Journal of Theoretical and Applied Information Technology
28th February 2022. Vol.100. No 4
2022 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# ADOCA: A NOVEL TECHNIQUE TO DEFRAUD CREDIT CARD USING AN OPTIMZED CATBOOST ALGORITHM

[1]KANEEZ ZAINAB, [2]NAMRATA DHANDA, [3]QAMAR ABBAS

[1]Department of Computer Science and Engineering,
Amity University, Lucknow Campus, Uttar Pradesh, India
[2] Department of Computer Science and Engineering,
Amity University, Lucknow Campus, Uttar Pradesh, India
[3] Department of Computer Science and Engineering,
Ambalika Institute of Technology & Management, Uttar Pradesh, India

E-mail: [1]kaneez_srm@yahoo.com, [2]ndhanda@lko.amity.edu, [3]qrat_abbas@yahoo.com

## ABSTRACT

Cashless economy has increased the demand of digital affairs. Online transactions using Credit cards is one of the most often used medium of digital transactions. Spike in recent years is seen in fraudulent transactions across the digital platform. The researchers have suggested many techniques in the past for detection of fraudulent transactions. But due to the key challenges like the changing profiles of both fraudulent and non-fraudulent transaction and data being unbalanced hinders technologies like data mining and major algorithm of machine learning (such as KNN, SVM, Random Forest and Decision Tree) and models of deep learning. Therefore, a novel proposal has been suggested for detecting the credit card fraud transaction using an optimized CatBoost Algorithm for determining that whether the transaction is legitimate or fraudulent by optimizing the Bayesian-based hyper parameter to tune the parameter of the CatBoost Algorithm. Hence, we suggest this novel approach as ADOCA (Anomaly Detection using an Optimized CatBoost Algorithm). Based on that, we compare our approach from the different binary classification algorithms that includes Logistic Regression, KNN, SVC, Decision Tree and Random Forest.

**KEYWORDS:** *Credit card fraud, Binary Classification, Machine Learning, CatBoost, Optimized CatBoost*

## I. INTRODUCTION

In the past decades one of the easiest payment method for e-commerce and communication was credit card transaction. With the successful credit card transaction, it is quite smooth to buy anything using digital payment through credit cards. As the customers of credit card increased, there was also a remarkable increase of fraudsters who purchased products using other's credit card without letting the owner of that card know, from that time it was called as fraudulent transaction. The fraudulent transaction was very less in amount but by those transaction billions of losses were transacted from the credit cards.

Thus the main concern is to identify those fake transactions by any means. Many researches were suggested for detecting those fraudulent transactions but none of them were able to detect the real time fraudulent transactions. So, a lot of Machine Learning and deep learning approaches were proposed, however these approaches are also not able to detect those fake transactions very clearly. So, we need a novel approach to identify the credit card fraud transaction and classify them that which transactions are legitimate and which transaction are fraudulent.

Hence, a novel approach has been proposed for credit card fraud detection to classify that which transactions are fraudulent and which one are legitimate. Our proposed novel technique is "A Novel Approach for Credit Card Fraud Detection using an Optimized CatBoost Algorithm". This novel technique is best suited for the credit card fraud detection on real time dataset. So, we suggest our novel technique as

ADOCA (Anomaly Detection using Optimized CatBoost Algorithm).

To implement our new approach, there was a requirement of real-world transaction dataset. So, we downloaded the credit card fraud transaction from UCI [22]. In which, we have total of 284,807 credit card transaction without missing values, that consists of 492 datasets as fraudulent and the rest of are legitimate.

In this novel technique first, we implement the dimensionality reduction algorithm that is PCA (Principal Component Analysis) on the dataset that has been downloaded from the UCI, following that we implement an Optimized CatBoost Algorithm. Here the enhancement is done by tuning the hyperparameter of the CatBoost Algorithm using Bayesian-based an intelligent method, after that applying K-Fold cross validation. We performed model evaluation of the implemented approach by calculating accuracy, precision, recall, f1 score and AUC score. After that we compared our proposed novel method from the existing method like Logistic Regression, K-NN, Decision Tree and Random Forest.

