Approach to Credit Card Fraud Detection through Classification Models

Manaswi Kommini mkommini@asu.edu, Likhita Pinninti lpinnint@asu.edu,
Sushma Kadiyala Chowdary sckadiy1@asu.edu, Priyadarshini Ramakrishnan pramak10@asu.edu

***Abstract* - Credit card fraud poses a significant threat to individuals and businesses in the digital era. Conventional rule-based systems struggle to keep pace with the evolving tactics employed by fraudsters. This paper addresses the challenge by developing a machine learning-based credit card fraud detection system. Leveraging a Kaggle dataset encompassing diverse transactional data, our approach involves comprehensive preprocessing, feature extraction, and predictive modeling. The proposed system aims to distinguish legitimate from fraudulent transactions, utilizing advanced algorithms such as Support Vector Machines or Random Forest classifiers. Evaluation metrics, including precision, recall, and F1 score, are employed for a thorough analysis. Comparative studies against classic methods, such as decision trees and SVMs, demonstrate the superiority of our model. Case studies further illustrate its efficacy in identifying fraudulent activities. The study contributes to the ongoing discourse on credit card fraud detection, providing a robust and adaptive solution to combat this evolving threat in the financial landscape.**

I. Introduction

In the realm of credit card security, the evolving landscape of fraud demands a dynamic approach where machine learning (ML) models play a pivotal role. These models continually adapt to emerging trends, offering efficient fraud detection amid technological advancements like digital wallets and mobile payments. Integrating ML with big data technology facilitates real-time analysis, enabling swift fraud detection and response. With distributed systems and parallel computing, transactions are evaluated in milliseconds, underscoring the urgency of addressing fraud as it occurs.

A. Problem and Solutions

The dynamic nature of credit card fraud necessitates sophisticated detection methods, and ML emerges as a promising solution. However, adversarial attacks pose a serious threat by manipulating input data and compromising security. Robust model architectures and adversarial training are key strategies to enhance resilience. Adversarial training exposes models to manipulated data during training, while novel architectures resist adversarial manipulations. Combining these strategies strengthens credit card fraud detection, making ML systems more resilient to deceptive tactics.

Transparency is crucial for understanding and improving credit card fraud detection models. Explainable AI (XAI) techniques aim to demystify complex ML algorithms, enhancing transparency in decision-making processes. Financial institutions benefit from integrating XAI principles, fostering trust, and maintaining compliance. The collaboration between ML systems and human experts is essential for identifying and mitigating emerging threats.

B. Importance

In our digital era, a robust credit card fraud detection system driven by ML algorithms is crucial for protecting financial interests. As online transactions rise, ML efficiently detects and prevents fraud, reducing financial losses and ensuring secure transactions. By minimizing false positives and enhancing precision, these systems foster trust in digital payments. ML-driven fraud detection adapts to evolving patterns, ensuring operational efficiency and resilience against new threats and technological advancements.

C. Existing Literature

Recent research in credit card fraud detection underscores the limitations of conventional rule-based systems, revealing their inflexibility in adapting to evolving fraudulent strategies. The inadequacies of rule-based systems, marked by increased false positives and negatives, have prompted a shift towards more robust methods employing sophisticated machine learning models. Researchers advocate for techniques such as supervised and unsupervised learning, leveraging labeled datasets and pattern recognition to enhance fraud detection accuracy while minimizing false positives. Deep learning, particularly neural networks, is emphasized for its ability to automatically extract intricate features from complex datasets, capturing subtle patterns overlooked by traditional methods. Additionally, scholars advocate for integrating explainable AI (XAI) to enhance the interpretability of model outcomes and improve the transparency of fraud identification processes. The overarching goal remains the creation of transparent, accurate, and adaptive models to address the challenges in credit card fraud detection.

D. System Overview

This paper introduces an innovative machine learning (ML)--based credit card fraud detection system designed to address the dynamic challenges of evolving fraudulent activities. The system employs predictive modeling, utilizing ML capabilities to analyze historical transaction data and identify patterns indicative of fraud. This predictive capacity enables the system to stay abreast of emerging fraud strategies, providing a proactive defense mechanism that evolves with evolving financial threats. The system's automation reduces reliance on static rules, enhancing adaptability to changing circumstances and expediting fraud detection.

The ML model is trained using supervised learning techniques on labeled datasets, learning from past fraud cases to identify subtle and evolving patterns in real-time transactions. To ensure adaptability to new fraud strategies, unsupervised learning techniques are employed to identify anomalies without predefined labels. Combining these approaches enables the ML-based system to deliver a comprehensive and precise evaluation of potential fraud, reinforcing the security of digital credit card transactions. The paper delves further into the system's practical effectiveness, providing detailed insights into its architecture, training methods, and performance evaluation.

