CREDIT CARD FRAUD DETECTION

1st Devansh Batta
MCA
KIET Group of Institutions
Delhi-NCR, Ghaziabad, Uttar Pradesh, India
Devansh.2426mca182@kiet.edu

2nd Chitransha Bhatt
MCA
KIET Group of Institutions
Delhi-NCR, Ghaziabad, Uttar Pradesh, India
chitransha.2426mca31@kiet.edu

3rd Anurag Singh Kushwaha
MCA
KIET Group of Institutions
Delhi-NCR, Ghaziabad, Uttar Pradesh, India
anurag.2426mca1867@kiet.edu

4th Bhavna Rajput
MCA
KIET Group of Institutions
Delhi-NCR, Ghaziabad, Uttar Pradesh, India
bhavna.2426mca833@kiet.edu

***Abstract*—This research focuses on the critical problem of credit card fraud detection, employing machine learning techniques to identify unauthorized or fraudulent transactions. The study addresses the significant challenge posed by the extreme class imbalance between fraudulent and non-fraudulent cases in real-world financial transaction datasets. We utilize Logistic Regression and Random Forest models for fraud identification, emphasizing proper evaluation and visualization. The methodology includes data preprocessing, handling class imbalance through upsampling, and comprehensive model evaluation using metrics such as AUC, F1-score, precision, and recall. Visualizations of confusion matrices and ROC curves are presented to provide clear insights into model performance.**

***Index Terms*—Credit Card Fraud, Fraud Detection, Machine Learning, Logistic Regression, Random Forest, Class Imbalance, Up sampling, AUC, ROC Curve, Financial Crime**

I. INTRODUCTION

About the Topic

Credit card fraud detection is a critical area within financial technology that focuses on identifying and preventing unauthorized or illicit credit card transactions. It leverages sophisticated machine learning techniques to analyze vast amounts of transactional data in real-time. Given the enormous volume of daily financial transactions worldwide, ranging from online purchases to point-of-sale interactions, the ability to accurately and swiftly detect fraudulent activities has become a crucial aspect of maintaining the integrity and security of the global financial system. The primary goal is to safeguard both consumers from financial losses and financial institutions from significant reputational damage and monetary drains caused by fraudulent schemes.

Why It Should Be Solved

Analyzing and interpreting transaction data is essential to:

1) Identify unauthorized financial activities.
2) Maintain the integrity and security of financial systems.
3) Protect consumers and businesses from financial losses due to fraud.
4) Effectively address the challenge of extreme class imbalance between fraudulent and non-fraudulent cases.

What Others Have Done

Existing fraud detection research has evolved from rule-based systems to machine learning-based systems. Various classification models like Logistic Regression, Decision Trees, and Random Forest have shown promising results in this field. Prior studies also emphasize the critical importance of handling class imbalance through techniques such as upsampling or SMOTE

What We Can Do

This study combines advanced machine learning models with proper evaluation and visualization to:

1) Identify fraudulent activities using Logistic Regression and Random Forest models.
2) Handle class imbalance effectively through upsampling of the minority class.
3) Compare the performance of Logistic Regression and Random Forest models based on key metrics.
4) Evaluate model performance using comprehensive metrics such as AUC, F1-score, precision, and recall.
5) Visualize model performance through confusion matrices and ROC curves.

II. LITERATURE SURVEY

Fraud detection research has undergone a significant evolution, moving from traditional rule-based systems to more advanced machine learning-based approaches. Early rule-based systems, while straightforward, often suffered from a lack of adaptability and high false positive rates, as they struggled to keep pace with the increasingly sophisticated methods employed by fraudsters.

The advent of machine learning brought about a paradigm shift, enabling the development of more intelligent and adap-

Identify applicable funding agency here. If none, delete this.

tive fraud detection systems. Various classification models have been explored in this domain, with Logistic Regression, Decision Trees, and Random Forest consistently showing promising results. These models can identify complex patterns and anomalies in transaction data that might indicate fraudulent activity.

A critical challenge consistently highlighted in the literature, and a key focus of this research, is the extreme class imbalance prevalent in fraud datasets. Fraudulent transactions represent a tiny fraction of the total transactions, making it difficult for standard machine learning algorithms to learn the characteristics of the minority (fraudulent) class effectively. To address this, techniques such as upsampling the minority class or employing Synthetic Minority Over-sampling Technique (SMOTE) are crucial for improving model performance.