## 2. RELATED WORK

Many studies and findings includes Machine Learning Techniques to classify the legitimate and illegitimate transactions. In recent years many financial institution used Machine Learning algorithms to identify phishing, website detection. Providing a reliable environment for their regular activities, the most commonly used field of Machine Learning were supervised and unsupervised learning. Majorly the supervised learning algorithms were used in many research and also unsupervised algorithms were used. The models were trained using historical data. The historical data are priory labelled as legitimate and illegitimate transactions after training the models. They were used to classify the legitimate and illegitimate transaction. The model results were analyzed by using various multidimensional result analysis techniques like precision, recall, accuracy, AUC and ROC Curves. Few of the algorithms used were logistic regression, KNN, SUM, Probabilistic Neural network and genetic programming. The model built via probabilistic neural network had better results as compared to other approaches whereas the accuracy of the BNN (Bayesian Neural Network) was found to be 90.3%, while the accuracy of decision tree was 73.6%. The data that was used for training the models were collected from 76 different companies of credit card. The data included 38 illegitimate transactions which were evaluated by the firms.

For the development of the model to classify unsupervised illegitimate transaction on credit card SOM (Self Organizing Map) was used. The advantage of using SOM is that it is independent of the historical data for training. It learns from it's improved transactions, thus SOM is a better model for prediction which does not depends on the pattern of the data that is trained. The newly developed and accepted field of machine learning is deep learning. Deep learning is used for modelling complex system by using concepts of neural network. Many researchers have used deep learning for classification of illegitimate credit card transaction.

Jurgousky et al [19] used deep learning based LSTM (Long term memory) model for prediction of illegitimate credit card transaction.

Fiore et al [21] proposed a generative adversarial network used synthetic making scheme instance for removing the skewness of the data.

Carcello et al [33] proposed a model using the unattended outliers rating and the supervised classifier based feature selection process, the model showed a promising result for the classification of the illegitimate credit card fraud detection. In yet another paper Carcillo implemented SCARFF (scalable real time fraud finder) which includes Big Data Analytics techniques like (CASSANDRA Kafka and SPARK) along with machine learning techniques for classifying the illegitimate Credit Card transactions [34]

Yuan [16] proposed a novel technique by mixing deep neural networks and spectral graph analysis, it also uses deep auto encoder for feature selection but the data was not processed for the class imbalance, hence the results would be better if the model was used after processing the class imbalance of data. The ensemble learners of supervised machine learning also known as lazy learners were used for the classification of illegitimate credit card transactions. The ensemble learners are classified into two fields of bagging and boosting algorithms like random forest and bagging classifier are part of the bagging techniques whereas the boosting has Ada boost, gradient boost, XG Boost, Vote boost. The bagging classifiers uses any of the classifiers along with the bagging technique, most often used classifier is decision tree. The Boosting algorithm is based on the concept of weak learners and decision stump (single node tree). Previously many papers used the ensemble learners based classifiers for the classification of illegitimate transactions. The optimization is an integral part of the machine learning. Many optimizers are available for deducing the optimal solution for the problem. One such optimizers are bio inspired Algorithms which is used for achieving regional solution to the optimization problems. Kamaruddin [47] combined auto Associative neural network with the particle Swarn optimizers to build a model for the classification of illegitimate credit card transactions

## 3. PROPOSED METHODOLOGY

The proposed novel approach ADOCA contains various stages of data pre-processing, training the model, and the multidimensional result analysis of the illegitimate credit card transactions. The Architecture diagram of the novel ADOCA approach is shown in the figure 1 and the various stages are discussed below.

This proposed novel method is of Machine Learning technique which is performed on Intel i3 processor system of 8GB of ram. Python programming is used for implementing and testing of the proposed novel technique on Jupiter notebook framework.

