International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

Probabilistic XGBoost Threshold Classification with Autoencoder for Credit Card Fraud Detection

D. Padma Prabha[1], C. Victoria Priscilla[2]
[1]Assistant Professor, Department of Computer Applications
Madras Christian College (Autonomous), Affiliated to University of Madras
Chennai, India.
padmaprabha@mcc.edu.in
[2] Head & Associate Professor, PG Department of Computer Science
Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women (Autonomous)
Affiliated to University of Madras, Chennai, India.
victoriapriscilla.c@sdnbvc.edu.in

Abstract— Due to the imbalanced data of outnumbered legitimate transactions than the fraudulent transaction, the detection of fraud is a challenging task to find an effective solution. In this study, autoencoder with probabilistic threshold shifting of XGBoost (AE-XGB) for credit card fraud detection is designed. Initially, AE-XGB employs autoencoder the prevalent dimensionality reduction technique to extract data features from latent space representation. Then the reconstructed lower dimensional features utilize eXtreame Gradient Boost (XGBoost), an ensemble boosting algorithm with probabilistic threshold to classify the data as fraudulent or legitimate. In addition to AE-XGB, other existing ensemble algorithms such as Adaptive Boosting (AdaBoost), Gradient Boosting Machine (GBM), Random Forest, Categorical Boosting (CatBoost), LightGBM and XGBoost are compared with optimal and default threshold. To validate the methodology, we used IEEE-CIS fraud detection dataset for our experiment. Class imbalance and high dimensionality characteristics of dataset reduce the performance of model hence the data is preprocessed and trained. To evaluate the performance of the model, evaluation indicators such as precision, recall, f1-score, g-mean and Mathews Correlation Coefficient (MCC) are accomplished. The findings revealed that the performance of the proposed AE-XGB model is effective in handling imbalanced data and able to detect fraudulent transactions with 90.4% of recall and 90.5% of f1-score from incoming new transactions.

Keywords- Autoencoder; XGBoost; Credit card; Fraud detection; Imbalance; Deep learning.

I. INTRODUCTION

Digital transaction has become the easiest way of money handling that are encouraged by bank to their customers. On the other hand, cyber frauds become a threat to the customers during their transactions. Hence the bank needs a perfect solution to overcome this problem. The Nilson Report one of the global newsletters providing statistical reports about the payment industry, predicted that in the year 2030 the increase in fraud loss is expected to be $ 49.32 billion [1]. Progressive research are been carried out predominantly in credit card fraud detection (CCFD) through machine learning (ML) and deep learning (DL) models. Many approaches [2]-[6] have been proposed for CCFD in the literature. A systematic survey paper on CCFD [2] is suggested to the readers for the detailed research works done with Machine learning methods. The literature identifies the challenges met by the researchers in CCFD are concept drift – change in behavioral patterns of the customers during their purchase [3]. Class imbalance - the nature of data is imbalanced with majority of legitimate transactions and minority of fraudulent transactions [4]. Misclassification of data - predicting the legitimate as fraudulent and vice versa leads to customer dissatisfaction [5], [6].

Handling imbalanced data is one of the major hurdles in CCFD. During the training phase, the data is biased towards the majority class which decreases the performance of the classifier [7]. Another drawback of CCFD is the curse of dimensionality consequences in overfitting [8]. Therefore important features are selected with different feature selection techniques by preventing irrelevant and noisy data [9]. Hence to increase the performance and computational efficiency, data preprocessing is an important phase before training the model [10]. The main focus of this work is to build a model that extracts meaningful features using an autoencoder that are subsequently employed in classification using a powerful ensemble model XGBoost. The main contribution of the proposed study are summarized as follows:

- In the proposed AE-XGB method, we developed an AE model that extracts the low-dimensional transactional data features from the high-dimensional credit card dataset and then depends on the XGBoost for classifying the data as legitimate and fraudulent.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

- By adopting XGB with probabilistic threshold the performance of AE-XGB can be enhanced. Therefore experiment was conducted to find the optimal threshold with prediction probabilities based on the scoring metric considering the nature of the problem.
- To evaluate the performance of AE-XGB, extensive experiments of various performance evaluation metrics are done by default (0.5) and the optimal threshold is then compared with other ML models.

To circumvent the aforementioned challenges of overfitting due to class imbalance and high dimensionality in CCFD, many works are contributed that are elaborated in the following section. The rest of this paper is outlined as follows: the related work is discussed in section 2, and the proposed AE-XGB is described with some preliminaries in section 3. The data preprocessing and performance analysis are elaborated in section 4. In section 5 the experimental results of AE-XGB and comparison with other methods are shown. Finally, section 6 concludes the study and discusses future work.

II. RELATED WORK

CCFD is one of the challenging ongoing research from the research community as the fraudsters change their pattern of conduct during the transaction. Hence, it is difficult for a bank to fix a solution since fraud is detected after the occurrence [2]. Another challenge confronted by researchers is the imbalance in data. The dataset has fewer fraudulent transactions than legitimate [11],[12]. Studies show that the performance of the model will decrease for an imbalanced dataset [13]. Various sampling techniques are contributed by different authors to balance the dataset like oversampling and undersampling [13]–[20]. Ahmad and Kasasbeh [22] proposed a clustering and similarity-based selection approach using Fuzzy C-means by grouping similar features to prevent the removal of important features during the sampling of instances. Itoo et al. [23] prepared a comparative study on different machine learning classifiers with an imbalanced dataset. In the preprocessing stage random under sampling is applied to balance the dataset and divided into three different ratios of training data before feeding to the classifier.

