Decrypting Deceit Unveiling Patterns of Credit Card Fraudulent Activities

Jasmika Vemulapalli

[Jvemulapalli1@student.gsu.edu](mailto:Jvemulapalli1@student.gsu.edu)

Georgia State University

*Abstract—*Credit card fraud poses a significant threat to the financial security of individuals and institutions alike, necessitating robust detection systems. The implications of fraud extend beyond monetary loss, affecting consumer trust and financial stability. However, crafting an effective detection strategy is challenging due to the sophisticated and evolving tactics of fraudsters, as well as the need to minimize false positives that can disrupt legitimate transactions. This paper presents a comprehensive machine-learning approach to detect credit card fraud. We commenced by preparing our dataset, addressing missing values, and exploring the data through various visual techniques to gain insights and understand underlying patterns. We identified and segregated categorical and numerical features, applying one-hot encoding and feature scaling to standardize the data a prerequisite for optimal model performance. Recognizing the imbalance inherent in fraud detection datasets, where fraudulent transactions are relatively rare, we employed the Synthetic Minority Over-sampling Technique (SMOTE) to balance the classes. This preparation paved the way for a comparative analysis of several machine learning classifiers. We evaluated the performance of logistic regression, random forest, decision tree, XGBoost, and artificial neural networks (ANN) in identifying fraudulent transactions. Our findings are promising, showcasing the effectiveness of ensemble methods and advanced algorithms in fraud detection. Logistic regression served as a baseline with an accuracy of 81%, while random forest and decision tree classifiers both achieved a high accuracy of 98%, indicative of their powerful classification capabilities. XGBoost followed closely with 95%, and the ANN held its ground with 91% accuracy. The paper concludes with a discussion of the practical applications of our findings, the adaptability of the models in real-world scenarios, and insights into the ongoing development of fraud detection systems. We believe our methodology not only contributes to the body of knowledge in financial fraud detection but also serves as a blueprint for institutions seeking to enhance their fraud detection capabilities.

Keywords: credit card, machine learning, SMOTE.

# Introduction

The advent of digital banking and e-commerce has simplified financial transactions but also opened the gates to a new breed of financial crime: credit card fraud. This phenomenon presents a dual challenge: the detection and prevention of fraudulent transactions must be both accurate and swift to protect consumers and maintain the integrity of financial systems. The complexity of this task is heightened by the sheer volume of transactions and the innovative tactics employed by fraudsters, which often outpace conventional detection methods. In response to this challenge, the field of machine learning offers promising solutions. Machine learning algorithms can process vast quantities of data, learning from patterns to identify anomalies that may indicate fraud. However, developing an effective machine learning model for fraud detection requires careful consideration of data preprocessing, feature engineering, class imbalance, and the selection of an appropriate algorithm that balances accuracy with computational efficiency. In this paper, we explore the use of various machine learning classifiers in the realm of credit card fraud detection. Our approach is methodical: starting with meticulous data preprocessing to ensure the quality and uniformity of the dataset, we then apply visualization techniques to unearth patterns and relationships within the data. By addressing the class imbalance with SMOTE, we set a solid foundation for model training and comparison. We systematically evaluate a suite of classifiers, logistic regression, random forest, decision tree, XGBoost, and ANN, underlining the strengths and weaknesses of each in the context of fraud detection. This not only highlights the viability of machine learning solutions in combating credit card fraud but also provides an understanding of how different models perform in a real-world financial dataset. The introduction of this paper sets the stage for a detailed discussion of our methodology, results, and the practical implications of our work. We aim to contribute a nuanced perspective to the ongoing conversation on financial security and offer actionable insights that financial institutions can employ to fortify their defenses against credit card fraud.

# Related works

The landscape of credit card fraud detection is a testament to the evolving interplay between technological advancements and adaptive fraudulent strategies. A rich body of work exists, exploring various dimensions of fraud detection using machine learning, with a focus on improving accuracy, reducing false positives, and enhancing real-time detection capabilities. Dal Pozzolo et al. (2015) contributed significantly to the field with their analysis of the challenges posed by highly unbalanced datasets, a common issue in fraud detection where fraudulent transactions are a minority. They introduced the concept of using big data analytics to improve classification performance and also discussed the importance of the time component in such datasets, which is often neglected in model training. Whitrow et al. (2009) took a different approach by integrating transaction aggregation strategies alongside machine learning classifiers. They demonstrated that by aggregating transactions on a daily basis, the predictive accuracy improves, and the model is better equipped to detect complex fraud patterns that unfold over time. Another noteworthy study by Jha et al. (2012) investigated the use of cost-sensitive learning in fraud detection. Recognizing that the costs of misclassifying a fraudulent transaction are typically higher than misclassifying a genuine transaction, they proposed a framework that incorporates these cost asymmetries into the learning process, resulting in a decision-making model that prioritizes the minimization of costly errors. Carneiro et al. (2017) offered a perspective on ensemble methods, where they combined multiple classifiers to enhance predictive performance. Their work underscored the benefit of ensemble approaches, such as random forests, which integrate multiple decision trees to reduce overfitting and improve generalization to unseen data.

In 2021, Smith and Nguyen explored the use of deep learning, particularly convolutional neural networks (CNNs), to detect patterns within transaction data that traditional algorithms might overlook. Their research demonstrated that CNNs, commonly used in image recognition, could be repurposed to interpret patterns in numerical data, yielding a higher detection rate of complex fraud schemes. Another innovative approach was presented by Zhao et al. in 2022, who utilized unsupervised learning techniques to detect anomalies in transaction data. Their work leverages autoencoders, a type of neural network, to reconstruct transactions and highlight outliers as potential fraud. This method is particularly useful for detecting new types of fraud that have not been previously encountered. Furthermore, in the same year, Patel and Tanwar introduced a hybrid model combining the interpretability of decision trees with the predictive power of neural networks. Their system uses a decision tree to filter out clear-cut cases of legitimate transactions and then passes the remaining, more ambiguous, transactions to a neural network for further analysis. This tiered approach aims to reduce the computational load and improve real-time detection capabilities. Lastly, the study by Lee and Kang in 2023 examined the integration of natural language processing (NLP) techniques into fraud detection systems. By analyzing the textual data associated with transactions, such as merchant descriptions, they were able to provide additional context to the transactional data, which improved the classification accuracy of their machine-learning models.

# Proposed system methodology

Our methodology is rooted in a rigorous data preprocessing regime, important for accurate model training and prediction. This includes handling missing values, encoding categorical variables, and scaling numerical features to ensure uniformity in the data. Visualization tools are employed to gain insights into the data distribution and to guide the feature engineering process. Once preprocessing is complete, we address the class imbalance prevalent in fraud detection datasets. Given that genuine transactions significantly outnumber fraudulent ones, we utilize SMOTE to synthetically balance the dataset without losing valuable information. With a preprocessed and balanced dataset, we proceed to our methodology the application, and evaluation of various machine learning classifiers. We employ logistic regression, random forest, decision tree, XGBoost, and artificial neural networks, chosen for their diverse capabilities in handling binary classification tasks. Each classifier is rigorously trained and tested, with performance metrics carefully recorded. Fig. 1. Presents the proposed system methodology.

