FMDB *Transactions on Sustainable Computing Systems*

Comparative Analysis of Machine Learning Algorithms for Credit Card Fraud Detection

Sonjoy Ranjon Das[1,*], Rejwan Bin Sulaiman[2], Usman Butt[3]

[1,2]Department of Computer Science and Technology, Northumbria University, London, United Kingdom.
[3]Faculty of Engineering and Information Technology, The British University in Dubai, Dubai, United Arab Emirates.
sonjoy.das@northumbria.ac.uk[1], rejwan.sulaiman@northumbria.ac.uk[2], usman.butt@buid.ac.ae[3]

Abstract: The issue of credit card fraud poses a significant concern for users of online transactions, necessitating the implementation of effective fraud detection mechanisms. Fraud detection is often done using machine learning algorithms. Practitioners can compare and analyse algorithms to get the best one for their credit card fraud detection scenario. This article describes a detailed study to find the best credit card fraud forecasting model. The study tests cutting-edge supervised machine learning methods on two datasets. Using eight algorithms improves credit card fraud detection accuracy and efficacy. Logistic Regression, Decision Trees, Random Forests, Multilayer Perceptions, Naive Bayes, XGBoost, KNN, and SVM are examples (SVM). Additionally, Principal Component Analysis (PCA) is used to reduce dimensionality and improve algorithm performance during experimentation. XGBoost has the maximum accuracy of 99.96 percent for the first dataset, while Random Forest has 99.92 percent for the second. Cross-validation with Logistic Regression, Decision Trees, Random Forests, and XGBoost proves Random Forests are better at credit card fraud detection. Random Forests excel at undersampling and oversampling. Thus, this paper proposes XGBoost and Random Forests as the most reliable credit card fraud detection algorithms.

Keywords: Comparative Analysis; Machine Learning Algorithms; Random Forests; Multilayer Perceptions; Naive Bayes; Credit Card Fraud Detection; Fraud Detection Mechanisms; Principal Component Analysis; Logistic Regression.

Received on: 09/05/2023, **Revised on:** 05/09/2023, **Accepted on:** 22/10/2023, **Published on:** 20/12/2023

Cite as: S. Ranjon Das, R. Bin Sulaiman and U. Butt, "Comparative Analysis of Machine Learning Algorithms for Credit Card Fraud Detection," *FMDB Transactions on Sustainable Computing Systems*., vol. 1, no. 4, pp. 225–244, 2023.

1. Introduction

As online shopping becomes increasingly prevalent, debit and credit cards offer a simple and secure way to embrace the digital age [13]. However, it is important to acknowledge the rising issue of credit card fraud (CCF) amidst the increasing popularity of these payment methods. CCF refers to the unauthorized use of someone else's credit or debit card for purchases or cash advances [3]. To combat credit card fraud (CCF) in the face of increasingly sophisticated digital techniques used by thieves to gain unauthorized access to credit card numbers and personal information, this research aims to construct a comparative analysis using state-of-the-art machine learning algorithms to effectively identify instances of fraudulent credit card use [11]. In pursuit of this objective, the study has developed skills in topic selection, literature review, data gathering, analysis, and preprocessing. These skills were honed using tools such as Summon, Catalogue, Article Database, and Google Scholar. Additionally, programming in Python and Jupyter Notebook was utilized. Through this research endeavor, the article aims to provide consumers and financial institutions with advanced tools to effectively combat the constantly changing challenges of credit card fraud (CCF) [15]. In today's digital landscape, it has become essential to utilize efficient machine learning algorithms to

*Corresponding author.

differentiate between legitimate and fraudulent transactions [19]. This empowers bank employees to take proactive measures in notifying cardholders, thereby ensuring the safeguarding of their finances [12].

Given the growing concerns among financial organizations regarding credit card fraud and the widespread use of e-commerce platforms, it has become crucial for practitioners and industries to prioritize the secure usage of credit cards. This involves conducting a thorough analysis to determine which algorithm can deliver dependable results with minimal computational costs in the context of Credit Card Fraud (CCF) [32]. This paper utilizes eight different algorithms in the study to compare and determine the most effective one through extensive experimentation. However, existing fraud detection methods often suffer from limitations, resulting in either false positives or false negatives. This can lead to financial losses for both cardholders and financial institutions [24]. Considering this, the present study aims to compare the performance of eight machine learning (ML) algorithms in detecting credit card fraud (CCF) using two datasets. This research has thoroughly examined Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Multilayer Perceptron (MLP), Naive Bayes (NB), XGBoost, KNN, and SVM algorithms to determine the model that produces the most accurate predictions and results for dataset 1 and 2. Furthermore, to improve the accuracy and effectiveness of CCF detection, this study has integrated Principal Component Analysis (PCA) for dimensionality reduction. This ensures that important data characteristics are preserved, thereby enhancing the overall detection process [13]. Throughout the research, various performance metrics such as accuracy, precision, F1-score, and recall will be calculated and analyzed for each model. The findings will be visually presented, compared with other models and research efforts, and evaluated based on their accuracy, precision, f1-score, and recall performance. To mitigate the risks of overfitting and underfitting, the paper will employ cross-validation, a widely recognized technique in data analysis [22]. This approach will provide valuable insights into which model outperforms others, guiding the identification of the most effective solution for CCF detection.

Therefore, the objectives of this study are to conduct a comprehensive evaluation of credit card fraud detection performance using dataset-1 and dataset-2 and to compare various methods in order to achieve higher precision in detection. The specific objectives include:

- Evaluating credit card-based machine learning detection algorithms to identify the most effective approach for detecting fraud.
- Utilizing high-quality imbalanced data and addressing the challenge of false positives to improve the accuracy and reliability of credit card fraud detection.
- Conducting a comparative analysis of multiple machine learning algorithms to determine which model produces the most accurate results in credit card fraud detection.
- Employing cross-validation, along with techniques to address overfitting and underfitting, to guarantee the accuracy and reliability of the selected model.

Through the application of supervised machine learning algorithms, this study demonstrates the effectiveness of using a combination of feature engineering, class balancing techniques, and cross-validation to detect credit card fraud with high accuracy [1]. By analyzing and comparing the performance of various algorithms, this paper emphasizes the significance of selecting the appropriate algorithm and fine-tuning its parameters to attain optimal results. Ultimately, this research contributes to the advancement of more reliable and efficient fraud detection systems.

2. Literature Review

In response to significant financial losses caused by fraudulent transactions, the banking industry has increasingly relied on machine learning (ML) algorithms to detect credit card fraud (CCF) [31]. In recent years, numerous research studies have been conducted to evaluate the efficacy of machine learning-based methods for CCF identification [19]. The purpose of this section is to present a comparative analysis of recent publications in order to identify the limitations of existing research. To ensure the highest quality of work, an extensive literature review has been conducted across reputable platforms such as DBLP, Scimagojr, Ieeexplore, Google Scholar, and other credible sources. This review helped propose a more effective machine-learning algorithm for the problem. Different machine learning algorithms, including Random Forest, KNN, Logistic Regression Classifier, and Naïve Bayes, have been investigated to determine their effectiveness in detecting credit card fraud. These algorithms have been proven to have accuracy rates of 97.50%, 99.96%, and 99.95%, respectively, but in their study, they didn't compare other algorithms and did not show the AUC and ROC curves [2]. Another study [26] that utilized the same supervised machine learning algorithms - logistic regression, random forest, and decision trees - discovered that the random forest classifier, when combined with the boosting technique, outperformed the logistic regression and decision trees methods. It achieved an accuracy of 96% (with an area under the curve value of 98.9%). However, this study had the drawback of not being able to identify the names of fraudulent and non-fraudulent transactions [27].

Using the same dataset as [28] and [26], KNN, Naive Bayes, and Logistic Regression were analyzed in [16]. The optimal accuracy for these classifiers was found to be 97.92%, 97.69%, and 54.86%, respectively. The study found that KNN performed better than Naive Bayes and Logistic Regression. In contrast, the RF Classifier was the best-performing algorithm in the previous study. However, in this project, eight supervised machine learning algorithms were compared, and the XGBoost algorithm performed exceptionally well, achieving an accuracy of 99.96%, precision of 99%, recall of 79%, and an R1-score of 89%.

Hassan et al. [19] conducted additional research on the detection of credit card fraud using machine learning (ML) algorithms. In a separate study, five algorithms, including Random Forest, Naïve Bayes, K-Nearest Neighbor, Logistic Regression, and Multilayer Perceptron, were assessed on a European dataset for both fraudulent and genuine transactions. The findings revealed that Random Forests demonstrated the highest accuracy rate of 99.7%. However, the study had certain limitations as it did not address hyperparameters, ensemble approaches, or other machine learning algorithms. To address these limitations, [2] conducted a comparison of different machine learning (ML) techniques for credit card fraud detection using hyperparameters and ensemble approaches. The results showed that Random Forests remained the best algorithm, with an accuracy rate of 99.77%. It was followed by Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), and Logistic Regression (LR).

Artificial Neural Networks (ANN) are more frequently used in the field of credit card fraud detection. Another study conducted by Bin Sulaiman et al. [25] explored the efficacy of Artificial Neural Networks (ANN), Federated Learning, Privacy-Preserving, and Blockchain in comparison to Machine Learning. Experiments have demonstrated that the utilization of deep learning algorithms yields effective results, and hybrid approaches are preferred for their models. In a similar way, [23] evaluated the influence of feature extraction and sample selection on CCF detection using a dataset with an unbalanced class distribution. They utilized oversampling and multiple classification techniques to enhance the performance of the classifier. However, their study did not include comparisons with existing methods and did not explore the scalability and computing demands of their proposed approaches, as discussed in the paper by Bin Sulaiman et al. [25]. Despite this limitation, their proposed hybrid solution, which incorporates oversampling, undersampling, and machine learning techniques, could serve as a foundation for future research in the field of financial services fraud detection.

However, another approach, which involves a systematic analysis of distinct machine learning (ML) methods, includes a structured comparison and evaluation of the performance of various ML algorithms on a specific task or problem. A study conducted by Izotova and Valiullin, [4] a systematic analysis of various ML approaches for CCFD and concluded that random forests were the most effective algorithms compared to logistic regression and SVM. However, deep learning techniques are becoming increasingly popular due to research conducted by Zhang et al., [30] proposed a new deep learning technique for CCFD. This technique achieved a high detection rate of 96.3% and a low false positive rate of 0.2%. However, the paper lacked a comprehensive discussion of the model's architecture and training procedure, as well as the ethical implications and limitations of deep learning for CCFD. In the present day, graph convolutional networks (GCNs) are also utilizing CCFD. The paper [33] developed a model to establish the relationships between consumers, retailers, and transactions for CCFD, achieving a 98.0% accuracy and a 0.90 area under the curve (AUC). This study did not address all ethical implications and limitations associated with the use of GCNs for CCFD [34].

It can be inferred from a review of the relevant literature that each paper has its strengths and weaknesses [35]. To overcome these limitations, this research utilized various ML algorithms and conducted a thorough comparative analysis [36]. Overall, this research contributes to the existing literature by providing a comprehensive approach to CCFD, addressing the limitations of previous studies, and filling the gaps in current knowledge [37]. The proposed models have the potential to significantly improve the effectiveness of credit card fraud detection systems and ultimately safeguard individuals and financial institutions from the harmful impacts of fraudulent activities.

