in imbalanced datasets," *Decision Support Systems*, Vol. 112, pp. 111-124, 2018

[34] M.A.H. Farquad and I. Bose, "Preprocessing unbalanced data using support vector machine," *Decision Support Systems*, Vol. 53, No. 1, pp. 226-233, 2012

[35] A. Dal Pozzolo, G.Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: a realistic modeling and a novel learning strategy," *IEEE transactions on neural networks and learning systems*, Vol. 29, No. 8, pp. 3784-3797, 2017

[36] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, Vol. 91, pp.216-231, 2019