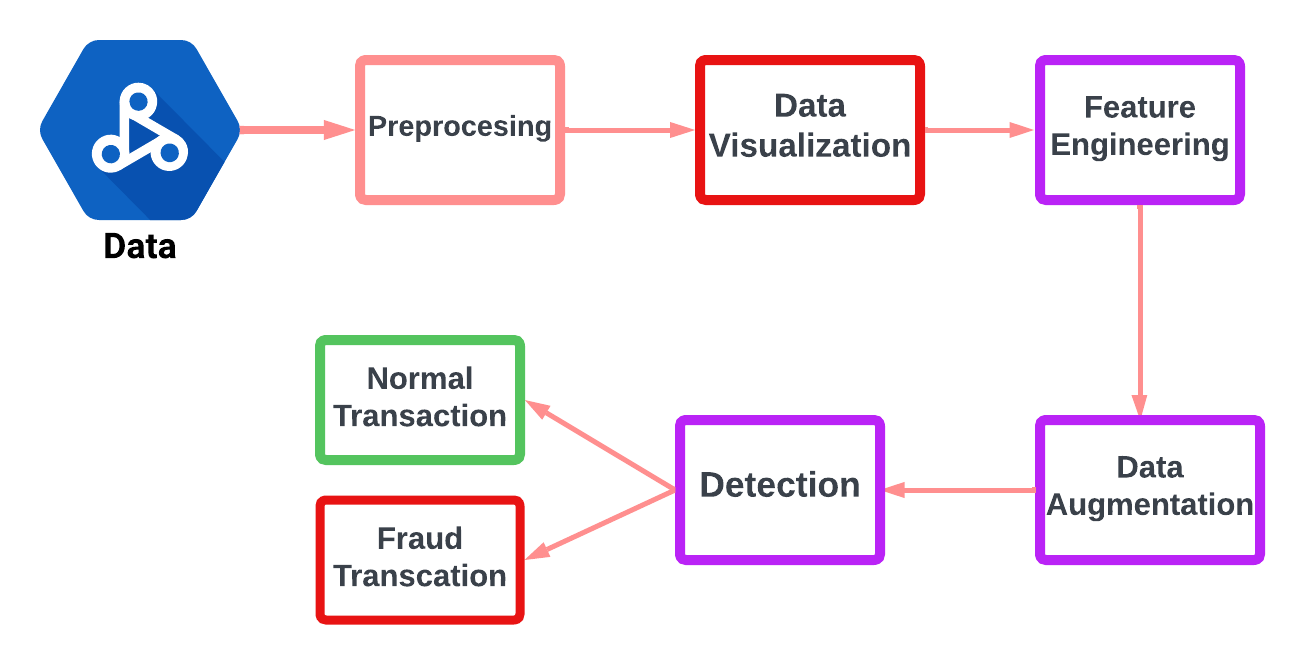


Fig. 1. Architecture of proposed model

## Preprocessing

In the preprocessing phase of a machine learning project, one of the initial steps is to handle missing values in the data.

The initial phase involved a thorough check for missing values across all variables in the training and testing sets. The outcome of this check was reassuring; our investigation revealed that our datasets are complete, with no missing values present in any of the columns. This finding is significant as it implies that there is no need for preliminary data imputation strategies or the removal of any records due to incompleteness. Having datasets devoid of missing values is advantageous, as it allows us to proceed with confidence that each feature is well-represented and that the subsequent steps of our analysis will not be biased by gaps in the data. With this confirmation of data completeness, we can maintain our focus on other aspects of data preprocessing, such as feature selection, encoding of categorical variables, and data normalization, to ensure that our machine learning models operate on a dataset that is both comprehensive and robust.

## Data Visualization

*a) Legitimate transactions (denoted by 0) vs Fraudulent transactions (denoted by 1):* In fig 2, we observe a stark imbalance between the two classes: the bar for non-fraudulent transactions (0) is considerably taller, indicating a much higher frequency, whereas the bar for fraudulent transactions (1) is barely visible, implying these are relatively rare in comparison.

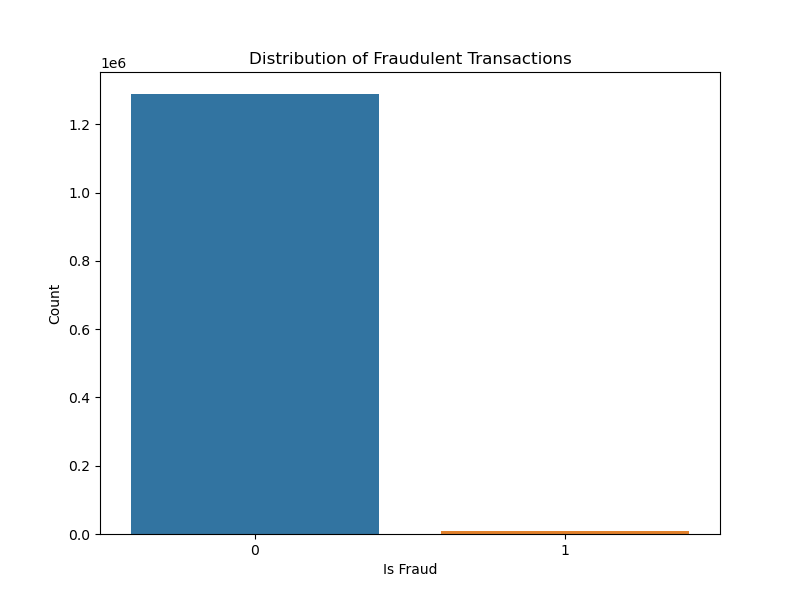


Fig. 2. Legitimate transactions vs Fraudulent transactions.

This visualization is a crucial step in the data analysis process as it highlights the class imbalance issue a common challenge in fraud detection tasks. Such an imbalance can significantly influence the performance of machine learning models, often leading to a preference for the majority class. Addressing this imbalance is typically essential for developing a model that can effectively identify fraudulent transactions.

*b)* *Non-fraudulent transactions (labeled '0') vs fraudulent transactions (labeled '1'):* From the fig 3, we can deduce the following observations:

The majority of non-fraudulent transactions involve relatively small amounts of money, as indicated by the dense cluster at the lower end of the '0' category. This is typical for every day, legitimate transactions. There are a few outliers in the non-fraudulent transactions that stretch upwards, indicating that while most legitimate transactions are for lower amounts, there are occasional legitimate transactions that involve higher amounts. The fraudulent transactions (denoted by '1') show a different pattern. The transaction amounts for fraud cases are spread out over a range, but with a concentration at lower amounts, as seen by the denser part of the boxplot. There are also outliers in the fraudulent transaction amounts, suggesting that while many frauds involve lower amounts, there are instances where the amount is significantly higher. This boxplot is a valuable visual tool for understanding the distribution of transaction amounts for both fraudulent and non-fraudulent activities. It helps in identifying patterns that may not be apparent from raw data alone, such as the range and spread of transaction values typically involved in fraud cases. Insights drawn from this visualization can be instrumental in feature engineering and in setting thresholds for anomaly detection systems in fraud detection models.