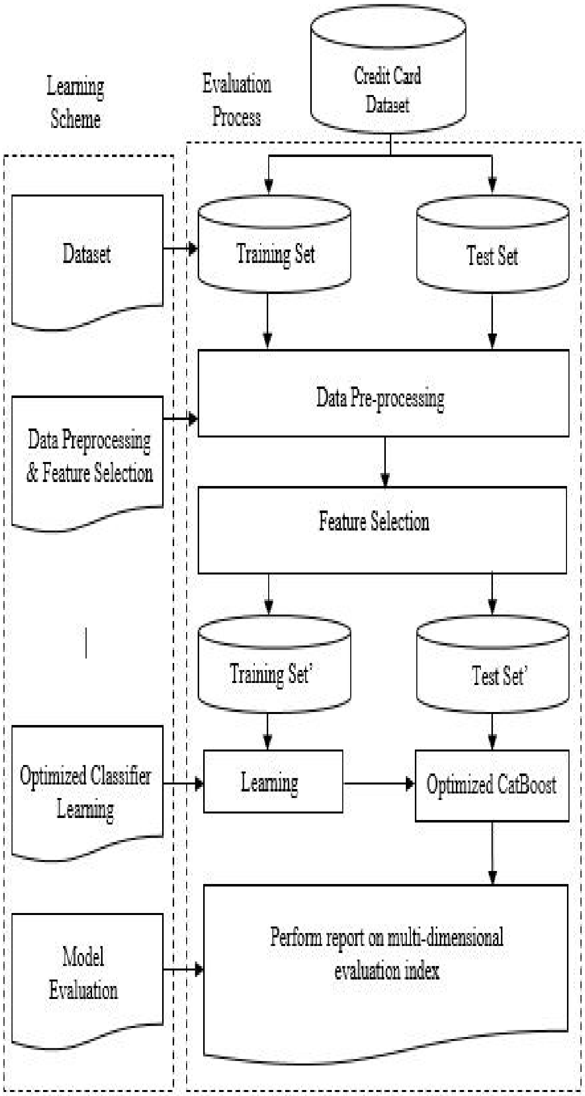

*Figure 1: ADOCA for credit card fraud detection*

### 3.1 Dataset and Pre-processing

A real time dataset is used for building the model and evaluation of the model. The credit card data set of users from Europe was collected in September 2013 which contains 2,84,807 transactions which includes both legitimate and illegitimate transactions. The dataset has 492 illegitimate transactions and 2,84,315 legitimate transactions. After applying the PCA algorithm only 31 important features are available in order to maintain the privacy of the client. The important features like time of the transaction and transaction amount feature and not changed by using dimension reduction algorithm

*Table I : Credit Card Dataset*

| Data Set | total no. of transactions | total no. of legitimate transactions | total no. of fraudulent transactions | No. of features | Ref. |
|---|---|---|---|---|---|
| Credit card data | 284,807 | 284,315 | 492 | 31 | [22] |

Table I summarizes the data set which includes the details related to the transactions, like total number of legitimate and illegitimate transactions. The table clearly indicate that the data is tuned imbalanced. The data skewness leads to either over fitted or under fitted model. For removing the skewness of the data set we use the random over sampler. The random over sampler will balance the number of legitimate and illegitimate transactions for training the model. We use the k-fold validation to avoid the overfitting of the model to obtain the most suitable parameters for the model. We used the grid search CV model. Here the average of all the parameters are tested and measured over the dataset [38]

### 3.2 Feature Selection

The Catboost Algorithm is a combination of categorical and boosting. This includes gradient boosting algorithm on the decision tree. Catboost algorithm creates extra features by combining the obtainable categorical features that has a high dependency. It is not feasible to hold all the combinations hence it uses greedy method to find the best combination. It uses the categorical data used by the tree along with the available categorical features in data for the combinations.

### 3.3 The Optimized Catboost Classifier

The Catboost algorithm is employed to identify illegitimate purchase of transaction on the credit card. The Catboost algorithm is an optimized advanced decision tree learning algorithm. The Catboost algorithm uses gradient boosting algorithm on the decision tree. Catboost is the Matrixnet algorithm's successor, that is widely used within the organization to ranking task, forecast and make recommendations. It's basic and can be extended across a variety of areas and variety of issues.

The Catboost Algorithm uses Bayesian based approach for hyper tuning optimization for the parameters. The high performance Catboost Algorithm can manage massive volume of data easily. It was developed by Yandex researchers and engineers.