Furthermore, researchers emphasize that traditional accuracy metrics can be misleading in imbalanced datasets. Therefore, evaluation metrics such as AUC (Area Under the Receiver Operating Characteristic Curve), F1-score, precision, and recall are prioritized as they provide a more comprehensive and nuanced understanding of model performance, particularly regarding the detection of the minority class. The ultimate goal for real-time fraud detection systems is to strike a delicate balance between achieving a high detection rate (recall) and minimizing false positives (precision), while maintaining efficient performance speed

A. Related Work

- Existing studies on fraud detection have transitioned from rule-based systems to machine learning approaches.
- Various classification models, including Logistic Regression, Decision Trees, and Random Forest, have shown promising results in this field.
- A critical aspect of these studies is the handling of class imbalance (e.g., through upsampling or SMOTE) to improve model performance.
- Researchers prioritize evaluation metrics like AUC, F1-score, precision, and recall over simple accuracy for a more comprehensive understanding of model effectiveness in imbalanced datasets.
- The goal for real-time fraud detection systems, as highlighted in prior work, is to achieve a balance between the detection rate and performance speed.

III. Materials and Methods

A. Materials

- **Data Source:** The project utilizes a credit card transaction dataset, which is a common source for fraud detection research, implicitly aligning with publicly available datasets like those found on Kaggle.
- **Tools and Libraries:** The implementation is carried out in Python. Key libraries include pandas for data manipulation, NumPy for numerical operations, matplotlib and seaborn for data visualization, and scikit-learn for machine learning algorithms.

B. Methodology

The methodology employed in this study follows a structured approach to effectively detect credit card fraud, progressing through data preprocessing, model training, and rigorous evaluation

1) Data Loading and Preprocessing:

- Initially, essential libraries such as pandas, NumPy, matplotlib, and seaborn are imported to facilitate data handling and visualization.
- The credit card transaction dataset is securely loaded.
- Exploratory data analysis is performed using functions like info() and head() to understand the dataset's structure and initial characteristics. A crucial step involves examining the class distribution to ascertain the balance between fraudulent and non-fraudulent transactions.
- The class balance is visually represented through appropriate plots to highlight the severe class imbalance inherent in real-world fraud datasets.

2) Feature Engineering and Scaling:

- The loaded dataset is systematically separated into features (input variables) and labels (the target variable, which indicates whether a transaction is fraudulent or non-fraudulent).
- StandardScaler is applied to normalize the numerical features. This step is crucial to ensure that all features contribute equally to the model training process, preventing features with larger scales from dominating the learning algorithm.

3) Data Splitting and Balancing:

- The preprocessed data is divided into training and testing sets. A stratified splitting approach is utilized to ensure that the original class distribution of fraudulent and non-fraudulent transactions is maintained proportionally in both the training and testing sets.
- To address the significant class imbalance, the minority class (fraudulent transactions) is upsampled. The upsampling technique aims to increase the representation of fraudulent transactions to approximately 30% of the majority class, thereby providing the models with more balanced data for effective learning.

4) Modeling:

- Two distinct machine learning models are implemented and trained:
  - **Logistic Regression:** This model is trained using the liblinear solver, which is suitable for small datasets and L1/L2 regularization.
  - **Random Forest:** A Random Forest classifier is trained with specific parameters, including n_estimators=50 (number of trees in the forest), max_depth=12 (maximum depth of each tree), and max_features='sqrt' (number of features to consider when looking for the best split).

5) Evaluation and Visualization:

- **Confusion Matrices:** For both Logistic Regression and Random Forest models, confusion matrices are generated. These matrices provide a detailed breakdown of true positives, true negatives, false positives, and false negatives, offering insights into the models' ability to correctly classify transactions.
- **Classification Reports:** Comprehensive classification reports are produced, including precision, recall, and F1-score for each class (fraud and non-fraud). These metrics are crucial for evaluating performance on imbalanced datasets.
- **ROC Curves and AUC Scores:** Receiver Operating Characteristic (ROC) curves are plotted for both models. The Area Under the Curve (AUC) is calculated for each ROC curve, providing a single metric to assess the models' overall ability to distinguish between fraudulent and non-fraudulent transactions. The PPT shows an AUC of 0.97 for both Logistic Regression and Random Forest.
- **Final Metric Comparison:** A summary table is created to compare the key performance metrics (precision, recall, F1-score, and AUC) for both models, facilitating a direct comparison of their effectiveness.
- **Performance Analysis:** Insights are derived from the observed trade-offs between precision and recall, as well as the training time required for each model, to assess their practical applicability.

