THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 3 (2024)
https://doi.org/ 10.62019/abbdm.v4i3.207

Identifying Credit Card Fraud with Machine Learning: Evaluation of Algorithms and Oversampling Techniques

Beenish Ahmed*, Sarfraz Hussain, Daniyal Shakir, Najeeb ur Rehman, Ghalib Nadeem

Chronicle

Article history

Received: Aug 20, 2024
Received in the revised format: Aug 29, 2024
Accepted: Sept 1, 2024
Available online: Sept 3, 2024

Beenish Ahmed is currently affiliated with Department of Computer Science Iqra University, Karachi, Pakistan.
Email: beenish.ahmed@iqra.edu.pk

Sarfraz Hussain is currently affiliated with Department of Electronic Engineering, Indus University, Karachi, Pakistan.
Email: narfrazjogi1996@gmail.com

Daniyal Shakir is currently affiliated with S & P Global, Karachi, Pakistan.
Email: daniyalshakir@gmail.com

Najeeb ur Rehman is currently affiliated with Agha Khan University, Karachi, Pakistan.
Email: najeeb.rahman@aku.edu

Ghalib Nadeem is currently affiliated with Department of Electrical & Computer Engineering Iqra University, Karachi, Pakistan.
Email: ghalibnadeem@iqra.edu.pk

Corresponding Author*

Abstract

The physical loss of a credit card or the theft of sensitive credit card data is referred to as credit card fraud. For detection, a variety of machine learning methods can be applied. This study presents several algorithms for identifying transactions as authentic or fraudulent. The study used the credit card fraud detection dataset. The SMOTE approach was utilized for oversampling due to the extremely unbalanced nature of the dataset. Additionally, a feature selection process was carried out, and the dataset was divided into training and test sets. The experiment employed eight machine learning models, including Random Forest, AdaBoost, Support Vector Classifier, Extreme Gradient Boost, and Logistic Regression algorithms. The findings indicate that few algorithms have a high degree of accuracy when detecting credit card fraud. The Random Forest model can be applied to find further anomalies. This approach has proven to be effective in accurately detecting fraudulent activity and reducing the number of false positives and negatives. The results of the experiment also highlight the importance of feature selection in improving the performance of the models by the highest accuracy score with 96%, precision of 100, Recall of 91, and F1Scroe of 95. By using a combination of different machine learning algorithms, financial institutions can enhance their fraud detection systems and better protect their customers from potential financial losses. The findings from this study demonstrate the potential for significant advancements in the field of fraud detection using machine learning techniques.

Keywords: Fraud Detection; K-NN; Multilayer Perceptron; Naive Bayes; Random Forest.

INTRODUCTION

Worldwide, the quantity of newly conventional businesses is growing (Ghauri, P et al., 2021). All of those businesses strive to give their clients the highest caliber of service possible. Every day, businesses investigate a lot of data to thrive in that. These data are accessible in various set-ups and invented from a multitude of sources. Furthermore, some of the most significant components of the business's forthcoming are limited in this data. As a result, businesses must handle, store, and most prominently securely store this data. Many pieces of data can be misused by other businesses or, worse yet, taken if security isn't maintained. Financial information is typically stolen, which can be harmful to an individual or the entire firm. Frauds come in a variety of forms. When someone forges

a check or pays with one meaningful there is not enough money, it is recognized as check fraud (Karpoff, J. M. et al., 2021). Online sales fraud refers to the sale of phony or counterfeit goods by con artists, as well as the taking of money without providing the promised goods. There are a few others, including identity theft, credit card fraud, debt cancellation, insurance fraud, and fraud involving aid. One of the most frequent types of fraud due to the growing acceptance of cashless transactions is credit card fraud when a credit card is used for fraudulent drives without the cardholder's knowledge, it is referred to as credit card fraud. In 2016, there were €1.8 billion worth of fraudulent transactions made with credit cards that were obtained all over the world (Priscilla, C. V., et al., 2020). Even though the volume of credit card transactions has increased dramatically, effective fraud detection systems have kept or reduced the quantity of fraud. On the other hand, identity thieves are always devising novel methods to pilfer data (Mittal, S., et al., 2020). Credit card fraud comes in two flavors. One involves physically stealing the card, while the other involves taking private information from the card, like the card number, CVV code, kind of card, and other details. A fraudster can steal significant sums of money or make large purchases using credit card information before the cardholder discovers it.

Because of this, businesses distinguish between fraudulent transactions and those that are not using various machine learning techniques. This research aims to investigate several machines learning algorithms, including Random Forest (RF), Logistic Regression (LR), and Extreme Gradient Boost (XGB), and identify which algorithm is most suited for detecting credit card fraud. The remaining portion of the paper is organized as follows: after the Introduction other section presents studies that address a particular issue; the section onward provides a synopsis of the dataset utilized in the experiment; and presents the findings. At last, concludes with a discussion of closing remarks and a bibliography.

LITERATURE REVIEW

The significant losses caused by fraudulent activity inspired researchers to develop a method for identifying and stopping fraud. Numerous approaches have already been put out and examined. A quick summary of a few of them is given below. Traditional methods that have shown useful include Gradient Boosting (GB), Support Vector Machines (SVM), Decision Trees (DT), LR, and RF. Utilizing GB, LR, RD, SVM, and a few more classifiers in combination, the research (Zhang, Y., et al., 2022) attained a high recall of more than 91% on a European dataset. Only after the dataset was balanced by under sampling did the data attain high precision and recall. A comparative analysis was conducted of the models LR, DT, and RF in a publication (Ghorbani, R., et al.,2020), which similarly used a European dataset. Random forest came first with 95.5% of accuracy among the models, followed by Linear Regression with accuracy of 90% and Decision Trees of 94.3% accuracy. The findings showed that Random Forest was the most accurate among the three models. (Ghorbani, R.,2020)

k-Nearest neighbors (KNN) and outlier detection methods can also be effective in fraud discovery, according to (Wang, B., et al., 2020). They have revealed value in dropping the number of untrue alarms and raising the amount of fraud detection. The KNN algorithm in the experimentation for publication (Wang, Y., et al., 2022) similarly did well when placed to the assessment and compared with other traditional techniques, A comparison of deep learning models with different classical algorithms was made in the

paper (Alquthami, T., et al., 2022) in contrast to the publications earlier stated. Each method placed to the test produced an accuracy rate of around 80%. The authors of the research (Ata, O., et al., 2020) used European dataset and compared the following algorithms: KNN, GB, LR, NB, DT, RF, SVM, XGBoost (XGB), MLP, and stacking classifier (a mixture of many machine learning classifiers). After preprocessing of the data carefully, each algorithm attained accuracy levels above 90%. The best and effective classifier was the stacking one, with 95% of accuracy and 95% of recall value. A neural network was evaluated using the European dataset in the paper (Brunner, C., et al., 2021). Backpropagation neural networks tuned using the Whale method were used in the experiment. Two input layers, twenty hidden levels, and two output layers made up the neural network. They obtained remarkable results on 500 test samples—96.40% accuracy and 97.83% recall—thanks to the optimization approach. Neural networks were utilized by the authors of papers (Panthakkan, A., et al.,2022) to show how using ensemble approaches improves results. Three datasets were utilized in the study to compare the Auto-encoder and Restricted Boltzmann Machine techniques. The results showed that MLP algorithms can be useful for detecting credit card fraud.