The shortcoming of the sampling technique is, that in oversampling duplication of fraudulent transactions is created which leads to overfitting of the model whereas in undersampling substantial legitimate transactions may be missed [21],[22]. Hence without altering the dataset, fine-tuning the threshold can tackle the hurdle of class imbalance considering the importance of type1 or type 2 errors for binary classification problems [26]. Lipton et al. [27] derived the best threshold with a probabilistic approach using f1-score and maintained a threshold not greater than 0.5. Thai-Nghe [28] found the optimal threshold with probabilities from the Bayesian classifier. The metrics for imbalanced data like f1-score, g-mean cohen kappa and balanced accuracy are applied to estimate the best threshold by maximizing the score [29]. Threshold shifting in the neural network model is established on the reconstruction error by the autoencoder [30], and a suitable threshold is set to find the fraudulent and legitimate transaction.

The ensemble methods combine the predictions of weak learners to a strong classifier constructed on weights. To reduce memory storage while pruning and increase efficiency, Yin et al. [31] proposed RotEasy algorithm. Raghuwanshi et al. [32] implemented kernalized ELM technique by assigning weight for the train data. The XGBoost classifier is an ensemble model advanced from gradient tree boosting [33]. Several boosting algorithms have been developed like LightGBM [34] and catboost [35] to avoid the drawbacks that existed in GBM. Subsequently, XGBoost a powerful model was selected for this study.

Among the traditional ML approaches, DL methods have excellent results in extracting lower dimensional features from the latent space (LS) representation [36]. The autoencoder, dimensionality reduction technique supports the model to perform better. Recently deep learning techniques are focused on by researchers to improve performance as it is a subset of machine learning. Oluwasanmi et al. [37] proposed a dual network with Luong's concatenation attention to the latent space to learn abnormal detection, the mean and standard deviation is normalized with variational AE and designed LSTM to train Gaussian distribution of sequential data analysis. Li et al. [38] developed a new model with a random forest for feature selection to reduce the dimensionality of data and to improve the performance of autoencoder with the combination of three layers. Pumsirirat and Yan [39] created a model with a deep autoencoder and Restricted Boltzmann machine (RBM) to find the anomalies. The reconstruction of error is done with backpropagation by finding the error signal and the performance metric AUC obtained for the European dataset are 0.9603 and 0.9505 respectively. Ng et al. [40] proposed an approach of dual autoencoder features with different activation methods to learn new features instead of original features. Tang et al. [30] developed a combined model of LightGBM and autoencoder. The LightGBM selects the important features and autoencoder is added to find the threshold from the reconstruction error. The model is compared with VAE and DAE but the proposed model generated 89.82% of accuracy. Lin and Jiang [41] built a related model AE-PRF as AE is employed for dimensionality reduction and RF to classify the data with probabilistic threshold. They experimented with resampling techniques to handle the imbalance in dataset and concluded that AE-PRF

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

with sampling methods is not much better but achieved AUC as 96.3 without sampling of dataset.

III. PROPOSED METHODOLOGY

In this research, we propose a AE-XGB method of extracting features using autoencoder and employed a classifier XGBoost to classify the fraudulent and legitimate transactions. The classification is based on the optimal threshold obtained from probabilistic score. The following sections elaborate the concept of AE and the formulation of XGBoost that is used in the proposed model.

A. Autoencoder

Autoencoder [42] is an unsupervised feed-forward backpropagation neural network, consisting of encoder and decoder operations. The encoder compresses the higher dimensional input to a lower dimensional latent space and then the decoder reconstructs the original input from the hidden representation. Figure 1 shows the working of a simple autoencoder used in this work.

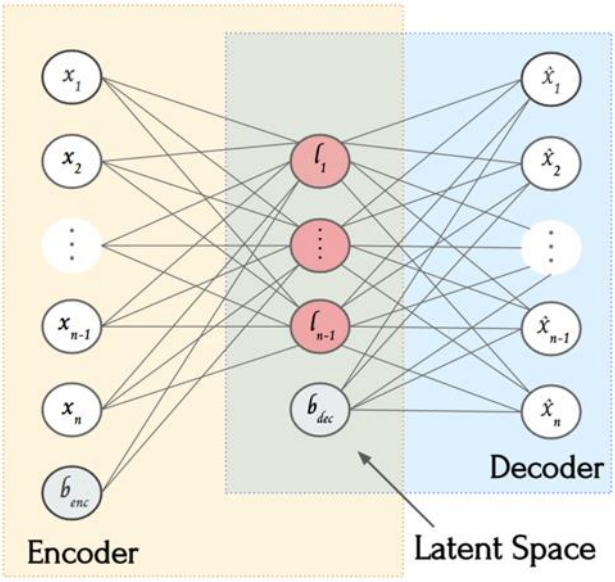


Figure 1. The general structure of autoencoder with encoding and decoding layers. The encoder converts the input vector x to the lower dimensional latent space representation; while the decoder reconstructs the latent space l to the output vector x̂.

In our model, all the layers are activated with a rectified linear unit (ReLU) [43]. In the encoder phase, each neuron takes input data vector x that is compressed to latent representation space as l with dimension $d < D$, and then the decoder reconstruct the output as $\hat{x}$.

$$\mathrm{l} = \sigma(\mathrm{x} * \mathrm{W} + \mathrm{b}_{\mathrm{enc}}) \quad (1)$$

In the above equation (1), the parameter W denotes the weight of the matrix D x d, σ denotes the activation function and b_{enc} represents the bias vector of the encoder. The decoder phase reconstructs the input vector x encoded as latent space l to the reconstructed vector $\hat{x}$.

$$\hat{\mathrm{x}} = \sigma(\mathrm{x} * \mathrm{W} + \mathrm{b}_{\mathrm{dec}}) \quad (2)$$

where W represents the weight matrix d x D and b_{dec} denotes the bias vector related to the decoder in equation (2). σ is the activation function of the dense layer in the decoder. The autoencoder is trained to minimize the error between the reconstructed output $\hat{x}$ and the original input x.