**Algorithm 1: Ordered Boosting**

```
Input : $\{ (x_k , y_k) \}^n_{k=1}$ , I;
$\sigma \leftarrow$ random permutation of [1,n];
$M_i \leftarrow 0$ for i = 1..n;
for t $\leftarrow$ 1 to I do
    for i $\leftarrow$ 1 to n do
        $r_i \leftarrow y_i - M_{\sigma(i)-1(xi)}$ ;
    for i $\leftarrow$ 1 to n do
     $\Delta M \leftarrow$
       Learn Model(($x_j, r_j$ ):
       $\sigma$ (j)<= i);
       $M_i \leftarrow M_i + \Delta M$;
return $M_n$
```

**Algorithm 2 : Building a tree in Catboost**

```
Input : M , $\{(x_i , y_i) \}^n_{i=1}$ , $\alpha$, L, $\{\sigma_i\}^s_{i=1}$, Mode
grad $\leftarrow$CalcGradient (L,M,y);
r $\leftarrow$ random (1,s) ;
If Mode= Plain then
    G $\leftarrow$ ($grad_r$ (i) for i=1..n);
If Mode = Ordered then
    G $\leftarrow$ ($grad_{r, \sigma (i) -1}$ (i) for i= 1..n) ;
T$\leftarrow$ empty tree;
for each step of top down procedure do
    for each candidate split c do
    $T_c \leftarrow$ add split c to T;
     If Mode = Plain then
        $\Delta(i) \leftarrow$ avg ($grad_r (p)$ for
        p : $leaf_r (p) = leaf_r (i)$ ) for i = 1..n;
      If Mode = Ordered then
        $\Delta(i) \leftarrow$ avg ($grad_{r ,\sigma r ( i ) -1}$ (p) for
        p : $leaf_r (p) = leaf_r (i)$ , $\sigma_r (p) < \sigma_r (i)$ )
        for i = 1..n;
      loss ($T_c$) $\leftarrow$ cos($\Delta$ , G)
   T $\leftarrow$ arg min $_{Tc}$ (loss ($T_c$))
If Mode= Plain then
     $M_{r'}$ (i) $\leftarrow$ $M_{r'}$ (i) – $\alpha$ avg ($grad_{r'}$ (p) for
     p : $leaf_{r'}$ (p) = $leaf_{r'}$(i)) for r'= 1..s , i = 1..n;
If Mode = Ordered then
    $M_{r' , j}$ (i) $\leftarrow$ $M_{r' , j}$ (i) – $\alpha$ avg ($grad_{r' , j (p)}$ for
    p : $leaf_{r'}$(p) = $leaf_{r'}$ (i) , $\sigma_{r'}$ (p) <= j ) for r'
    =1..s,
    i=1..n , j>=$\sigma_{r'}$(i) – 1 ;
return T, M
```

In the Catboost algorithm [39] contains plain and ordered boosting methods. The Catboost Algorithm adds an independent random permutation as a feature in the training data. The $\sigma_0$ to $\sigma_s$ permutation is employed for evaluation of splits possible in the tree. $\sigma_0$ is used to select the leaf value $b_j$ of the trees obtained i.e. if the given permutation contains a short history then a high variance can be observed between target statistics (TS) and the predictions made and used by the ordered boosting. If we use only one permutation then there are chances of increase in the variance, hence we increase the count of permutation in order to check that the variance is not abruptly increased.

The Catboost algorithm make use of decision tree as base predictors. The decision tree makes use of a single split benchmark for the complete tree that would give a balanced tree which will be less prone towards overfitting and potentially boost the execution time of prediction. The pseudo code of the algorithm is mentioned in the second figure. The supporting model $M_{ij}$ is retained in the ordered boosting boosting mode throughout the process of learning. The $M_{ij}$ represents the prediction for the $i^{th}$ input based on the $1^{st}$ $J^{th}$ sample of the permutation $\sigma_j$ for building a tree $T_i$ we will randomly select a random permutation from $\{\sigma_i \ldots \sigma_n\}$ in each iteration. The permutation greatly affect the tree learning process and also it affect the Target Statistics.

The gradients $grad_{r,j}(i)$ are computed based on the supporting model $(M_{r,j})$. While constructing the tree we consider the approximation of the gradient G, using the cosine similarity cos (. ,.) , for calculating the gradient for current sample we use the previous sample permutation. For calculating the split value of the leaf node, we take the average of gradients of the earlier example which are prior to the current leaf nodes. The Target Statistics is highly impacted by the permutation of the feature set and the leaf node is also impacted highly because of the permutation. All the supporting models are boosted by the tree built by using the permutation. It is observed that a single common tree is used for all the models. But only the structure is same and contains different leaf nodes for individual supporting models based on the input feature set. The pseudo code 2 corresponds the same.