Deep neural network-based fraud detection of transactions is the subject of many articles. These models work better on larger datasets, but they are computationally expensive (QUINTIN-JOHN, S. M. I. T. H., et al., 2021). As numerous articles have shown, this strategy may produce excellent results, but what if the same or even better outcomes can be obtained with fewer resources? Our main objective is to demonstrate that, with the right preprocessing, several machine learning algorithms may produce respectable results. The majority of the aforementioned papers' authors employed under sampling techniques, which served as justification for adopting an alternative strategy: oversampling techniques. The authors of this research choose to compare the applicability of LR, RF, NB, and MLP for credit card fraud detection in light of the available information. To make that happen, an experiment was carried out.

PROBLEM: CREDIT CARD FRAUD DETECTION

Context

For credit card firms to safeguard their clients and stop unlawful payments, they must be able to identify fraudulent credit card transactions. The extremely uneven nature of the data—the bulk of transactions are valid, and a very small percentage are fraudulent—makes this problem a classic machine learning challenge.

Objective

Create a machine learning model that can reliably detect fraudulent transactions while reducing the number of false positives (regular transactions mistakenly identified as fraudulent) and false negatives (fraudulent transactions missed)

Dataset Description

The dataset includes credit card transactions performed by European cardholders in September 2013. Out of the 284,807 transactions that took place over two days, 492 are fake.

Dataset Features

- Time: The number of seconds that passed between this transaction and the dataset's initial transaction.
- V1 - V28: Principal components acquired by Principal Component Analysis (PCA) to safeguard sensitive data and user identities.
- Amount: Transaction amount.
- Class: Response variable, 0 denotes a valid transaction and 1 denotes a fraudulent transaction.

Challenges

- Data Imbalance: Just 0.172% of all transactions are in the fraud class.
- Principal Component Interpretation: Direct interpretation is challenging because features V1 through V28 are the outcome of a PCA transformation.
- Performance Evaluation: The accuracy of the confusion matrix is insufficient because of the imbalance. The Area Under the Precision-Recall Curve (AUPRC) is a suggested metric for assessing model performance.

PROPOSED METHODOLOGY

Exploratory Data Analysis (EDA)

- Recognize how the variables "Time" and "Amount" are distributed.
- Examine how well the main components are correlated.
- Show how genuine and fraudulent transactions are distributed.

Preprocessing

- Deal with any missing values.
- Standardize and normalize the variables "Time" and "Amount."
- To remedy the imbalance, apply balancing techniques such as under sampling or oversampling (SMOTE), as shown in Figure 1 about the imbalance data classes.

Modeling

- Test different machine learning algorithms:
- Logistic Regression
- Random Forest
- Gradient Boosting
- XGB
- AdaBoost
- Decision Tree
- CatBoost
- kNN
- Use cross-validation to ensure model robustness.

2. Evaluation

- To assess performance, use relevant metrics such as AUPRC in addition to precision, recall, and F1-score.
- Examine the confusion matrix to find common errors.

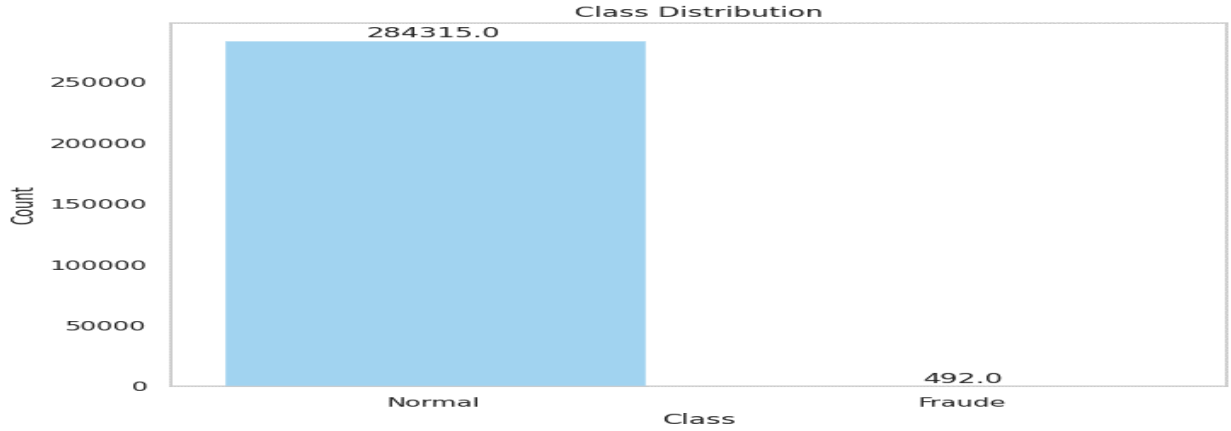

Figure 1.
Majority and Minority classes

Examining the timing of each transaction and the distribution of transaction values in both legitimate and fraudulent scenarios as the distribution of transaction values and timestamps among legitimate and fraudulent operations is thoroughly examined in this investigation. We aim to relate these distributions to find any trends or irregularities that can point to fraud. For example, fraudulent transactions may need unusual quantities as compared to typical transactions, or during exact periods of the day they may cluster. This phase is vital for understanding the important properties of the data and making detection models that performs well. We divide the dataset into variables for testing and training. In this case, we trained the model using a balanced subset of the dataset, saving the whole dataset for final testing.

For usage of this technique, the dataset must be separated first into two different sets: one for model training and the other for performance evaluation. However, we address the problem of imbalance class which frequently exists in datasets, especially in detecting frauds situations, rather than using the entire dataset together for training and testing, we used a balanced data subset for the training stage, making genuine transactions and fake transactions equal. This technique of balancing helps in the model's effective learning without preferring the mainstream class. We evaluate the model's performance on the complete, imbalanced dataset once it has been trained on this balanced sample. It offers assessment of the model's performance more accurately in real-world circumstances—where deceitful transactions are uncommon—the final testing stage is vital which ensures the effective model work for both classes—identifying fraud accurately while minimizing false positives on genuine transactions—is the goal.