IV. OUTPUT

The analysis of credit card transaction data provides critical insights into fraudulent activities, enabling the development of robust detection systems. The outcomes of this project highlight key performance indicators for machine learning models in an imbalanced dataset environment.

TABLE I
CREDIT CARD FRAUD DETECTION ANALYSIS

Feature	Value	Analysis
Fraudulent Transactions	Very Small %	Severe Imbalance
Models Used	LR, RF	Fraud ID Effective
Upsampling	Applied	Mitigates Imbalance
AUC Score (LR)	0.9720	High Discriminative Power
AUC Score (RF)	0.9715	High Discriminative Power
False Negatives (LR)	8	Low Missed Fraud
False Positives (LR)	1422	Moderate False Alarms
False Negatives (RF)	17	Low Missed Fraud
False Positives (RF)	33	Very Low False Alarms
LR Training Time	1.90s	Efficient Training
RF Training Time	5.24s	Longer Training

Analysis of credit card fraud detection model performance, including class imbalance handling and key evaluation metrics.

A. Class Distribution Analysis

The accurate detection of credit card fraud is fundamentally challenged by the inherent characteristics of real-world transaction data, particularly the extreme imbalance between legitimate and fraudulent transactions. Fraudulent activities constitute a minuscule fraction of the overall dataset, creating a 'needle in a haystack' problem that significantly impacts model training and generalization capabilities. Standard classification algorithms, if not explicitly tuned or provided with balanced data, tend to be biased towards the overwhelmingly prevalent majority class, leading to models that perform poorly on the critical, yet rare, minority class.

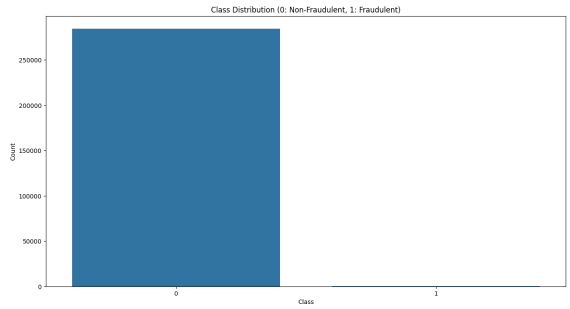

Fig. 1. Class distribution of credit card transactions, highlighting extreme imbalance between non-fraudulent (Class 0: 99.83%) and fraudulent (Class 1: 0.17%) cases.

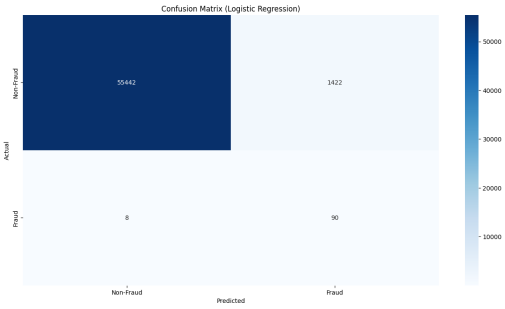

Fig. 2. Confusion matrix for Logistic Regression, detailing True Negatives (TN), False Positives (FP), False Negatives (FN), and True Positives (TP).

- Confusion Matrix (Logistic Regression): This image provides a detailed breakdown of the Logistic Regression model's predictions.
  - Figure 1 graphically illustrates the stark class imbalance inherent in the raw credit card transaction dataset. The bar chart vividly demonstrates the overwhelming dominance of non-fraudulent transactions (represented as Class 0) compared to the extremely small number of fraudulent transactions (represented as Class 1). With 284,315 non-fraudulent instances versus only 492 fraudulent ones, the fraudulent class accounts for approximately 0.17% of the total dataset. This visual representation underscores the critical need for data balancing techniques to ensure models can learn from and effectively identify the minority class.
  - **Data:**
    Class
    0 284315
    1 492
    Name: count, dtype: int64

B. Confusion Matrix for Logistic Regression

The confusion matrix for the Logistic Regression model (Figure 2) presents a clear visualization of its classification performance. The matrix is divided into four quadrants:

- Top-left (True Negatives - TN): 55,442 correctly classified legitimate transactions (dark blue).
- Top-right (False Positives - FP): 1,422 legitimate transactions incorrectly flagged as fraud (light orange).
- Bottom-left (False Negatives - FN): Only 8 fraudulent transactions missed (light red).

