Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Enhanced Fraud Detection in Credit Card Transactions With Data Balancing and XGBoost

Mst. Sirazum Munira Mim
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
sirazummuniramim@gmail.com

Md. Munem Shahriar
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
munemshahriar223@gmail.com

Mobassir Ahmmed
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
mobassirahmmeds@gmail.com

Md. Muktar Hossain
Computer Science and Engineering
Rajshahi University of
Engineering and Technology
Rajshahi, Bangladesh
mmuktar997@gmail.com

A.S.M Delwar Hossain
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
delwar.hossain.vu@gmail.com

Tanver Ahmed
Computer Science and Engineering
Varendra University
Rajshahi, Bangladesh
tanverahmed.cse@gmail.com

***Abstract*—Credit card fraud detection presents a significant challenge in the financial sector due to the rarity of fraudulent transactions and the need for real-time accurate classification. Fraudulent transaction detection using traditional manual approaches is ineffective and time-consuming. This research introduced a model that combines XGBoost and an oversampling method to solve imbalanced classification in fraud detection. Using a labeled credit card fraud dataset, the performance of this model is compared to other machine learning models. This evaluation brings out the impact of techniques such as feature engineering, class imbalance handling, and parameter tuning in improving results and demonstrates the proposed model's superiority over other approaches. According to comparison findings, XGBoost with oversampling other models such as Logistic Regression (LR), Support Vector Machine (SVM), and Decision Tree (DT). The results show that the SVM, DT, and LR classifiers have overall accuracies (OA) of 0.9941, 0.9992, and 0.9992, respectively, while XGBoost classifiers have an OA of 0.9996.**

***Index Terms*—Imbalanced Classification, Oversampling, Support Vector Machine, Logistic Regression, Decision Tree, XGBoost, Accuracy, Precision, Recall, F1-Score, Kappa Score.**

I. Introduction

Financial fraud is a rising threat to businesses and the financial industry, including involvement in illegal acts for monetary benefit. Credit card transactions have grown as a result of the growth of internet technology, leading to an increase in fraud. This fraud can be classed as internal (collusion between cardholders and banks using fraudulent identities) or external (illegal use of stolen credit cards). [1].

XGBoost is a popular and potent machine learning algorithm that is well-known for its accuracy and efficiency while processing a wide range of data types and complexity [2]. It belongs to the class of methods for algorithm learning known as gradient boosting frameworks, which combine the ability to predict several models to provide a more reliable prediction overall [3]. Gradient boosting is the foundation of XGBoost, which improves the technique by highlighting regularization [4]. It iteratively fixes the mistakes created by the previous models by adding choices to an array one after the other. Through this iterative method, XGBoost is able to gradually enhance its predictive performance. The goal of XGBoost is to identify the best possible combination of weak learners that together yield a powerful prediction model [5]. This procedure measures the difference between actual and expected values by decreasing a loss function.

Algorithmic approaches have been used as a result of research on detecting external fraud, which makes up 90% of credit card fraud examples. Transactions have been categorized to be genuine or fraudulent using data mining techniques as Support Vector Machines (SVM), Decision Trees (DT), XGBoost(XGB), AdaBoostRegressor, Random Forest, and Logistic Regression (LR) [6]. Data imbalance, feature selection, along suitable performance metrics, represent some of the problems that credit card fraud detection must overcome. This study compares XGBoost against SVM, decision trees, Naïve Bayes, AdaBoost Regression, and logistic regression on highly imbalanced credit card fraud data using oversampling, evaluating accuracy, precision, recall, F1-score, and Kappa score.