3. Research Methodology

For the purpose of defining an operational procedure or methodology for the identification of credit card fraud (CCF), a list of methods has been compiled. These methods have been summarized in Figure 1, as shown below:

Data Collection and Preprocessing: During the initial phase, this paper obtained two datasets for the CCFD task from the internet. One dataset consisted of approximately 284,807 data points [16], while the other dataset was composed of simulated credit card transactions generated using Sparkov [6]. It is crucial to highlight that both datasets were unbalanced, which presented a distinct challenge for the research.

Exploratory Data Analysis (EDA): This research paper utilized a comprehensive Exploratory Data Analysis (EDA) technique to delve into the data and uncover its inherent characteristics. The analysis was conducted using Python Streamlit-based EDA software. This approach facilitated the visualization and analysis of the datasets, allowing for the identification of patterns,

correlations, and associations among the different features. Consequently, this exploration has provided valuable insights for anyone interested in the data.

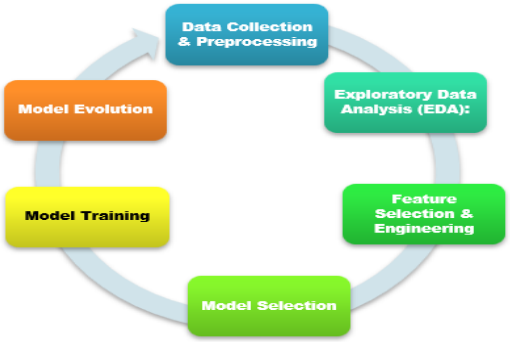


Figure 1: Cyclic Process of CCFD

Feature Selection and Engineering: Feature selection and engineering in credit card fraud detection involve the process of selecting relevant features from the available dataset and creating new features that enhance the performance of fraud detection models [13]. The goal is to identify the most informative and discriminative features that can accurately distinguish between fraudulent and non-fraudulent transactions [38]. The research paper utilized two methods for feature selection and one method for feature engineering in the field of credit card fraud detection [39].

Dimensionality Reduction: Techniques such as Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) can reduce the dimensionality of the feature space while preserving the most important information [10]. This can help reduce computational complexity and eliminate irrelevant or redundant features [40].

Cross-Validation: Proper validation techniques, such as k-fold cross-validation, can help evaluate the performance of different feature sets and engineering strategies. This ensures the selection of the most effective features for fraud detection [7].

Anomaly Indicators: The implementation of anomaly indicators is vital for effective credit card fraud detection, as they detect abnormal patterns for investigation [5]. These indicators encompass various factors, such as unusual transaction amounts, geographic anomalies, atypical transaction times, uncommon merchant categories, rapid successions of transactions, and unfamiliar account activity [30]. By integrating these indicators with advanced machine learning algorithms, the study aims to improve the precision of credit card fraud detection systems. This will allow for the detection of potentially fraudulent activities by identifying deviations from normal behavior.

Model Selection: Subsequently, this research proceeded with a meticulous selection process to identify the machine learning model that was most suitable for the available data [41]. Given that the problem involved binary classification, we thoroughly explored various well-established methods, including LR, RF, DT, SVM, KNN, and Naive Bayes, among others. This comprehensive evaluation aimed to determine the optimal model that would produce the most accurate and reliable results for the research at hand [42].

Model Training: After preprocessing the data and selecting the appropriate machine learning model, this paper proceeded to the training phase by utilizing Python packages such as sci-kit-learn [43]. The dataset was divided into training and testing sets, with 70% of the data allocated for training purposes and the remaining 30% for testing [44]. This division ensured a thorough evaluation of the model's performance, allowing for a strong assessment of its effectiveness [45].

Model Evaluation: Upon completing the model training process, this paper proceeded to evaluate its performance using eight different machine learning algorithms. Multiple metrics, including accuracy, precision, recall, and F1-score, were employed for the evaluation [20]. The findings from these assessments were meticulously recorded in the dedicated "Results and Visualization" section, offering a comprehensive overview of the model's performance and providing valuable insights for further analysis and interpretation [46].

3.1. Exploring Algorithms Utilized in Credit Card Fraud Detection

Logistic Regression: This classification method makes predictions on the likelihood of binary outcomes of the CCF Dataset, such as fraud or non-fraud, which is the target column in the dataset [9].

$$p = 1 / (1 + exp (-(b0 + b1*x1)))$$

Decision Trees: With this method, a model is constructed by iteratively subdividing the data depending on the characteristics that are the most important. This model has been utilised in the identification of fraudulent activity in this research dataset to recognise trends in the transaction data [14].

$$Entropy = -p1 * log2(p1) - p2 * log2(p2) - ... - pn * log2(pn)$$

Random Forest: The Random Forest method of ensemble learning is a way to increase accuracy and prevent overfitting by combining many decision trees into a single model. Since it can manage large datasets with many features, this model is commonly used in the detection of credit card fraud. This is one of the reasons why the authors have chosen to use it in this study [21].

$$y = mode (y1, y2, yk) \text{ where } y1, y2,..,yk \text{ are the decision tree models used in the random forest algorithm}$$

Support Vector Machines (SVM): This technique locates the hyperplane that most effectively separates the data into distinct classes. Due to its ability to manage high-dimensional data, that is why this paper utilised it in fraud detection [3].

$$y = b0 + (sum(aiyixi)) \text{ for } i \text{ in range } (1, n) \text{ [where } n \text{ is the number of independent variables, } xi \text{ is the } i\text{-th independent variable, } yi \text{ is the } i\text{-th dependent variable, } ai \text{ is the weight of the support vectors, and } b0 \text{ is the intercept]}$$

K-Nearest Neighbours (KNN): This algorithm classifies data according to its resemblance to its nearest neighbours. It is frequently used to detect credit card theft since it can identify fraudulent transactions that are like previously identified fraud cases [29].

$$d (x, y) = sqrt (sum (xi - yi) ^2) \text{ where } d (x, y) \text{ is the distance between } x \text{ and } y, \text{ and } xi \text{ and } yi \text{ are the ith features of } x \text{ and } y, \text{ respectively.}$$

Naive Bayes: Naive Bayes is a straightforward yet effective classification technique that is utilised in machine learning applications. It is predicated on Bayes' theorem and works under the assumption that the characteristics that are utilised to categorise the data are unrelated to one another. The following is an example of how the formula for the Naive Bayes algorithm might be expressed [29]:

$$P (y \mid x) = P (x \mid y) * P(y) / P(x)$$

- P (y | x) is the posterior probability of the class y given the predictor x
- P (x |y) is the likelihood probability of the predictor x given the class y
- P(y) is the prior probability of the class y
- P(x) is the probability of the predictor x

Gradient Boosting: The effective machine learning approach known as Gradient Boosting can be utilised for classification and regression work, respectively. It is a type of statistical technique known as an ensemble approach, and what it does is produce a more robust model by combining the results of several less accurate models [8]. The following is an expression that can be used to formulate the gradient-boosting algorithm.

$$F\_m(x) = F\_\{m-1\} (x) + h\_m(x)$$

- F m(x) represents the model at iteration m.
- F m-1(x) is the prior model at iteration m-1,
- while h m(x) is the new model added at iteration m to improve the accuracy of the forecast

MLP: Multilayer Perceptron (MLP) is a technique for classification and regression applications that is based on neural networks. Several layers of interconnected nodes (neurons) undertake sophisticated calculations to convert input data to output labels [20]. The MLP formula can be written as follows

$$y = f (W\_2 * f (W\_1 * x + b\_1) + b\_2)$$

x is the input vector, y is the output vector, f is the activation function applied to the output of each neuron, W_1 and W_2 are the weights of the first and second layer, respectively, b_1 and b_2 are the biases of the first and second layer, respectively.

4. Results and Discussions

4.1. Experimental Settings

In order to conduct the experimental methods, Python Jupyter Notebook and Google Colab were used, considering the computer's specifications, which include an Intel Core i3 CPU clocked at 2.30GHz and 8GB of RAM. For this research, the Scikit-Learn machine learning (ML) framework was used, while the Sparkov tool was employed for data generation purposes.

4.2. Fraud Detection Machine Learning Algorithms

This study elucidates the eight-leading machine-learning algorithms used for detecting CCF. These include Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, SVM, MLP, Naive Bayes, and KNN. Experimenting with various algorithms is crucial to identify the most optimal approach for CCF detection.

4.3. Discussion on the Datasets

Credit card fraud detection is crucial for preventing consumers from being billed for unauthorized transactions. In this research, two datasets are utilized. Dataset-1 [17] contains information on card transactions from the European banking sector dating back to September 2013. Out of a total of 284,807 transactions over two days, the data company identified 492 instances of fraudulent activity, resulting in only 0.172% of transactions being classified as fraudulent. Both datasets hold significant reliability and popularity among researchers [47]. This project conducted an extensive Exploratory Data Analysis (EDA) specifically for both datasets. The EDA process was carried out in a step-by-step manner, meticulously describing each step along with relevant illustrations and visualizations [48]. This comprehensive analysis aimed to gain deeper insights into the datasets and uncover valuable patterns, trends, and relationships within the data [49].

Dataset 1 presents a challenge due to its severe class imbalance. To address this challenge, the research implemented resampling techniques to create a balanced dataset [50]. Resampling involves either oversampling the minority class or undersampling the majority class. Oversampling techniques include randomly duplicating instances from the minority class or generating synthetic samples using methods such as SMOTE (Synthetic Minority Over-Sampling Technique) [51]. These techniques aim to mitigate the issue of class imbalance and enhance the model's ability to predict the minority class accurately. In addition, to protect the privacy of the clients, the dataset does not include the original characteristics and data background, except for the "time" and "amount" values, which have not been transformed into PCA format [52]. The numerical input variables have undergone PCA transformations, with the primary components labeled as V1, V2, and so on, up to V28. The "time" variable represents the number of seconds elapsed between the start of data collection and the following transaction, while the "amount" variable can be utilized for learning that is sensitive to cost and specific to each instance [53].