To handle the class imbalance in the dataset, we performed the random undersampling technique in this phase. By cautiously eliminating examples from the majority class, this method balanced the dataset and lowered the chances of overfitting. In this instance, we found that the dataset comprises 492 fake transactions. We select at random the

similar quantity of genuine transactions (492) from the majority class to balance the dataset (Li, Z., et al., 2021). We make it assure that our model sees an equal number of samples from both classes by doing this, which is necessary for it to correctly absorb to differentiate among fraudulent and non-fraudulent actions. To make the order of transactions random, we shuffled the data after selecting the balanced subset. This is a critical stage because it preserves the model to pick up any accidental patterns from the data's order, letting it to focus only on the ultimate patterns that distinguish authentic transactions from deceitful ones. We used this method to sidestep the overfitting traps that often arise when dealing with tremendously imbalanced datasets, instead we created a model that is precise, robust, and compatible to new data, as shown in Figure 2 about the normal and Fraudulent class after sampling method.

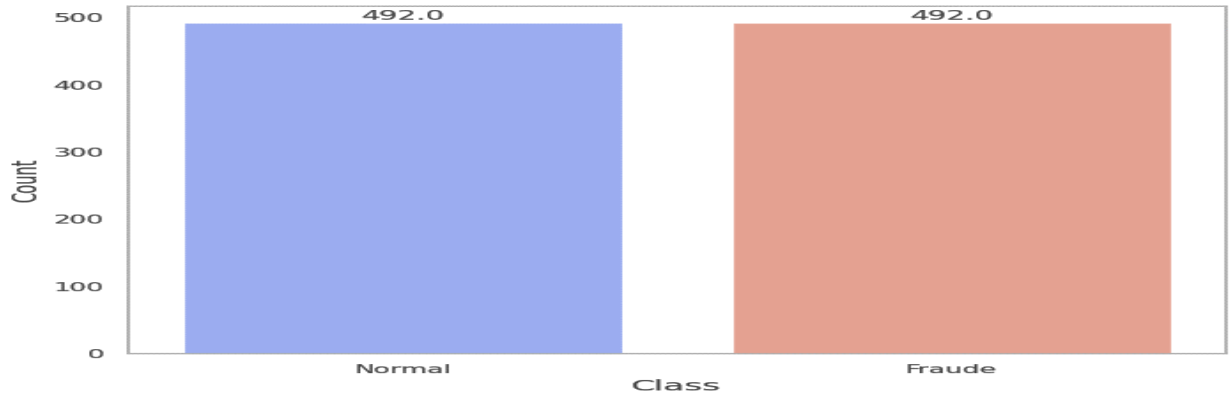

Figure 2.
Distribution of Normal and Fraudulent Transitions.

We created a new count plot to evaluate and visually endorse the usefulness of the data balancing process to assess if the dataset has been balanced successfully between genuine and fake transactions, this phase requires making a new count plot. We can make sure that there are an equal number of examples in each class by visualizing the class distributions. This is important since it will help when training a machine learning model to effectively distinguish between the two classes. We can easily confirm that the balancing technique—whether through under sampling, oversampling, or another method—has been performed correctly thanks to the count plot's clear visual representation. Before starting the model training process, this is a crucial step since it reduces the possibility that the model will be biased in favor of the majority class, which will improve the model's performance on the minority class (Gao, Y., et al.,2022).

The correlation matrix is calculated to identify variables that are more sensitive to fraudulent transactions as shown in figure 3. To accurately detect these relationships, it's essential to use the balanced dataset we have generated we may learn more about the relationships between these variables and, more crucially, how they affect the probability of a transaction being fraudulent by examining the correlation matrix. We make sure the analysis isn't biased by the overrepresentation of non-fraudulent transactions by utilizing the balanced dataset, which could obscure the importance of important variables linked to fraud. We will also contrast this with the correlation matrix that was created using the initial unbalanced dataset. This assessment will show how challenging it is to find the most significant factors when there is a significant imbalance in the data. Data balancing is critical for effective feature analysis because, in an

unbalanced situation, the overwhelming number of valid transactions makes it difficult to classify patterns associated with fraud.

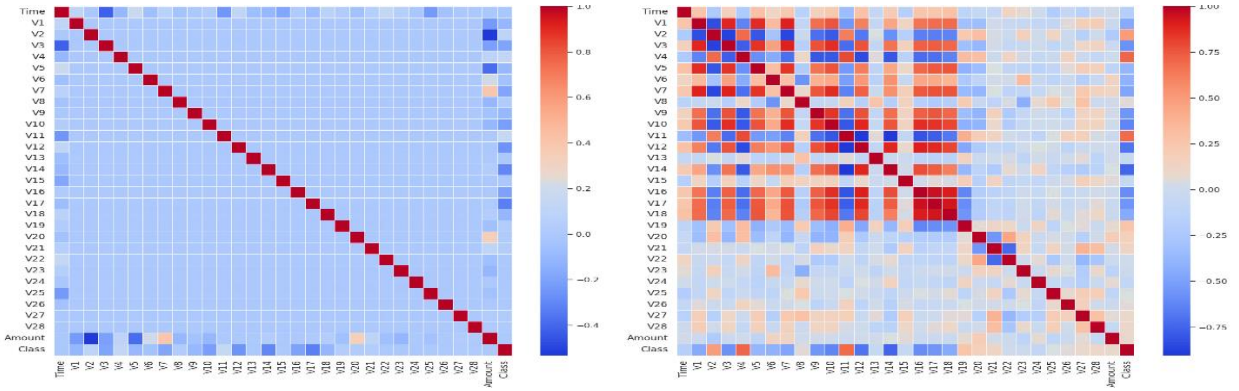

Figure 3.
Before and After balance data.

The unbalanced dataset's variables did not correlate well with the targeted output variable, that makes it problematic to discover any important designs linked to fraudulent transactions. The networks between the target class and variables, though considerably extra apparent in the balanced dataset, signifying that balancing the dataset takes revealed important relatives that were formerly unseen (Darwish, S. M. et al., 2022). Particularly, there are noteworthy harmful relations among variable quantity V3, V10, V12, and V14 and the target class. This suggests that the chance of a fraudulent transaction cultivates as the morals of these variable quantities fall. The occurrence of a higher negative value in this variable quantity is a noteworthy pointer of fraud, henceforth these potentials are vital for detecting fraudulent dealings (Rahmani, F., et al., 2022).