II. Related Work

In this modern era people are using different methods to pay money instead of hand cash when purchasing something. Payments are often made through mobile banking or credit/debit cards. In the field of detecting credit card fraud, numerous studies have been completed. To provide rapid and precise outcomes for the real-time application for recognizing credit card fraud, machine learning algorithms are employed. Data mining classification problems like credit card fraud detection aim to correctly categorize credit card transactions as valid or fraudulent. [7] E.A. Amusan et al. [8] firstly tried to balance the unbalanced data set using under-sampling. To increase prediction accuracy, he then investigated a number of machine learning classification models, such as Random Forest, KNN, logistic regression, and decision trees. In this paper the random forest algorithm shows accuracy of 95% where the other algorithm shows accuracy above 90%. To differentiate between legitimate and fraudulent transactions, E.A.M. Suresh Kumar [9] initially used the decision tree method and then the random forest algorithm. In this method, the decision

tree method is used for classification in a supervised learning algorithm. The accuracy of this model was 90%.

III. Methodologies

This section outlines the dataset and the four classifiers evaluated in the experiments: XGBoost, SVM, decision trees, and logistic regression. The classifier development process consists of multiple stages: collection of data, preprocessing, analysis, training, and evaluation. Preprocessing involves formatting and sampling the data, with random oversampling intended to balance positive cases. Principal Component Analysis (PCA) is used to select features and reduce dimensionality during analysis. Then Random oversampling is used, which is a non-heuristic strategy for balancing class distributions that involves randomly reproducing minority target instances. Classifiers are then trained on the processed data, and their performance is measured using metrics like True Positive, True Negative, False Positive, and False Negative. The classifiers are compared using accuracy, precision, recall, F1-score, and Kappa score.

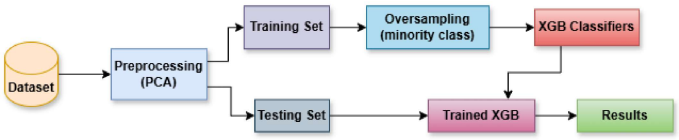

Fig. 1: Steps to detect credit card fraud.

A. Dataset Description

The dataset is collected from the ULB Machine Learning group, and its description is found in [10]. The dataset contains 284807 transactions that occurred in two days, and the dataset was created by European cardholders in September 2013. The dataset contains 0.172% positive class(Fraud case) and 99.828% false case(Normal transactions) of total transactions. The dataset is extremely imbalanced, with only numerical input variables obtained by PCA, providing 28 main components and 30 features in total. Details about the features are unavailable due to confidentiality concerns. The dataset has three features: 'time' (elapsed seconds), 'amount' (transaction value), and 'class' (a binary target of 1 for fraud and 0 for non-fraud) [11].

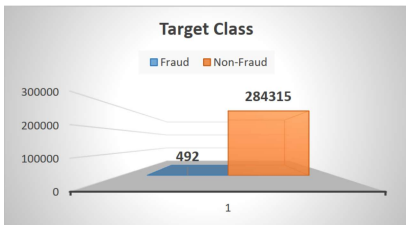

Fig. 2: Distribution of target classes.

B. Oversampling Minority Class

Random oversampling is a non-heuristic strategy for balancing class distributions that involves randomly reproducing minority target instances.

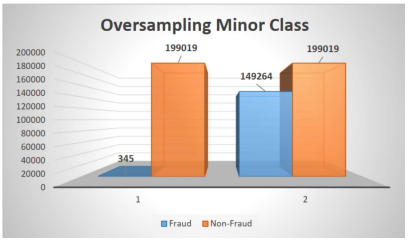

Fig. 3: Class target distribution after oversampling method.

C. Support Vector Machine

The Support Vector Machine (SVM) seeks to identify a hyperplane $w^T x + b = 0$ that optimally divides data points into two distinct classes with the greatest margin [12]. The hyperplane is defined as:

$$w^T x + b = 0$$

where:

- w: Weight vector
- b: Bias term
- x: Input data points

The kernel technique, which uses a function $\phi(x)$ to transfer the input data x into a higher-dimensional space, is used for non-linear data. The goal is to minimize the following in order to maximize the margin, $\frac{2}{\|w\|}$:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \varsigma_i$$

Subject to the constraints:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \varsigma_i$$

where y_i are the class labels ($+1$ or -1). Slack variables ς_i are added to account for possible mistakes or miscalculations if these points cannot be linearly separated. C is a cost parameter > 0 is associated with these errors [13].