After the completion of forward computation, the difference between the predicted output data and input data is calculated with mean absolute error (MAE) [44]. The error loss updates the model's parameters as the subsequent predictions produce improved outputs. The MAE is computed using the following equation (3).

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}|\mathrm{y_i} - \mathrm{x_i}| \quad (3)$$

where MAE is the sum of the absolute difference between the prediction x_i and actual value y_i for the total number of data points n.

B. XGBoost

The eXtreme Gradient Boosting (XGBoost) is a powerful ensemble learning algorithm for classification and regression. This algorithm was developed by Chen and Guestrin [45] to avoid overfitting a model with the regularization technique. It is a progressive form of gradient boosting algorithm to improve the speed and model efficiency [33]. It creates decision trees parallel and based on the residual, then build a new tree. Tree boosting is an ensemble algorithm that transforms weak learners into strong classifiers for better performance classification [13]. Figure 2 shows the general structure and functionality of XGBoost.

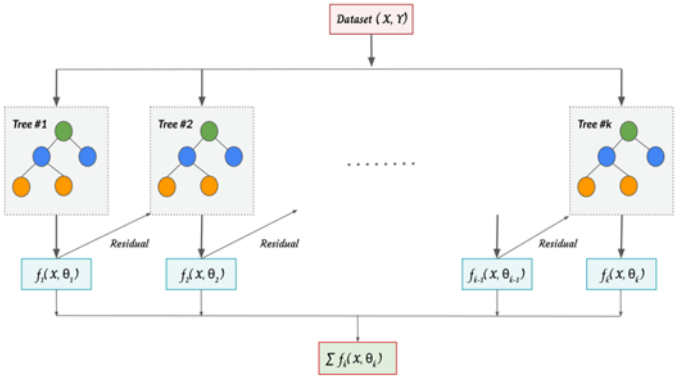


Figure 2. The general architecture of XGBoost represents the performance of each decision tree based on residuals of the previous one to obtain the additive function $\sum f_k(x, \theta_k)$.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

Mathematical description of XGBoost

The objective function of XGBoost (O) is to find the fitness of training data features x_i and the target is represented as y_i for an ensemble tree with K trees is expressed as

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F} \quad (4)$$

in equation (4) f is expressed as the functional space and the possible set of classification and regression trees (CART) is given as $\mathcal{F}$. The regularized objective equation is

$$O(\theta) = \sum_i^n l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (5)$$

where in equation (5), l is the convex loss function, that measures the difference between the predicted value $\hat{y}_i$ and real value y_i for the additive training t. The greedy method learns f_i and minimizes the objective function. Let $\hat{y}_i^{(t)}$ the prediction value for each t is as follows:

$$O^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (6)$$

Taylor series of loss function applied in equation (6) where the constant value is removed and the final expression is given as

$$O^{\sim(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) \quad (7)$$

where g_i and h_i in equation(7) are expressed as $g_i = \partial_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right)$ and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right)$.

Let $I_j = \{i | q(x_i) = j\}$ the index number to the j^{th} leaf node is rewritten as $\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$ in (6) where γ and λ are normalizing coefficients. Hence, the best objective function of j^{th} leaf with weight w_j^*, the optimized objective function stated in equation (8) as

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (8)$$

$$O^{\sim(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (9)$$

equation (9) is used as a score function to measure the good tree structure. The Greedy algorithm split a single leaf into two leaves and the score is calculated and used to split candidates. The left and right nodes are represented as I_L and I_R after the split. Thus, the following equation (10) expresses the loss reduction after a split.

$$O_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_j} h_R + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda}\right] + \gamma \quad (10)$$

XGBoost algorithm regularizes the objective function to prevent overfitting. Hence it is added to build the model to classify the arrival of a new transaction as fraudulent or legitimate.

C. Proposed AE-XGB method

The proposed autoencoder with XGBoost (AE-XGB) model is shown in figure 3. The CCFD data initially undergo some preprocessing methods to clean the data. The number of attributes is high, hence feature selection techniques are applied to select meaningful important attributes for model training. But still, it has a huge dimension with 214 attributes, therefore AE can be used as a dimensionality reduction method to extract features from transaction data in the first stage. The data is partitioned as the training dataset of 70 % data and the testing dataset of 30% data.

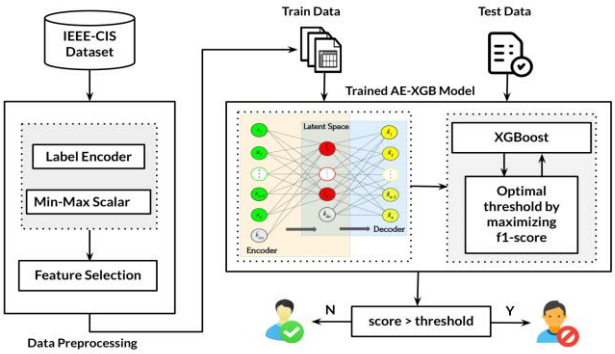


Figure 3. Architecture of the proposed AE-XGB.

Autoencoder can generate low dimensional latent space data transformation, which can be decoded to the original features as codes by adjusting the weights with error backpropagation and gradient descendent methods. In the second stage, the obtained codes of training data are applied to train the XGBoost model to classify the data as fraudulent or legitimate with the associate classification threshold value. To determine the optimal threshold the data is trained in the XGBoost model to classify data with the best F1-score metric. Finally, the performance of the trained model AE-XGB along with the optimal threshold is verified for every test data to detect if it is fraudulent or legitimate.