In contrast, there is a positive suggestion among variables V2, V4, and V11 with the mark class. Higher levels of these features in these circumstances are related to a higher accidental that the deal is fraudulent. These factors are similarly significant because raised values indicate conceivable fraud. The imbalanced and balanced dataset changes highlight how critical data preprocessing is, especially when the target class is underrepresented (Wang, L., et al., 2021). These decisive associations could stay covered in the nonappearance of balance, which would outcome in a model that completes poorly at recognizing fraud as shown in Figure 4. By utilizing the knowledge gathered from these crucial variables, we may develop a model that more precisely detects fraudulent transactions by revealing these linkages through balancing.

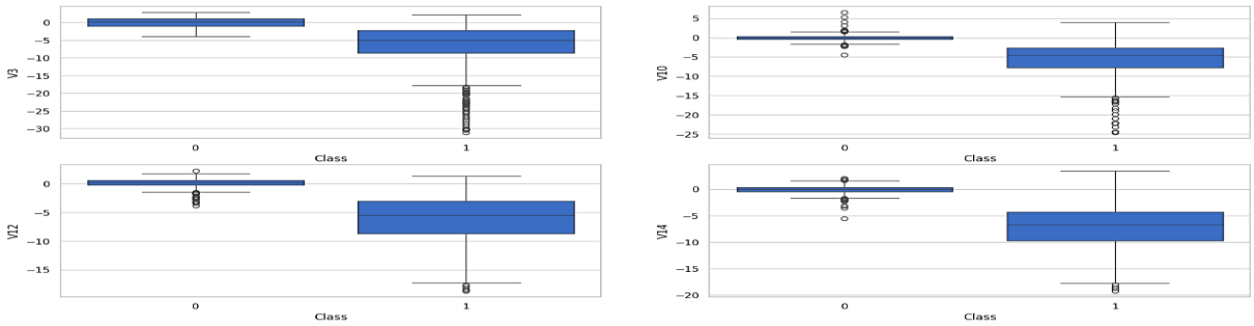

Figure 4.
Box plot of Majority and minority class.

The distribution of values for four important variables (V3, V10, V12, and V14) between two classes—fraudulent transactions (Class 0) and non-fraudulent transactions (Class 1)—is shown in the box plots. In the previous correlation matrix study, these variables were found to have substantial relationships with the target variable (Class). This is a thorough analysis:

1. V3

• **Distribution in Non-Fraudulent Transactions (Class 0):** There is minimal fluctuation in the values of V3, which are closely grouped around zero in non-fraudulent transactions.
• **Distribution of Fraudulent Transactions (Class 1):** Compared to non-fraudulent transactions, the values of V3 in fraudulent transactions are more widely dispersed adversely, with a substantially lower median value.
• **Interpretation:** This shows a clear reflection of the negative correlation found in the correlation matrix. V3 is a critical signal for fraud detection because lower values are substantially linked to fraudulent transactions.

2. V10

• **Distribution in Non-Fraudulent Transactions (Class 0):** For non-fraudulent transactions, the values of V10 are centered with a small dispersion around zero, just like in V3.
• **Distribution in Fraudulent Transactions (Class 1):** Compared to non-fraudulent transactions, V10 shows a substantially wider range and a lower median value, indicating a negative skew.
• **Interpretation:** The negative connection previously observed is supported by the box plot for V10. The probability of a transaction being fraudulent rises as V10 falls.

3. V12

• **Distribution in Non-Fraudulent Transactions (Class 0):** V12 displays a small number of outliers and a compact distribution for non-fraudulent transactions around zero.
• **Data Distribution in Fraudulent Transactions (Class 1):** V12 has a lower median and a wider distribution with more negative values in fraudulent transactions.
• **Interpretation:** Lower values of this variable are suggestive of fraud, as confirmed by the pattern that V12 closely resembles that of V3 and V10.

4. V14

• **Distribution in Non-Fraudulent Transactions (Class 0):** As with the other variables, the values of V14 for non-fraudulent transactions are almost entirely centered around zero.
• **Distribution in Fraudulent Transactions (Class 1):** V14 displays a larger distribution with a noticeable negative skew and a lower median for fraudulent transactions.
• **Analysis:** V14 supports the pattern that shows a greater correlation between fraudulent transactions and negative values.

Overall Interpretation

• **Clear Separation:** There is a noticeable difference in the distributions of these variables in fraudulent and non-fraudulent transactions, as shown by the box plots. While fraudulent transactions are linked to much lower (more negative) values, non-fraudulent transactions consistently display values close to zero.

• **Significance of Negative Correlations:** Because of their robust negative correlations with the target class, these variables (V3, V10, V12, and V14) are essential for detecting fraudulent transactions. The likelihood of a fraudulent transaction increases with the negative values of these variables.

• **Implications for Modeling:** Models used for fraud detection benefit greatly from these insights. By concentrating on these factors, the model may be trained to identify the trends connected to fraudulent activity, improving the accuracy of its predictions.

This research emphasizes the significance of these characteristics in differentiating between legitimate and fraudulent transactions and provides additional validation for the correlation matrix findings as shown in Figure 5.

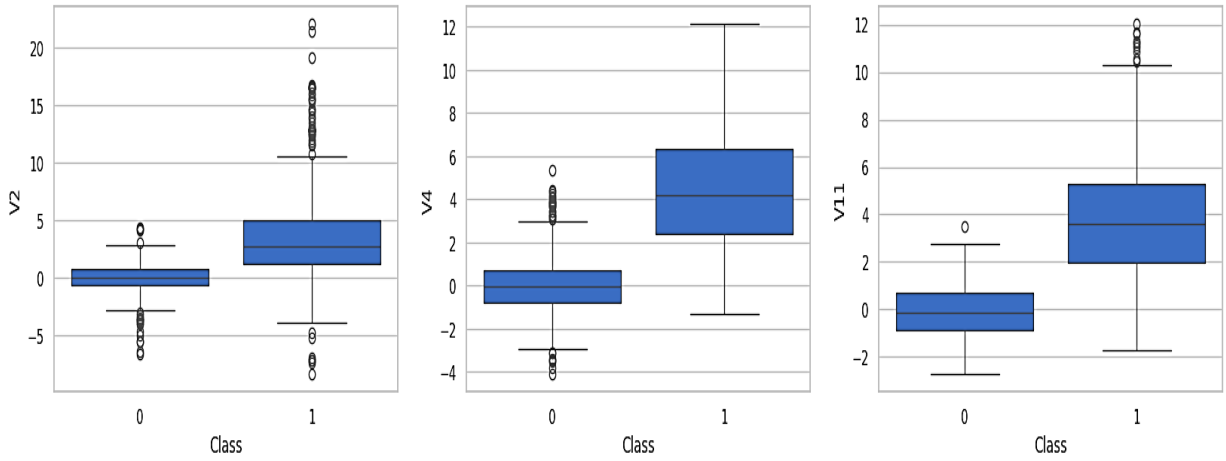

Figure 5.
Box plot of positively correlated Majority and minority class.