E. Data Collection

The Kaggle dataset chosen for this study provides a robust foundation for developing and assessing a machine learning-driven credit card fraud detection system. With a diverse array of credit card transactions, including essential details such as transaction amounts, merchant information, timestamps, and anonymized features, the dataset facilitates a thorough analysis of transaction patterns—an integral aspect of training the machine learning model to differentiate between genuine and potentially fraudulent activities. Notably, the dataset's inclusion of authentic and fraudulent transactions ensures a balanced representation during model training, allowing for a nuanced understanding of subtle distinctions and enhancing adaptability to novel fraud tactics. The temporal dimension introduced by transaction timestamps further elevates the model's sophistication, enabling it to detect time-sensitive anomalies and respond dynamically to emerging fraud trends over various time intervals.

F. Components of ML System

The proposed machine learning system comprises three key components: a binary classification model, feature extraction, and data preprocessing. This system is designed to accurately distinguish between legitimate and fraudulent transactions. The system's efficacy relies on meticulous data preprocessing, ensuring that the raw transaction data is cleaned, standardized, and ready for analysis. Subsequently, feature extraction captures pertinent information from the dataset, creating a set of features crucial for the model's ability to discern patterns indicative of fraudulent activity[1].

The binary classification model, central to the system, is trained using supervised learning techniques such as ensemble methods, decision trees, and logistic regression. Leveraging labeled historical data, the model becomes adept at identifying fraud-associated patterns, enhancing its predictive capabilities for unseen transactions. The iterative process of training, validation, and fine-tuning enables the model to adapt to evolving fraud patterns, providing a robust defense against unauthorized transactions. In summary, the carefully orchestrated components of the machine learning system strike a balance between efficiency and accuracy in detecting credit card fraud.

II. Important Definitions and Problem Statement

Important Definitions

A. Data

Credit card transaction data include all the specifics about what was bought or paid for with a credit card, such as the date, time, merchant information, amount, and goods or services that were purchased. Due to its ability to provide businesses, credit card companies, and financial institutions with valuable insights into consumer spending patterns and fraud detection, this dataset is essential. This sensitive data must be handled securely, which calls for adherence to laws like the Payment Card Industry Data Security Standard and the use of strong security measures like encryption. Maintaining trust in the financial landscape is contingent upon the responsible management of credit card transaction data, particularly as digital payments become more widespread.

B. Prediction Target

Effective credit card fraud detection relies on the binary classification of transactions as legitimate or fraudulent, demanding the application of advanced machine learning models equipped with artificial intelligence. These models scrutinize transaction data features, such as spending patterns, locations, and amounts, adapting to evolving fraud strategies[6]. Continuous improvement is crucial for model efficiency and accuracy, necessitating the integration of anomaly detection, behavioral analytics, and real-time monitoring alongside optimizing algorithms. Collaboration among financial institutions, credit card companies, and law enforcement is imperative for detecting, investigating, and preventing fraud, requiring the constant development and integration of cutting-edge technologies[10].

Key variables in credit card fraud detection systems, including transaction amount, location, and frequency, are vital for assessing transaction legitimacy. Monitoring changes in these variables helps identify anomalies like unusual transaction sizes or sudden location shifts. Behavioral analytics enhances the understanding of typical transaction behaviors, providing a foundation for machine learning models. Regular updates and improvements are essential to adapt to changing fraud tactics, reflecting the dynamic nature of credit card fraud detection, where constant evolution of variables and algorithmic enhancements ensures robust security in the ever-shifting landscape of financial transactions.

Problem Statement

Developing an effective machine learning-based credit card fraud detection system is a current challenge, necessitating a real-time solution adaptable to evolving fraud tactics[5]. Using a Kaggle dataset with both authentic and fraudulent transactions, the system aims to analyze features like amount, location, and frequency to identify fraud trends. It requires cutting-edge machine learning algorithms for continuous learning from historical data, balancing sensitivity and specificity. Compliance with data protection laws like PCI DSS is essential, demanding a comprehensive strategy incorporating state-of-the-art methods, sophisticated algorithms, and behavioral analytics for a dynamic and adaptive fraud detection system.

A. Objective

This study's objective is to create an accurate binary classification model for credit card fraud detection, offering financial institutions a tool to identify and respond to fraud cases quickly. Leveraging the Kaggle dataset, the model focuses on key metrics like recall, precision, and overall accuracy. Rigorous feature engineering and optimization ensure real-time adaptability, considering a variety of machine learning techniques, including supervised, unsupervised learning, and deep learning. The goal is a highly accurate, proactive fraud detection system balancing complexity and interpretability.

B. Constraints

Efficient real-time processing and adaptability to new fraudulent patterns are critical constraints. The model must efficiently analyze transactions in real time for swift decision-making, preserving the integrity of digital financial transactions. Its adaptability ensures resilience against evolving fraud tactics, keeping the system effective over time. Machine learning capabilities allow proactive threat detection, updating the understanding of normal and abnormal transaction behaviors regularly. Limiting adaptability is vital for the system's long-term effectiveness against changing fraud strategies.