Specifically, the XGBoost model with optimal threshold classification is used to classify data as fraudulent with optimal threshold θ. The threshold depends on the objective of the problem therefore, different threshold values of AE-XGB produce different classification outcomes. The pseudocode of proposed AE-XGB method with optimal threshold is shown below as Algorithms 1 and 2.

Algorithm 1: Optimal Threshold (D_{Train}, D_{Test}, μ)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
Article Received: 28 April 2023 Revised: 19 June 2023 Accepted: 30 June 2023

Input: *D_{Train} with $x_i \in X$, target $y_i \in \{0,1\}$*
$\mu \in$ { XGBoost }
Output: *Finding the best threshold μ by comparing all values in terms of metric f1-score (F)*
Mo ← build model(D_{Train}, μ)
predictionScore ← (Mo,D_{Test}, μ)
maxScore ← maximum(predictionScore)
BestF ← 0; bestThresh ← 0
for each currentThresh ∈ maxScore do
currentF ← calculate F using currentThresh
if (currentF > bestF) then
bestF ← currentF
bestThresh ← currentThresh
return bestThresh

Algorithm 2: AE-XGB

Input: *Training data (D_{Train}), Testing data (D_{Test}), and metric f1-score (F)*
Output: *Classifying legitimate and fraudulent data (D_{Test}), based on a dynamic threshold (θ)*
Train AE with D_{Train}
$T \leftarrow AE(D_{Train})$
Train XGBoost with T
Find the best threshold by calling Optimal Threshold (T, D_{Test}, θ)
$V \leftarrow AE(D_{Test})$
for each v in V do
$q \leftarrow XGBoost(v)$
if $q > \theta$ then [v] ← 1 //Fraudulent
else [v] ← 0 //Legitimate
return output [v]

IV. EXPERIMENTAL ANALYSIS

To validate the performance of the proposed methodology, we compare it with other state of art machine learning algorithms. Specially we selected the ensemble algorithms to verify the benefit of optimal threshold instead of the default for the same dataset. We also compare the AE-XGB model with default and optimal threshold to evaluate the classification performance. The dataset used for the experiment, the preprocessing methods, performance analysis and the experimental results found are discussed in this section.

A. Dataset and Data preprocessing

The CCFD dataset used in this experiment is from IEEE-CIS [46] a real-world financial dataset from Kaggle. The total transaction data is 590540. The dataset is broken into two files namely transaction and identity includes 392 transaction features and 42 identity features. The two files are merged as a single entity with 'TransactionID' as the primary key. The dataset is highly imbalanced with 3.5% of fraudulent class of total data is appropriate for any binary classification problem. The target attribute 'isFraud' is binary data, where 1 represents fraudulent transaction and 0 otherwise. Now the dimension of the dataset has become 434 including the target feature that needs attention over some preprocessing steps.

Firstly data cleaning is done by removing missing data, and then the categorical features are transformed into numerical features using label encoder. Since autoencoder is used in the framework, feature scaling is adopted to normalize the features of the dataset. The rescaled numeric feature ranges [0, 1], eliminating different scale values between features. The general equation of min-max ranging from 0 to 1 is given as:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (11)$$

where x and x' in (11) are the actual data and the scaled value respectively. Secondly, as the dataset suffers due to curse of dimensionality, A hybrid feature selection technique [47] is applied to identify relevant and important 214 features from the data before it is fed to the AE-XGB framework for training. As stated earlier, the dataset has huge dimensionality, autoencoder model is used as a dimensionality reduction technique to improve the performance of the XGBoost in classifying fraudulent and legitimate transactions.

B. Performance Metrics

CCFD is an imbalanced classification problem with less frequency of positive class than the negative class. Therefore, accuracy is not a good metric that simply predicts fraud as legitimate giving high accuracy [2]. To address this issue, other evaluation metrics like Precision, Recall, F1-score and MCC are also accessed. In IEEE-CIS dataset the positive class is represented as fraudulent and negative as legitimate. From the confusion matrix shown in Table 1, the possible outcomes of classification are True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN).

TABLE I. CONFUSION MATRIX

Actual	**Predicted**	
	Fraudulent	***Legitimate***
Fraudulent	TP	FN (Type 2 Error)
Legitimate	FP (Type 1 Error)	TN

TP is the number of positive transactions which are actually positive. TN is the number of transactions that are predicted rightly as negative. FP is the number of transactions that are classified as positive but truly negative. FN is the number of transactions predicted as negative but actually positive. Therefore, the following metrics given in equation (12),(13)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

and (14) are applied to evaluate the efficiency of models used in this experiment.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

$$\text{F1} - \text{score} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

A comprehensive Mathew correlation coefficient (MCC) metric is appropriate for balanced and imbalanced dataset [34]. The values of MCC range between -1 and +1, where the value obtained from equation (15) ranging +1 indicates perfect predictions and -1 specifies contradictory predictions.

$$\text{MCC} = \frac{\text{TP X TN} - \text{FP X FN}}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \quad (15)$$