The distribution of values for three variables (V2, V4, and V11) between two classes fraudulent transactions (Class 1) and non-fraudulent transactions (Class 0—is depicted in the box plots below. Robust positive correlations among these variable quantities and the target (Class) originated. This is a thorough analysis:

1. V2

- **Distribution in Non-Fraudulent Transactions (Class 0):** The standards of V2 in non-fraudulent connections are positioned about zero, with a comparatively tight delivery and a few outliers on the inferior end.
- **Distribution in Fraudulent Transactions (Class 1):** For fake communications, the values of V2 are usually higher, with the middle value meaningfully superior to that of non-fraudulent dealings. The distribution also shows a wider feast with more extreme positive values.
- **Interpretation:** The positive association observed for V2 income that as the value of V2 upsurges, the likelihood of the deal being fraudulent also upsurges. This makes V2 a key pointer for detecting fraud.

2. V4

- **Distribution in Non-Fraudulent Transactions (Class 0)**: The values of V4 for non-fraudulent transactions are typically placed around zero, with negligeable difference.

- **Distribution in Fraudulent Transactions (Class 1)**: In difference, the standards of V4 for fraudulent transactions are meaningfully advanced, with a wider variety and a much-advanced median.
- **Interpretation**: The box plot for V4 strengthens the positive association realized in the correlation matrix. Higher values of V4 are indicative of a higher likelihood of fraud.

3. V11

- **Distribution in Non-Fraudulent Transactions (Class 0):** V11 values for non-fraudulent transactions are closely packed around zero, with very little deviation.
- **Distribution in Fraudulent Transactions (Class 1):** Fraudulent transactions exhibit higher values of V11, with a distribution that is shifted upwards compared to non-fraudulent transactions.
- **Interpretation:** Similar to V2 and V4, V11 shows that higher values are associated with fraudulent transactions, making it an important feature for distinguishing between fraudulent and non-fraudulent activities as shown in Figure 6.

Overall Interpretation

• **Clear Separation:** There is a noticeable difference between the distributions of these variables in fraudulent and non-fraudulent transactions, as seen by the box plots. While fraudulent transactions frequently have values substantially higher than zero, non-fraudulent transactions usually show values close to zero.

• **Significance of Positive Correlations:** Because of their robust positive correlations with the target class, these variables (V2, V4, and V11) are essential for detecting fraudulent transactions. Raised values of these factors supplement the probability of a fraudulent transaction.

• **Implications for Modeling:** These results highpoint how vital it is to integrate this definitely related variable quantity into fraud detection replicas. By emphasizing these characteristics, the model can be trained to more precisely distinguish designs linked to fraudulent dealings, which will upsurge forecast exactness in the end. This investigates proposals a full sympathetic of the rudiments swaying the probability of fraud in the dataset, completing previous answers on damagingly linked variable quantity. By combination these examines, a strong fraud discovery model that takes benefit of the advantages of both definitely and damagingly connected features is created. We can see that there are distinguished outliers for every mutable from the box plots. Since they present noise and bias into the training procedure, these outliers have the possible to alter the presentation of organization models.

Therefore, to assurance the accuracy and pliability of our models, we will methodically analyze these outliers to better comprehend their influence and, if obligatory, move forward with their removal (Adler, D. A., et al., 2021). This version highpoints the significance of leading detailed research beforehand acting and offers a more detailed discussion of the possible glitches carried on by outliers. It also stresses the essential of upholding the accuracy and heftiness of the classification models. To train the model on numerous data subsets, we employed cross-validation. This technique assurances that the model achieves well in the nonappearance of data and allows us to measure its recital more thoroughly (Li, T., et al., 2020). Cross-validation reduces overfitting and

harvests a more precise valuation of the model's true accuracy by separating the dataset into frequent folds as shown in Figure 6. Cross-validation dividers the data into numerous equal-sized "folds." After training on a exact amount of these folds, the model is verified on the remaining fold or folds. To get a final recital number, this procedure is done several times, using each fold as the test set precisely after. The results are then averaged. By exploiting frequent training and validation situations, this method not only recovers accuracy but also delivers insights into the constancy and solidity of the model across numerous data ruptures. Its assistance in classifying recital variances and guarantees that the model functions properly across different data subsets (Cabitza, F., et al., 2021). Cross-validation can too be beneficial in hyperparameter alteration, which is the process of precisely comparing numerous model outlines to see which one produces the best simplification performance. Because of this, it's a critical tool for emerging dependable machine learning models.