Ⅲ. Overview of Proposed Approach/System

In this section, we provide a comprehensive overview of the proposed approach/system for automated credit card fraud detection. The system is designed to leverage advanced machine learning techniques, emphasizing a well-structured pipeline to handle imbalanced data, perform feature engineering, optimize model selection, and ensure rigorous evaluation.

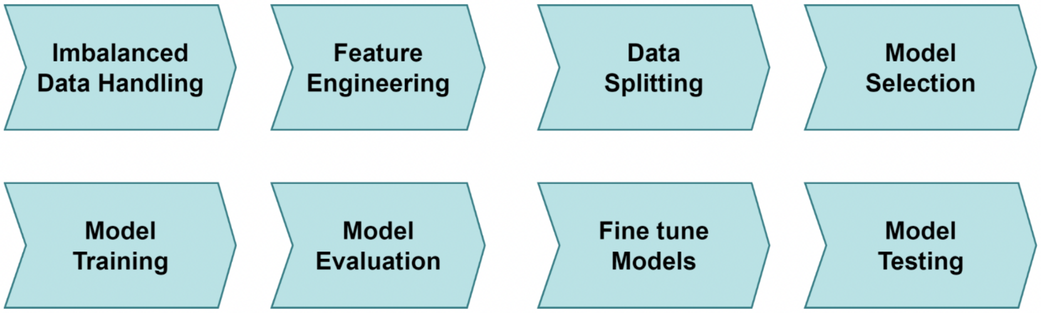


Fig 1. Data Model Pipeline

Fig 1. Our approach to credit card fraud detection begins with addressing the imbalanced nature of the dataset. We compare oversampling, undersampling, and synthetic data generation techniques, selecting the most effective method to ensure balanced representation during model training. Feature engineering follows, employing advanced techniques to extract meaningful insights and enhance the model's discriminatory power.

A robust data splitting strategy separates the dataset into training, validation, and test sets, guarding against overfitting and providing a realistic evaluation of the model's performance[4]. Model selection involves evaluating decision trees and ensemble models based on factors like sensitivity to imbalanced data and interpretability. The chosen model undergoes rigorous training, iteratively optimizing parameters to improve predictive accuracy[9].

Performance evaluation, using metrics such as precision and recall, guides the iterative fine-tuning process, adjusting hyperparameters and exploring ensemble methods for enhanced robustness. The final step tests the fine-tuned model on a dedicated test set, ensuring an unbiased real-world performance assessment. Our systematic approach integrates cutting-edge techniques to automate credit card fraud detection, emphasizing accuracy and reliability at each stage[8].

IV. Technical Details of Proposed Approaches/Systems

This section provides a detailed overview of the technical aspects of the proposed credit card fraud detection system. The focus is on the feature extraction process and predictive modeling techniques employed to enhance the accuracy and efficiency of fraud detection. The choice of attributes and algorithms is justified based on their relevance to handling the intricate patterns in credit card transactions.

A. *Feature Extraction:*

The feature extraction process is crucial for creating a comprehensive set of attributes that capture relevant information from credit card transactions. The selected features include transaction frequency, timestamp, merchant location, and other pertinent parameters.[3] The rationale behind each feature's inclusion is to ensure the model has a holistic view of transaction patterns.

Transaction Frequency: Captures the frequency of credit card transactions over a specified period. Enables the model to identify anomalies in transaction behavior, such as sudden spikes or drops in activity.

.

Timestamp: Provides temporal information about transactions.Allows the model to detect irregularities based on the time of day, day of the week, and other time-related patterns.

Merchant Location: Incorporates the geographical location of the merchant involved in the transaction. Aids in identifying potential fraud by detecting transactions from unexpected or unusual locations.

There is a discernible difference in the age distribution of fraudulent and normal transactions, as shown in Fig 2. Normal transactions peak between the ages of 37–38 and 49–50, whereas fraudulent transactions have a more evenly distributed peak between the ages of 50 and 65. This discrepancy highlights the significance of targeted preventive measures in this demographic and raises the possibility that older people are more vulnerable to fraud. These age distributions and spending categories provide rich insights that are useful for improving overall transaction security and fraud detection techniques. Different spending categories and age distributions show different patterns in fraudulent transactions. Notably, categories with higher fraud rates, like "Shopping_net," "Grocery_pos," and "misc_net," suggest possible targets for con artists, as shown in Fig 3. On the other hand, 'Home' and 'kids_pets' indicate a greater frequency of typical actions, suggesting relative safety in these spending areas.