If the feature set includes any categorical values then it uses the concept of supporting models. If there are no categorical value it will use the simple gradient boost decision tree.

The Catboost algorithm is a parametrized machine learning model. It contains many parameters like depth of the tree, the co-efficient, the cos function, the learning rate of the gradient technique, bagging temperature for the Bayesian bootstrap, border count which is directly related to the training time, scale_POS_weight. All the parameter values can be optimized, by tuning the hyper parameter making use of the grid search CV model. Many of these parameters can lead to over fitted or under fitted model if not optimized.

### 3.4 Model Evaluation Using Performance Metrics

The dataset that is obtainable for training the model is unbalanced hence the result analysis cannot be done by dividing the dataset in 80:20 ratio in which 80 percent of data is to be used for training purpose and rest 20 percent for testing. Traditionally the rest of 20 percent is predicted by the model and the result is analysed with the actual outcome. There are greater chances that the testing data would be biased towards any outcome. Hence you will have a statistical approach i.e. K-Fold Cross Validation. We have used K- Fold of 5 for the result analysis. The k-fold of 5 will form five sub dataset out of the original data sets where each subset contains 20 percent of original dataset. By doing so we have equal chance of randomly selecting dataset with legitimate and illegitimate as an outcome.

Four of the sub datasets are employed for training the model and rest of the dataset is predicted based on that, accuracy is being calculated. The process is repeated for five times and the accuracy obtained in five times is the final accuracy of the model.

Further many tests are conducted during the K-Fold cross validation which includes the confusion matrix which contains the number of correctly predicted legitimate and illegitimate transaction and also the number of incorrectly predicted legitimate and illegitimate transaction based on the attributes of the confusion matrix, further the precision of model, recall of the model accuracy of the model, AUC and F1 scores are calculated as.

The Confusion Matrix uses the following terms for measuring credit card fraud detection performance:

TP(i.e. True Positive) applies to the no of fatalities Credit card Transactions are classified appropriately.

FP(i.e. False Positive) represents the number of fraudulent purchases in credit card fraud classified as fraud.

FN(i.e. False Negative) signifies fake credit card numbers transactions which are marked as natural.

TN(i.e. True Negative) corresponds to the amount of daily, correctly categorized credit card purchases.

The measures, which are used for evaluating the performance are as follows:

Accuracy= (TP+TN) / (TP+TN+FP+FN)

Precision= TP/ (TP+FP)

Recall= TP/ (TP + FN )

F1 – Score = 2* (Precision * Recall) / (Precision + Recall )

Precision and recall are the measure of performance of the model for every class of the outcome. Accuracy results in general performance of the dataset. If the dataset is skewed then there are possibilities that the accuracy might be high for an outcome and low for another. Precision measures the correctly predicted outcomes a high precision indicates the model is fitted perfectly for both the biased as well as unbiased dataset. The recall indicates the error rate of the model for all the possible outcome. A high recall means the model is under fitted it could be due to improper processing of the data set or due to skewed dataset. Generally the values of precision and recall are plotted for a visual representation and better understanding of the effectiveness of the model. Also the AUC score of the model can be calculated if the score is near to 1 that means the model is perfectly fitted for the dataset.

## 4. EXPERIMENTAL RESULTS

The result analysis of the discussed model for prediction of illegitimate transaction is done by using a 5-fold cross validation is used on the dataset. Using the Bayesian-based hyper parameter optimisation method, the suggested methodology is equipped with tailored parameters. The concept of k-fold cross validation is adopted as a strategy to plan and verify the testing and evaluation of the dataset. The experimental data were uniformly divided into k subsets for k-fold cross testing. One subset is used as the control sample of every trial, and the remainder of the k-1 subsets are used as training sets. On a limit of k experiments, each subset being utilized as a study range for one sub-set for a time. The model's performance is measured as its average findings obtained from studies with k.

The accuracy of classification is usually used to check whether the machine learning model used for prediction of illegitimate transaction is perfectly fitted or under/over fitted model. Because of Class, Credit card transaction unbalance data collection, consistency the success evaluation as discussed in the introduction is inadequate by itself. For evaluation and comparison of our comprehensive approach, we executed the following tests.