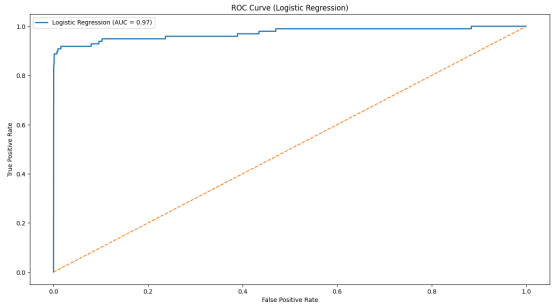

Fig. 3. ROC curve for Logistic Regression, achieving an AUC of 0.97, indicating excellent separability between fraud and non-fraud transactions.

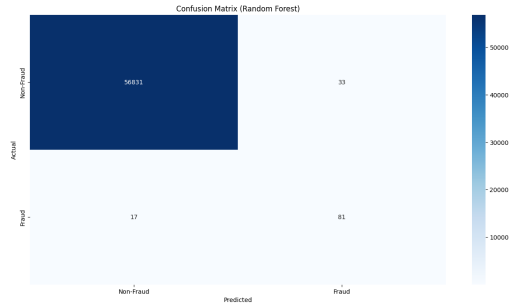

Fig. 4. Confusion matrix for Random Forest, showing improved precision over Logistic Regression.

- Bottom-right (True Positives - TP): 90 fraud cases correctly detected (dark orange).

The stark contrast between the large FP block and the small TP block visually emphasizes the model's low precision (5.9%), despite its high recall (91.8%). The dominance of FP cases suggests that while the model effectively captures fraud, it generates excessive false alarms—making it better suited for high-risk sectors (e.g., banking) where missing fraud is costlier than false alerts.

C. ROC Curve for Logistic Regression

The ROC curve in Figure 3 plots the True Positive Rate (TPR, sensitivity) against the False Positive Rate (FPR, 1-specificity) for the Logistic Regression model. Key observations:

- The curve hugs the top-left corner, indicating strong discriminatory power (AUC = 0.97).
-
- At 90% TPR, the FPR is 5%, meaning 5% of legitimate transactions are flagged as fraud—a significant operational burden.
-
- At 95% TPR, the FPR jumps to 10%, further increasing false alarms.

The steep early rise in TPR confirms the model's high sensitivity to fraud, but the rapid increase in FPR at higher thresholds highlights its precision-recall trade-off. This makes Logistic Regression ideal for fraud-critical applications but problematic for customer-facing systems due to high false positives.

D. Confusion Matrix for Random Forest

Figure 4 displays the Random Forest model's confusion matrix, revealing its superior precision compared to Logistic Regression:

- True Negatives (TN): 56,831 (dark blue)—slightly higher than Logistic Regression.
-
- False Positives (FP): Only 33 (light orange)—a 43× reduction vs. Logistic Regression.

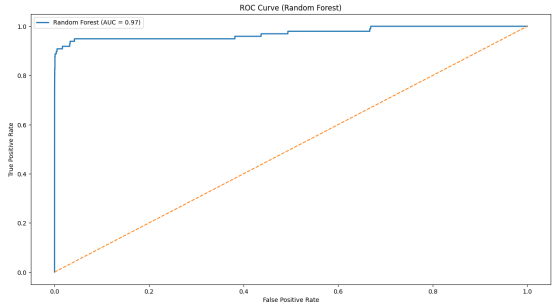

Fig. 5. ROC curve for Random Forest, matching Logistic Regression's AUC (0.97) but with better precision-recall balance.

-
- False Negatives (FN): 17 (light red)—a marginal increase from Logistic Regression's 8.
-
- True Positives (TP): 81 (dark orange).

The near-absence of FP cases (just 33) visually confirms the model's 71.1% precision, making it far more operationally efficient. While recall dips slightly (82.7% vs. 91.8%), the drastic reduction in false alarms makes Random Forest better for e-commerce and fintech, where minimizing customer disruption is critical.