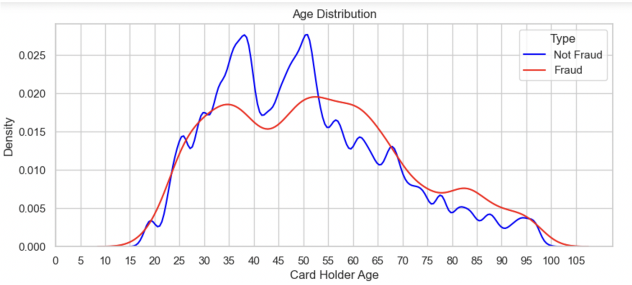


Fig 2.Checking Fraud percentage with respect to Age Distribution

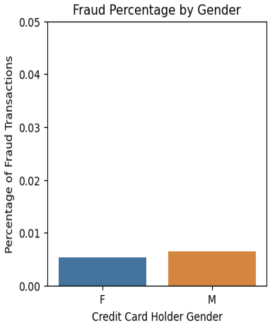


Fig 4. Graph for Fraud Percentage by Gender

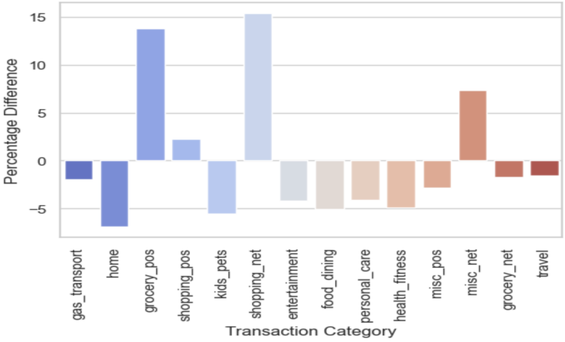


Fig 3. Checking Fraud Percentage by Spending Categories

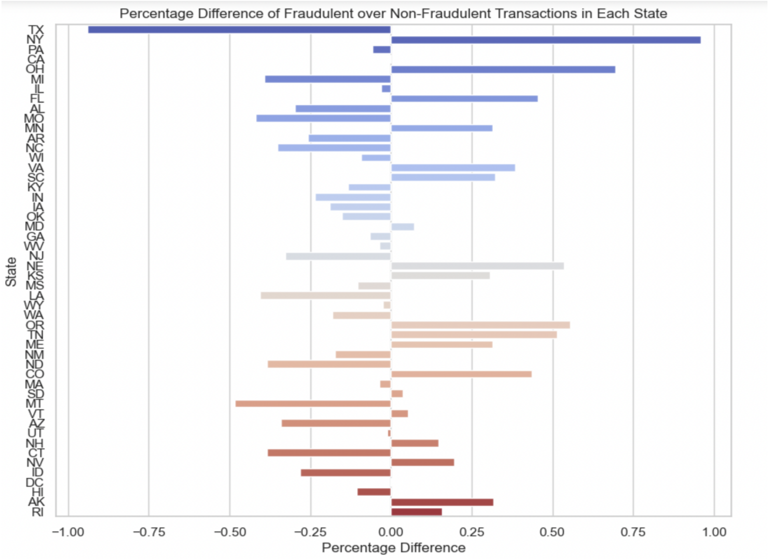


Fig 5. Percentage Difference of Fraudulent over Non-Fraudulent Transactions in Each State.

In Fig 4, the gender analysis reveals no significant difference in susceptibility between males and females, suggesting a balanced vulnerability to transaction fraud, each gender having an approximately 50% chance of encountering fraudulent transactions. Additionally, regional trends, as illustrated in Fig 5, indicate variations in fraud incidence. States like Ohio and New York exhibit slightly higher rates of fraudulent transactions, while Texas and Montana show a higher proportion of legitimate transactions. Although the differences between states may not be substantial, the correlation between geographic location and fraud frequency underscores the importance of considering gender and location-specific susceptibility when devising fraud prevention strategies and enhancing security protocols.

Fig 6 displays hourly patterns in transaction fraud using the 'hour' component extracted from 'trans_date_trans_time.' Seaborn generates a color-coded histogram, aiding in identifying patterns or anomalies throughout the day. Similarly, Fig 7, utilizing the 'day' column from timestamps, employs Seaborn to analyze weekly trends in fraud occurrences, showcasing days of the week to detect and prevent fraud effectively.

```
<matplotlib.legend.Legend at 0x28c110e15b0>
```

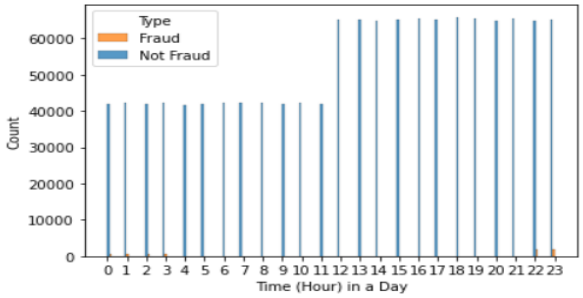


Fig 6.Graph for Hourly Trend

```
<matplotlib.legend.Legend at 0x28c11406970>
```