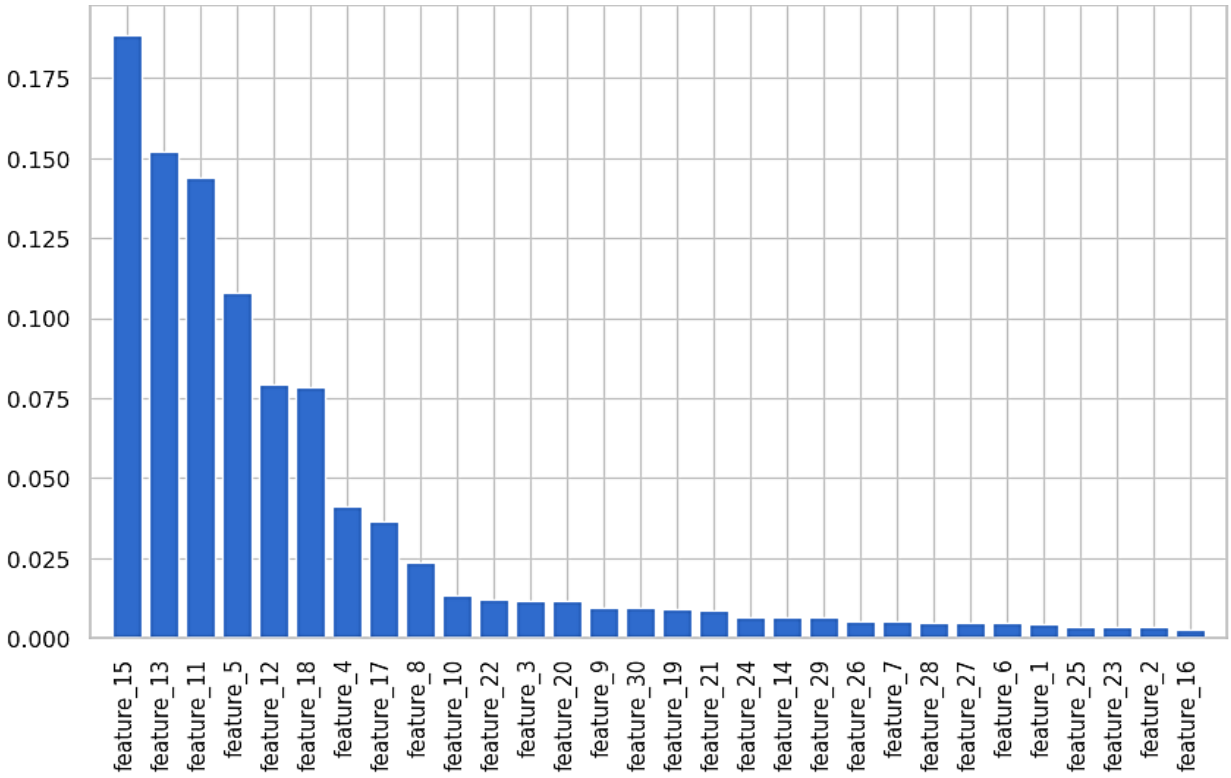

Figure 6.
Tenure-based attributes' importance.

Model Metrics

To thoroughly measure each model's presentation, we calculate the classification metrics in this phase, which comprise accuracy, ROC-AUC, F1-score, precision, and recall. The target is to establish which model works finest for our specific trial by investigative these procedures. To regulate which model is the greatest precise and best achieves the trade-offs among false positives and false negatives (Nadeem, G., et al., 2023), (Nadeem, G., et al., 2024). The presentation of the numerous models remained also inspected in figure 7, this complete analysis assistances to select the finest model for our classification tricky.

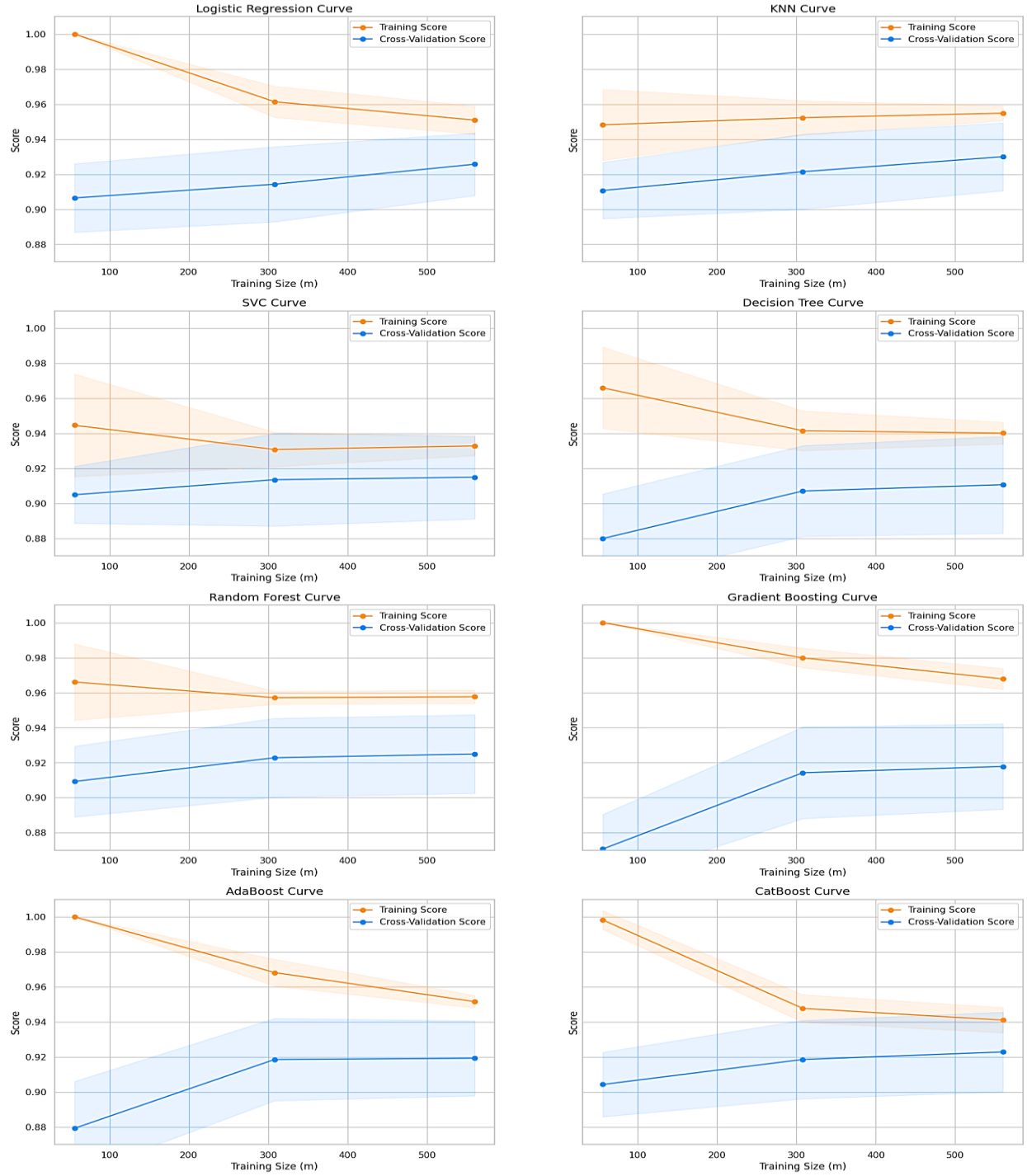

Figure 7.
Training and validation score of multiple models.

Analysis of Learning Curves

The analysis of the learning curves presented in the provided graphs offers insights into the performance of the models as the training set size increases. Here is a detailed analysis:

1. Logistic Regression Curve

- **Training Score:** The Logistic Regression performance on the training set is nearly perfect but slightly decreases as the training set size increases.
- **Cross-Validation Score:** The cross-validation performance gradually improves as the training set size increases, suggesting that the model benefits from more training data. However, the gap between the training and validation curves suggests possible overfitting, where the model fits the training set very well but does not generalize as well to new data.

2. KNN Curve

- **Training Score:** The training score for KNN is high but slightly decreases as the training set size increases.
- **Cross-Validation Score:** The cross-validation performance improves with the increased training size but remains significantly lower than the training performance, indicating that the model may be overfitting the training data.

