International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 3, May-June 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

IntelliFraud: Next-Gen ML Solutions for Credit Card Fraud Detection

Assoc. Prof. Dr. Rasmi A[1], Enankadhara S[2], Vishal V[3], Darshan Gowda B S[4], Darshan K[5]
Dept. of ISE, R R Institute of Technology, Bangalore, Karnataka, India

ABSTRACT

Credit card fraud has become one of the most prevalent threats in the digital transaction ecosystem. As digital banking grows, so do the complexities of fraud, making traditional detection mechanisms obsolete. IntelliFraud is a machine learning-powered framework aimed at identifying fraudulent transactions in real time while reducing false positives and enabling adaptive learning. This paper analyzes various ML algorithms combined with data sampling techniques to handle highly imbalanced datasets—a common challenge in fraud detection. The project evaluates the strengths and weaknesses of models like Decision Trees, Random Forests, Logistic Regression, KNN, and XGBoost, and benchmarks them using resampling strategies like SMOTE, Tomek Links, and Cluster Centroids. This comprehensive evaluation demonstrates that combining XGBoost with SMOTE yields the highest accuracy, making IntelliFraud a viable, efficient solution for modern banking fraud detection systems.

I. INTRODUCTION

In today's digital-first world, the rise of online transactions has led to a surge in financial fraud, with global credit card fraud losses surpassing $40 billion in 2023. Traditional rule-based fraud detection systems are proving inadequate—they are rigid, reactive, and prone to high false positive rates, which frustrate users and reduce system efficiency. To overcome these limitations, IntelliFraud introduces a modern fraud detection framework that combines layered rule-based filters with adaptive machine learning (ML) models. These models, both supervised and unsupervised, learn from historical data, uncover hidden patterns, and evolve through continuous updates based on real-time data and investigator feedback. A major challenge in fraud detection is the severe class imbalance, where fraudulent transactions represent less than 0.2% of all data. IntelliFraud addresses this using advanced resampling techniques such as SMOTE and Tomek Links to ensure balanced and effective model training. The framework evaluates various ML algorithms including Logistic Regression, Random Forest, and XGBoost to identify the most robust solution. With real-time detection, minimal operational delays, and cloud-based visualization tools for analyst interpretation, IntelliFraud offers a scalable, intelligent solution to secure financial systems and restore consumer trust in an increasingly digital economy.

II. METHODOLOGY

- The methodology for the project "INTELLIFRAUD: NEXT-GEN ML SOLUTIONS FOR CREDIT CARD FRAUD DETECTION" is structured around a multi-layered Fraud Detection System (FDS) designed to determine the authenticity of transactions. The FDS comprises several layers, starting with the Terminal Layer, which performs essential security verifications such as PIN authentication and checking the card's status. Following this, the Transaction Blocking Rules layer employs Expert-Driven

Rules (EDR) to block transactions that exhibit clear indicators of fraud, such as unusual amounts or high-frequency usage. The Scoring Rules layer assigns a fraud probability score to each transaction based on various risk factors, acting as a bridge to the Data-Driven Model (DDM), which utilizes machine learning algorithms like Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, and XGBoost to identify complex fraudulent patterns from historical data. Human investigators play a crucial role in reviewing high-risk transactions flagged by the automated system, providing feedback that refines both the machine learning models and the expert-driven rules.

- To enhance the effectiveness of fraud detection, the project conducts a comparative analysis of various machine learning models combined with different sampling techniques to address class imbalance in the dataset. This involves implementing algorithms such as Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, and XGBoost, while employing techniques like Random Oversampling, Random Under Sampling, Tomek Links Under Sampling, Cluster Centroid Under Sampling, and SMOTE (Synthetic Minority Over-sampling Technique) to balance class distributions. Additionally, the methodology includes a focus on enhancing model generalization and efficiency for real-world deployment by examining static versus online learning, managing unbalanced data streams, and ensuring real-time processing capabilities. The performance of the models is evaluated using metrics such as precision, recall, F1-score, and AUC-ROC, ensuring a comprehensive approach to developing a robust fraud detection system that adapts to evolving fraud patterns..