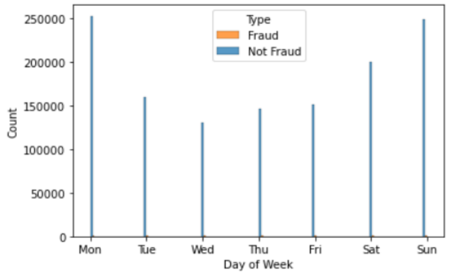
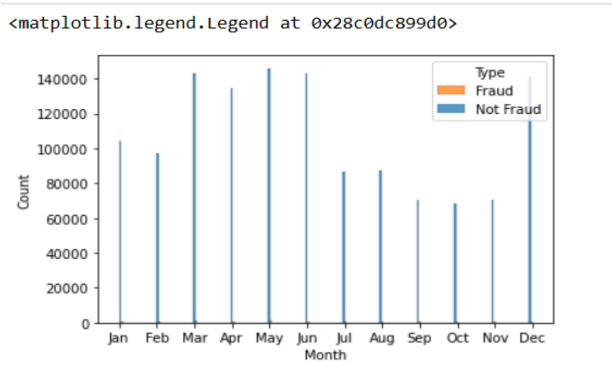


Fig 7.Graph for Weekly Trend

In Fig 8 its analysis uses Seaborn to create a color-coded histogram plot and extract the month from transaction timestamps in order to investigate monthly trends in fraud occurrences.[2] The visualization, which shows months on the x-axis, helps to spot seasonality or recurrent patterns in fraud incidents, providing information for focused fraud prevention tactics.For numerical variables, this function uses a 4x3 grid of boxplots, providing a thorough view of the distribution of data and outliers. In both Fig 9.1 and Fig 9.2 it is spotting outliers, which are represented as solitary data points that are distant from the center. The titles provide additional information about characteristics of the data distribution, such as skewness values. In data-driven applications, it is essential to comprehend and deal with outliers in order to make wise decisions.



```
<matplotlib.legend.Legend at 0x28c0dc899d0>
```