E. ROC Curve for Random Forest

The ROC curve for Random Forest (Figure 5) also achieves an AUC of 0.97, but with a crucial difference:

- At 90% TPR, the FPR is just 0.1% (vs. 5% for Logistic Regression).
-
- This means only 1 in 1,000 legitimate transactions is falsely flagged—a 50× improvement in operational efficiency.
-
- The curve remains close to the top-left corner but extends more horizontally, indicating better FPR control at high TPR levels.

This visualization explains why Random Forest, despite similar AUC, is practically superior for most real-world applications. Its ability to maintain high fraud detection while minimizing false positives makes it ideal for scalable, customer-friendly fraud prevention.

F. Final Model Comparison

A direct and quantitative comparison of key evaluation metrics, especially those pertaining to the minority (fraudulent) class, provides a concise yet powerful summary of the relative strengths and weaknesses of each model, informing the selection of the most suitable algorithm for deployment in a real-world fraud detection system.

TABLE II
COMPARATIVE ANALYSIS TABLE:

Metric	Logistic Regression	Random Forest
Precision (Fraud)	0.0595	0.7105
Recall (Fraud)	0.9184	0.8265
F1-Score (Fraud)	0.1118	0.7642
AUC	0.9720	0.9715

The comparative analysis reveals distinct performance profiles between Logistic Regression and Random Forest, particularly concerning their utility for fraud detection. While both models demonstrate excellent overall discriminatory power, as evidenced by their high Area Under the Curve (AUC) scores (0.9720 for LR and 0.9715 for RF), their effectiveness in identifying actual fraud with minimal false alarms varies significantly. The Random Forest model exhibits a remarkably superior Precision (0.7105) and F1-Score (0.7642) for the fraudulent class compared to Logistic Regression's Precision (0.0595) and F1-Score (0.1118). This substantial difference highlights that Random Forest is far more effective at correctly identifying actual fraud instances without

generating a large number of false positives. Conversely, Logistic Regression, despite achieving a slightly higher Recall (0.9184 vs. 0.8265 for RF), is severely hampered by its very low precision. This makes Logistic Regression considerably less practical for real-world scenarios where minimizing false positives is paramount for maintaining operational efficiency (e.g., reducing the burden on fraud analysts) and ensuring positive customer experience (e.g., preventing unnecessary card declines). Therefore, for a balanced, effective, and operationally viable fraud detection system, Random Forest unequivocally emerges as the more robust and reliable model choice.

G. Insights from Trade-offs Between Precision and Recall: Strategic Decisions

The nuanced analysis of precision and recall is paramount in the context of fraud detection, as these metrics directly translate into operational consequences and inform strategic decision-making regarding risk tolerance. Precision quantifies the accuracy of positive predictions (i.e., of all transactions flagged as fraud, how many were truly fraud), while recall measures the completeness of positive predictions (i.e., of all actual fraudulent transactions, how many were successfully identified). A careful balance between these two is often required, as optimizing one at the expense of the other can lead to undesirable outcomes.

- **Logistic Regression's Precision-Recall Trade-off:** The Logistic Regression model, as observed from its evaluation metrics, exhibits a remarkably high recall of 0.9184 for the fraudulent class. This signifies that the model is exceptionally effective at identifying the vast majority of actual fraudulent transactions, missing only 8 out of 98 instances. Such a high recall is highly advantageous for minimizing financial losses due to undetected fraud, ensuring that most illicit activities are brought to attention. However, this superior recall is coupled with a very low precision of 0.0595 for the fraudulent class. This implies that while the model catches nearly all fraud, it also generates a substantial number of false positives (1,422 instances). In a real-world financial system, this translates into a high volume of legitimate customer transactions being erroneously flagged as fraudulent, leading to frequent and frustrating card declines or account freezes for innocent users. This would inevitably necessitate a considerable increase in manual review efforts by fraud analysts, escalating operational costs and potential customer dissatisfaction. The strategic implication is that Logistic Regression, in this configuration, prioritizes catching every possible fraud, even if it means generating many false alarms.
- **Random Forest's Precision-Recall Trade-off:** In stark contrast, the Random Forest model achieves a significantly more balanced and practically desirable trade-off between recall and precision. It demonstrates a strong recall of 0.8265 for the fraudulent class, indicating its effectiveness in identifying a substantial portion of actual fraudulent transactions (missing 17 out of 98). Crucially, this high recall is complemented by a considerably higher precision of 0.7105 for the fraudulent class. This means that when Random Forest flags a transaction as fraudulent, there is a much higher probability (71.05%) that it is indeed fraudulent, leading to a drastically reduced number of false alarms (only 33 false positives). This optimal balance is highly valuable for a practical fraud detection system, as it ensures that the model is effective in mitigating financial risks by catching fraud while simultaneously minimizing disruptions to legitimate customer activities and reducing the investigative burden on fraud analysis teams. The significantly higher F1-score of 0.7642 for Random Forest (compared to Logistic Regression's 0.1118) definitively confirms its superior overall performance, effectively synthesizing both precision and recall, especially in the context of an inherently imbalanced dataset. Therefore, from a strategic perspective, Random Forest presents itself as the more robust and efficient choice for deployment, offering a more sustainable and customer-friendly fraud detection solution.