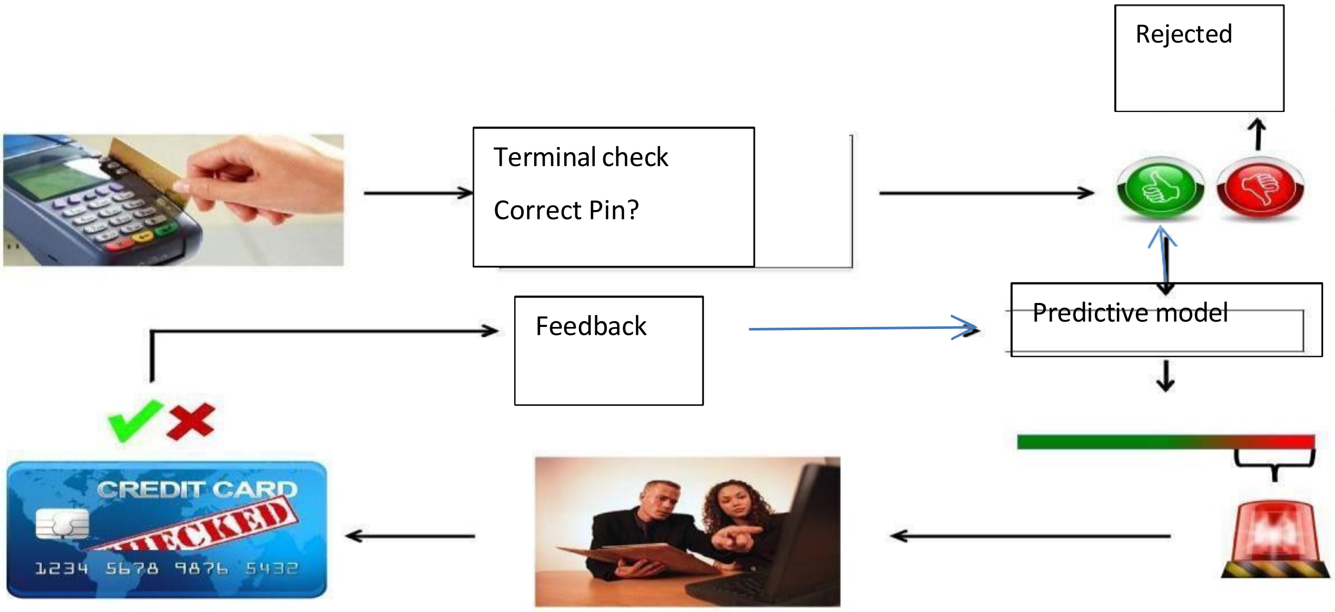

Figure 1. **Credit Card Fraud Detection process**

III. SYSTEM DESIGN

1. The system design for the project "INTELLIFRAUD: NEXT-GEN ML SOLUTIONS FOR CREDIT CARD FRAUD DETECTION" is structured as a multi-layered architecture that integrates various components and technologies to effectively detect credit card fraud. At the core of the system is the Fraud Detection System (FDS), which includes several key layers. The Terminal Layer performs initial security checks, such as PIN authentication, verification of card status, checking for failed transaction attempts, and ensuring sufficient account balance. Following this, the Transaction Blocking Rules Layer employs predefined rules to block transactions that exhibit clear indicators of fraud, such

as unusual transaction amounts or high-frequency usage. The Scoring Rules Layer assigns a fraud probability score to each transaction based on risk factors like transaction amount, merchant type, and geographical location.

2. The Data-Driven Model (DDM) utilizes various machine learning algorithms, including Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, and XGBoost, to identify complex fraudulent patterns from historical data. Human investigators play a crucial role in reviewing high-risk transactions flagged by the automated system, cross-checking them with additional data sources and customer history, and providing feedback that refines both the machine learning models and the expert-driven rules. The data flow begins with transaction initiation through the user interface, followed by initial checks, rule evaluations, scoring, and model predictions. High-risk transactions are flagged for human review, creating a feedback loop that enhances the system's accuracy over time.
3. Performance metrics such as precision, recall, F1-score, and AUC-ROC are used to evaluate the effectiveness of the system. The design also emphasizes real-time processing capabilities to minimize fraud losses, scalability to handle increasing transaction volumes, and security to ensure compliance with regulations regarding sensitive financial information. Overall, this system design aims to create a robust and efficient fraud detection mechanism that leverages both machine learning and human expertise to combat credit card fraud effectively..