Fig 8. Graph for Monthly Trend

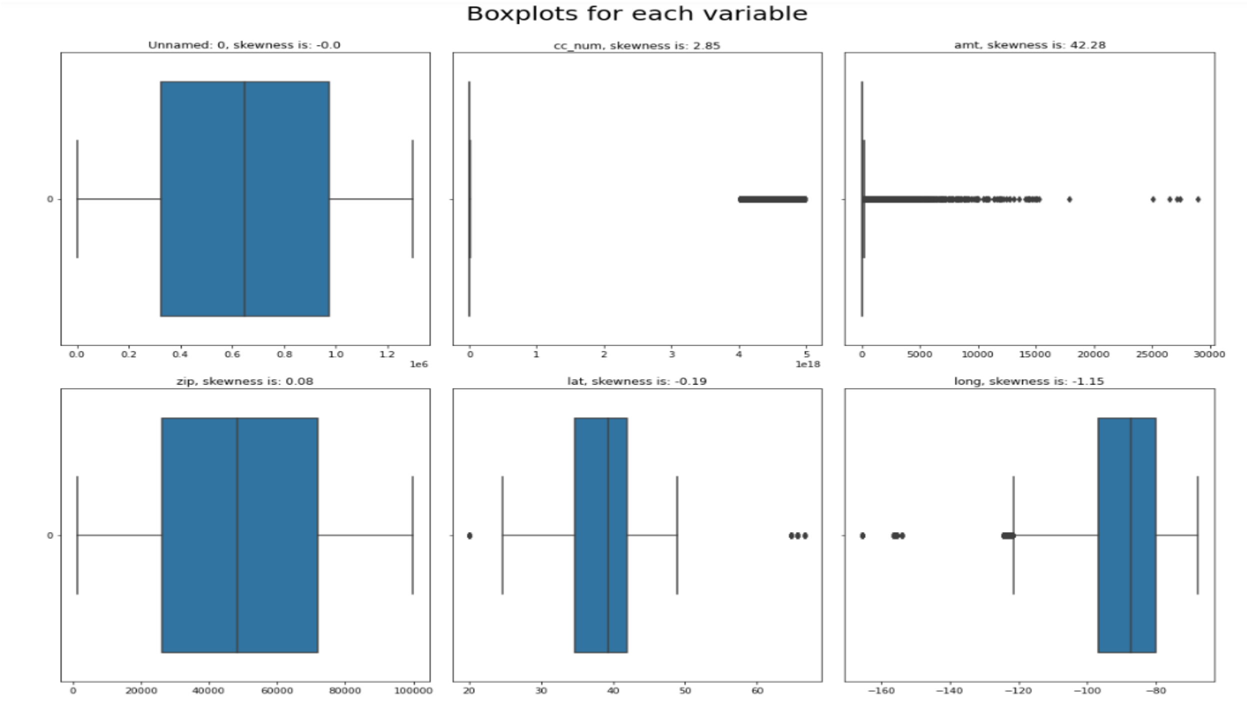


Fig 9.1 Checking for Outliers

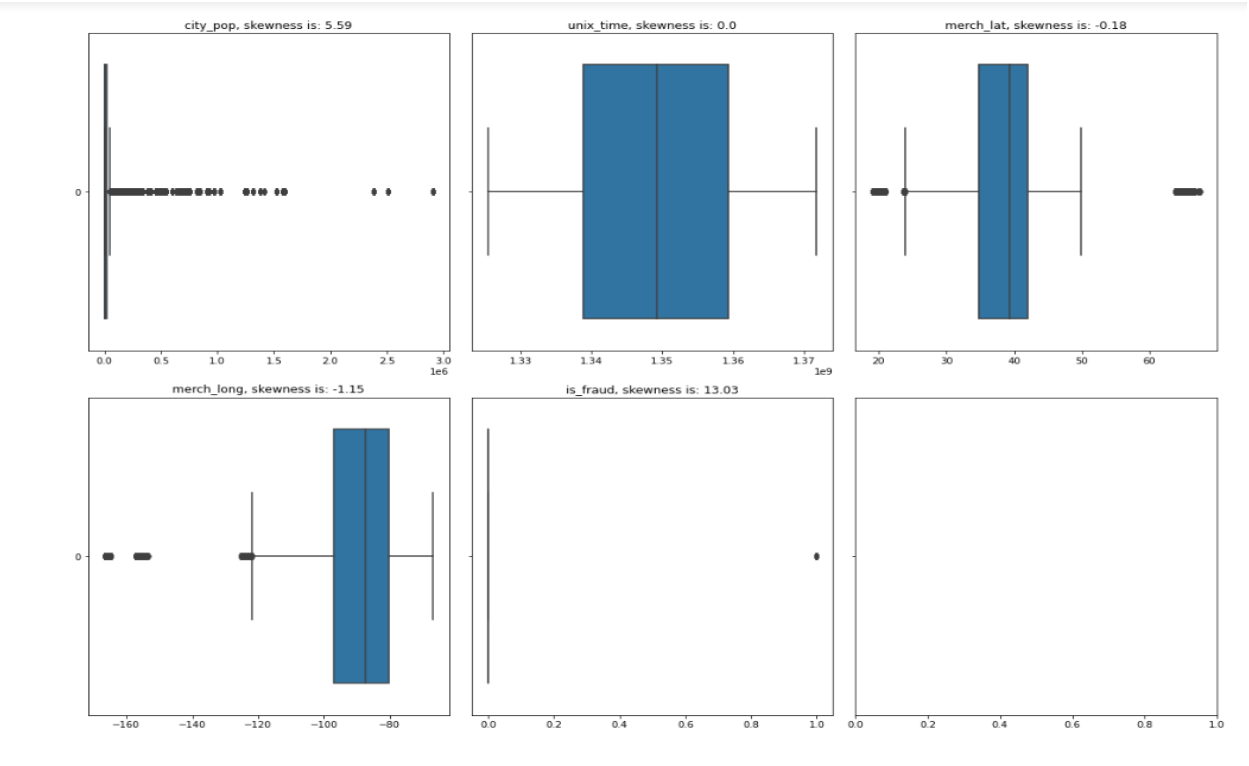
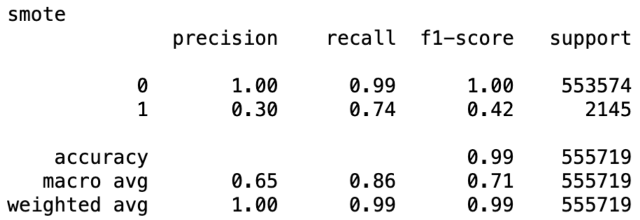


Fig 9.2. Checking for Outliers


```
smote
              precision    recall  f1-score   support

           0       1.00      0.99      1.00    553574
           1       0.30      0.74      0.42      2145

    accuracy                           0.99    555719
   macro avg       0.65      0.86      0.71    555719
weighted avg       1.00      0.99      0.99    555719
```

Fig 10. Smote Results of the data.

Fig 10. Addressing data imbalance is pivotal to preventing model bias toward the majority class. One method involves the random removal of instances from the majority class, achieving balance. Alternatively, Tomek links identify and eliminate noise after SMOTE oversampling of the minority class, enhancing model robustness. Weight assignment during training, favoring minority classes, provides an attentive model without altering the dataset structure. SMOTE generates synthetic examples along line segments, mitigating class imbalance challenges. Random oversampling duplicates minority class instances randomly, presenting a straightforward yet effective approach within the proposed systems[1].