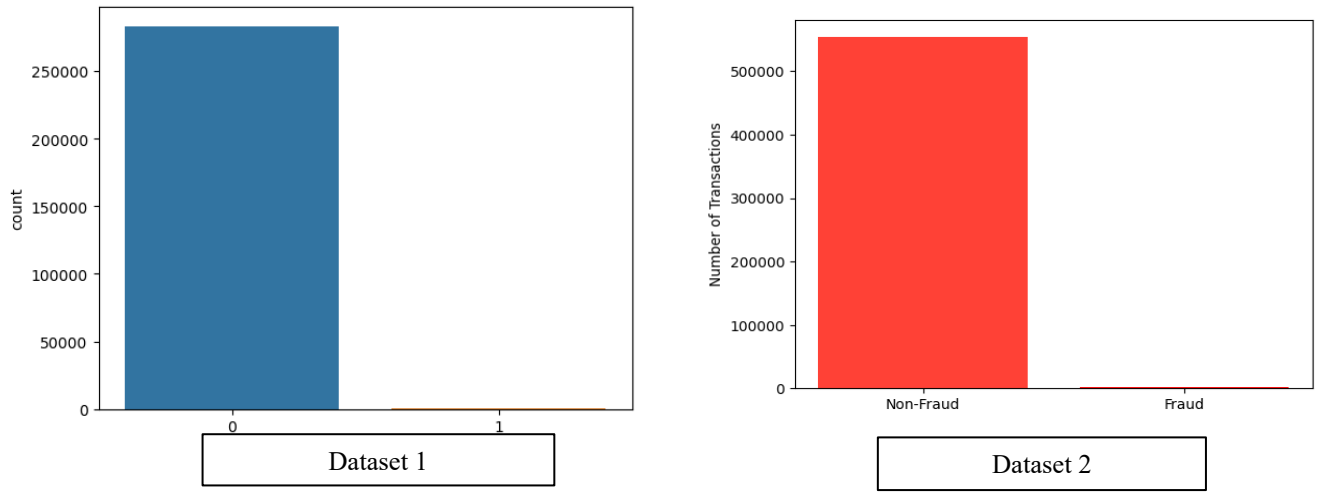
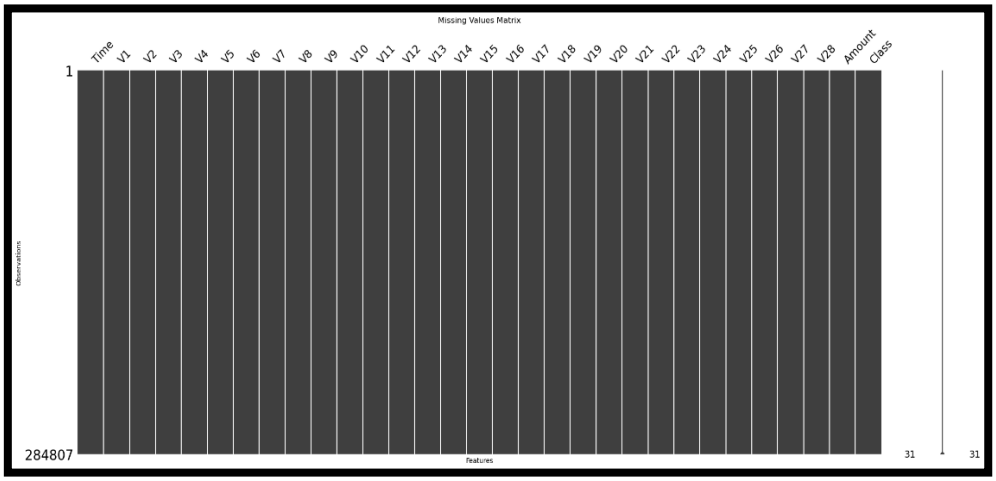


Figure 2: Fraudulent vs non-fraudulent Transaction

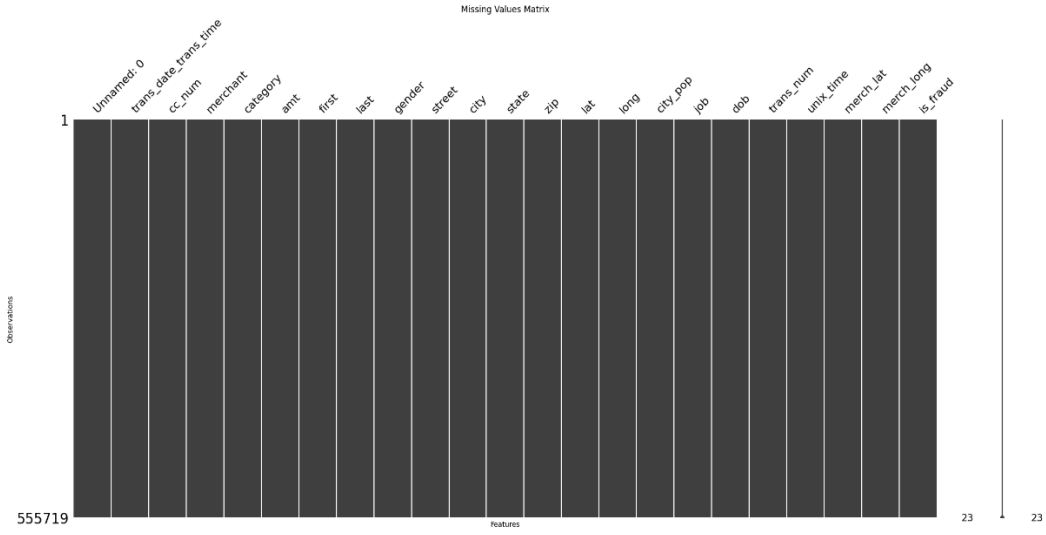
Dataset-2 [15] utilized in this study is a simulated credit card transaction dataset, encompassing both legitimate and fraudulent transactions that occurred between January 1, 2019, and December 31, 2020. This dataset involves credit cards belonging to 1000 customers who conducted transactions with a pool of 800 merchants. The dataset was generated using the Sparkov Data Generation tool developed by Brandon Harris, and it covers the specified duration. The individual files from this simulation

were merged and converted into a standardized format. In this dataset, the target column is "isfraud". Based on this target column, the paper focused on conducting evaluations, similar to dataset 2, using eight machine learning algorithms. The goal was to determine which algorithm performed the best when compared to the algorithms used in Dataset 1. In Dataset 1, the class label is assigned as 1 when a fraud has occurred and 0 when it has not. Similarly, Dataset 2 also includes a column named "isfraud," where 1 represents fraudulent transactions and 0 represents non-fraudulent transactions. This crucial information is depicted in Figure 2. Understanding these class labels is of utmost importance in developing robust machine-learning algorithms capable of accurately detecting fraudulent transactions and strengthening the security of credit card transactions.

This paper conducted a comprehensive analysis of datasets 1 and 2 using the "isnull ()" function to detect any missing values. This paper is pleased to inform you that both datasets are free of any null values, as depicted in Figure 3. To gain a better understanding of the pattern of missing values, this paper created a matrix plot using "msno. matrix () " [30] in Figure 4 visually depicts the isnull pattern in the data. In this plot, the missing values in the data frame are represented as white cells, while the darkness of the cell indicates the degree of missing values. This matrix plot provides valuable insights into the distribution and magnitude of missing values in the dataset, helping us effectively handle them in the data analysis and modeling processes.