D. Logistic Regression

Based on one or more input features, logistic regression estimates the probability of a binary outcome using a functional method. It defines the optimal parameters for the sigmoid function, a nonlinear function [11]. The sigmoid function is defined in equation (1), whereas equation (2) represents the input x as a weighted sum of the feature values (z), where the coefficients w are multiplied by each corresponding feature element and added to generate a single result. This value is used to classify the target class. If the sigmoid value exceeds 0.5, the output is classed as 1, otherwise as 0.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

$$x = w_1 z_1 + w_2 z_2 + \cdots + w_n z_n \tag{2}$$

Where σ is the sigmoid function, and w are the best-fit coefficients for the input data vector z.

E. Decision Tree

A decision tree splits records into subsets based on attribute values, starting with the root node. The tree arises by recursively splitting nodes until no significant splits are possible or the node size is insufficient. Splitting algorithms like ID3, C5.0, and CART make use of metrics such as information gain, gain ratio, and Gini coefficient. Pruning eliminates superfluous nodes to prevent overfitting. New records are classified by traversing the tree from root to leaf, with the class determined by the label of the leaf [14].

F. XGBoost

An ensemble learning technique called Extreme Gradient Boosting (XGB) combines the outputs of several decision trees to create a powerful model that improves prediction accuracy. Because of its versatility and strong architecture, XGB handles imbalanced datasets quite well. It emphasizes the minority class and support for evaluation requirements that are better suited to skewed data. Its iterative boosting approach corrects biases resulting from class imbalance, while regularization avoids overfitting to the dominant class. XGB, with its scalability and adaptable hyperparameters, performs well even on extremely skewed datasets [15].

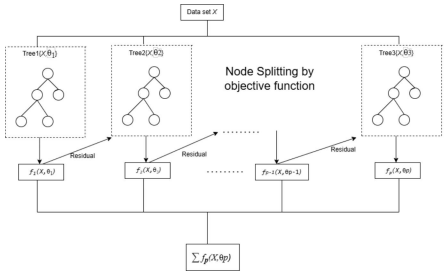

Fig. 4: XGBoost algorithm working procedure [15].

G. Evaluation Metrics

A wide range of metrics, including accuracy, precision, recall, F1-score, and kappa score, are used to assess the Extreme Gradient Boosting (XGB) model's performance. These indicators offer a comprehensive evaluation of the model's credit card fraud detection capabilities. The formulas for overall accuracy, precision, recall, F1-score, and kappa score are displayed in Equations 1, 2, 3, 4, and 5, respectively:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1-Score = \frac{2*Precision*Recall}{Precision+Recall} \tag{4}$$

$$k = \frac{p_o - p_e}{1-p_e} \tag{5}$$

where,

- True Positive(TP): The number of cases accurately anticipated as positive.
- False Positive(FP): The number of cases that were wrongly anticipated as positive.
- False Negative(FN): The number of cases that were wrongly anticipated as negative.
- True Negative(TN): The number of cases that were accurately anticipated as negative.
- p_o means the overall accuracy of the model
- p_e means the measure of the agreement between model predictions and actual class values as if they occurred by random.

IV. Experimental Result and Analysis

In this experiment, four classifier models based on Decision Tree, Support Vector Machine, Logistic Regression, XGBoost are developed with the oversampling technique. To evaluate the models the dataset is distributed in a 43:57 ratio, where random oversampling is used for the 43:57 distribution. Random oversampling is a non-heuristic strategy for balancing class distributions that involve randomly reproducing minority target instances. Accuracy, Precision, Recall, F1-score, and kappa score are the performance metrics. The Accuracy, Precision, Recall, F1-score, kappa score for the test size of 30% and 50% are present in the Table I and II respectively.