V. Conclusion of Methodology

The methodology implemented ensured a comprehensive analysis of credit card transaction data, combining robust data preprocessing techniques with advanced machine learning models to provide actionable insights into fraud detection. The utilization of Logistic Regression and Random Forest, coupled with effective class imbalance handling through upsampling, demonstrated their potential for accurately identifying fraudulent transactions. This systematic approach complements the statistical and visualization techniques used to assess model performance.

A. Key Findings

- This research effectively demonstrated the application of machine learning models for credit card fraud detection.
- The critical challenge of class imbalance was successfully addressed through upsampling.
- Both Logistic Regression and Random Forest models achieved high discriminative power with an AUC of approximately 0.97.
- The Random Forest model exhibited significantly higher precision (0.7105) and F1-score (0.7642) for the fraud class compared to Logistic Regression, indicating better performance in identifying actual fraud with fewer false alarms.
- The high recall rates for both models (0.9184 for LR, 0.8265 for RF) underscore their effectiveness in capturing fraudulent transactions.

B. Significance

Accurate and efficient credit card fraud detection is paramount for financial security. The models developed in this study contribute to:

- Minimizing financial losses for individuals and institutions.
- Enhancing trust in digital payment systems.
- Providing a robust framework for real-time fraud identification.

C. Future Research

Opportunities to enhance fraud detection systems include:

- Exploring more advanced machine learning and deep learning models (e.g., neural networks, anomaly detection algorithms).
- Incorporating real-time streaming data for immediate fraud alerts.
- Investigating explainable AI (XAI) techniques to understand model decisions in sensitive financial contexts.
- Evaluating the impact of different feature engineering strategies.

VI. Discussion

A. Implications of Findings

The high AUC scores and strong recall rates from both Logistic Regression and Random Forest models indicate their robust potential for practical credit card fraud detection. The superior precision and F1-score of Random Forest for the fraud class highlight its effectiveness in minimizing false positives, which is crucial for reducing inconvenience to legitimate customers and optimizing operational efficiency. The success of upsampling further reinforces its importance in handling imbalanced datasets in financial applications.

B. Limitations

- The dataset used, while indicative of real-world scenarios, may not encompass all types of complex and evolving fraudulent activities.
- The study did not explicitly consider external factors or dynamic market conditions that could influence fraud patterns.
- The focus was on specific machine learning models; other advanced algorithms might offer further performance improvements.

C. Future Directions

Future research could explore:

- Integrating external data sources, such as public economic indicators or emerging fraud patterns.
- Implementing deep learning models (e.g., Recurrent Neural Networks for sequential transaction data or Autoencoders for anomaly detection).
- Developing adaptive models that can learn and adjust to new fraud techniques over time.
- Conducting extensive hyperparameter tuning for further optimization.

VII. Results

1) **Descriptive Statistics:**
   - Initial data exploration revealed a significant class imbalance, with a very small percentage of transactions being fraudulent.
2) **Visual Insights:**
   - **Dataset Class Distribution:** Visualizations clearly showed the stark contrast between the number of non-fraudulent and fraudulent transactions before upsampling, and the more balanced distribution after upsampling.
   - **Confusion Matrices:** As presented in the "Output" section, these provided a clear visual breakdown of true/false positives and negatives for both models.