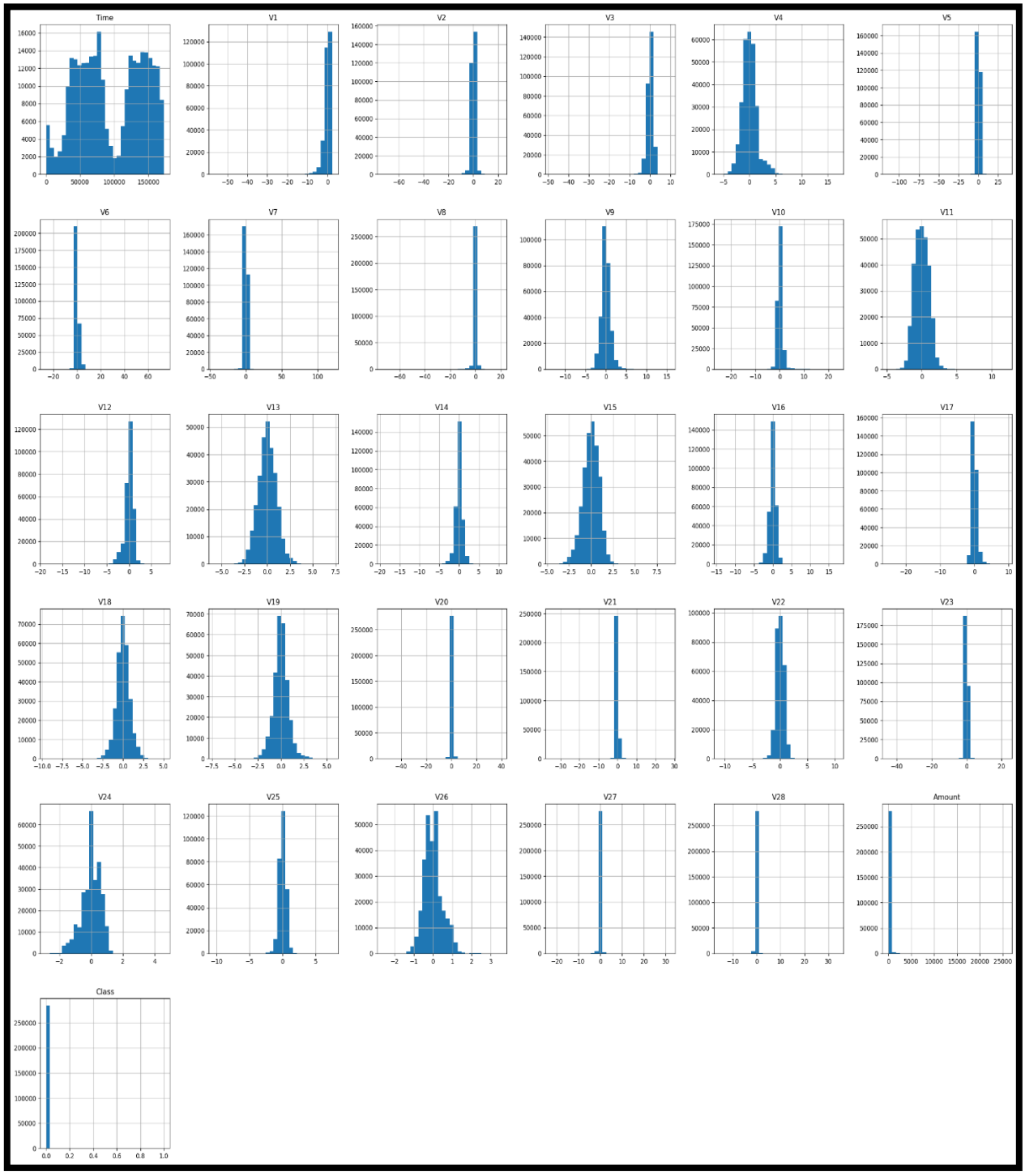
Dataset 1



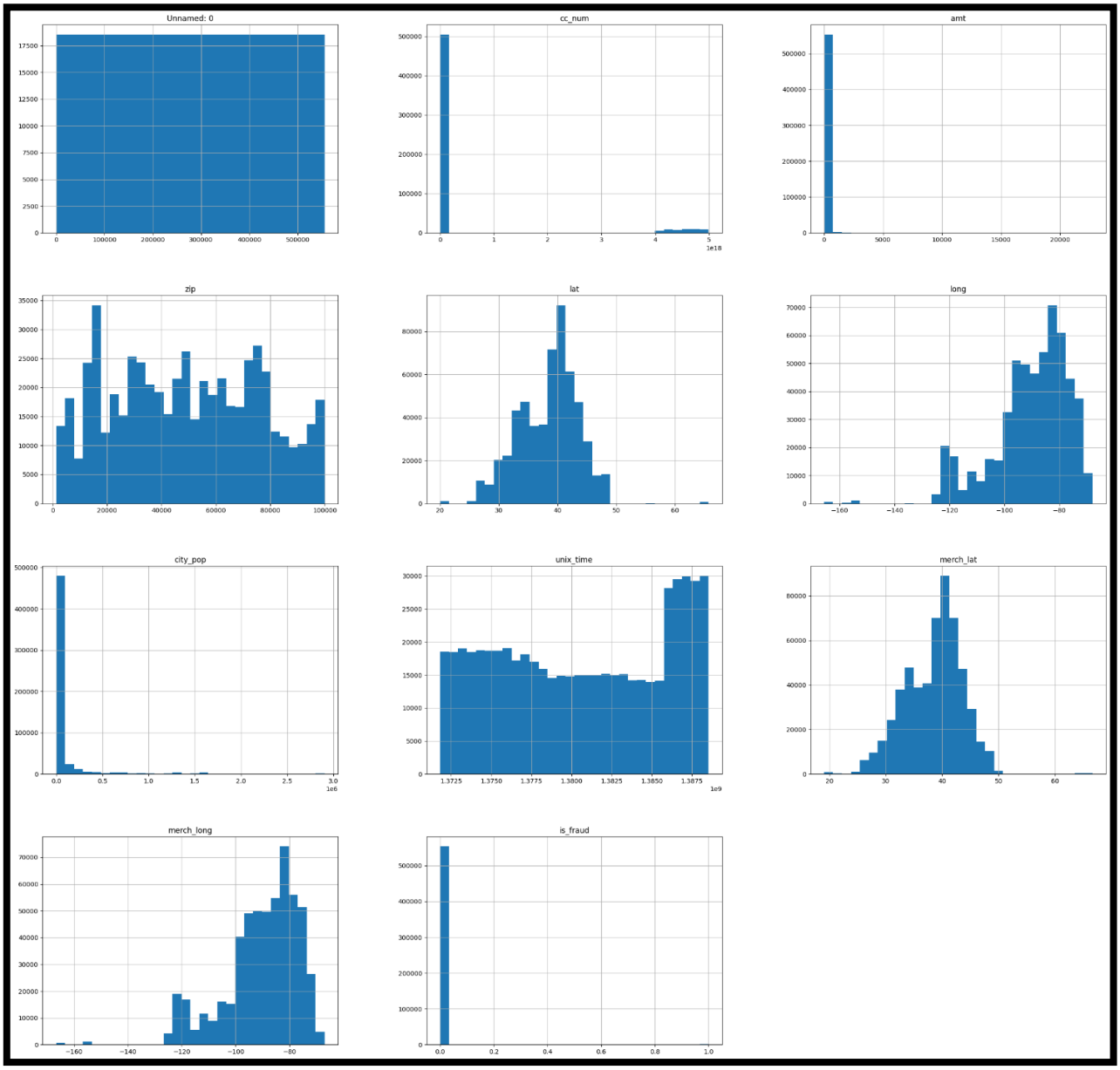
Dataset 2

Figure 3: Checking the missing values using the MSNO matrix

For a better understanding of the dataset, histograms are an effective visual tool that offers valuable insights into data distribution by displaying the frequency of data occurring within different ranges or bins [13]. In the context of credit card fraud detection, histograms play a crucial role in identifying and preventing fraudulent transactions associated with specific transaction amounts [13]. Figure 4 showcases this aspect, highlighting the significance of histograms in detecting and mitigating fraudulent activities for both datasets. Figure 5 is used to examine the occurrence of fraudulent transactions within a specific period. By generating a scatter plot that compares transaction time with transaction amount, it provides a visual representation of both legitimate and fraudulent transactions. This graphical representation facilitates intuitive analysis of the data, aiding in the identification and understanding of both valid and fraudulent transactions.



Dataset 1



Dataset 2

Figure 4: Histogram of Credit Card Fraud Detection

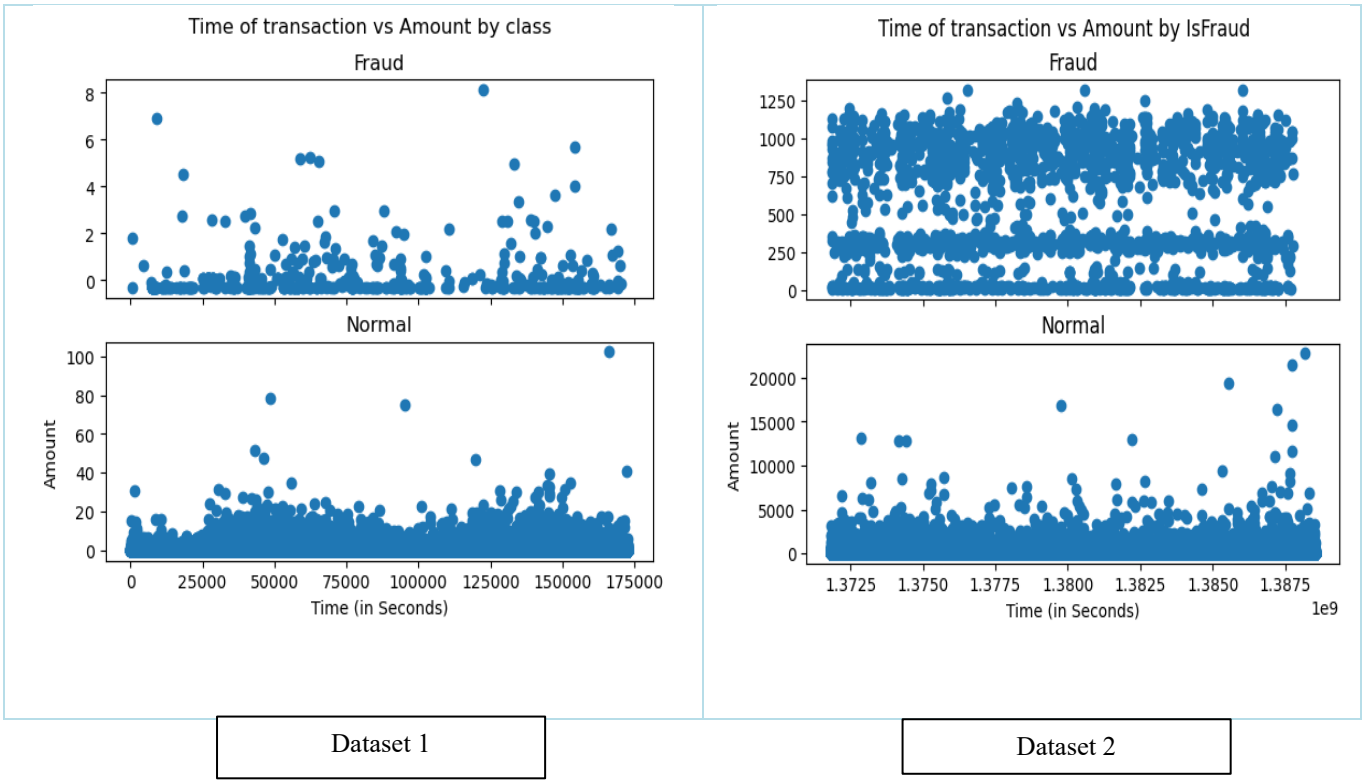


Figure 5: Scatter Plot of Transaction Time Vs. Amount

4.4. Comparison of Algorithm Accuracy for Dataset 1 and 2

In this phase, the researcher of this paper has categorized eight commonly used machine learning classification models. Although there are other models available, these eight are widely utilized. The implementation of these models utilizes algorithms from the sci-kit-learn package. The results of the applied machine learning algorithms are presented in Table 1. Furthermore, Figure 6 and Figure 7 showcase bar charts depicting various performance metrics. Notably, XGBoost exhibits the highest accuracy for Dataset 1, whereas RF achieves the highest accuracy for Dataset 2. Additionally, Figure 8 illustrates separate bar diagrams for these two algorithms.

Table 1: Accuracy, Precision, Recall, and F1-Score of different ML Algorithms for Dataset 1 and 2

Models Name	Accuracy Score		Precision Score		Recall Score		F1 Score	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2
LR	99.86 %	99.62 %	61 %	100 %	56%	100%	59%	100%
DT	99.90 %	99.81 %	72%	72.88%	81%	79.38%	76%	75.99%
RF	99.95 %	99.92 %	97%	94.40%	77%	84.22%	86%	89.02%
SVM	99.82 %	99.62 %	55%	0%	56%	0%	58%	0%
MLP	99.84 %	99.62 %	43%	0%	17%	0%	27%	0%
KNN	99.83 %	99.88 %	100%	79.73%	51%	93.44%	97%	86.04%
XGBoost	99.96 %	99.65 %	99%	98.08%	79%	7.99%	88%	14.74%
NB	99.30 %	99.62 %	15%	0%	63%	0%	24%	0%