While excellent accuracy is exhibited by all models, it does not accurately represent how well fraud is actually identified. A clearer picture is provided by metrics like F1-score and Kappa, which demonstrate how effectively fraud detection and misclassification mistakes are balanced by XGBoost to surpass other models.

On 30% test set, XGBoost outperforms other classifiers in every metric, with outstanding overall accuracy (0.9996), precision (0.9611), F1-Score (0.9432), and Kappa (0.8864), indicating its durability on the oversampled data. While SVM has the highest recall (0.9351), it suffers from poor accuracy (0.5848) and F1-Score (0.6403), making it less reliable overall. DT and LR perform relatively well in terms of accuracy (0.9992 for both) and balanced metrics, but they fall short of XGBoost's total performance. XGBoost is the obvious pick for this fraud detection because of its excellent balance across various evaluation metrics.

On the 50% test set, XGBoost is superior to other classifiers in overall accuracy (0.9995), precision (0.9657), F1-Score (0.9262), and Kappa (0.8524), emphasizing its durability and balance. While SVM has the highest recall (0.9292), its poor accuracy (0.6024) and F1-Score (0.6643) make it less reliable than other models. Decision Tree and Logistic Regression perform relatively well across metrics, but XGBoost consistently outperforms them, making it the most effective classifier on this dataset.

The tables (30% and 50% test sizes) compare SVM, LR, DT, and XGB on five criteria. XGB regularly has the best overall accuracy (0.9996, 0.9995), precision (0.9611, 0.9657), F1-score (0.9432, 0.9262), and Kappa (0.8864, 0.8524). SVM improves across all metrics with a 50% test size, including improved accuracy (0.6024) and Kappa (0.3296), but still lags overall. LR and DT remain consistent with only minor changes to accuracy and recall. XGB excels at balancing precision and recall, making it the best classifier by resolving data imbalances and enhancing

TABLE I: Classifier Performance on 30% Test Data with Oversampling

	Classifiers			
Metrics	SVM	LR	DT	XGB
accuracy	0.9924	0.9992	0.9992	**0.9996**
precision	0.5848	0.9420	0.8990	**0.9611**
Recall	**0.9351**	0.8129	0.8638	0.9257
F1-Score	0.6403	0.8663	0.8806	**0.9432**
Kappa	0.2824	0.7072	0.7612	**0.8864**

TABLE II: Classifier Performance on 50% Test Data with Oversampling

	Classifiers			
Metrics	SVM	LR	DT	XGB
accuracy	0.9941	0.9992	0.9991	**0.9995**
precision	0.6024	0.9387	0.8819	**0.9657**
Recall	**0.9292**	0.7962	0.8599	0.8930
F1-Score	0.6643	0.8536	0.8705	**0.9262**
Kappa	0.3296	0.7072	0.7411	**0.8524**

classifier performance.