3. SVC Curve

- **Training Score:** SVC shows an almost perfect performance on the training set, with a slight drop as the training set size increases.
- **Cross-Validation Score:** The cross-validation score slightly increases with the training set size, but the gap between the training and cross-validation scores suggests that the model is overfitting.

4. Decision Tree Curve

- **Training Score:** Decision Tree's performance on the training set is very high and remains almost constant as the training set size increases.
- **Cross-Validation Score:** The cross-validation score is significantly lower and does not improve much with an increased training size. This indicates that the decision tree may be overfitting the training set and being unable to generalize well to new data.

5. Random Forest Curve

- **Training Score:** Random Forest also shows very high performance on the training set, with a slight drop as the training set size increases.
- **Cross-Validation Score:** The cross-validation score improves slightly with the increased training size, but the gap still suggests slight overfitting.

6. Gradient Boosting Curve

- **Training Score:** Gradient Boosting shows high performance on the training set, with a slight drop as the training set size increases.
- **Cross-Validation Score:** The cross-validation score improves with the increased training set size, indicating that the model benefits from more data to generalize better.

7. AdaBoost Curve:

- **Training Score:** The training performance is high but decreases as the training set size increases.
- **Cross-Validation Score:** The cross-validation score improves with the increased training set size, but the gap between the curves suggests that the model may be overfitting.

8. CatBoost Curve:

- **Training Score:** CatBoost shows good performance on the training set, with a slight drop as the training set size increases.
- **Cross-Validation Score:** The cross-validation score improves with the increased training set size, indicating that the model may benefit from more data.

GENERAL DISSCUSSION

- **Overfitting:** Most models, except perhaps Gradient Boosting and CatBoost, show signs of overfitting, where the training score is significantly higher than the cross-validation score.
- **Benefits of More Data:** Most models improve with an increased training set size, suggesting that more data could help improve the model's generalization.
- **Hyperparameter Tuning:** Models like Decision Tree and KNN may need additional hyperparameter tuning or regularization techniques to reduce overfitting.

Based on these analyses, it is advisable to focus on adjusting the models to improve generalization and explore regularization methods or the collection of more data to increase overall performance.

Overfitting and Generalization

Even with their impressive performance, there is still a discernible difference between the cross-validation and training scores, especially in the initial phases of training (Lomboy, K. et al., 2021). This suggests that even if SVC and logistic regression are useful, they may still be somewhat prone to overfitting. On the other hand, the comparatively smaller gap in comparison to other models implies that they have achieved a better balance between training data fitting and fresh data generalization.

Model Stability

The cross-validation scores of SVC and logistic regression converge towards the training scores as the training size rises. This behavior suggests that these models can produce more consistent and dependable predictions with enough data, which makes them excellent candidates for implementation in a production setting.

Potential for Improvement

Since the hopeful outcomes of SVC and logistic regression, hyperparameter alteration and added optimization may be necessary to more recover their success. By using

approaches like Grid Search CV or Randomized Search CV, these models' strength be better and the change among the training and validation notches might be abridged smoothly.

Table 1.
Accuracy Metrics of Machine Learning models.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	76	67	92	77
SVC	93	96	88	92
KNN	96	98	92	95
Decision Tree	94	95	91	93
Random Forest	96	100	91	95
Extreme Gradient Boost	95	100	90	92
CatBoost	94	98	88	93
AdaBoost	95	99	89	94

We can see that logistic regression works better by looking at the recall score. SVC wins out when the other measures are taken into account, though. From here on, we use the SMOTE method, which creates artificial points to balance the dataset. SMOTE generates synthetic locations along these distances by estimating the separation between the minority class's closest neighbors. When compared to under-sampling, this method typically yields more accuracy; but, because of the artificial data's complexity, training the model takes longer (Brandt, J., et al., 2021), (Feng, S., et al., 2022). Using SMOTE can enhance the model's generalization capacity by offering a more balanced training dataset, particularly when the minority class is underrepresented.

- Accuracies: 96
- Precision: 100
- Recall: 91
- F1: 95
- AUC: 0.95

These renewed answers clearly show how abundant the logistic regression model has progressive. The model's volume to notice patterns linked to fraudulent transactions has been wired and its skill to simplify has been certain by the use of SMOTE to balance the dataset. The model is now far more accomplished of precisely forecasting together positive and negative cases, as seen by the upsurges in precision, recall, and F1 score. Moreover, the AUC value indicates better overall performance, representative the improved ability of the logistic regression model to favor between fraudulent and non-fraudulent dealings. This improvement as shown in Table 2 highlights how vital data balancing plans are to making dependable forecast models, chiefly when training with unbalanced datasets.

Table 2.
Random Under sampling and SMOTE Oversampling on Machine Learning Models.

Model	Technique	Score	
0	Logistic Regression	Random Undersampling	0.943182
1	KNN	Random Undersampling	0.931818
2	SVC	Random Undersampling	0.931818
3	Decision Tree	Random Undersampling	0.903409

Model	Technique	Score	
4	Random Forest	Random Undersampling	0.926136
5	Gradient Boosting	Random Undersampling	0.931818
6	AdaBoost	Random Undersampling	0.937500
7	LightGBM	Random Undersampling	0.937500
8	XGBoost	Random Undersampling	0.937500
9	CatBoost	Random Undersampling	0.931818
10	Logistic Regression	Oversampling (SMOTE)	0.920455
11	KNN	Oversampling (SMOTE)	0.920455
12	SVC	Oversampling (SMOTE)	0.920455
13	Decision Tree	Oversampling (SMOTE)	0.920455
14	Random Forest	Oversampling (SMOTE)	0.920455
15	Gradient Boosting	Oversampling (SMOTE)	0.920455
16	AdaBoost	Oversampling (SMOTE)	0.920455
17	LightGBM	Oversampling (SMOTE)	0.920455
18	XGBoost	Oversampling (SMOTE)	0.920455
19	CatBoost	Oversampling (SMOTE)	0.920455

CONCLUSION

Several noteworthy implications are formed since the model presentation investigation: With a groove of nearly 0.9659, the Random Forest model with random undersampling had the maximum accuracy. This proposes that Random Forest is the finest model for the specific cataloging tricky when trained with a stable dataset. Its extraordinary recital proves how well it can accomplish the dataset's delicacies. When trained with random undersampling, mutually the kNN and Random Forest models established imposing performance, procurement correctness notches of about 96%. These demonstrations that models are fairly hardy and that, when secondhand with undersampling, they can grip tilted data well. It's stimulating to letter that out of all the models inspected, the Logistic Regression model with random undersampling conventional the deepest accuracy notch of 77% and in undersampling of 0.94. This discovery suggests that, in difference to other models, Logistic Regression may not be as fruitful in handling imbalanced datasets with undersampling, contempt its extensive use. It highlights how vital it is to select a model dependent on the specific structures of the data.