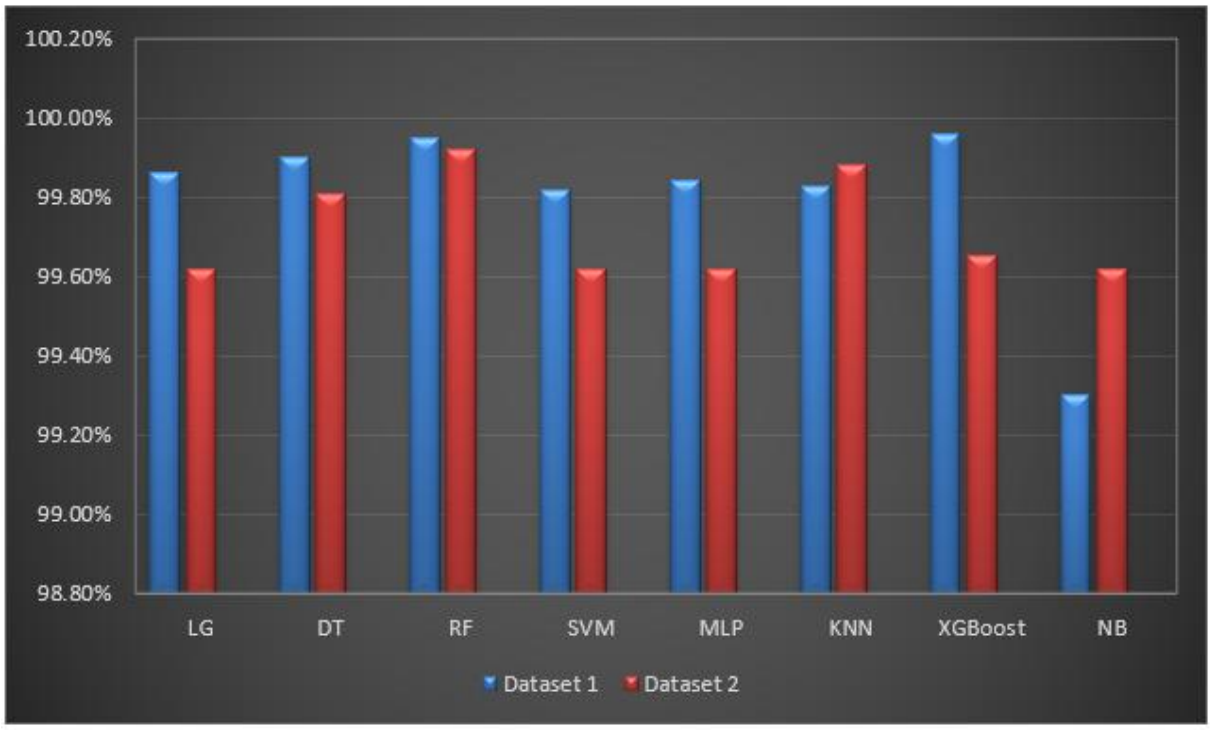


Figure 6: Accuracy Score of Different ML Algorithms using Bar Diagram

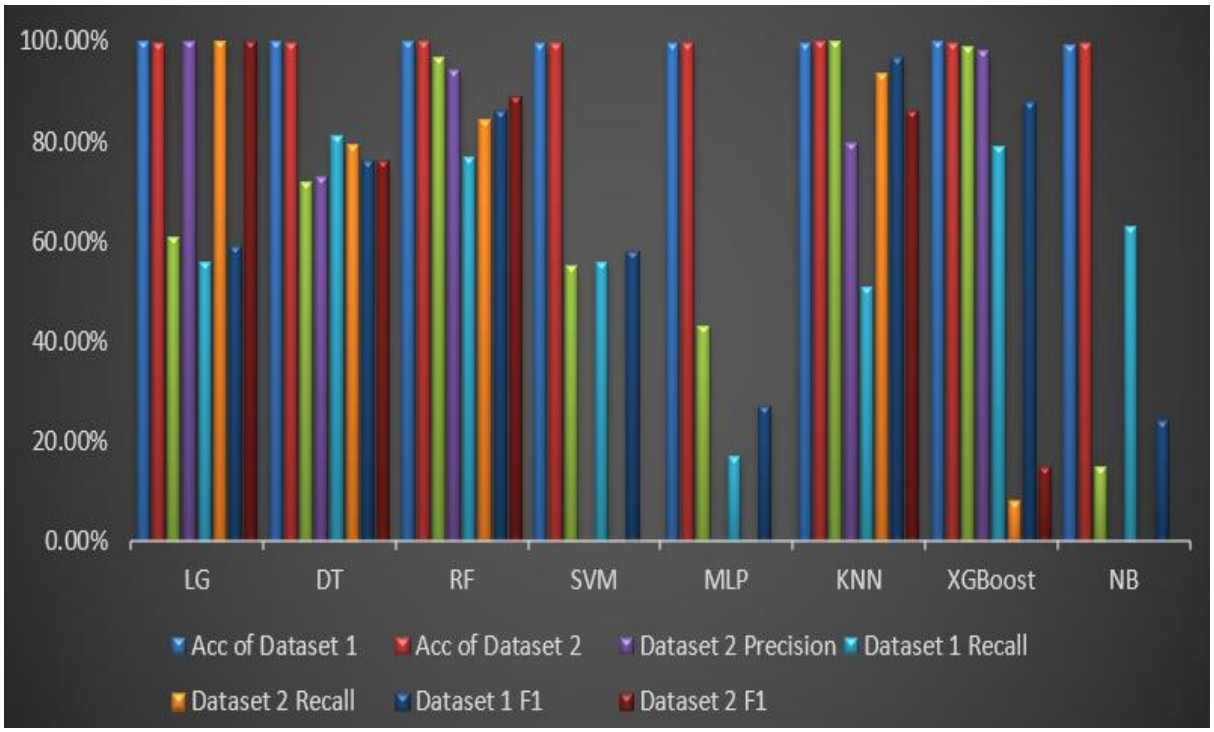


Figure 7: Accuracy, Precision, Recall, and F1-Score of Different ML Algorithms using Bar Diagram

The visual representation provided in Figure 10 showcases the scores of accuracies, precision, F1, and recall obtained from the top two machine learning algorithms, RF and XGBoost, for datasets 1 and 2.

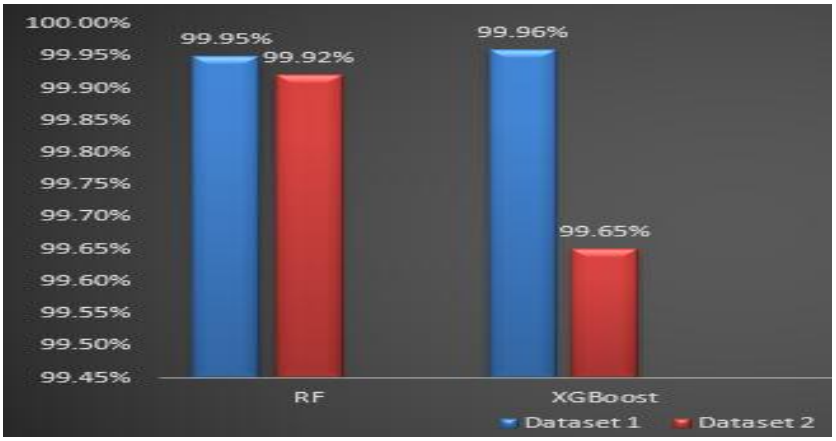


Figure 8: Comparative Analysis of Top Two Algorithm Accuracy

Notably, the XGBoost algorithm demonstrates a higher level of efficacy compared to the other algorithms in the case of dataset 1, as evidenced by this illustrative analysis. At the same time, RF represents the highest accuracy in terms of dataset 2 (Fig. 9).

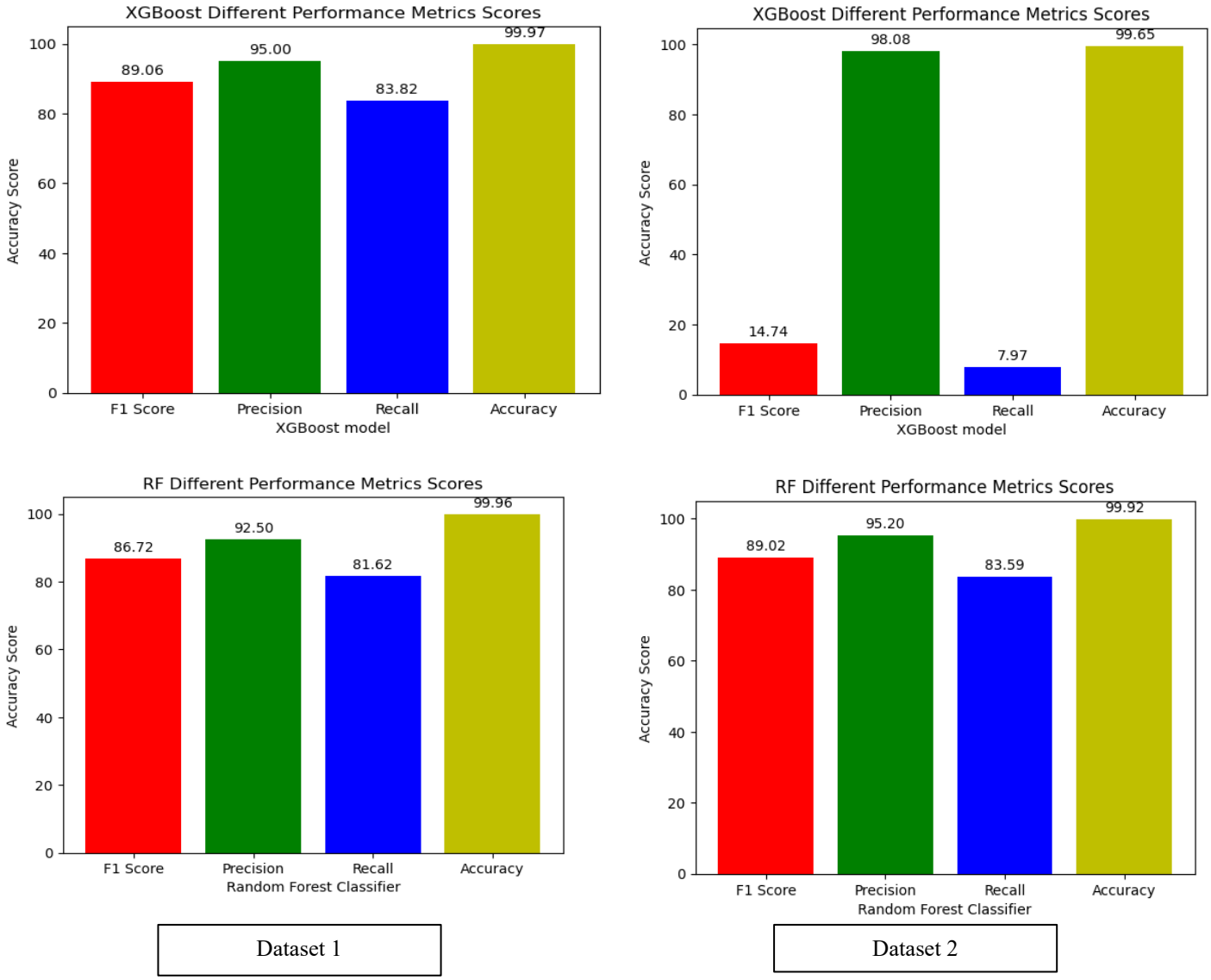


Figure 9: Comparative Analysis of Top two algorithms Accuracy, Precision, Recall and F1-Score

4.5. The Confusion Matrix and Classification Report for Credit Card Fraud Detection

Confusion metrics are used to evaluate the effectiveness of a fraud detection model [24]. These metrics include True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Precision, which measures the accurate identification of fraudulent transactions; recall, which gauges the successful identification of fraudulent transactions; and F1 Score, which is a weighted average of precision and recall, are used to assess the performance of the model in this project [18]. In Figure 10, the confusion matrices and classification reports are displayed for the RF and XGBoost algorithms, which demonstrate the highest accuracy for both datasets. The y-axis represents the actual labels, while the x-axis represents the predicted labels. The matrix compartments contain heatmaps representing the counts of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) instances for each algorithm. For RF on dataset 1, the confusion matrix values are as follows according to Figure 13: TP = 85298.00, TN = 111.00, FP = 25.00, and FN = 9.00. Similarly, for dataset 2, the values are TP = 166049.00, TN = 535.00, FP = 105.00, and FN = 27.00.

The classification report for RF on dataset 1 is as follows according to the figure 13:

- For Class 0: Precision = 1.00, Recall = 1.00, F1-score = 1.00, and Support = 85307.00.
- For Class 1: Precision = 0.93, Recall = 0.82, F1-score = 0.87, and Support = 136.00.

For RF on dataset 2, the classification report values are:

- For Class 0: Precision = 1.00, Recall = 1.00, F1-score = 1.00, and Support = 166076.00.
- For Class 1: Precision = 0.95, Recall = 0.84, F1-score = 0.89, and Support = 640.00.

The accuracy scores are the same for both datasets, but the macro averages differ. For dataset 1, the macro average values are Precision = 0.96, Recall = 0.91, F1-score = 0.93, and Support = 85443.00. For dataset 2, the macro average values are Precision = 0.98, Recall = 0.92, F1-score = 0.94, and Support = 166716.00.