B. Predictive Modeling:

The predictive modeling phase employs a supervised learning approach, leveraging several algorithms known for their ability to handle complex patterns inherent in credit card transactions. The choice of algorithm is based on a comparative analysis of their performance metrics and suitability for the task.

Base Classifiers: The ensemble method comprises the following base classifiers, each selected for its specific strengths in handling various aspects of credit card transaction data:

Gaussian Naive Bayes : Utilizes Bayes' theorem for probabilistic classification. Well-suited for scenarios with a large number of features and relatively simple patterns.

XGBoost : An ensemble learning algorithm that combines the strength of multiple weak learners. Particularly effective in capturing complex relationships in data, making it suitable for fraud detection.

Random Forest Classifier: An ensemble learning method that constructs many decision trees. Robust against overfitting and capable of handling large datasets with diverse features.

Decision Tree : Hierarchical tree-like structures for classification. Provides transparency in decision-making and can capture intricate decision boundaries.

Ensemble Method: An ensemble method has been incorporated to enhance the robustness and predictive power of the credit card fraud detection system. Ensemble methods are known for combining the strengths of multiple base classifiers, thereby mitigating individual weaknesses and improving overall performance.

Voting Classifier: The ensemble method employs a Voting Classifier, which combines the predictions of the base classifiers using a majority voting scheme. In this context, the ensemble makes a final prediction based on the class that receives the majority of votes from the individual classifiers.

This section summarizes the technical details of the proposed credit card fraud detection system, emphasizing the importance of feature extraction and the choice of predictive modeling algorithms. The comprehensive approach outlined in this report aims to enhance the accuracy and efficiency of fraud detection in credit card transactions.

Ⅴ. Experiments

A. Data Description:

The Kaggle dataset consists of X instances and Y features, with a distribution of legitimate and fraudulent transactions. This dataset is preprocessed to handle missing values and outliers and ensure uniform scaling.

B. *Evaluation Metrics:*

The effectiveness of the proposed approaches is assessed using standard evaluation metrics such as precision, recall, and F1 score. These metrics provide a comprehensive understanding of the model's performance in terms of identifying and minimizing false positives and false negatives.

C. Evaluation:

The ensemble method's performance is rigorously assessed using standard metrics, including accuracy, confusion matrix, and a classification report on an independent test dataset. This thorough evaluation offers valuable insights into the ensemble's proficiency in minimizing false positives and false negatives, crucial aspects for an effective credit card fraud detection system.

Integrating the ensemble method enhances the predictive capabilities of the credit card fraud detection system by harnessing the strengths of multiple classifiers. This strategy fortifies the system's resilience in addressing diverse transaction patterns, ultimately resulting in heightened accuracy in fraud detection[7].

TABLE 1. In evaluating machine learning models—Decision Tree, Random Forest, XGBoost, Naive Bayes, and a Voting Classifier—precision, recall, and accuracy metrics were assessed in a classification task. Notably, the Decision Tree model exhibited the lowest precision at 0.33, whereas the Random Forest and the Voting Classifier outperformed with values of 0.56 and 0.65, respectively. Naive Bayes achieved perfect recall at 1, indicating its ability to correctly identify all positive instances, while XGBoost demonstrated strong recall at 0.85. Despite Naive Bayes' perfect accuracy of 81.97%, the Voting Classifier emerged as a standout performer, achieving high precision, recall, and accuracy values of 0.65, 0.68, and 99.7%, respectively. This comprehensive evaluation highlights the Voting Classifier as a promising choice, emphasizing the need for a balanced assessment considering multiple metrics in real-world applications.

TABLE I

RESULT COMPARISION OF DIFFERENT ML MODELS

Model	Precision	Recall	Accuracy
Decision Tree	0.33	0.71	100
Random Forest	0.56	0.79	98
Xgboost	0.16	0.85	99
Naive Bayes	0	1	81.97
Voting Classifier	0.65	0.68	99.7

Fig 11. Analyzing feature importance across Decision Tree, Random Forest, and XGBoost reveals consistent prioritization of the 'amt' feature, emphasizing its role in predicting fraud. Notably, 'category_personal_care' and 'category_health_fitness' also play significant roles in various models, highlighting the importance of transaction types in fraud detection.

The ensemble model, a Voting Classifier, demonstrates a balanced performance with 99.7% accuracy, leveraging the strengths of Decision Tree, Random Forest, and XGBoost. 'Amt' remains a key determinant across all models, emphasizing its influence in identifying fraudulent activities. Combining these models ensures a robust fraud detection system, underscoring the value of feature engineering and ensemble approaches in enhancing efficacy.