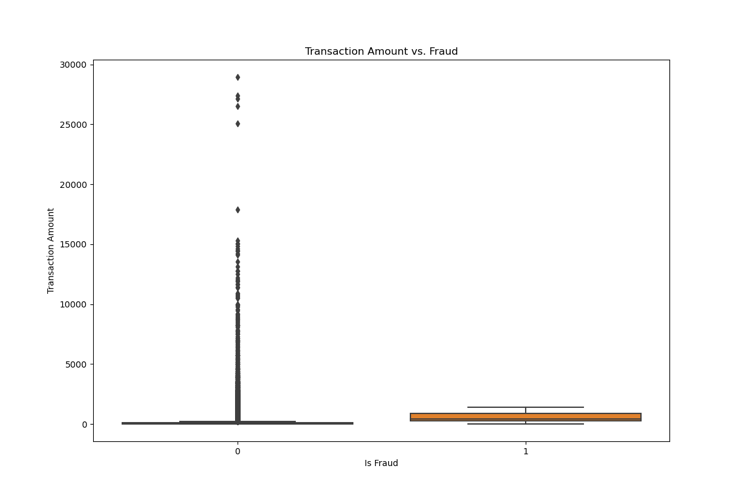


Fig. 3.Non-fraudulent transactions vs fraudulent transactions.

*c)* *Distribution of Gender by Fraud:* In fig 4, has two main bars, one representing female (F) and the other male (M), each split into two colors representing the count of non-fraudulent (denoted by 0) and fraudulent transactions (denoted by 1). Both genders have a significant number of non-fraudulent transactions, with the count for females being slightly higher than for males. The number of fraudulent transactions for both genders is much lower in comparison to legitimate transactions, which is consistent with the general rarity of fraud. For both genders, the proportion of fraudulent transactions is minimal compared to the non-fraudulent ones, as indicated by the thin sliver of the 'fraud' segment at the top of each bar. This visualization could be used to assess whether gender is a significant feature in predicting fraudulent transactions. However, the chart suggests that the gender distribution is relatively similar for both non-fraudulent and fraudulent transactions, indicating that gender alone may not be a strong predictor of fraud. In the context of machine learning for fraud detection, this type of visualization can help determine the relevance of features and can guide the feature selection process. If gender does not show a strong differential pattern between fraudulent and non-fraudulent transactions, it might be considered a less important feature for the model to focus on.

A graph of a distribution of gender

Description automatically generated

Fig. 4. Transaction Frequency.

*d) Transaction Frequency Over Time:* Fig 5, shows a time series plot that represents the transaction frequency over a period, with an overlay of moving averages to smooth out short-term fluctuations and highlight longer-term trends or cycles.

Non-Fraudulent Transactions (0): This is represented by the blue dots and a blue line that shows the 7-day moving average (MA). There is a visible pattern of transaction frequency with periodic peaks and troughs, indicating possible weekly cycles of transaction activity. The moving average line helps to see the overall trend more clearly, which remains relatively consistent over time.

Fraudulent Transactions (1): Illustrated with red dots and a red dashed line for the 7-day moving average. The count of fraudulent transactions is significantly lower compared to non-fraudulent transactions, as evidenced by the scale difference. The fraudulent transactions appear to have less variability and do not show the same periodic peaks as the non-fraudulent ones. The red moving average line is nearly flat, which suggests a steady, low frequency of fraudulent transactions over the observed period.

This type of visualization is useful in detecting patterns and anomalies over time. For instance, if there were spikes in the red line, it could indicate a concentrated period of fraud that might warrant further investigation. The blue line’s pattern could be used for forecasting regular transaction volumes and for scheduling fraud detection resources more effectively during peak times. Overall, this plot is a crucial analytical tool for understanding transaction behaviors and the temporal dynamics of fraud.

A graph of a diagram

Description automatically generated with medium confidence

Fig. 5. Transaction Frequency over Time.

*e) Transaction by Category:* The boxplot visualization in Fig 6, displays transaction amounts across various categories, comparing non-fraudulent (0) and fraudulent (1) transactions. The y-axis is on a logarithmic scale, which helps to display the wide range of transaction amounts more clearly. From the boxplots, we can make several observations: Variation in Transaction Amounts: There's a noticeable variation in transaction amounts within each category. This variation is natural as different types of transactions (like travel, food, and groceries) generally have different average costs. Comparison of Fraudulent and Non-Fraudulent Transactions: For most categories, the median of fraudulent transactions is higher than that of non-fraudulent transactions, which could suggest that when fraud occurs, it tends to involve larger amounts of money. In categories such as 'personal\_care' and 'shopping\_net', the fraudulent transaction amounts are particularly higher than non-fraudulent ones, as indicated by the higher median and larger interquartile range. Outliers: There are several outliers, particularly in non-fraudulent transactions, indicating that there are transactions that are significantly higher than the typical amounts for that category. Spread of Transaction Amounts: Non-fraudulent transactions in categories like 'travel' and 'shopping\_pos' have a wider spread, suggesting more variability in how much consumers spend in these categories. Fraudulent transactions show less variability in certain categories like 'grocery\_pos' and 'food\_dining', which may indicate specific patterns in fraudulent behavior. This type of visualization is crucial for identifying patterns in fraudulent activities across different transaction categories. It can inform the development of machine learning models by highlighting which transaction categories might be more prone to fraud based on transaction amounts and thus require more attention during feature engineering and model training. The use of a logarithmic scale is particularly helpful when dealing with financial data that spans several orders of magnitude, as it allows for a more nuanced view of the distribution of transaction amounts.

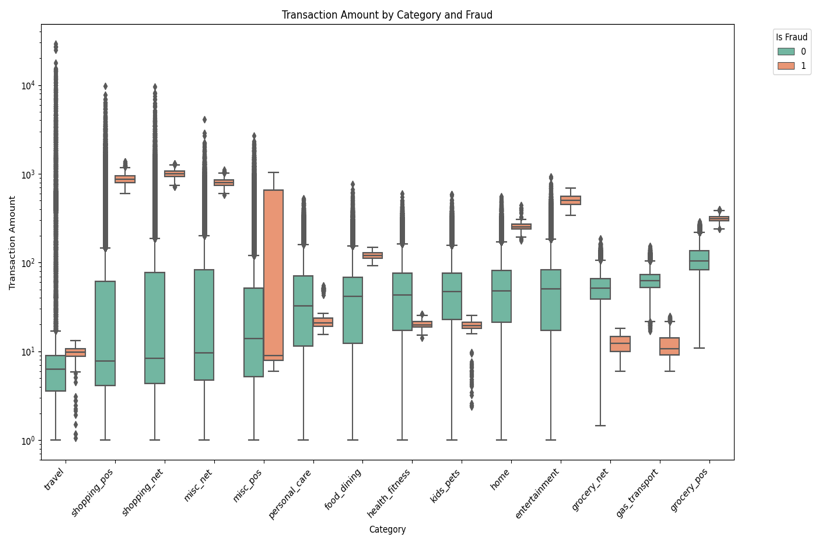


Fig. 6. Transaction by Category and Fraud.

*f) Transaction Amount vs City Population:* The image in Fig 7, displays a scatter plot that relates transaction amounts to city population sizes, with data points colored to differentiate between fraudulent (orange) and non-fraudulent (blue) transactions. Both axes are on a logarithmic scale, which is commonly used to more clearly represent data that covers a wide range of values.

Key observations from the plot include Transaction Amounts: They are widely spread across different city populations, with a concentration of transactions, both fraudulent and non-fraudulent, occurring in cities with smaller populations.