(i) Evaluate the results against a set of indicators including accuracy, precision, recall, F1-score and AUC to thoroughly evaluate ADOCA algorithm classification performance;

(ii) Using the freely accessible accuracy tests as the validation criterion to check that the ADOCA algorithm is superior, relative to other comparison algorithms.

According to the confusion matrix shown in Table II, evaluation indicators for instance accuracy, precision, recall, F1-score and AUC are set.

*Table II: Comparison between ADOCA and other algorithms with comprehensive indicators*

| Algorithm | Accuracy | Precision | Recall | F1_Score | AUC |
|---|---|---|---|---|---|
| LR | 0.9682 | 0.0015 | 0.0101 | 0.0030 | 0.7020 |
| KNN | 0.9690 | 0.1122 | 0.1498 | 0.1284 | 0.5930 |
| SVM | 0.9706 | 0.0020 | 0.0106 | 0.0014 | 0.4780 |
| DT | 0.9550 | 0.4583 | 0.0799 | 0.1375 | 0.6630 |
| RF | 0.9787 | 0.2521 | 0.6547 | 0.3642 | 0.8690 |
| Proposed ADOCA | 0.9996 | 0.9452 | 0.7931 | 0.8624 | 0.9801 |

The above Table II provides a brief insight of the performance of various models compared to the ADOCA model. The standard models like LR, KNN, SVM, DT, and RF were trained employing the same dataset and there result is shown in the table. The same data set is employed to train the ADOCA model, the data set was treated for the class imbalance utilizing the over sampler so that the class with less number of data would not make our model under fitted or over fitted.

The experimental results from the following aspects are analysed in Table II. First of all, the influence of the 5-fold cross verification approaches on the effect is not conclusive for the CatBoost introduced in this paper and the suggested ADOCA algorithm, but the ADOCA is more influenced than the CatBoost Algorithm which indicates a more reliable performance by the ADOCA algorithm. In addition, by comparing the effect of machine learning algorithms like LR, KNN, SVM, DT, RF and ADOCA algorithms and the improvement provided by ADOCA using oversampling plays a significant role in this. Finally, when contrasting the performance of the ADOCA algorithm with the algorithms in Comparison, It can be observed that the maximum precision is obtained by the ADOCA algorithm.

*Table III. Performance comparison of the proposed approach with other methods utilizing the accuracy metric*

| Approach | Accuracy |
|---|---|
| Concept Drifts Adaption [31] | 80% |
| Local Outlier Factor [4] | 97% |
| Isolation Forest [29] | 95% |
| Random Forest [45] | 95.5% |
| ANN [46] | 92.86% |
| Proposed Approach | 99.96 % |

In the Table III we have listed the various techniques and models used by different authors for prediction of illegitimate credit card transactions and their respective accuracy. We can clearly see that our proposed model has the highest accuracy of 99.96% whereas the model built by using the Concept Drifts Adaption [31] had the least accuracy of 80%. It can be observed the that the proposed method has the maximum accuracy thus it outperforms other models.

## 5. CONCLUSION

An abrupt spike is witnessed in digital transaction in the recent years because of the ease and the hassle free transactions and the governments are also boosting the cashless economy. A gradual rise is seen in illegitimate transactions. Now a days the machine learning models are used in each and every field owing to its ability to provide a better performance according to the change of dataset. Previously many authors have tried various technique to handle such illegitimate transactions. There were many key challenges like the dataset generally available is skewed and pre-processing of the dataset was not done in an efficient way. Our key goal in the work is to build a model that can predict accurately the illegitimate transactions.

We have proposed a new technique that make use of Optimised CatBoost algorithm for predicting the illegitimate transactions. Several studies were performed utilizing real-world data sets. The dataset is processed for the lost values and the skewness by using scalars, handling the missing values and over samplers. Further we have selected best features by using the feature selection. The model was hyper tuned by using the random search CV. Then to measure the performance of the model it was compared with various standard models like Logistic Regression, SVM, KNN, Decision Trees, and Random Forest Classifiers. The performance was measured by using a multidimensional result analysis. The result clearly shows a greater improvement in the accuracy of the model, for a better result analysis cross validation is also used. The proposed technique is better compared to previous approaches used by various authors in the past.