The metric area under receiver operating characteristic curve (AUC) is the trade-off between the true positive rate (TPR) and false positive rate (FPR) based on the threshold values adopted by maximizing F1-score of the classifier. If the AUC value is 1 then the classifier is considered perfect. AUC is commonly used metric in CCFD to evaluate the performance of models used. Precisely, the two important metrics recall and MCC are most prominent in detecting fraudulent transactions. When higher is the recall then fraud catching rate is more, which is the foremost objective of CCFD. While comparing the performance of model the metric MCC is considered, as it takes all the outcomes of confusion matrix as its parameter. As the dataset is imbalanced geometric mean (g-mean) given in equation (16) is a good indicator to evaluate the presence of bias in the model. G-mean is the balance between the TPR and TNR computed using the formula

$$\text{G} - \text{mean} = \sqrt{\text{TPR X TNR}} \quad (16)$$

V. RESULTS AND DISCUSSION

This study proposes a fraud detection system built on autoencoder for dimensionality reduction and XGBoost classifier for classifying fraudulent and legitimate transactions on the best threshold identified by the classifier. The experimental analysis is performed with the IEEE-CIS dataset from Kaggle. To evaluate the effectiveness of the proposed AE-XGB, the performance metrics precision, recall,f1-score, MCC and ROC are included. Instead of simply classifying the transaction using XGBoost, a probabilistic classification is built to classify data as fraud with probability p and legitimate with probability $1-p$ where $0 \leq p \leq 1$. Then the classification outcome from AE-XGB is classified as fraudulent when probability $p > \theta$ and legitimate otherwise. Definitely, for different threshold θ the performance of classification will lead to different values. Therefore, finding the best threshold is confronted to shift the threshold to produce better results in terms of f1-score. In this experiment, different threshold values of ROC curve ranging from 0 to 1 in step interim are applied to the classifier to test the data. The optimal threshold is found by maximized f1-score compared with the best score and current score on each data. Then the threshold is shifted from the default (0.5) to the optimal threshold. The optimal threshold recommended by the ROC curve for AE-XGB is 0.3, which has the best f1-score of 0.9057. The AE-XGB model produced good precision of 0.9066 and recall of 0.9048 for the optimal threshold of 0.3 are plotted in the PR- curve and to test overfitting of data, ROC curve is plotted with FPR on x-axis and TPR on y-axis as shown in figure 4. The AUC of training and testing data are found to be 0.99 and 0.98 respectively. Thus AE-XGB is suitable for dealing with imbalanced data.

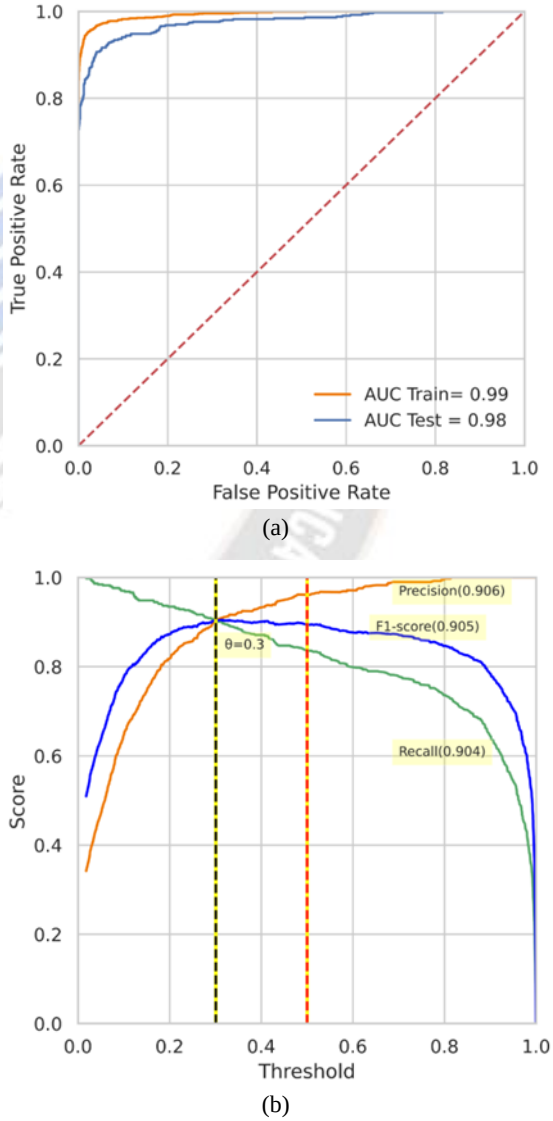
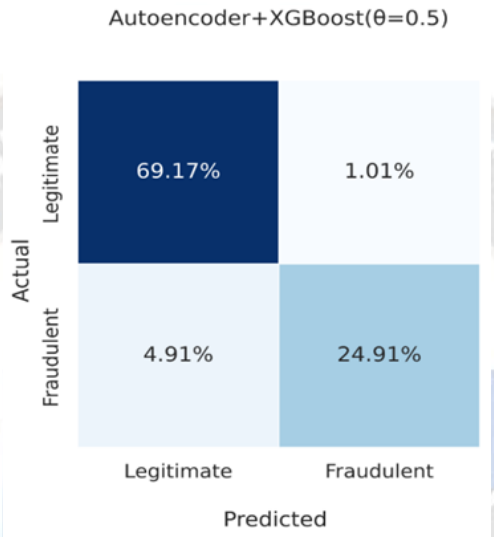


Figure 4. (a) ROC with AUC train and test values of 0.99 and 0.98 respectively. (b) PR curve representing the highest precision, recall and F1-score for the tuned threshold θ=0.3.