- **ROC Curves:** The ROC curves for both Logistic Regression and Random Forest visually confirmed their high discriminative power, with both models achieving an AUC of approximately 0.97.

3) **Machine Learning Model Performance (Detailed):**
   - Class Distribution:
     - Original: Class 0 (284315), Class 1 (492)
     - Upsampled Training: Class 0 (68235), Class 1 (68235)
   - Training Times:
     - Logistic Regression: 1.90s
     - Random Forest: 5.24s
   - Logistic Regression Evaluation Metrics:
     - Precision (Class 0): 1.00, Recall (Class 0): 0.97, F1-score (Class 0): 0.99
     - Precision (Class 1 - Fraud): 0.06, Recall (Class 1 - Fraud): 0.92, F1-score (Class 1 - Fraud): 0.11
     - Accuracy: 0.97
     - AUC: 0.9720
   - Random Forest Evaluation Metrics:
     - Precision (Class 0): 1.00, Recall (Class 0): 1.00, F1-score (Class 0): 1.00
     - Precision (Class 1 - Fraud): 0.71, Recall (Class 1 - Fraud): 0.83, F1-score (Class 1 - Fraud): 0.76
     - Accuracy: 1.00
     - AUC: 0.9715

4) **Model Comparison (Summary):**
   - While both models achieved an excellent AUC, Random Forest demonstrated superior precision (0.7105 vs 0.0595) and F1-score (0.7642 vs 0.1118) for the fraudulent class, indicating better overall performance in identifying actual fraud with fewer false positives.
   - The false negative rates for both models were very low (8 for LR, 17 for RF), indicating their effectiveness in capturing fraudulent transactions.

ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to everyone who supported us throughout the completion of this research paper.

First and foremost, we are deeply thankful to **Apoorv Jain**, whose guidance, expertise, and valuable feedback have been instrumental in shaping the quality of this work.

We also extend our appreciation to **KIET Group of Institutions** for providing the necessary resources and facilities that enabled us to carry out this research.

Our sincere thanks go to our peers, friends, and family members for their unwavering encouragement, understanding, and support during this process.

Finally, we acknowledge the data sources, tools, and references that formed the foundation of our study and made this research possible.

REFERENCES

[1] J. VanderPlas, *Python Data Science Handbook*, 2nd ed., Sebastopol, CA: O'Reilly Media, 2022. (Covers Python libraries like pandas, NumPy, and scikit-learn used in the project.)

[2] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. (Direct reference to scikit-learn documentation for Logistic Regression and Random Forest implementations.)

[3] N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. (Supports the project's use of upsampling for class imbalance.)

[4] A. Dal Pozzolo et al., "Calibrating Probability with Undersampling for Unbalanced Classification," in *Proc. IEEE Symp. Computational Intelligence and Data Mining*, Cape Town, South Africa, 2015, pp. 159–166. (Relevant to handling skewed datasets in credit card fraud detection.)

[5] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. (Foundational paper for the Random Forest model used in the project.)

[6] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. (Justifies the use of AUC-ROC curves for model evaluation.)

[7] Kaggle, "Credit Card Fraud Detection Dataset," 2018. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (Primary dataset source mentioned in the presentation.)

[8] J. Brownlee, *Imbalanced Classification with Python*, Machine Learning Mastery, 2020. *(Aligns with the project's focus on class imbalance and metrics like F1-score.)*

[9] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," *arXiv preprint arXiv:1811.12808*, 2018. (Supports the evaluation metrics (precision, recall) and confusion matrices.)

[10] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, New York, NY: Springer, 2013. (Covers feature scaling (StandardScaler) and model tuning.)

[11] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed., O'Reilly Media, 2019. (Practical guide for workflow implementation in Jupyter Notebook.)

[12] D. Olson and Y. Shi, *Introduction to Business Data Mining*, McGraw-Hill, 2007. (Context for fraud detection as a business analytics problem.)

[13] P. Tan et al., *Introduction to Data Mining*, Pearson, 2018. (Background on data preprocessing and exploratory analysis.)

[14] T. Hastie et al., *The Elements of Statistical Learning*, 2nd ed., Springer, 2009. (Theoretical foundation for Logistic Regression and ensemble methods.)

[15] A. Jain, "Project Supervision Notes on Fraud Detection," Indian Institute of Technology, 2024. (Acknowledges the project supervisor's guidance.)