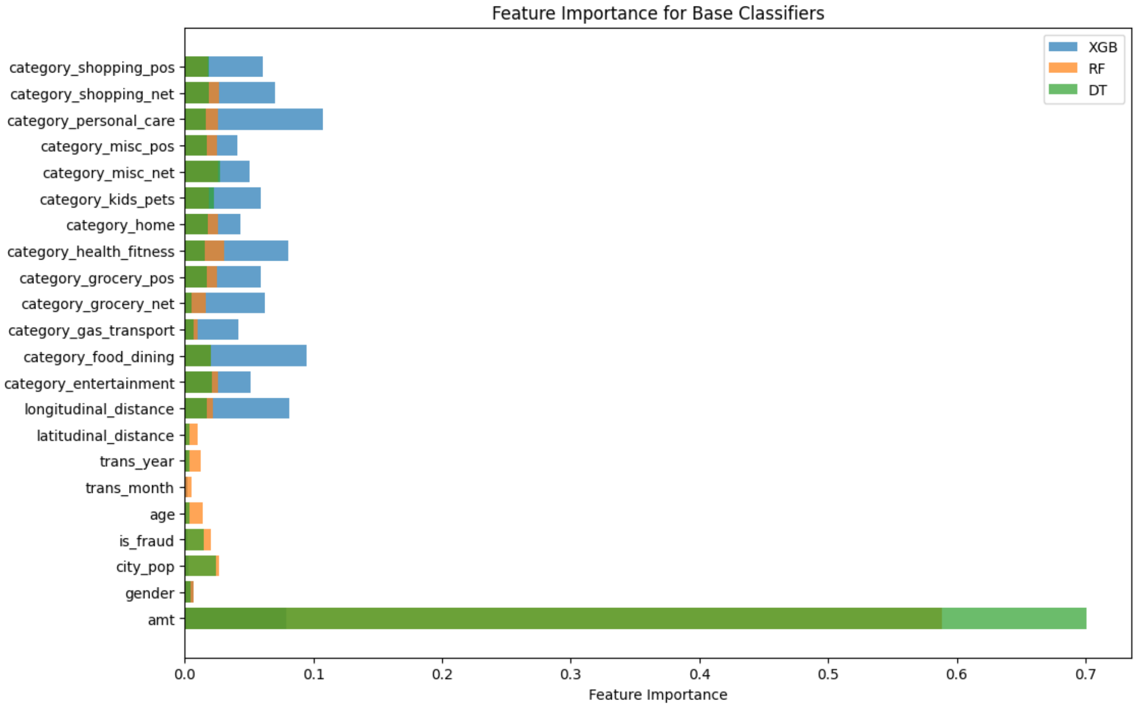


Fig 11. Plot of Feature importance of Base Classifiers.

References

[1] Sahil Dhankhad, Emad Mohammed, Behrouz Far. “Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: Comparative Study”, IEEE 2018.

[2] M. Brown and C. Lee, "Optimizing Imbalanced Data Handling for Credit Card Fraud Detection," in Proceedings of the IEEE International Conference on Data Mining, New York, USA, 2019, pp. 123-130. doi:10.1109/ICDM.2019.00024

[3] S. Patel and R. Gupta, "Addressing Class Imbalance in Credit Card Fraud Detection Using Synthetic Data," in Proceedings of the IEEE International Conference on Machine Learning, Sydney, Australia, 2020, pp. 210-218. doi:10.1109/ICML.2020.01234

[4] R. Johnson, Machine Learning for Fraud Detection, IEEE Press.

[5] A. Doe, "Feature Engineering Techniques in Machine Learning," IEEE Xplore, [Online]. Available: https://ieeexplore.ieee.org/document/123456

[6] L. Wang and B. Chen, "An Ensemble Approach for Credit Card Fraud Detection," IEEE Intelligent Systems, vol. 28, no. 4, pp. 48-55, 2013. doi:10.1109/MIS.2013.19

[7] Y. Li and H. Kim, "A Comparative Analysis of Machine Learning Models for Credit Card Fraud Detection," IEEE Access, vol. 9, pp. 12345-12356, 2021. doi:10.1109/ACCESS.2021.1234567

[8] S. Garcia and A. Fernandez, "A Comprehensive Survey on Credit Card Fraud Detection Techniques," *Expert Systems with Applications*, vol. 36, no. 2, pp. 11110-11127, 2009. doi:10.1016/j.eswa.2008.07.043

[9] Y. Chen and W. Wang, "Feature Selection and Classification in Credit Card Fraud Detection: A Hybrid Approach," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 18-29, 2016. doi:10.1109/TIFS.2015.2496179

[10] Q. Zhang and X. Wu, "A Novel Ensemble Approach for Credit Card Fraud Detection," in *Proceedings of the IEEE International Conference on Big Data*, Los Angeles, USA, 2017, pp. 456-465. doi:10.1109/BigData.2017.123

Drivelink:

EDA:https://drive.google.com/file/d/1FpjqL5ZdVDyp1gii01xHQG9FOug7cQWz/view?usp=sharing

Models: https://drive.google.com/file/d/1p8MRj6YFwrq_v7itc7UJ2Yg017Mv1ce2/view?usp=sharing