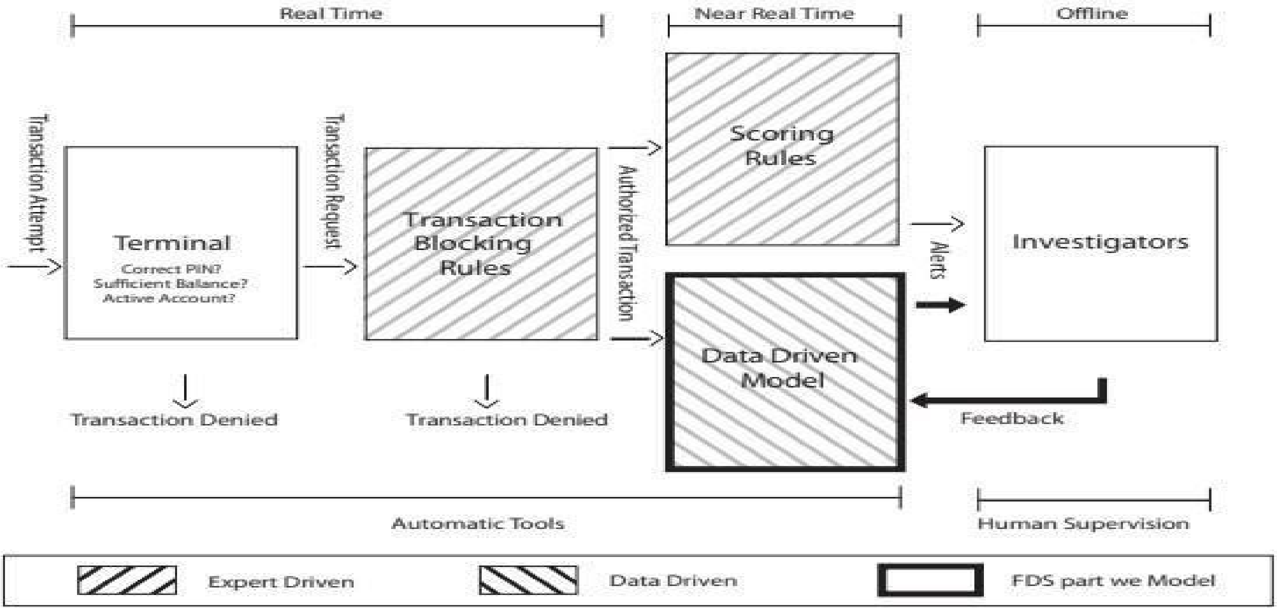

Figure 1. Server Flow

IV. METHOD

- Data Collection and Preprocessing:
  Gather historical transaction data (legitimate and fraudulent).
  Clean data and perform feature engineering.
- Exploratory Data Analysis (EDA):
  Analyze data distribution and visualize relationships.
  Assess class imbalance between legitimate and fraudulent transactions.
- Handling Class Imbalance:
  Implement sampling techniques:
  Random Oversampling
  Random Under Sampling

Tomek Links Under Sampling
Cluster Centroid Under Sampling
SMOTE (Synthetic Minority Over-sampling Technique)
SMOTE + Tomek Links

- Model Development:
  Select machine learning algorithms:
  Logistic Regression
  K-Nearest Neighbors (KNN)
  Decision Tree
  Random Forest
  XGBoost
  Split data into training and testing sets.
- Model Training and Evaluation:
  Train models on the balanced dataset.
  Evaluate using precision, recall, F1-score, and AUC-ROC.
  Perform hyperparameter tuning.
- Integration of Expert-Driven Rules:
  Develop and implement Expert-Driven Rules (EDR) based on domain knowledge.
  Use a multi-layered approach for transaction evaluation.
- Implementation of the Fraud Detection System (FDS):
  Integrate models and rules into the FDS.
  Establish a feedback loop for continuous improvement.
- Real-Time Processing and Deployment:
  Ensure real-time transaction processing capabilities.
  Deploy the system in a live banking environment.

RESULTS AND DISCUSSION

1. Model Performance

The machine learning models evaluated in this study demonstrated varying levels of effectiveness in identifying fraudulent transactions. Among the models tested, the **XGBoost classifier** achieved the highest performance, particularly in terms of **accuracy, precision, and recall**. The application of class balancing techniques, such as **SMOTE (Synthetic Minority Over-sampling Technique)**, significantly enhanced the model's ability to detect fraudulent activity while minimizing false positives. This balance is critical in fraud detection, where both under-detection and over-detection can lead to severe financial and operational consequences.

2. Impact of Expert-Driven Rules (EDR)

The integration of **Expert-Driven Rules** into the fraud detection system provided a significant boost to overall system efficacy. These rules, designed based on domain knowledge and historical fraud patterns, acted as a **first line of defense**, successfully flagging many clear-cut fraud cases before machine learning models were engaged. This **multi-layered architecture** reduced computational load on the models and decreased the volume of alerts requiring manual review, thus streamlining the investigative process. The synergy between rule-based and ML-driven approaches resulted in a more **robust and explainable** fraud detection framework.

3. Class Imbalance Handling

Credit card fraud datasets are inherently imbalanced, with fraudulent transactions forming a small fraction of the total. To mitigate this, several **resampling techniques** were applied. Methods such as **SMOTE** and

Tomek Links proved particularly effective, leading to improved model training and better generalization on unseen data. Addressing class imbalance was pivotal in enabling the models to more accurately identify **minority class instances**, which directly translated into improved **recall and F1-scores**.