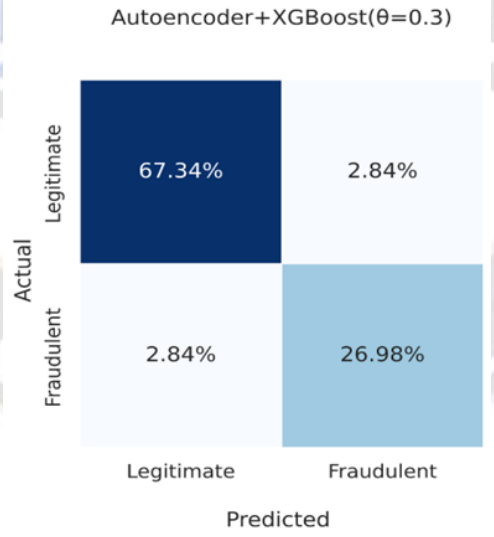
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

The same is compared with the default threshold of 0.5 and we obtained a precision of 0.9612 and recall of 0.8353, here we can observe an increase in precision and decrease in recall. The objective of the problem is to increase the recall rate as it denotes the fraudulent class that is predicted correctly. Similarly, for the tuned threshold other metrics are also considered to find the best MCC, g-mean, AUC and kappa. To investigate the performance of autoencoder, we designed a framework without autoencoder and assessed the performance only with XGBoost. The optimal threshold attained is 0.245 with the maximized f1-score as 0.7325 and recall as 0.6258. However, the performance of proposed method has produced better results. Table 2 shows the performance of all the above stated metrics for the optimal and default threshold values.

The confusion matrix in figure 5 shows the TP, TN, FP and FN detection rate for the default and optimal threshold.



(a)



(b)

Figure 5. Confusion matrix for test set with (a) default threshold(θ=0.5) and (b) optimal threshold(θ=0.3).

TABLE II. PERFORMANCE OUTCOME OF AE-XGB.

Model	Precision	Recall	f1_Score	G-mean	Cohen (K)	MCC
XGBoost(θ=0.5)	0.9714	0.5215	0.6786	0.7220	0.6723	0.7065
XGBoost(θ=0.245)	0.8831	0.6258	0.7325	0.7901	0.6723	0.7065
AE-XGB(θ=0.5)	**0.9612**	0.8353	0.8938	0.9074	0.8531	0.8571
AE-XGB(θ=0.3)	0.9066	**0.9048**	**0.9057**	**0.9322**	**0.8656**	**0.8656**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

When θ =0.3, TP has been increased by 2.07% where the model performed well in catching fraudulent transactions. At the same time, FP also increased by 1.83% which has to be taken care since, an increase in the false positive rate may lead to customer dissatisfaction as legitimate transactions are assumed to be fraudulent.

To compare the performance of AE-XGB with other ensemble algorithms for the optimal and default threshold in terms of f1-score are shown in figure 6.

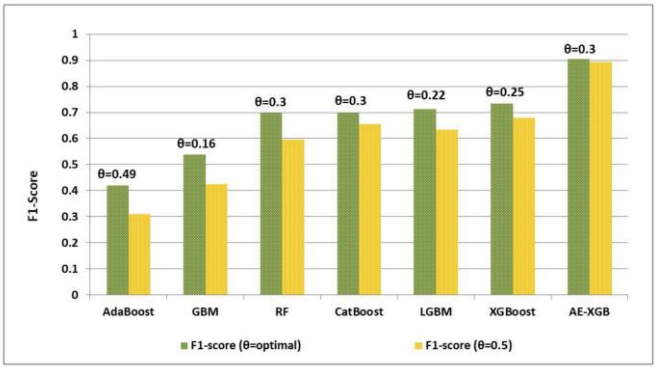


Figure 6. F1-score comparison of AE-XGB and other boosting models with default (θ=0.5) and optimal threshold attained by each model.

Table 3 presents the performance comparison of the proposed AE-XGB with other previous works done with the same dataset and similar approaches. It can be observed that AE-XGB outperforms the other works scoring high almost in all metrics. The proposed AE-XGB with default threshold (θ=0.5), had the highest precision of 96% and recall of 84% achieved. Thus, threshold tuning was implemented using AE-XGB to attain optimal threshold θ=0.3 to produce a precision score of 91% and the highest recall of 90%. The objective of credit card fraud detection is to achieve the highest recall possible while attaining a satisfactory f1-score of 91%. Hence AE-XGB with optimal threshold θ=0.3 is chosen to be the best.

TABLE III. PERFORMANCE COMPARISON OF AE-XGB AND RELATED METHODS FOR IEEE-CIS DATASET.

Research	Methods	Precision	Recall	F1-score
Esraa Faisal Malik et al. [48]	AdaBoost+XGB	0.94	0.59	0.73
Siddharth Vimal et al. [49]	Deep-Q NR	0.48	0.35	0.41
Ruangsakorn et al. [50]	XGBoost	0.95	0.57	0.72
This research	AE-XGB (θ=0.5)	**0.96**	0.84	0.89
This research	AE-XGB (θ=0.3)	0.91	**0.90**	**0.91**

VI. CONCLUSION

The study proposes AE-XGB method for digital fraud detection. Autoencoder is employed to reduce the dimensionality of data by extracting data attributes. Later the acquired features are fed to XGBoost which is utilized with probabilistic threshold classification to classify fraudulent and legitimate transactions with a related probability. The final classification of fraudulent class by AE-XGB depends on the associated probability more than the determined threshold. The IEEE-CIS fraud detection dataset was applied to evaluate the performance of AE-XGB. Moreover, the dataset is extremely imbalanced, we opted ensemble learning method XGBoost to handle class imbalance in the dataset using regularization technique. The experimental outcome indicates that the AE-XGB is suitable for tackling imbalanced dataset without resampling data. The IEEE-CIS fraud detection dataset is partitioned into training dataset of 70% data and testing dataset of 30% data for determining the performance of AE-XGB. However, to test the robustness and efficiency of AE-XGB, we need to implement the model with different datasets. To analyse the performance of AE-XGB method, it is compared with other machine learning algorithms such as AdaBoost, GBM, Random Forest, CatBoost, LGBM and XGBoost. The outcome of the proposed model exhibits a promising f1-score of 91%. The experimental evaluation results of AE-XGB were compared with other related methods such as AdaBoost+XGB [48], Deep-Q NR [49] and XGBoost [50] of same dataset. The comparison result shows that AE-XGB with θ=0.3 had the good precision score of 91% and high recall of 90% with a highest f1-score of 91%.