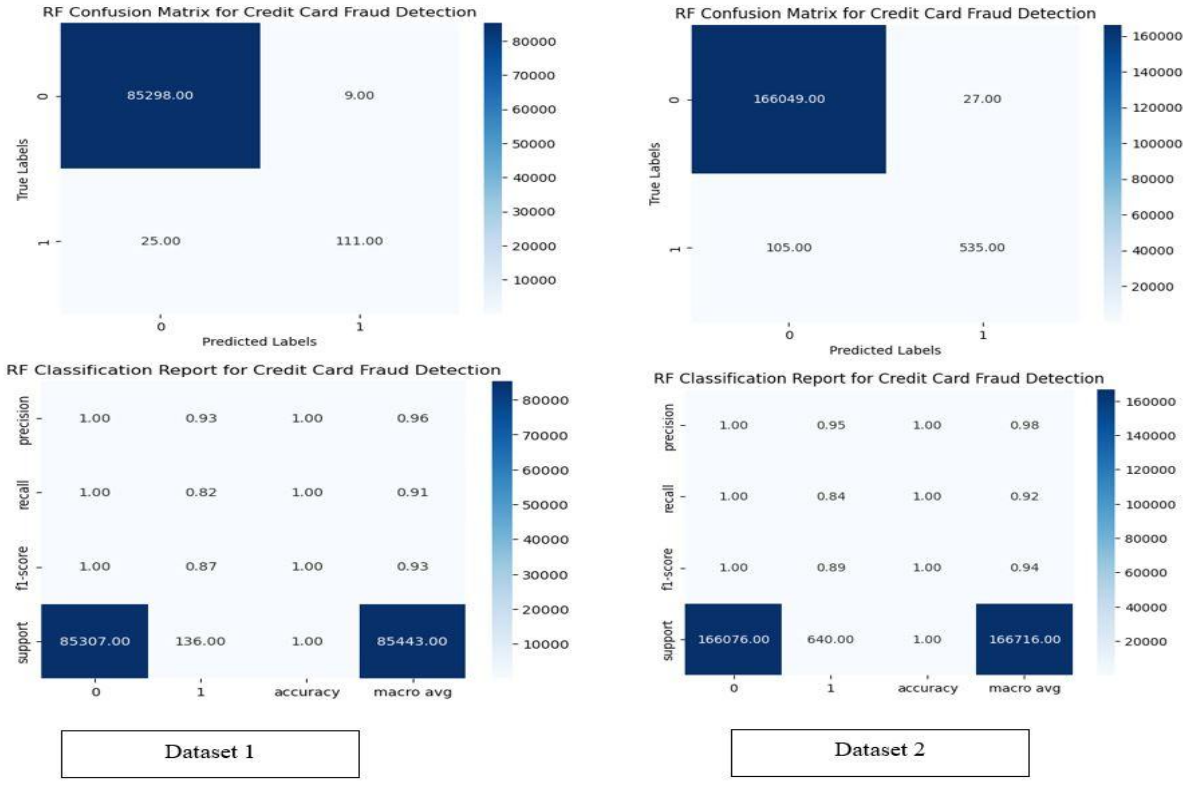


Figure 10: Illustration of Random Forest Algorithm Confusion Matrices and Classification Report for both datasets

In Figure 11, this paper presents the confusion matrix values for XGBoost applied to Dataset 1 and Dataset 2. The corresponding values for dataset 1 are as follows: True Positives (TP) = 85,000, True Negatives (TN) = 1.1, False Positives (FP) = 22, and

False Negatives (FN) = 6. On the other hand, for dataset 2, we observe the following values: TP = 170,000, TN = 51, FP = 590, and FN =1.

The classification report for XGBoost on dataset 1 is as follows:

- For Class 0: Precision = 1.00, Recall = 1.00, F1-score = 1.00, and Support = 8.5e+04.
- For Class 1: Precision = 0.95, Recall = 0.84, F1-score = 0.89, and Support = 1.4e+02.

For XGBoost on dataset 2, the classification report values are:

- For Class 0: Precision = 1.00, Recall = 1.00, F1-score = 1.00, and Support = 166076.00.
- For Class 1: Precision = 0.98, Recall = 0.08, F1-score = 0.15, and Support = 640.00.

The accuracy scores are the same for both datasets, but the macro averages differ. For dataset 1, the macro average values are Precision = 0.97, Recall = 0.92, F1-score = 0.95, and Support = 8.5e+04. For dataset 2, the macro average values are Precision = 0.99, Recall = 0.54, F1-score = 0.57, and Support = 166716.00.

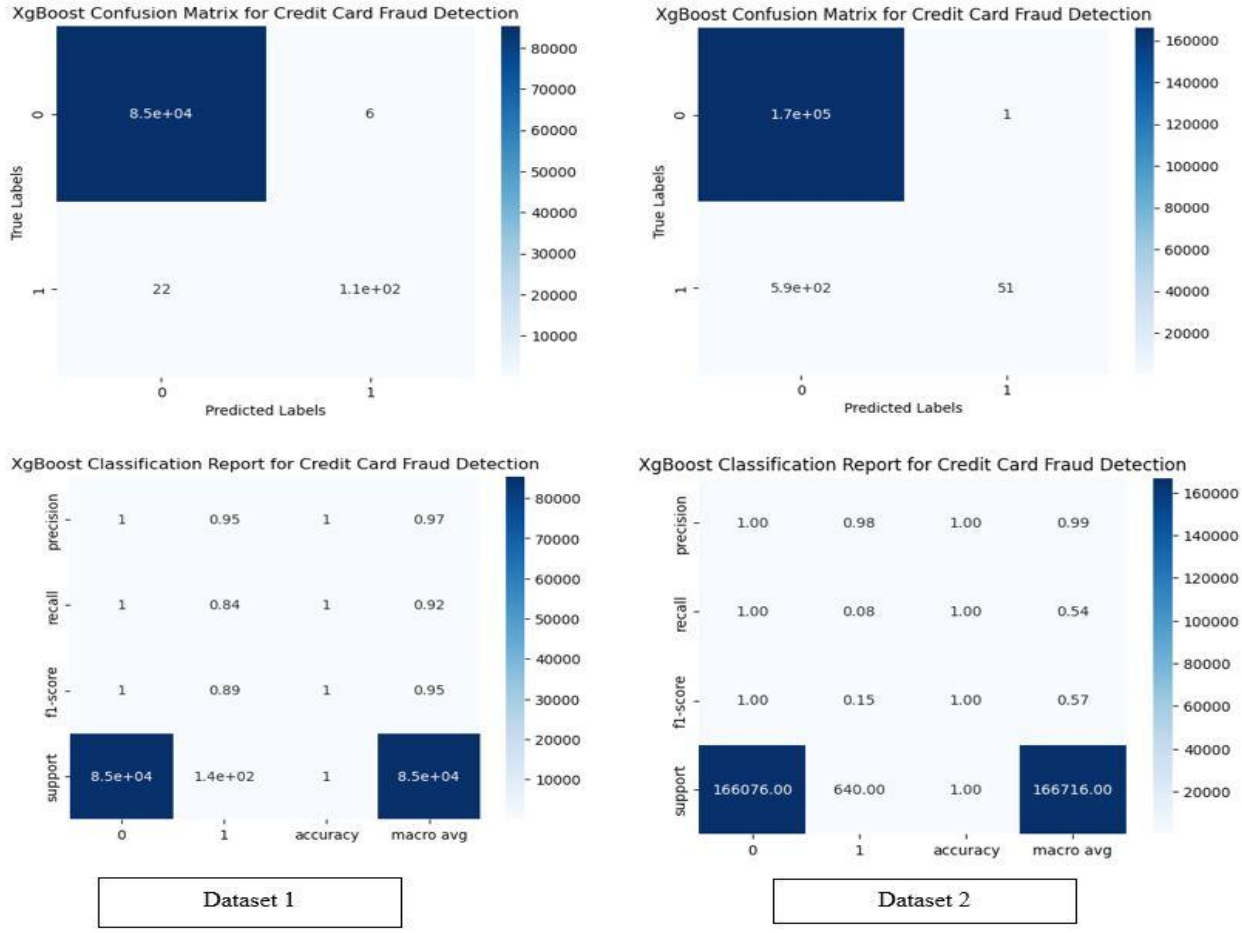


Figure 11: Illustration of XGBoost Algorithm Confusion Matrices and Classification Report for both datasets

4.6. Discussion of the results

The rationale behind employing two datasets is to assess whether our selected algorithms exhibit consistent accuracy across both datasets. Upon analyzing these two distinct datasets, it becomes evident that there are variations in the algorithm's outcomes, which were out of one goal. This research analysis observed that XGBoost performed exceptionally well on dataset-1, achieving the highest accuracy among all the tested algorithms. However, when this paper introduced the simulated dataset 2, the random forest algorithm emerged as the top performer in terms of accuracy, followed by KNN and decision tree. This implies that the attributes of dataset 2, which are hypothetical and simulated, could potentially provide an advantage to the

random forest algorithm, leading to its higher accuracy. It is worth noting that while the random forest algorithm ranked second for Dataset 1 with an accuracy of 99.95%, it achieved a slightly lower accuracy of 99.92% in Dataset 2, as shown in Figure 12. Despite this slight difference, both Random Forest and XGBoost demonstrate remarkable accuracy in credit card fraud detection. Therefore, considering the findings of previous researchers and the comparable performance of random forest in both datasets, this paper suggests that XGBoost and random forest are suitable choices for Credit Card Fraud Detection (CCFD).

However, it is important to note that when considering overall performance metrics such as precision, recall, F1 score, and accuracy, both datasets exhibit higher values compared to the rankings of individual algorithms. This suggests that regardless of the specific algorithm's ranking, the overall performance of the models is superior in both datasets.

5. Evaluating Results of Undersampling and Oversampling Techniques

In the domain of class imbalance, addressing skewed class distributions in datasets often requires the use of under-sampling and oversampling techniques [22]. To enhance our understanding of dataset-1, this paper also utilized Random undersampling and SMOTE (Synthetic Minority Over-sampling Technique) oversampling methods in conjunction with four prominent algorithms. The objective was to determine which algorithm performed the best in addressing the issue of class imbalance and improving the overall analysis of Dataset 1. By comparing the results of these techniques, valuable conclusions could be drawn regarding the most effective approach for handling class imbalance in the dataset.

Undersampling involves reducing the number of samples from the majority class, while oversampling entails increasing the number of samples from the minority class [19]. For the analysis of credit card fraud detection using different ML, the author has employed four algorithms - LR, DT, XGBoost, and RF - utilizing both under-sampling and oversampling techniques. Undersampling is implemented by selectively reducing the number of samples in the majority class, which results in a more balanced distribution among classes. By constraining the samples in the majority class, the classifier is compelled to focus more on the minority class, which could potentially lead to improved performance.

In Table 2 below, it is evident that after applying under-sampling techniques, the Random Forest Classifier exhibits higher accuracy with an overall performance of 93.68%. On the other hand, Logistic Regression (LR) and Decision Tree (DT) demonstrate a similar accuracy of 91.57% after under-sampling. Based on these results, it can be concluded that the Random Forest (RF) model is the optimal choice for this dataset after applying it to undersampling.

5.1. Exploring Under-Sampling Techniques and Impact on Scoring

In this project, the author implemented four machine learning (ML) algorithms and compared their performance to determine whether Random Forest (RF) or XGBoost is more effective in the case of under-sampling. After analyzing the datasets, it was determined that RF performed better in both datasets. It achieved accuracy rates of 93.68% for dataset 1 and 94.40% for dataset 2. Based on these results, this project recommends RF as the superior model for Credit Card Fraud Detection (CCFD) when under-sampling techniques are employed. These techniques are further illustrated in Table 2 and Figure 12, under Sampling Using Bar Diagram for dataset 1 and 2.