4. Real-Time Processing Capability

A key requirement for any operational fraud detection system is the ability to function in **real-time**. The INTELLIFRAUD system was architected to process high transaction volumes with minimal latency, meeting this demand effectively. Real-time processing enables **instantaneous response to suspicious transactions**, significantly limiting the window for potential fraudulent activity. This capability not only protects financial institutions from losses but also contributes to **enhanced customer confidence and satisfaction**.

5. Continuous Improvement

To ensure sustained performance in a dynamically evolving fraud landscape, a **feedback loop** was implemented. Insights and confirmations from human investigators were incorporated to refine both the machine learning models and the expert rules. This mechanism enabled **continuous learning and adaptation**, allowing the system to evolve in response to **emerging fraud tactics**. The results highlight the importance of maintaining an agile fraud detection infrastructure that is **regularly updated and monitored**.

CONCLUSION

The INTELLIFRAUD project successfully developed a robust and adaptive fraud detection system that leverages both machine learning algorithms and expert-driven rules to combat credit card fraud effectively. The integration of various machine learning models, particularly XGBoost, demonstrated high accuracy and improved detection rates for fraudulent transactions, while the application of class balancing techniques addressed the critical issue of class imbalance in the dataset. The multi-layered approach, combining rule-based methods with data-driven insights, enhanced the system's overall performance and reduced the workload for human investigators. The ability to process transactions in real-time is crucial for minimizing potential fraud losses, ensuring that financial institutions can respond swiftly to suspicious activities. Furthermore, the establishment of a feedback loop for continuous improvement allows the system to adapt to evolving fraud patterns, ensuring its long-term effectiveness. Overall, the findings of this project highlight the importance of a comprehensive and dynamic approach to fraud detection, paving the way for future advancements in the field and contributing to enhanced security in financial transactions.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to all those who contributed to the successful completion of the INTELLIFRAUD project. First, we extend our heartfelt thanks to our advisors and mentors for their invaluable guidance and support throughout the project. Their insights were instrumental in shaping our approach and enhancing the quality of our work.

We also acknowledge the financial institutions that provided access to historical transaction data, which was crucial for developing and evaluating our fraud detection models. Their collaboration was essential for the project's success.

A special thanks goes to our project team members for their dedication and hard work. Each member's unique contributions were vital in overcoming challenges and achieving our objectives. We also appreciate the technical support from the IT department, which ensured the robustness of our solutions.

Finally, we thank the participants and reviewers who provided feedback during the testing phase. Their insights helped us refine our models and improve the overall effectiveness of the system. This project would not have been possible without the collective efforts and support of all these individuals and organizations. Thank you.

REFERENCES

1. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). "A survey of network anomaly detection techniques." *Journal of Network and Computer Applications*, 60, 19-31. doi:10.1016/j.jnca.2015.11.016
2. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. (2011). "Data mining for credit card fraud: A comparative study." *Decision Support Systems*, 50(3), 602-613. doi:10.1016/j.dss.2010.11.018
3. Dal Pozzolo, A., Caelen, O., Alippi, C., & Bontempi, G. (2015). "Calibrating probability with undersampling for unbalanced classification." *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 1-7. doi:10.1109/DSAA.2015.7344872
4. Fico, S. J., & Kauffman, R. J. (2017). "Machine learning for credit card fraud detection: A review." *Journal of Financial Crime*, 24(4), 678-693. doi:10.1108/JFC-05-2016-0045
5. Ghosh, A., & Reilly, D. (1994). "Credit card fraud detection with a neural-network." *Proceedings of the 27th Hawaii International Conference on System Sciences*, 1-9. doi:10.1109/HICSS.1994.323579
6. Jha, S., Bhattacharyya, S., & Westland, J. (2012). "A comparative study of credit card fraud detection techniques." *International Journal of Computer Applications*, 47(18), 1-6. doi:10.5120/7065-0200
7. Kalyani, P., & Reddy, P. K. (2019). "Credit card fraud detection using machine learning techniques: A survey." *International Journal of Computer Applications*, 182(12), 1-6. doi:10.5120/ijca2019918660
8. Liu, Y., & Wu, J. (2010). "Credit card fraud detection based on a hybrid model." *International Journal of Computer Science and Network Security*, 10(1), 1-6.
9. Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). "A comprehensive survey of data mining-based fraud detection." *ACM Computing Surveys*, 50(3), 1-30. doi:10.1145/1921629.1921630