In future, for verifying the efficiency and robustness of AE-XGB we plan to apply similar CCFD datasets to the proposed model. Furthermore, to investigate the applicability of AE-XGB, the proposed model can be applied to different applications.

REFERENCES

[1] Nilson Report, “The Nilson Report Newsletter Archive,” The Nilson Report, 2019. https://nilsonreport.com/publication_newsletter_archive_issue.php?issue=1146.

[2] C. V. Priscilla and D. P. Prabha, "Credit Card Fraud Detection: A Systematic Review," in Springer, Cham, 2020, pp. 290–303. https://doi.org/10.1007/978-3-030-38501-9_29

[3] A. D. Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," IEEE Trans. Neural Networks Learn. Syst., vol. 29, no. 8, pp. 3784–3797, 2018. https://doi.org/10.1109/TNNLS.2017.2736643

[4] H. Wang, P. Zhu, X. Zou, and S. Qin, "An Ensemble Learning Framework for Credit Card Fraud Detection Based on Training Set Partitioning and Clustering," in 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCom/IoP/SCI 2018, 2018, pp. 94–98.

[5] Abdallah, M. A. Maarof, and A. Zainal, "Fraud detection system: A survey," J. Netw. Comput. Appl., vol. 68, pp. 90–113, 2016. https://doi.org/10.1016/j.jnca.2016.04.007

[6] S. Sorournejad et al., "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective," CoRR, vol.abs/1611.0,November,2016. https://doi.org/10.48550/arXiv.1611.06439

[7] A. A. Taha and S. J. Malebary, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," IEEE Access, vol. 8, no. February,pp.25579–25587,2020. https://doi.org/10.1109/ACCESS.2020.2971354

[8] D. Elavarasan, P. M. Durai Raj Vincent, K. Srinivasan, and C. Y. Chang, "A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling," Agric., vol. 10, no. 9, pp. 1–27, 2020. https://doi.org/10.3390/agriculture10090400

[9] N. Barraza, S. Moro, M. Ferreyra, and A. de la Peña, "Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study," J. Inf. Sci., vol. 45, no.1,pp.53–67,2019. https://doi.org/10.1177/0165551518770967

[10] H. Jeon and S. Oh, "Hybrid-Recursive Feature Elimination for Efficient Feature Selection," Appl. Sci., vol. 10, no. 9, p. 3211, 2020. https://doi.org/10.3390/app10093211

[11] S. Wang, C. Liu, X. Gao, H. Qu, and W. Xu, "Session-Based Fraud Detection in Online E-Commerce Transactions Using Recurrent Neural Networks," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, pp. 241–252. https://doi.org/10.1007/978-3-319-71273-4_20

[12] Kaur, P., Gosain, A. Issues and challenges of class imbalance problem in classification. Int. j. inf. tecnol. 14, 539–545 (2022). https://doi.org/10.1007/s41870-018-0251-8

[13] Y. Zhang, G. Liu, L. Zheng, and C. Yan, "A hierarchical clustering strategy of processing class imbalance and its application in fraud detection," Proc. - 21st IEEE Int. Conf. High Perform. Comput. Commun. 17th IEEE Int. Conf. Smart City 5th IEEE Int. Conf. Data Sci. Syst. HPCC/SmartCity/DSS 2019, pp. 1810–1816, 2019.

[14] F. F. Noghani and M.-H. Moattar, "Ensemble Classification and Extended Feature Selection for Credit Card Fraud Detection," J. AI Data Min., vol. 5, no. 2, pp. 235–243, 2017.

[15] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," Proc. IEEE Int. Conf. Comput. Netw. Informatics, ICCNI 2017, vol. 2017-Janua, pp. 1–9, 2017.

[16] C. Wang and D. Han, "Credit card fraud forecasting model based on clustering analysis and integrated support vector machine," Cluster Comput., vol. 0123456789, pp. 1–6, 2018. https://doi.org/10.1007/s10586-018-2118-y

[17] J. Jurgovsky et al., "Sequence classification for credit-card fraud detection," Expert Syst. Appl., vol. 100, pp. 234–245, 2018. https://doi.org/10.1016/j.eswa.2018.01.037

[18] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, Random forest for credit card fraud detection," ICNSC 2018 - 15th IEEE Int. Conf. Networking, Sens. Control, pp. 1–6, 2018.

[19] G. Rushin, C. Stancil, M. Sun, S. Adams, and P. Beling, "Horse race analysis in credit card fraud—deep learning, logistic regression, and Gradient Boosted Tree," in 2017 Systems and Information Engineering Design Symposium (SIEDS), 2017, pp. 117–121.

[20] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, "Deep learning detecting fraud in credit card transactions," in 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, pp. 129–134.

[21] J. Akosa, "Predictive accuracy: A misleading performance measure for highly imbalanced data," in Proceedings of the SAS Global Forum, 2017.

[22] Ahmad, H., Kasasbeh, B., Aldabaybah, B. et al. Class balancing framework for credit card fraud detection based on clustering and similarity-based selection (SBS). Int. j. inf. tecnol. (2022). https://doi.org/10.1007/s41870-022-00987-w

[23] Itoo, F., Meenakshi & Singh, S. Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. Int. j. inf. tecnol. 13,1503–1511 (2021). https://doi.org/10.1007/s41870-020-00430-y

[24] C. Zhang and X. Zhang, "An effective sampling strategy for ensemble learning with imbalanced data," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10363 LNAI, pp. 377–388, 2017.