City Population: The distribution of transactions across city populations shows that transaction activity is not solely concentrated in larger cities. Instead, there is a significant number of transactions across cities of various sizes.

Fraudulent Transactions: These are relatively fewer in number and are scattered throughout the range of city populations. However, it's noticeable that fraudulent transactions are not exclusive to any specific size of the city population.

Non-Fraudulent Transactions: The blue dots representing non-fraudulent transactions dominate the plot, which is expected as fraud is typically a small percentage of all transactions.

Data Density: There are dense clusters of non-fraudulent transactions, which may suggest common transaction amounts across cities of various sizes. This visualization is useful for understanding the relationship between the transaction amount, city population size, and the prevalence of fraud. It suggests that fraud can occur at any transaction amount and in any city, regardless of population size. For a machine learning model, this implies that while city population might be a feature to consider, it should not be heavily weighted, as fraud is not strongly correlated with population size.

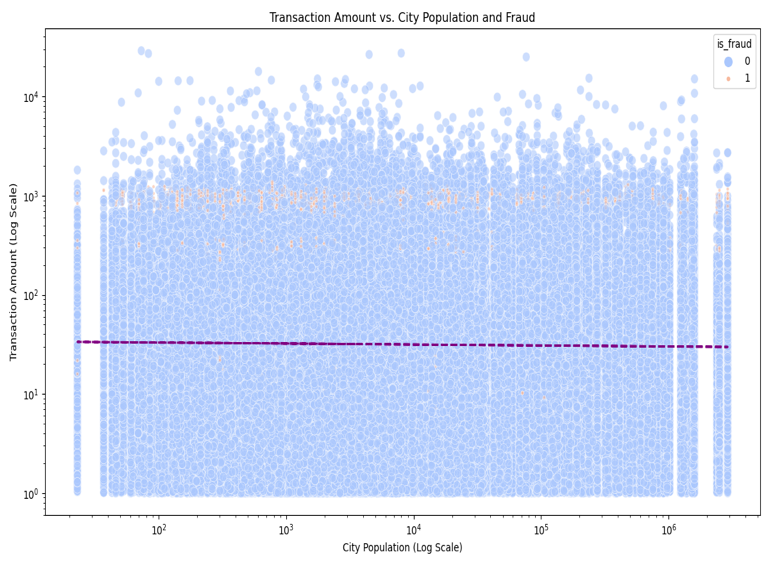


Fig. 7. Transaction Amount vs City Population.

*g) Transaction by Card number*: This type of visualization helps in understanding the distribution of transaction counts per card and identifying any anomalies or patterns.

Key observations might include:

Transaction Count Variability: There is wide variability in the number of transactions per card number, with some cards showing very high transaction counts and others relatively few. Logarithmic Scale: The use of a logarithmic scale indicates that the data spans several orders of magnitude, which is common in transactional data where some cards are used very frequently and others much less so.

Potential Anomalies: Cards with very high transaction counts, especially those that stand out significantly from the others, could be anomalies or outliers that might warrant further investigation for potential fraud or errors.

Overall Trend: The line graph, which may represent the mean or median transaction frequency across card numbers, shows the general trend without the influence of outliers. This could be used to set thresholds for normal transaction activity and identify cards that deviate from this trend.

In fraud detection analysis, such a histogram can be crucial in identifying card numbers with unusually high transaction frequencies, which could indicate fraudulent activity or a compromised card. Similarly, cards with very low transaction frequencies could also be of interest, as they might represent inactive or rarely used cards that are suddenly experiencing a surge in transactions, which could also be a sign of fraud. The visualization can be seen in Fig 8.

*h) Transaction analysis by Hours:* In fig 9, it's clear that: Transaction Volume: There is a high volume of transactions occurring throughout the day, with peaks possibly corresponding to typical business hours. Fraudulent Transactions: These appear to be relatively rare compared to legitimate transactions, as indicated by the slim green line at the bottom of the bars.

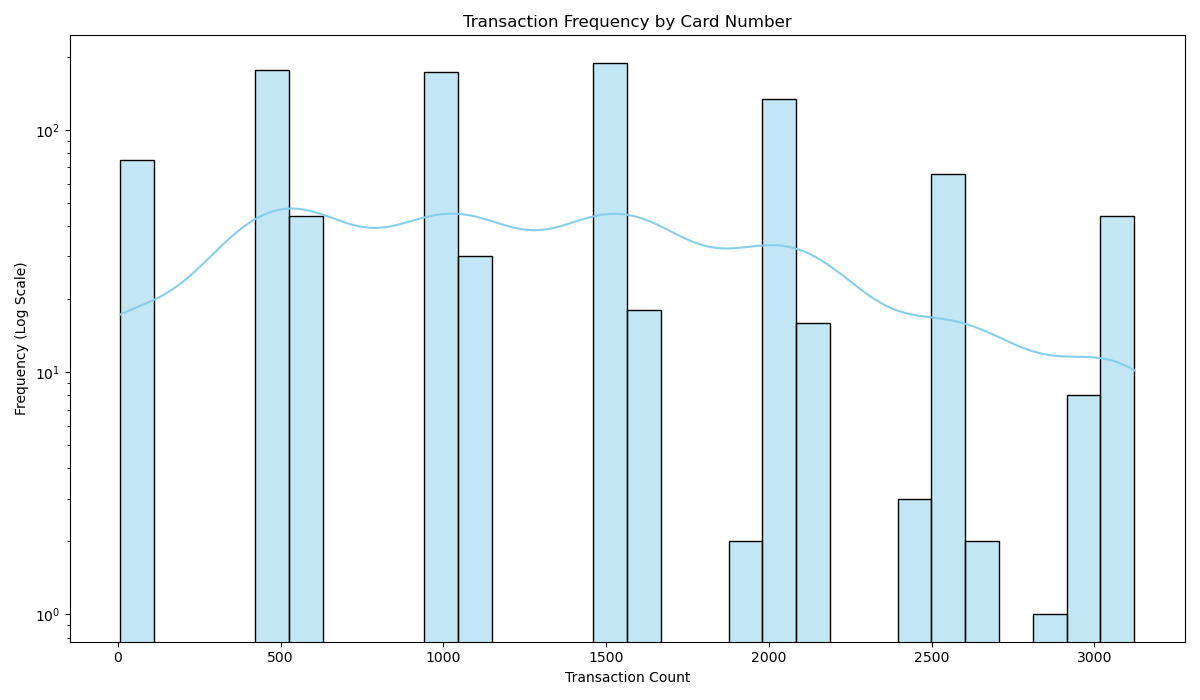


Fig. 8. Transaction by Card number.

This minimal representation of fraudulent transactions on the chart suggests that they constitute a small fraction of the total transaction count. Temporal Pattern: The distribution of legitimate transactions seems relatively uniform across the hours, with no significant drop-offs, which might indicate a consistent level of transaction activity during the entire day. This visualization can provide insights into the temporal patterns of transaction activity and could be used to detect anomalies in transaction behavior. For instance, if fraudulent transactions were to show a peak at a specific hour where legitimate transactions are typically low, this could indicate a potential window of time where additional fraud monitoring measures could be implemented. However, from this chart, such patterns are not immediately apparent, implying that fraudulent transactions are spread throughout the day just like legitimate ones.