In view of security issues related to confidentiality and sensitiveness of the data we don't have different datasets available for the research process. Further due to frequent changes in the methods used by the intruders for performing the illegitimate transactions we need updated dataset so that we can learn that whether the model is easily adopting to the pattern changes of the data or not.

## REFERENCES

[1] 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,'' *Inf. Sci.*, May 2019. Accessed: Jan. 8, 2019.

[2] N. Carneiro, G. Figueira, and M. Costa, "A data mining based system for credit-card fraud detection in e-tail,"*Decis. Support Syst.*, vol. 95, pp. 91_101, Mar. 2017.

[3] B. Lebichot, Y.-A. Le Borgne, L. He-Guelton, F. Oblé, and G. Bontempi, "Deep-learning domain adaptation techniques for credit cards fraud detection," in *Proc. INNS Big Data Deep Learn. Conference*, Genoa, Italy, 2019, pp. 78_88.

[4] H. John and S. Naaz, "Credit card fraud detection using local outlier factor and isolation forest," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 4, pp. 1060_1064, Sep. 2019.

[5] C. Phua, R. Gayler, V. Lee, and K. Smith-Miles, "On the communal analysis suspicion scoring for identity crime in streaming credit applications," *Eur. J. Oper. Res.*, vol. 195, no. 2, pp. 595_612, Jun. 2009.

[6] R. Bolton and D. Hand, ``Statistical fraud detection: A review,'' *Stat. Sci.*, vol. 17, no. 3, pp. 235_249, Aug. 2002.

[7] P. A. Dal, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: A realistic modeling and a novel learning strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3784_3797,
Sep. 2017.

[8] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602_613, Feb. 2011.

[9] N. Sethi and A. Gera, "A revived survey of various credit card fraud detection techniques," *Int. J. Comput. Sci. Mobile Comput.*, vol. 3, no. 4, pp. 780_791, Apr. 2014.

[10] A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," *Int. J. Syst. Assurance Eng. Manage.*, vol. 8, no. S2, pp. 937_953, Nov. 2017.

[11] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *Proc. ICCNI*, Lagos, Nigeria, Oct. 2017, pp. 1_9.

[12] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero, "BankSealer: A decision support system for online banking fraud analysis and investigation," *Comput. Secur.*, vol. 53, no. 1, pp. 175_86, Sep. 2015.

[13] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, "Detection of _nancial statement fraud and feature selection using data mining techniques," *Decis. Support Syst.*, vol. 50, no. 2, pp. 491_500, Jan. 2011.

[14] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent _nancial statements," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 995_1003, May 2007.

[15] D. Olszewski, "Fraud detection using self-organizing map visualizing the user pro_les," *Knowl.-Based Syst.*, vol. 70, pp. 324_334, Nov. 2014.

[16] J. T. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1721_1732, Nov. 2008.

[17] V. Zaslavsky and A. Strizhak, "Credit card fraud detection using self organizing maps," *Inf. Secur.*, vol. 18, p. 48, Jan. 2006.

[18] L.Wang, T. Liu, G.Wang, K. L. Chan, and Q. Yang, "Video tracking using learned hierarchical features," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1424_1435, Apr. 2015.

[19] J. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He-Guelton, and O. Caelen, "Sequence classi_cation for creditcard fraud detection," *Expert Syst. Appl.*, vol. 100, pp. 234_245, Jun. 2018.

[20] M. Kraus and S. Feuerriegel, "Decision support from _nancial disclosures with deep neural networks and transfer learning," *Decis. Support Syst.*, vol. 104, pp. 38_48, Dec. 2017.

[21] U. Fiore, A. D. Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classi_cation effectiveness in credit card fraud detection," *Inf. Sci.*, vol. 479, pp. 448_455, Apr. 2019.

[22] *Credit Card Fraud Dataset*. Accessed: Sep. 4, 2019. [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud/data

[23] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429_449, 2002.

[24] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Syst. Appl.*, vol. 51, pp. 134_142, Jun. 2016.

[25] R. D. Kumar, "Statistically identifying tumor suppressors and oncogenes from pan-cancer genome sequencing data," *Bioinformatics*, vol. 31, no. 22, pp. 3561_3568, 2015.