Table 2: Under Sampling and its Score for dataset-1 and 2

Models Name	Accuracy Score		Precision Score		Recall Score		F1 Score	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2
LR after Undersampling	91.57%	52.68%	96.73%	0%	87.25%	0%	91.75%	0%
DT after Undersampling	91.57%	92.31%	93%	87.12%	91.17%	98.27%	92.07%	92.36%
RF after Undersampling	93.68%	94.40%	97.87%	89.60%	90.19%	99.75%	93.87%	94.40%
XGBoost after Undersampling	92.63%	89.51%	94%	83.05	92.16%	97.78	93.07%	89.82%

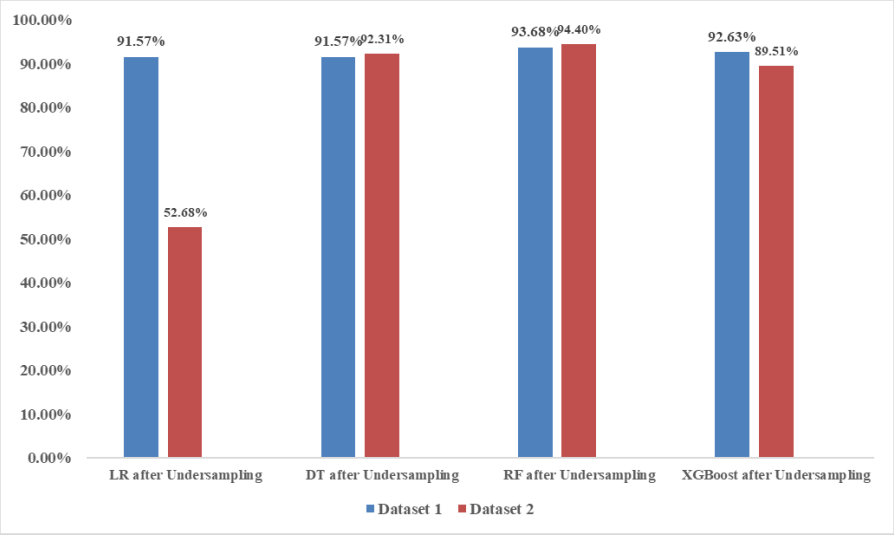


Figure 12: Under Sampling Using Bar Diagram for dataset 1 and 2

5.2. Exploring Over Sampling Techniques and their Impact on Scoring for dataset-1

The comparison between Dataset 1 and Dataset 2 using four different algorithms: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and XGBoost is presented in Table 3 and Figure 13. Oversampling Using Bar Diagram for dataset 1 and 2 in this project. By applying the oversampling technique, it becomes evident that Random Forest outperforms the other algorithms. The results indicate that Random Forest achieves an accuracy rate of 99.99% for Dataset 1 and 93.59% for Dataset 2. Based on these findings, this study recommends Random Forest as the optimal model for credit card fraud detection when using the oversampling technique.

Table 3: Oversampling Using Smooth and its Score for dataset 1 and 2

Models Name	Accuracy Score		Precision Score		Recall Score		F1 Score	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2	Dataset 1	Dataset 2
LR after Over Sampling	97.15%	52.68%	98.09%	0%	96.20%	0%	97.14%	0%
DT, after Over Sampling	99.86%	92.66 %	99.81%	88.37 %	99.91%	97.29 %	99.86%	92.61 %
RF after Over Sampling	99.99%	93.59%	99.98%	88.74%	100%	99.01%	99.99%	93.59 %
XGBoost, after Over Sampling	93.16%	88.11%	94.95%	81.67%	92.16%	96.55%	93.53%	88.49 %

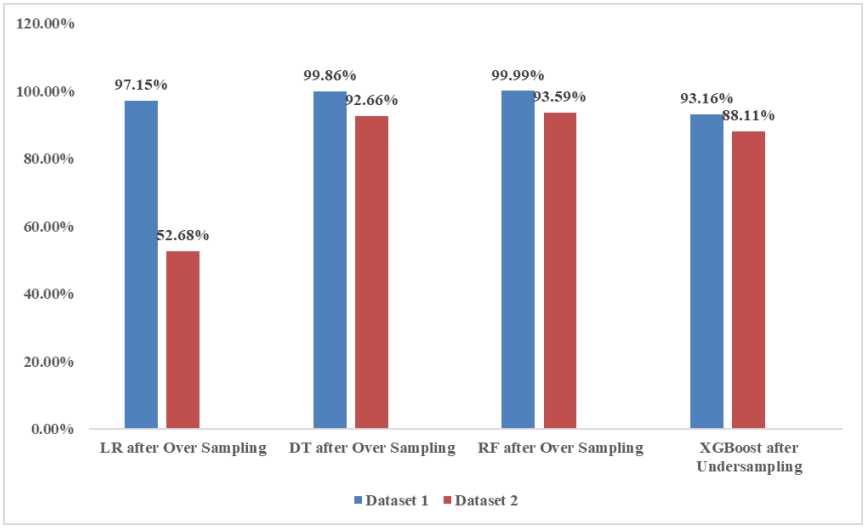


Figure 13: XG Boost after Oversampling Using Bar Diagram for dataset 1 and 2

5.3. Exploring Cross-Validation as a Technique to Mitigate Underfitting and Overfitting

Credit card fraud detection using machine learning algorithms can be affected by underfitting or overfitting, depending on how they are implemented. Underfitting occurs when the model fails to capture the underlying data patterns, while overfitting occurs when the model captures noise in addition to the patterns. Underfitting is more commonly observed.

Table 4: Cross Validation of Datasets 1 and 2

Models Name	Accuracy Score	
	Dataset 1	Dataset 2
Logistic Regression	99.91%	50.0%
Decision Tree Classifier	99.88%	44.94 %
Random Forest	99.95%	45.85%
XGBoost	99.96%	43.80%

To address these issues, this paper employed strategies such as cross-validation and regularization in datasets 1 and 2, as shown in Table 4 and Figure 14. The accuracy results, along with an illustrative table and figure, are presented above.

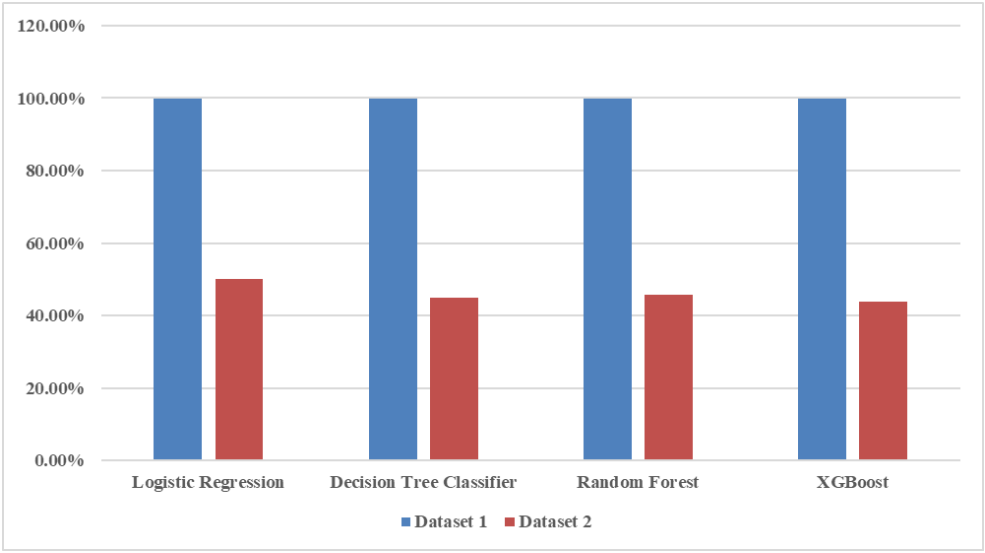


Figure 14: Comparison Accuracy, Precision, and Recall by LR, DT, and RF using Cross-Validation Methods for dataset 1 and 2

5.4. Why dataset-2 accuracy is not good as dataset-1?

In this research paper, the author explores various factors contributing to the lower accuracy of dataset-2 in comparison to dataset-1. Firstly, dataset-2 is constructed entirely from hypothetical data derived from dataset-1, rendering it less authentic, as dataset-1 primarily consists of imaginary data. Additionally, dataset-2 contains a smaller volume of data when compared to the larger dataset-1, which encompasses a more extensive dataset. Lastly, the application of distinct data preprocessing techniques in dataset 2 led to the removal of a substantial amount of data, further differentiating it from the well-preprocessed dataset 1. Despite these challenges, it is noteworthy that the random forest (RF) algorithm consistently outperformed other algorithms in both cases. In future research projects, the author intends to explore and implement updated methods to enhance the accuracy of dataset 2.

6. Conclusion and Future Work

In summary, this research paper explores the practical applications of several machine learning algorithms, such as Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, Support Vector Machines (SVM), Multilayer Perceptron, Naive Bayes, and K-Nearest Neighbors (KNN). Through a systematic comparison of their accuracies, it was discovered that XGBoost and Random Forest exhibited outstanding performance. Moreover, the study extensively evaluated the precision, recall, F1-score, and accuracy of all eight algorithms, facilitating a comprehensive assessment of their performance. The

primary objective of this research was to thoroughly analyze various machine learning algorithms and determine the best performer on the given dataset. The detailed explanations presented above successfully covered the two specific algorithms that performed exceptionally well, effectively achieving the main objective of this paper. The use of machine learning to detect credit card fraud is an ongoing research topic. Future research will focus on investigating new features or incorporating more advanced machine learning techniques to detect additional patterns or characteristics of credit card fraud using unsupervised learning. Deep learning, artificial neural networks, convolutional neural networks, ensemble methods, and anomaly detection algorithms may offer improved detection precision and performance. Most credit card fraud detection systems currently process transactions in batches at predetermined intervals. However, in the future, this paper will assess the effectiveness of XGBoost and Random Forest models using an unsupervised machine learning approach. By doing so, this research aims to produce more precise results and increase confidence in the practicality of these algorithms. The main challenges with credit card fraud detection include the absence of real-time datasets, imbalanced datasets, constantly evolving fraud patterns, concerns regarding data privacy and security, false positives and negatives, and the interpretability of advanced models.

Acknowledgment: I extend my profound gratitude to Dr. Rejwan Bin Sulaiman, my research project supervisor, and the resources provided by Northumbria University, which played a pivotal role in the successful completion of this research endeavor. I would also like to express sincere appreciation to my wife and my four-year-old daughter for their unwavering support during the challenging phases of this research. The dedication to this project is deeply rooted in the memory of my younger sister, who tragically succumbed to lung cancer on December 1, 2023.

Data Availability Statement: The datasets produced and scrutinized during this study can be accessed at https://www.kaggle.com/. Embracing principles of transparency and open science, this project invites fellow researchers to utilize, reproduce, and extend the project's data for further exploration. For inquiries regarding access to the data, please contact Sonjoy Ranjon Das via email at sonjoyict@gmail.com.

Funding Statement: This research was self-funded, as the author independently pursued this project out of personal interest during the research project proposal course at the university.

Conflicts of Interest Statement: The authors declare that no conflicts of interest exist concerning the publication of this paper. This project affirms that the research was conducted with integrity, free from any external influence that might compromise the objectivity of the study or the reporting of its results. Additionally, all references from which the data have been collected are duly provided in this project.

Ethics and Consent Statement: This study adhered to the ethical standards and guidelines set forth by Northumbria University, London. All research protocols involving human subjects were reviewed and approved by the ethics committee of the university and supervisor. Informed consent was obtained from all participants involved in the study, and they were assured of the confidentiality and anonymity of their information. The ethical conduct of this research aligns with the principles outlined in the research involving human subjects.