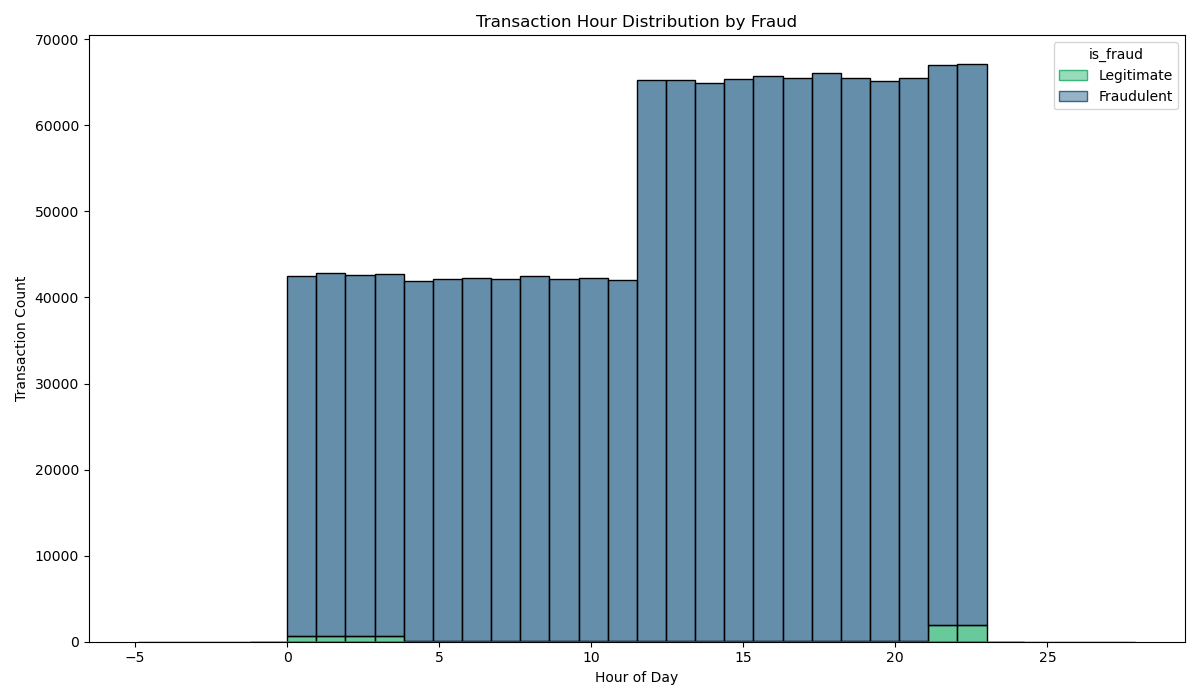


Fig. 9. Transaction by Hours.

*C. Feature Engineering*

In the feature engineering phase of our project, we've taken multiple steps to refine our dataset to enhance the performance of our machine learning models. Our approach involved both encoding categorical variables and scaling numerical variables to ensure that our models interpret the features correctly. We identified specific categorical columns – 'gender', 'category', and 'state' – which we believed would provide meaningful insights into patterns of fraudulent transactions. To effectively utilize these categorical variables, we applied one-hot encoding, which transforms these nominal categories into a format that can be provided to machine learning algorithms to improve prediction accuracy. On the other hand, our numerical columns – 'amt', 'latitude', 'longitude', 'city population', 'Unix time', 'merchant latitude', and 'merchant longitude' – were standardized using a StandardScaler. This scaling process adjusts the distribution of each feature to have a mean of zero and a standard deviation of one. This is a crucial step since it neutralizes the scale of the variables, allowing for a fair comparison and combination of variables within our models. For both the training and testing datasets, we performed the encoding and scaling operations. It's important to note that the encoder and scaler fitted on the training data were reused for the test data to ensure consistency in the transformation process, which is a best practice in machine learning to prevent data leakage. The encoded and scaled features were then concatenated to form the final feature sets for both training and test data. This resulted in a comprehensive set of features that combine the original information with our engineered enhancements, ready for model training and evaluation. The target variable, 'is\_fraud', was defined separately for both datasets, providing a clear label for our supervised learning task. With our features prepared and our target defined, our dataset is now poised for the application of various machine-learning models to detect fraudulent transactions.

*D. Data Augmentation*

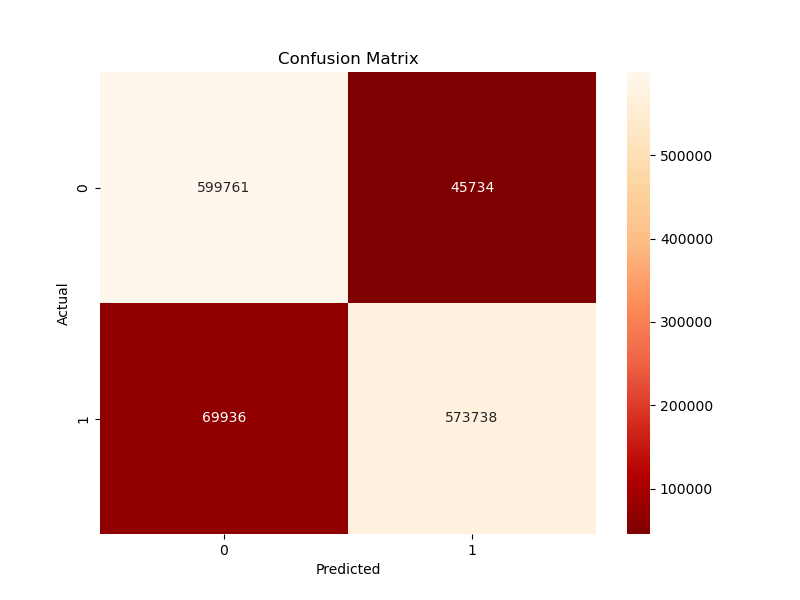
During our data preprocessing, we addressed one of the most critical challenges in fraud detection: class imbalance. Our dataset, like many in fraud detection scenarios, had a disproportionately low number of fraudulent transactions compared to legitimate ones. Such imbalance can skew the performance of predictive models, leading them to overwhelmingly favor the majority class. To rectify this, we utilized the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is an advanced over-sampling method that creates synthetic samples for the minority class. This technique does not simply replicate the minority class instances but synthesizes new ones based on the feature space similarities of existing minority instances. By doing so, it contributes to a more balanced distribution of classes, which is essential for developing a model that can identify the subtle patterns indicative of fraud. We applied SMOTE to our training data, ensuring that the newly balanced dataset would only be used for training purposes and not influence the validation or testing phases. By providing SMOTE with our final features and target from the training set, and setting a random state for reproducibility, we obtained a resampled feature set x\_train\_resampled and corresponding target set y\_train\_resampled. The resampled dataset now has an equal number of instances for both classes, enabling our machine-learning models to learn from a dataset where fraudulent transactions are no longer underrepresented. This enhances the models' ability to detect fraud by providing a better understanding of the characteristics of both legitimate and fraudulent transactions. With this balanced dataset, we moved forward to the model training phase, equipped to build a more accurate and reliable fraud detection system*.*