Smearing SMOTE oversampling, each model gotten a reliable accuracy groove of 0.92045. This consistency amongst models suggests that SMOTE balanced the dataset well, but it might have also lessened the variance in performance between the models, which would have resulted in an accuracy plateau. Even though SMOTE is an effective method for managing imbalance, using it in this situation did not significantly outperform random undersampling. Similar outcomes were obtained by the SMOTE oversampling and random undersampling approaches, which both effectively and highly accurately identified fraudulent transactions. This result shows that, depending on the particular requirements and project restrictions, either technique can be applied to this dataset

with effectiveness. For some models, such as Random Forest, random undersampling produced marginally better results, whereas SMOTE offered a reliable and consistent performance across all models.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor of research and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Adler, D. A., Tseng, V. W. S., Qi, G., Scarpa, J., Sen, S., & Choudhury, T. (2021). Identifying mobile sensing indicators of stress-resilience. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, *5*(2), 1-32.

Alquthami, T., Zulfiqar, M., Kamran, M., Milyani, A. H., & Rasheed, M. B. (2022). A performance comparison of machine learning algorithms for load forecasting in smart grid. *IEEE Access*, *10*, 48419-48433.

Ata, O., & Hazim, L. (2020). Comparative analysis of different distributions dataset by using data mining techniques on credit card fraud detection. *Tehnički vjesnik*, *27*(2), 618-626.

Brandt, J., & Lanzén, E. (2021). A comparative review of SMOTE and ADASYN in imbalanced data classification.

Brunner, C., Kő, A., & Fodor, S. (2021). An autoencoder-enhanced stacking neural network model for increasing the performance of intrusion detection. *Journal of Artificial Intelligence and Soft Computing Research*, *12*(2), 149-163.

Cabitza, F., Campagner, A., Soares, F., de Guadiana-Romualdo, L. G., Challa, F., Sulejmani, A., ... & Carobene, A. (2021). The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Computer Methods and Programs in Biomedicine*, *208*, 106288.

Darwish, S. M. (2020). An intelligent credit card fraud detection approach based on semantic fusion of two classifiers. *Soft Computing*, *24*(2), 1243-1253.

Feng, S., Keung, J., Zhang, P., Xiao, Y., & Zhang, M. (2022). The impact of the distance metric and measure on SMOTE-based techniques in software defect prediction. *Information and Software Technology*, *142*, 106742.

Gao, Y., Zhu, Y., & Zhao, Y. (2022). Dealing with imbalanced data for interpretable defect prediction. *Information and software technology*, *151*, 107016.

Ghauri, P., Strange, R., & Cooke, F. L. (2021). Research on international business: The new realities. *International Business Review*, *30*(2), 101794.

Ghorbani, R., & Ghousi, R. (2020). Comparing different resampling methods in predicting students' performance using machine learning techniques. *IEEE access*, *8*, 67899-67911.

Hasan, B., Shaikh, S. A., Khaliq, A., & Nadeem, G. (2024). Data-Driven Decision-Making: Accurate Customer Churn Prediction with Cat-Boost. The Asian Bulletin of Big Data Management, 4(02), Science-4.

Karpoff, J. M. (2021). The future of financial fraud. *Journal of Corporate Finance*, 66, 101694.

Li, T., Levina, E., & Zhu, J. (2020). Network cross-validation by edge sampling. *Biometrika*, *107*(2), 257-276.

Li, Z., Huang, M., Liu, G., & Jiang, C. (2021). A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. *Expert Systems with Applications*, *175*, 114750.

Long, R. (2021). Fairness in machine learning: Against false positive rate equality as a measure of fairness. *Journal of Moral Philosophy*, *19*(1), 49-78.

Lomboy, K. E. M. R., & Hernandez, R. M. (2021). A comparative performance of breast cancer classification using hyper-parameterized machine learning models. *International Journal of Advanced Technology and Engineering Exploration*, *8*(82), 1080.

Mittal, S., & Tyagi, S. (2020). Computational techniques for real-time credit card fraud detection. *Handbook of Computer Networks and Cyber Security: Principles and Paradigms*, 653-681.

Nadeem, G., Ali, S., Hasan, B., Yasin, S., & Hasan, B. (2024). Decoding Bovine Behavior: A Machine Learning Analysis of Disease and Event Detection. The Asian Bulletin of Big Data Management, 4(1), Science-4.

Nadeem, G., & Anis, M. I. (2024). Investigation of bovine disease and events through machine learning models.Pakistan Journal of Agricultural Research,37(2), 102-114.

Nadeem, G., Rehman, Y., Khaliq, A., Khalid, H., & Anis, M. I. (2023, March). Artificial Intelligence-basedprediction system for General Medicine. In2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)(pp. 1-6). IEEE.

Panthakkan, A., Valappil, N., Appathil, M., Verma, S., Mansoor, W., & Al-Ahmad, H. (2022, December). Performance Comparison of Credit Card Fraud Detection System using Machine Learning. In *2022 5th International Conference on Signal Processing and Information Security (ICSPIS)* (pp. 17-21). IEEE.

QUINTIN-JOHN, S. M. I. T. H., & Valverde, R. (2021). A perceptron based neural network data analytics architecture for the detection of fraud in credit card transactions in financial legacy systems. *WSEAS Transactions on Systems and Control*, *16*.

Rahmani, F., Valmohammadi, C., & Fathi, K. (2022). Detecting fraudulent transactions in banking cards using scale-free graphs. *Concurrency and Computation: Practice and Experience*, *34*(19), e7028.

Wang, B., & Mao, Z. (2020). A dynamic ensemble outlier detection model based on an adaptive k-nearest neighbor rule. *Information Fusion*, *63*, 30-40.

Wang, L., Han, M., Li, X., Zhang, N., & Cheng, H. (2021). Review of classification methods on unbalanced data sets. *Ieee Access*, *9*, 64606-64628.

Wang, Y., Cao, X., & Li, Y. (2022). Unsupervised outlier detection for mixed-valued dataset based on the adaptive k-nearest neighbor global network. *IEEE Access*, *10*, 32093-32103.

Zhang, Y., Liu, J., & Shen, W. (2022). A review of ensemble learning algorithms used in remote sensing applications. *Applied Sciences*, *12*(17), 8654.