References

1. A. Abd El-Naby, E. E.-D. Hemdan, and A. El-Sayed, “An efficient fraud detection framework with credit card imbalanced data in financial services,” Multimed. Tools Appl., vol. 82, no. 3, pp. 4139–4160, 2023.
2. A. Bhanusri et al., “Credit card fraud detection using Machine learning algorithms,” In Quest Journals Journal of Research in Humanities and Social Science, vol. 8, no. 2, pp. 631-641, 2020.
3. 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,” J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 1, pp. 145–174, 2023.
4. A. Izotova and A. Valiullin, “Comparison of Poisson process and machine learning algorithms approach for credit card fraud detection,” Procedia Comput. Sci., vol. 186, pp. 721–726, 2021.
5. A. Rb and S. K. Kr, “Credit card fraud detection using artificial neural network,” Global Transitions Proceedings, vol. 2, no. 1, pp. 35–41, 2021.
6. A. Srivastava, A. Kundu, S. Sural, and A. K. Majumdar, “Credit card fraud detection using hidden Markov model,” IEEE Trans. Dependable Secure Comput., vol. 5, no. 1, pp. 37–48, 2008.
7. A. Z. Mustaqim, S. Adi, Y. Pristyanto, and Y. Astuti, “The effect of recursive feature elimination with cross-validation (RFECV) feature selection algorithm toward classifier performance on credit card fraud detection,” in 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), Yogyakarta, Indonesia, 2021.

8. B. Bayram, B. Koroglu, and M. Gonen, “Improving fraud detection and concept drift adaptation in credit card transactions using incremental gradient boosting trees,” in 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 2020.
9. G. Sandhya, M. Abishek, S. Gunal Kumar, and R. S. Jisenthira Kumar, “Credit card fraud detection using machine learning algorithms,” in ICT Systems and Sustainability, Singapore: Springer Nature Singapore, pp. 313–320, 2023.
10. H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, “Credit card fraud detection based on machine and deep learning,” in 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2020.
11. H. Paruchuri, “Credit card fraud detection using machine learning: A systematic literature review,” ABC J. Adv. Res., vol. 6, no. 2, pp. 113–120, 2017.
12. J. Femila Roseline, G. Naidu, V. Samuthira Pandi, S. Alamelu alias Rajasree, and D. N. Mageswari, “Autonomous credit card fraud detection using machine learning approach☆,” Comput. Electr. Eng., vol. 102, no. 108132, p. 108132, 2022.
13. J. K. Afriyie et al., “A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions,” Decision Analytics Journal, vol. 6, no. 100163, p. 100163, 2023.
14. J. R. Gaikwad, A. B. Deshmane, H. V. Somavanshi, S. V. Patil, and R. A. Badgujar, “Credit Card Fraud Detection using Decision Tree Induction Algorithm,” International Journal of Innovative Technology and Exploring Engineering, vol.4, no. 6, pp. 92-95, 2014.
15. K. Chaudhary, J. Yadav, and B. Mallick, “A review of Fraud Detection Techniques: Credit Card,” In International Journal of Computer Applications, vol. 45, no. 1, pp. 485–496, 2012.
16. K. Shenoy, “Credit card transactions fraud detection dataset.” 05-Aug-2020, https://www.kaggle.com/datasets/kartik2112/fraud-detection. [Accessed: 04-Mar-2023].
17. Kaggle, “Machine Learning Group-ULB-Credit Card Fraud Detection.” 23-Mar-2018. https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. [Accessed: 04-Mar-2023].
18. M. A. D. Pawar, P. P. N. Kalavadekar, and M. S. N. Tambe, “A survey on outlier detection techniques for credit card fraud detection,” IOSR J. Comput. Eng., vol. 16, no. 2, pp. 44–48, 2014.
19. M. H. Ahmed, “A review: Credit card fraud detection in banks using machine learning algorithms,” ScienceOpen Preprints, 2023, Press.
20. M. K. Mishra and R. Dash, “A comparative study of Chebyshev functional link artificial neural network, multilayer perceptron and decision tree for credit card fraud detection,” in 2014 International Conference on Information Technology, Las Vegas, NV, USA, 2014.
21. M. R. Dileep, A. V. Navaneeth, and M. Abhishek, “A novel approach for credit card fraud detection using decision tree and random forest algorithms,” in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021.
22. P. Gupta, A. Varshney, M. R. Khan, R. Ahmed, M. Shuaib, and S. Alam, “Unbalanced credit card fraud detection data: A machine learning-oriented comparative study of balancing techniques,” Procedia Comput. Sci., vol. 218, pp. 2575–2584, 2023.
23. P. M. Dechow, L. A. Myers, and C. Shakespeare, “Fair value accounting and gains from asset securitizations: A convenient earnings management tool with compensation side-benefits,” J. Account. Econ., vol. 49, no. 1–2, pp. 2–25, 2010.
24. P. Sharma, S. Banerjee, D. Tiwari, and J. C. Patni, “Machine learning model for credit card fraud detection- A comparative analysis,” Int. Arab J. Inf. Technol., vol.18, no.6, pp. 789-796, 2021.
25. R. Bin Sulaiman, V. Schetinin, and P. Sant, “Review of machine learning approach on credit card fraud detection,” Hum-Cent Intell Syst, vol. 2, no. 1–2, pp. 55–68, 2022.
26. S. Dhankhad, E. Mohammed, and B. Far, “Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study,” in 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 2018.
27. S. Han, “Credit Card Fraud Detection with Machine Learning”, 2020, Ulb.ac.be. [Online]. Available: http://mlg.ulb.ac.be. [Accessed: 04-Mar-2023].
28. S. Khatri, A. Arora, and A. P. Agrawal, “Supervised machine learning algorithms for credit card fraud detection: A comparison,” in 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2020, Noida, India.
29. S. Kiran, J. Guru, R. Kumar, N. Kumar, D. Katariya, and M. Sharma, “Credit card fraud detection using Naïve Bayes model based and KNN classifier,” International Journal of Advance Research Ideas and Innovations in Technology, vol. 4, no. 3, pp. 44–47, 2018.
30. S. Zhang, H. Tong, J. Xu, and R. Maciejewski, “Graph convolutional networks: a comprehensive review,” Comput. Soc. Netw., vol. 6, no. 1, pp. 1-23, 2019.
31. V. N. Dornadula and S. Geetha, “Credit card fraud detection using machine learning algorithms,” Procedia Comput. Sci., vol. 165, pp. 631–641, 2019.

32. Y. Jain, N. Tiwari, S. Dubey, and S. Jain, “A comparative analysis of various credit card fraud detection techniques,” International Journal of Recent Technology and Engineering, vol. 7, no. 5S2, pp. 402–407, 2019.
33. Z. Zhang and S. Huang, “Credit card fraud detection via deep learning method using data balance tools,” in 2020 International Conference on Computer Science and Management Technology (ICCSMT), Shanghai, China, 2020.
34. A. Charmchian Langroudi, M. Charmchian Langroudi, F. Arasli, and I. Rahman, “Challenges and strategies for knowledge transfer in multinational corporations: The case of hotel ‘Maria the great,’” Journal of Hospitality & Tourism Cases, 2024, Press.
35. A. Kumar, “A study of significant characteristics of e-payment regime in India. MERC Global’s,” International Journal of Management, vol. 7, no. S1, pp. 168–174, 2019.
36. A. Kumar, “Demographic & geographic inclination towards store design: A study of shopping mall customers in Maharashtra-State, India,” India. Asian Journal of Management, vol. 2, no. 3, pp. 87–93, 2011.
37. A. Kumar, “Store image - A critical success factor in dynamic retailing environment: An investigation,” Asian Journal of Management, vol. 3, no. 1, pp. 1–5, 2012.
38. B. Priyanka, Y. Rao, B. Bhavyasree, and B. Kavyasree, “Analysis Role of ML and Big Data Play in Driving Digital Marketing’s Paradigm Shift,” Journal of Survey in Fisheries Sciences, vol. 10, no. 3S, pp. 996–1006, 2023.
39. B. Rashi, Y. S. Kumar Biswal, N. Rao, and D. Ramchandra, “An AI-Based Customer Relationship Management Framework for Business Applications,” Intelligent Systems and Applications In Engineering, vol. 12, pp. 686–695, 2024.
40. D. D. Kumar Sharma Kuldeep, “Perception Based Comparative Analysis of Online Learning and Traditional Classroom-Based Education Experiences in Mumbai,” Research Journey, Issue, vol. 330, no. 2, pp. 79–86, 2023.
41. E. Groenewald and O. K. Kilag, “E-commerce Inventory Auditing: Best Practices, Challenges, and the Role of Technology,” International Multidisciplinary Journal of Research for Innovation, Sustainability, and Excellence (RISE), vol. 1, no. 2, pp. 36–42, 2024.
42. K. M. Nayak and K. Sharma, “Measuring Innovative Banking User’s Satisfaction Scale,” Test Engineering and Management Journal, vol. 81(1), pp. 4466–4477, 2019.
43. K. Sharma and P. Sarkar, “A Study on the Impact of Environmental Awareness on the Economic and Socio-Cultural Dimensions of Sustainable Tourism,” International Journal of Multidisciplinary Research & Reviews, vol. 3, no. 1, pp. 84–92, 2024.
44. K. Sharma and S. Poddar, “An Empirical Study on Service Quality at Mumbai Metro-One Corridor,” Journal of Management Research and Analysis, vol. 5, no. 3, pp. 237–241, 2018.
45. K. Vora, “Factors Influencing Participation of Female Students in Higher Education w.r.t Commerce Colleges in Mumbai,” International Journal of Advance and Innovative Research, vol. 5, no. 3, pp. 127–130, 2018.
46. K. Vora, Sharma Kuldeep, and P. Kakkad, “Factors Responsible for Poor Attendance of Students in Higher Education with respect to Undergraduate - Commerce Colleges in Mumbai. BVIMSR’s,” Journal of Management Research, vol. 12, no. 1, pp. 1–9, 2020.
47. M. Farheen, “A Study on Customer Satisfaction towards traditional Taxis in South Mumbai,” Electronic International Interdisciplinary Research Journal, vol. 12, no. 1, pp. 15–28, 2023.
48. N. Bhat, M. Raparthi, and E. S. Groenewald, “Augmented Reality and Deep Learning Integration for Enhanced Design and Maintenance in Mechanical Engineering,” Power System Technology, vol. 47, no. 3, pp. 98–115, 2023.
49. P. Kakkad, K. Sharma, and A. Bhamare, “An Empirical Study on Employer Branding To Attract And Retain Future Talents,” Turkish Online Journal of Qualitative Inquiry, vol. 12, no. 6, pp.7615, 2021.
50. R. K. Gupta, “Adoption of mobile wallet services: an empirical analysis,” Int. J. Intellect. Prop. Manag., vol. 12, no. 3, p. 341, 2022.
51. S. Bhakuni, “Application of artificial intelligence on human resource management in information technology industry in India,” The Scientific Temper, vol. 14, no. 4, pp. 1232–1243, 2023.
52. T. N. Srinivasarao and N. G. Reddy, “Small and Medium Sized Enterprises Key Performance Indicators,” IOSR Journal of Economics and Finance, vol. 11, no. 4, pp. 1–06, 2020.
53. Y. Priyanka, B. Rao, B. Likhitha, and T. Malavika, “Leadership Transition In Different Eras Of Marketing From 1950 Onwards,” Korea Review Of International Studies, vol. 16, no. 13, pp. 126–135, 2023.