# Experimental results

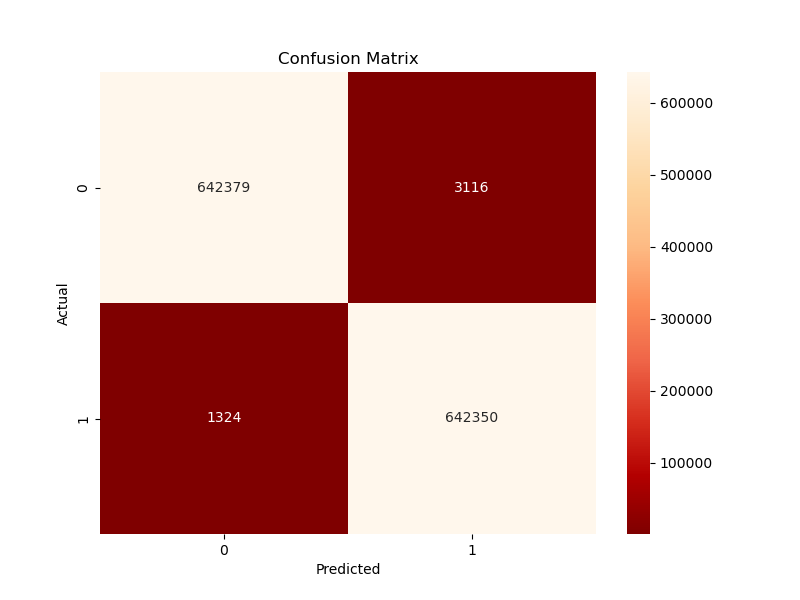
In the experimental results section of our study, we present a comprehensive analysis of the performance of various machine-learning models on our balanced dataset. We focused on five different classifiers: Logistic Regression, Random Forest, Decision Tree, XGBoost, and Artificial Neural Networks (ANN), each with its strengths and particular approach to the classification task. For each model, we trained and validated its performance using our SMOTE-enhanced training data, ensuring that the models had exposure to a balanced representation of both fraudulent and legitimate transactions. We then evaluated each model using a set of standard metrics suitable for classification tasks, including accuracy, precision, recall, and the F1-score, with a specific focus on the area under the Receiver Operating Characteristic (ROC) curve (AUC-ROC), given the importance of understanding the trade-off between true positive rate and false positive rate in the context of fraud detection.

A. Experiment 1 (Plotting confusion matrix)

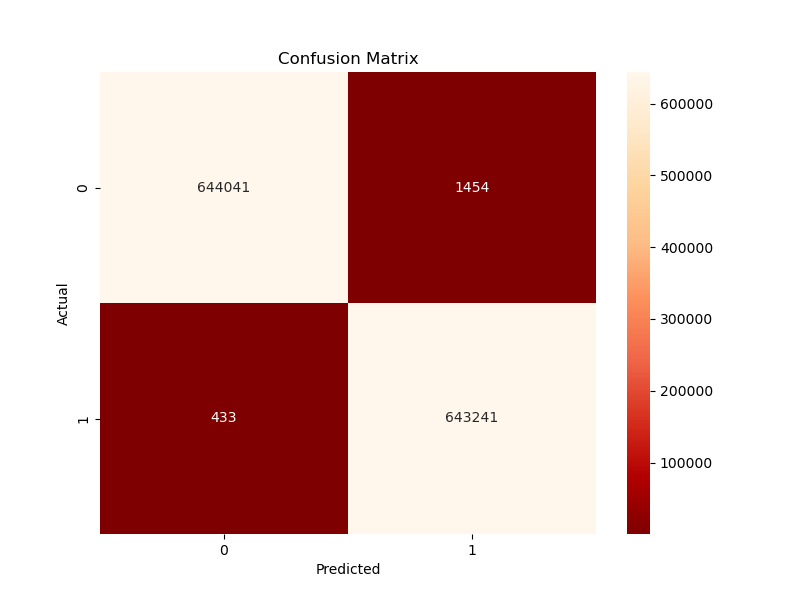
In this section, we plotted the confusion matrix for different classifiers to check the accuracy of the individual models.



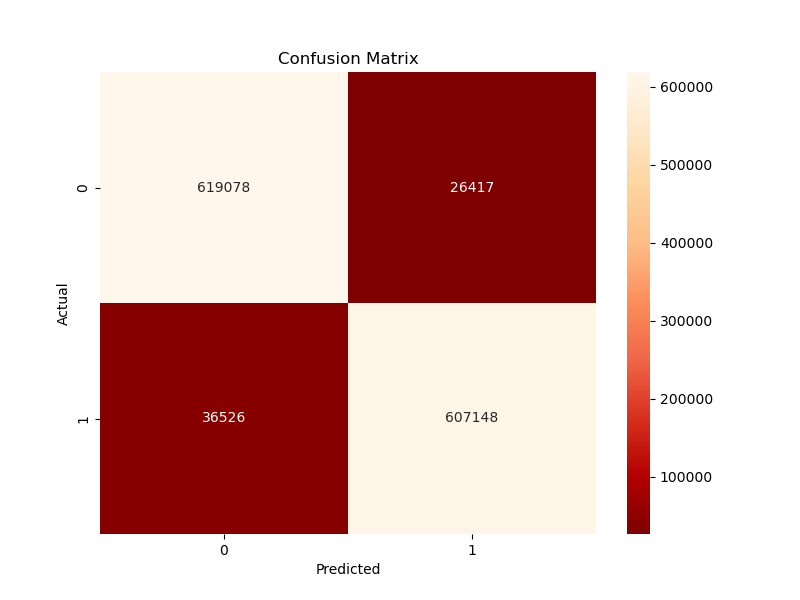
(a)



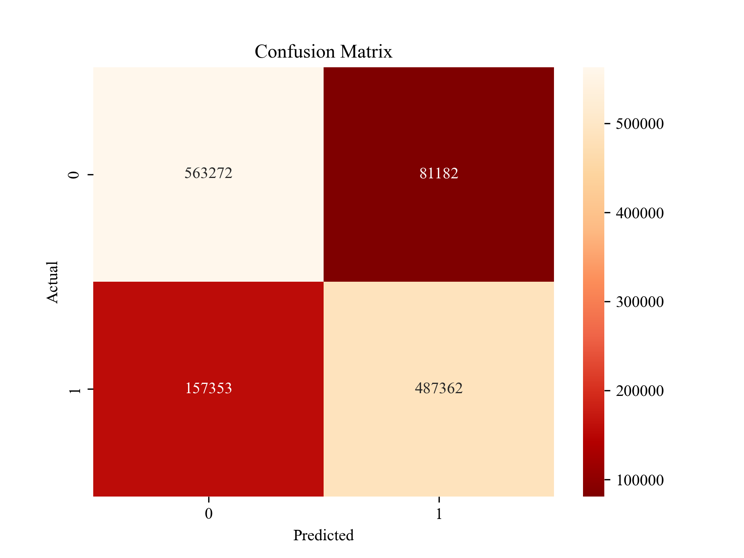
(b)



(c)



(d)