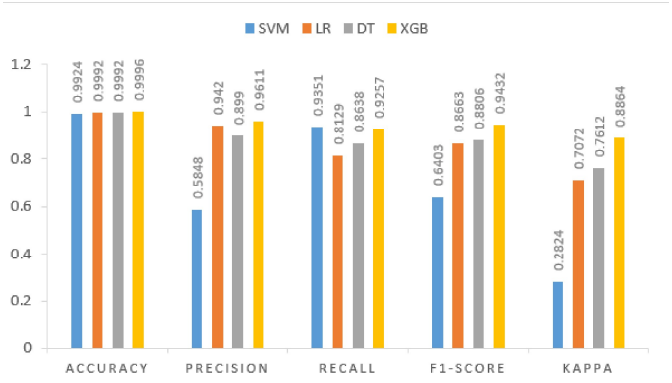

Fig. 5: Classifier Performance on 30% Test Data with Oversampling

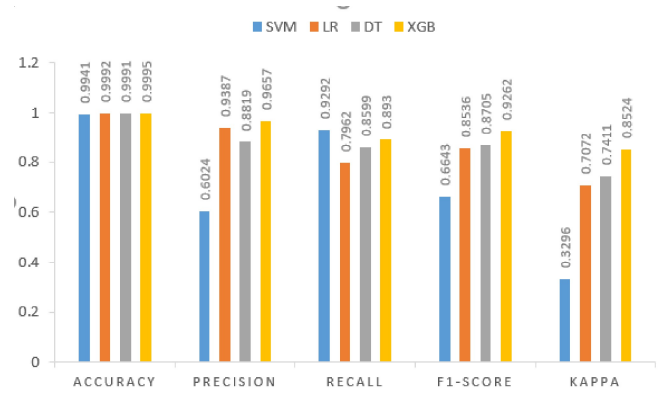

Fig. 6: Classifier Performance on 50% Test Data with Oversampling

V. Conclusion

The research demonstrates that XGBoost consistently outperforms other investigated classifiers, such as Decision Tree, Support Vector Machine, and Logistic Regression, across all evaluation metrics, including accuracy, precision, F1-score, and Kappa score, for both 30% and 50% test sizes. Its excellent results demonstrate its ability to handle imbalanced datasets with oversampling, particularly for tasks such as fraud detection where balancing precision and recall is important. Although SVM has the maximum recall, it has a low accuracy and F1-score, making it less reliable overall. Decision Tree and Logistic Regression perform consistently but not extraordinarily, with minimal differences depending on test size. XGBoost consistently delivers balanced and reliable results across varying test sizes, making it the most effective classifier for this dataset and oversampling method.

References

[1] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 1, pp. 145–174, 2023.

[2] A. Maulana, F. R. Faisal, T. R. Noviandy, T. Rizkia, G. M. Idroes, T. E. Tallei, M. El-Shazly, and R. Idroes, "Machine learning approach for diabetes detection using fine-tuned xgboost algorithm," *Infolitika Journal of Data Science*, vol. 1, no. 1, pp. 1–7, 2023.

[3] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning. frontiers of computer science," 2020.

[4] E. Al Daoud, "Comparison between xgboost, lightgbm and catboost using a home credit dataset," *International Journal of Computer and Information Engineering*, vol. 13, no. 1, pp. 6–10, 2019.

[5] H. Li, Y. Cao, S. Li, J. Zhao, and Y. Sun, "Xgboost model and its application to personal credit evaluation," *IEEE Intelligent Systems*, vol. 35, no. 3, pp. 52–61, 2020.

[6] T. Pranckevičius and V. Marcinkevičius, "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification," *Baltic Journal of Modern Computing*, vol. 5, no. 2, p. 221, 2017.

[7] K. Seeja and M. Zareapoor, "Fraudminer: A novel credit card fraud detection model based on frequent itemset mining," *The Scientific World Journal*, vol. 2014, no. 1, p. 252797, 2014.

[8] E. Amusan, O. Alade, O. Fenwa, and J. Emuoyibofarhe, "Credit card fraud detection on skewed data using machine learning techniques," *Lautech Journal of Computing and Informatics*, vol. 2, no. 1, pp. 49–56, 2021.

[9] M. S. Kumar, V. Soundarya, S. Kavitha, E. Keerthika, and E. Aswini, "Credit card fraud detection using random forest algorithm," in *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*. IEEE, 2019, pp. 149–153.

[10] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *2015 IEEE symposium series on computational intelligence*. IEEE, 2015, pp. 159–166.

[11] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *2017 international conference on computing networking and informatics (ICCNI)*. IEEE, 2017, pp. 1–9.

[12] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1–27, 2011.

[13] 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. 93 010–93 022, 2019.

[14] Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, 2011, pp. 1–6.

[15] D. T. Tran, T. Onjaipurn, D. R. Kumar, W. Chim-Oye, S. Keawsawasvong, and P. Jamsawang, "An extreme gradient boosting prediction of uplift capacity factors for 3d rectangular anchors in natural clays," *Earth Science Informatics*, pp. 1–15, 2024.