[26] I. Mekterovi¢, L. Brki¢, and M. Baranovi, "A systematic review of data mining approaches to credit card fraud detection," *WSEAS Trans. Bus. Econ.*, vol. 15, p. 437, Jan. 2018.

[27] M. A. Al-Shabi, "Credit card fraud detection using auto encoder model in unbalanced datasets," *J. Adv. Math. Comput. Sci.*, vol. 33, no. 5, pp. 1_16, 2019.

[28] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, Philadelphia, PA, USA, 2006, pp. 233_240.

[29] A. D. Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915_4928,
Aug. 2014.

[30] H. A. El Bour, Y. Oubrahim, M. Y. Ghoumari, and M. Azzouazi, "Using isolation forest in anomaly detection: The case of credit card transactions," *Periodicals Eng. Natural Sci.*, vol. 6, no. 2, pp. 394_400, 2018.

[31] A. Jog and A. A. Chandavale, "Implementation of credit card fraud detection system with concept drifts adaptation," in *Proc. ICICC*, Singapore, 2018, pp. 467_477.

[32] 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.

[33] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Inf. Sci.*, to be published.

[34] F. Carcillo, A. D. Pozzolo, Y.-A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "SCARFF: A scalable framework for streaming credit card fraud detection with spark," *Inf. Fusion*, vol. 41, pp. 182_194, May 2018.

[35] R. Saia and S. Carta, "A frequency-domain-based pattern mining for credit card fraud detection," in *Proc. 2nd Int. Conf. Internet Things, Big Data Secur.*, 2017, pp. 386_391.

[36] S. Yuan, X. Wu, J. Li, and A. Lu, "Spectrum-based deep neural networks for fraud detection," in *Proc. ACM Conf. Inf. Knowl. Manage. (CIKM)*, 2017, pp. 2419_2422.

[37] R. Saia, "A discrete wavelet transform approach to fraud detection," in *Proc. Int. Conf. Netw. Syst. Secur.* Cham, Switzerland: Springer, 2017, pp. 464_474.

[38] J.West and M. Bhattacharya, "Intelligent _nancial fraud detection:Acomprehensive review," *Comput.Secur .*, vol. 57, pp. 47_66, Mar. 2016.

[39] *UCSD: University of California, San Diego Data Mining Contest 2009*. Accessed: Jan. 14, 2019. [Online]. Available: https://www.cs.purdue. edu/commugrate/data/credit_card/

[40] S. Russel and P. Norvig, *Arti_cial Intelligence: A Modern Approach*, 3rd ed. London, U.K.: Pearson, 2016.

[41] G. Ke, Q. Meng, T. Finley, T.Wang,W. Chen,W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly ef_cient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146_3154.

[42] S. Dhankhad, E. Mohammed, and B. Far, “Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study,” in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, Jul. 2018, pp. 122_125.

[43] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, “Credit card fraud detection and
concept-drift adaptation with delayed supervised information,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*,
Jul. 2015, pp. 1_8.[44] A. Jovi¢, K. Brki¢, and N. Bogunovi¢, “A review of feature selection methods with applications,” in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, 2015, pp. 1200_1205.

[45] S. V. S. S. Lakshmi and S. D. Kavilla, “Machine learning for credit card fraud detection system,” *Int. J. Appl. Eng. Res.*, vol. 13, no. 24, pp. 16819_16824, 2018.

[46] A. Rohilla, “Comparative analysis of various classi_cation algorithms in the case of fraud detection,” *Int. J. Eng. Res. Technol.*, vol. 6, no. 9, pp. 1_6, 2017.

[47] S. Kamaruddin and V. Ravi, “Credit card fraud detection using big data analytics: Use of PSOAANN based one-class classi_cation,” in *Proc. Int. Conf. Informat. Anal.*, 2016, pp. 1_8.

[48] K. Zainab and N. Dhanda, "Big Data and Predictive Analytics in Various Sectors," International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2018, pp. 39-43, doi: 10.1109/SYSMART.2018.8746929.

[49] Zainab K., Dhanda N., Abbas Q. (2021) Analysis of Various Boosting Algorithms Used for Detection of Fraudulent Credit Card Transactions. In: Kaiser M.S., Xie J., Rathore V.S. (eds) Information and Communication Technology for Competitive Strategies (ICTCS 2020).