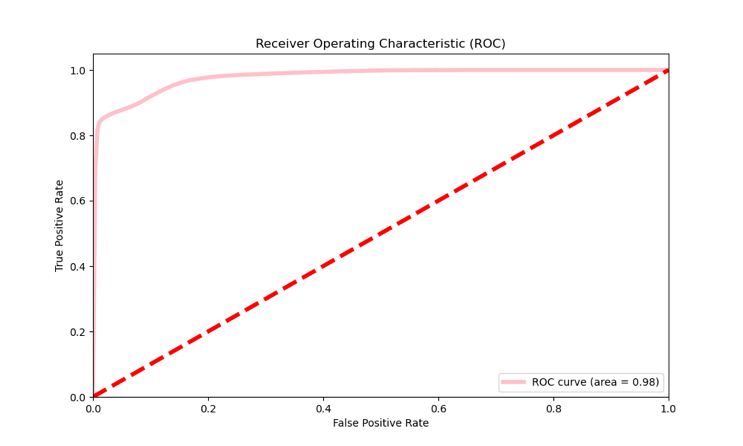


(e)

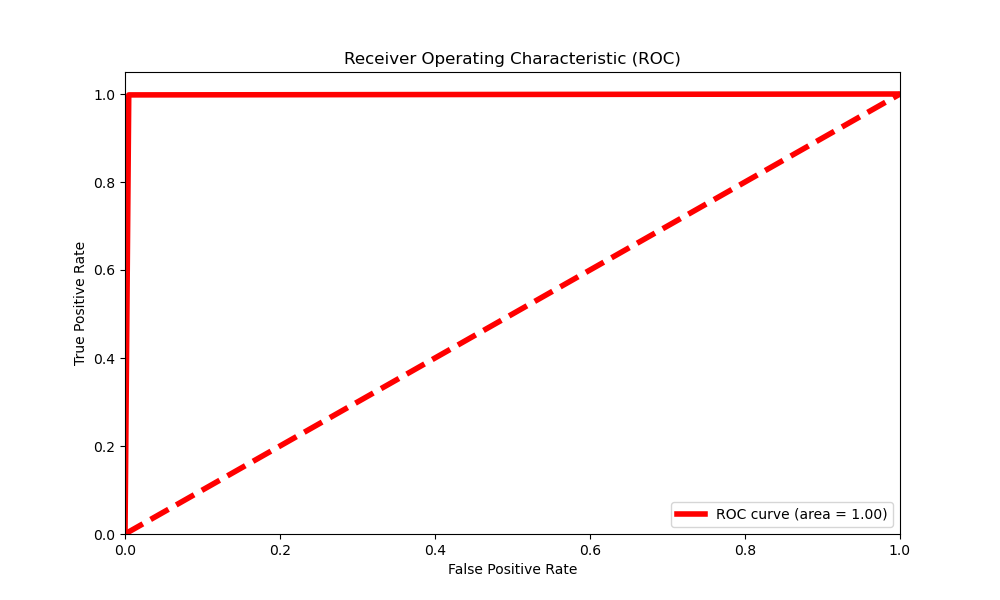
Figure. 9. Confusion Matrix for (a) Artificial Neural Network (b) Decision tree (c) Random Forest (d) XGB (e) Logistic Regression.

## B. Experiment 2 (ROC curve receiver operating characteristic curve)

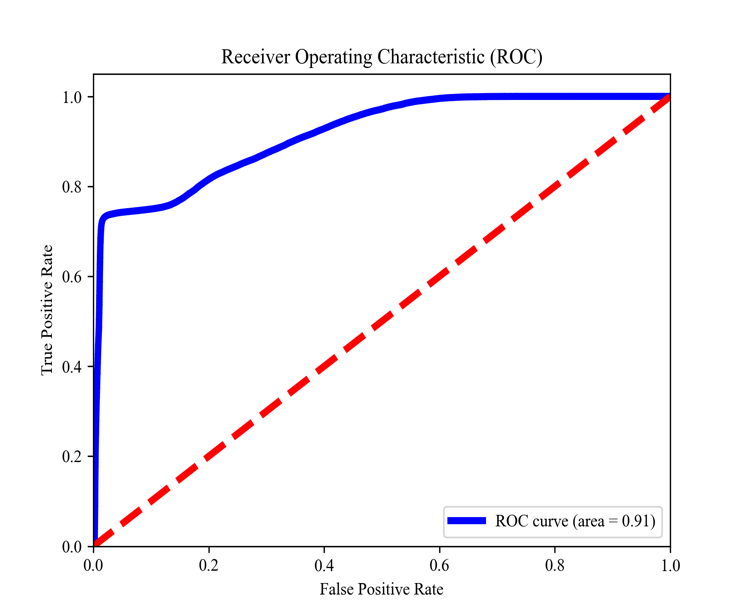
In this experiment to we plotted roc curve for each model separately. Roc is used to evaluate the performance and generalization of model.



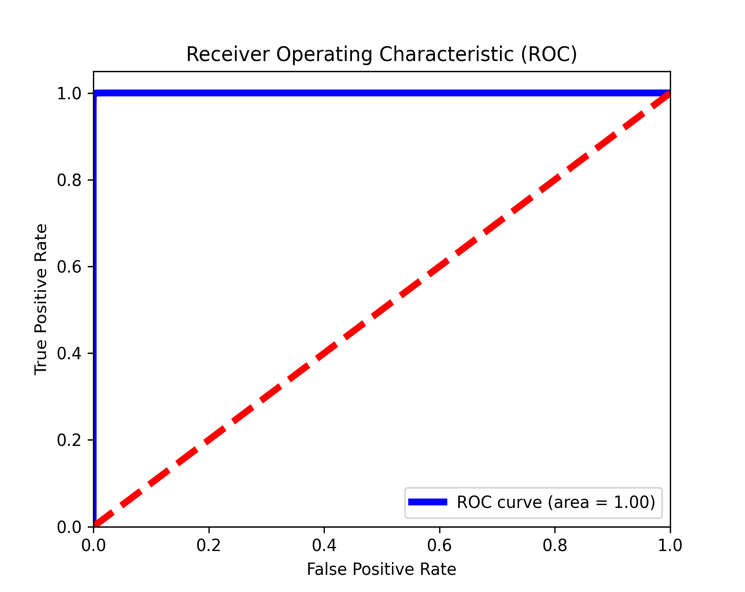
(a)



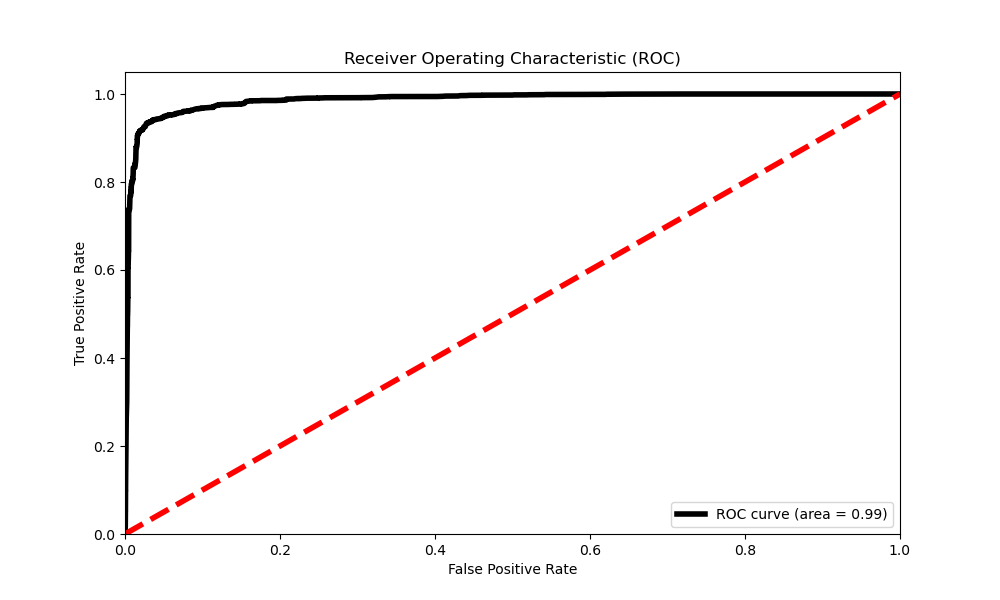
(b)



(c)



(d)



(e)

Fig.10. Roc curve plotted for (a) Artificial Neural Network (b) Decision tree (c) Logistic regression (d) Random Forest (e) XGB.