[25] G. Rekha, A. K. Tyagi, and V. Krishna Reddy, "A novel approach to solve class imbalance problem using noise filter method," Adv. Intell. Syst. Comput., vol. 940, pp. 486–496, 2020. https://doi.org/10.1007/978-3-030-16657-1_45

[26] G. Collell, D. Prelec, and K. R. Patil, "A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data," Neurocomputing, vol.275,pp.330–340,2018. https://doi.org/10.1016/j.neucom.2017.08.035

[27] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal thresholding of classifiers to maximize F1 measure," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8725 LNAI, no. PART 2. pp. 225–239, 2014.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7234
***Article Received**: 28 April 2023 **Revised:** 19 June 2023 **Accepted:** 30 June 2023*

[28] N. Thai-Nghe, "Learning optimal threshold on resampling data to deal with class imbalance," Proc. IEEE RIVF …, 2010.

[29] C. Esposito, G. A. Landrum, N. Schneider, N. Stiefl, and S. Riniker, "GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning," Journal of Chemical Information and Modeling, vol. 61, no. 6. pp. 2623–2640, 2021. https://doi.org/10.1021/acs.jcim.1c00160

[30] C. Tang, N. Luktarhan, and Y. Zhao, "SS symmetry An Efficient Intrusion Detection Method Based on," pp. 1–16, 2020.

[31] Q. Y. Yin, J. S. Zhang, C. X. Zhang, and N. N. Ji, "A novel selective ensemble algorithm for imbalanced data classification based on exploratory undersampling," Math. Probl.Eng.,vol.2014,no.ii,2014. https://doi.org/10.1155/2014/358942

[32] B. S. Raghuwanshi and S. Shukla, "Class imbalance learning using UnderBagging based kernelized extreme learning machine," Neurocomputing, vol. 329, pp. 172–187, 2019. https://doi.org/10.1016/j.neucom.2018.10.056

[33] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Ann. Stat., pp. 1189–1232, 2001. https://doi.org/10.1214/aos/1013203451

[34] G. Ke et al., "Lightgbm: A highly efficient gradient boosting decision tree," in Advances in neural information processing systems, 2017, pp. 3146–3154.

[35] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," in Advances in neural information processing systems, 2018, pp. 6638–6648.

[36] D. Chushig-muzo, C. Soguero-ruiz, P. De Miguel-bohoyo, and I. Mora-jim, "Interpreting clinical latent representations using autoencoders and probabilistic models ," vol. 122, 2021. https://doi.org/10.1016/j.artmed.2021.102211

[37] A. Oluwasanmi, M. U. Aftab, E. Baagyere, Z. Qin, M. Ahmad, and M. Mazzara, "Attention Autoencoder for Generative Latent Representational Learning in Anomaly Detection,"pp. 1–14, 2022. https://doi.org/10.3390/s22010123

[38] X. Li, W. Chen, Q. Zhang, and L. Wu, Computers & Security Building Auto-Encoder Intrusion Detection System based on random forest feature selection," Comput. Secur., vol.95,p.101851,2020. https://doi.org/10.1016/j.cose.2020.101851

[39] A. Pumsirirat and L. Yan, "Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine," Int. J. Adv. Comput. Sci. Appl., vol. 9, no.1,pp.18–25, 2018.

[40] http://dx.doi.org/10.14569/IJACSA.2018.090103

[41] W. W. Y. Ng, G. Zeng, J. Zhang, D. S. Yeung, and W. Pedrycz, "Dual autoencoders features for imbalance classification problem," Pattern Recognit., vol. 60, pp. 875–889, 2016. https://doi.org/10.1016/j.patcog.2016.06.013

[42] T. Lin and J. Jiang, "Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest," pp. 4–15, 2021. https://doi.org/10.3390/math9212683

[43] A. Alazizi and A. Habrard, "Dual Sequential Variational Autoencoders for Fraud Detection," vol. 2, pp. 14–26, 2020. https://doi.org/10.1007/978-3-030-44584-3_2

[44] A. F. M. Agarap, "Deep Learning using Rectified Linear Units(ReLU),"no.1,pp.2–8. https://doi.org/10.48550/arXiv.1803.08375

[45] Weijie Wang and Yanmin Lu., "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model," 2018. https://doi.org/10.1088/1757-899X/324/1/012049

[46] Dharmesh D, Natural Language Processing for Automated Document Summarization , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[47] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785

[48] Kaggle, "IEEE-CIS Fraud Detection." Online]. Available: https://www.kaggle.com/c/ieee-fraud-detection/data.

[49] C. V. Priscilla and D. P. Prabha, "A two-phase feature selection technique using mutual information and XGB- RFE for credit card fraud detection," Int. J. Adv. Technol. Eng. Explor.,vol.8,no.85,2021. https://doi.org/10.19101/IJATEE.2021.874615

[50] E. F. Malik, K. W. Khaw, B. Belaton, and W. P. Wong, "Credit Card Fraud Detection Using a New Hybrid Machine Learning architecture," 2022. https://doi.org/10.3390/math10091480

[51] S. Vimal, Application of Deep Reinforcement Learning to Payment Fraud, vol. 1, no. 1. Association for Computing Machinery. https://doi.org/10.48550/arXiv.2112.04236

[52] T. R. B and S. Yu, "A Study on Comparative Evaluation of Credit Card Fraud Detection Using Tree-Based," vol. 1, pp. 212–219. https://doi.org/10.1007/978-3-030-70639-5_20