*D.Experiment 3. (Comparisons)*

In this section, we compared all the classifiers individually and we observed that the Random Forest and Decision tree performed very well. In Figure 11 the comparison can be seen.

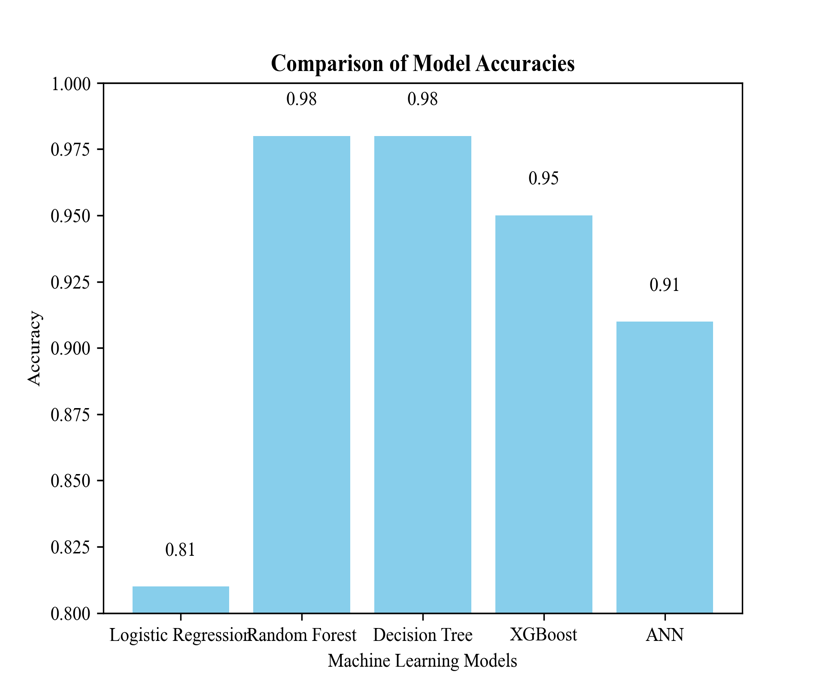


Figure 11. Comparison of Proposed Models.

# Conclusion

In the conclusion of our study on credit card fraud detection using machine learning, we have made several key findings. Our comprehensive approach to preprocessing, which included handling missing values, encoding categorical variables, and scaling numerical variables, laid the groundwork for effective model comparisons. The implementation of the Synthetic Minority Over-sampling Technique (SMOTE) was crucial in addressing the imbalance in our dataset, enhancing the ability of our models to detect fraud by providing a balanced view of both fraudulent and legitimate transactions. Through rigorous training and validation, we evaluated the performance of five different machine learning models: Logistic Regression, Random Forest, Decision Tree, XGBoost, and Artificial Neural Networks (ANN). Each model had its strengths, with ensemble methods like Random Forest and advanced algorithms like XGBoost showing particularly high accuracy. However, accuracy alone does not tell the full story. The precision, recall, and F1-scores of the models also played an essential role in understanding their performance, especially in a domain where the cost of false positives and false negatives can be high. In conclusion, our study contributes valuable insights into the application of machine learning in the detection of credit card fraud. We found that while more complex models like Random Forest and XGBoost offered high accuracy, simpler models such as Logistic Regression provided a strong baseline and could be preferable due to their interpretability and speed. The findings from this research can inform financial institutions and stakeholders in developing fraud detection systems. We advocate for a balanced approach, considering both the performance of models and their operational efficiency.

*Future Work*

* + Future research may explore the integration of these models into real-world systems, continuous learning from incoming data, and the deployment of models in active monitoring systems. Investigate advanced feature engineering techniques to extract more relevant information from the data and improve model performance. Explore anomaly detection methods to identify new and evolving fraud patterns that traditional classification models may not capture. Implement techniques for explaining the decisions made by complex models like neural networks to enhance transparency and trust in the fraud detection process. Continuously refining and fine-tuning existing machine learning models to improve their accuracy, efficiency, and scalability. This could involve exploring novel algorithms, optimizing hyperparameters, and leveraging advanced techniques such as deep learning.

Our research illustrates the potential of machine learning in combating credit card fraud and underscores the importance of continual innovation in the face of evolving fraud tactics.

##### References

1. R. Sailusha, V. Gnaneswar, R. Ramesh and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2020, pp. 1264-1270, doi: 10.1109/ICICCS48265.2020.9121114.
2. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," in *IEEE Access*, vol. 10, pp. 39700-39715, 2022, doi: 10.1109/ACCESS.2022.3166891.
3. V. Filippov, L. Mukhanov and B. Shchukin, "Credit card fraud detection system," *2008 7th IEEE International Conference on Cybernetic Intelligent Systems*, London, UK, 2008, pp. 1-6, doi: 10.1109/UKRICIS.2008.4798919.
4. S. Bonkoungou, N. R. Roy, N. H. A. -E. Ako and U. Batra, "Credit Card Fraud Detection using ML: A Survey," *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bengaluru, India, 2023, pp. 732-738, doi: 10.1109/IITCEE57236.2023.10091035.
5. I. Vejalla, S. P. Battula, K. Kalluri and H. K. Kalluri, "Credit Card Fraud Detection Using Machine Learning Techniques," *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)*, Nagpur, India, 2023, pp. 1-4, doi: 10.1109/PCEMS58491.2023.10136040.
6. S. K. S, K. K. Shah, K. Kumar, K. K. Patel and A. R. Sah, "Credit Card Fraud Detection Using Machine Learning Model," *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, Mysuru, India, 2022, pp. 1-7, doi: 10.1109/MysuruCon55714.2022.9972647.
7. A. S. Rathore, A. Kumar, D. Tomar, V. Goyal, K. Sarda and D. Vij, "Credit Card Fraud Detection using Machine Learning," *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, MORADABAD, India, 2021, pp. 167-171, doi: 10.1109/SMART52563.2021.9676262.
8. J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," *2017 International Conference on Computing Networking and Informatics (ICCNI)*, Lagos, Nigeria, 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782.
9. S. Negi, S. K. Das and R. Bodh, "Credit Card Fraud Detection using Deep and Machine Learning," *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2022, pp. 455-461, doi: 10.1109/ICAAIC53929.2022.9792941.
10. T. R. Pillai, I. A. T. Hashem, S. N. Brohi, S. Kaur and M. Marjani, "Credit Card Fraud Detection Using Deep Learning Technique," *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, Subang Jaya, Malaysia, 2018, pp. 1-6, doi: 10.1109/ICACCAF.2018.8776797.