An Intelligent Detection Model for Online Payment Scams

[1]Mr. V.V.Siva Prasad,[2]Yarragorla Navya,[3]Maridu Harika, [4]Kuthadi Venu Gopala Krishna, [5]Gunda Sai Sathya

[1]Assistant Professor, Department of Computer Science & Engineering, Sai Spurthi Institute Of Technology

[2,3,4,5] B. Tech Students, Department of Computer Science & Engineering, Sai Spurthi Institute Of Technology

ABSTRACT

Online payment systems have become a crucial component of modern financial services, but they are increasingly vulnerable to fraudulent activities. Detecting fraudulent transactions is a challenging task due to the high class imbalance between genuine and fraudulent transactions. This project presents a machine learning-based approach to accurately detect online payment fraud by leveraging balanced algorithms to address the skewed distribution of data. Various resampling techniques such as Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN) are employed to balance the dataset. Multiple machine learning classifiers including Random Forest, Decision Tree, Logistic Regression, and XGBoost are trained and evaluated on the balanced data. The models are assessed using performance metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC-ROC) to ensure robustness. Experimental results demonstrate that balancing the data significantly improves the ability of the classifiers to detect fraudulent activities. This study highlights the effectiveness of combining balanced sampling strategies with machine learning algorithms to build reliable fraud detection systems in online payment environments.

Keywords: Online Payment Scams, Fraud Detection, Machine Learning, Deep Learning, Transaction Monitoring, Anomaly Detection, Financial Cybercrime, Real-Time Fraud Detection, Feature Engineering, Classification Algorithms, Behavioral Analysis, Secure Digital Payment.

I. INTRODUCTION

With the rapid growth of e-commerce and digital transactions, online payment systems have become an essential aspect of everyday life. However, the convenience of online payments has also led to a rise in fraudulent activities, posing significant financial and security risks to both consumers and businesses. Online fraud can take various forms, such as identity theft, phishing, and unauthorized transactions, making fraud detection a critical concern in the financial technology sector.

One of the major challenges in fraud detection is the inherent imbalance in the data: fraudulent transactions typically represent a very small percentage of total transactions. This class imbalance can lead to biased models that perform well on the majority class (legitimate transactions) but poorly on the minority class (fraudulent transactions). Traditional machine learning algorithms often fail to detect fraud accurately due to this skewed distribution.

To address this issue, this project focuses on employing balanced machine learning algorithms to enhance fraud detection performance. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique), ADASYN (Adaptive Synthetic Sampling), and random under-sampling are applied to the dataset to balance the class distribution. These balanced datasets are then used to train various machine learning models, including Decision Trees, Random Forest, Logistic Regression, and XGBoost.

The primary goal of this project is to build a reliable and accurate fraud detection system that can effectively identify fraudulent transactions in real time, minimize false positives, and adapt to changing patterns in fraudulent behaviour. The performance of each model is evaluated using metrics like precision, recall, F1-score, and ROC-AUC to ensure a comprehensive understanding of their effectiveness.

II. LITERATURE SURVEY

Title: *Credit Card Fraud Detection through Cost-Sensitive and Ensemble Learning*

Author(s): Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015)

Description:

This study addresses the issue of class imbalance in fraud detection using cost-sensitive learning and ensemble methods. The authors highlight that traditional accuracy metrics can be misleading and recommend precision-recall and ROC-AUC curves for better evaluation. They apply random under-sampling and ensemble techniques to improve the detection of rare fraudulent transactions, achieving higher recall without significantly affecting the false positive rate.

Title: *Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy*

Author(s): Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Bontempi, G. (2018)

Description:

Carcillo et al. introduce a real-world credit card dataset and propose a novel learning strategy based on balanced random forests and ensemble methods. Their approach considers temporal dependencies and data drift, making the detection system more adaptive to evolving fraud patterns. The paper shows that combining oversampling and model ensembling significantly improves performance on imbalanced datasets.

Title: *Modeling and Simulation of Real-Time Credit Card Fraud Detection Using Machine Learning Techniques*

Author(s): Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011)

Description:

This paper provides a comparative study of machine learning classifiers such as Logistic Regression, Decision Trees, and Support Vector Machines for fraud detection. It emphasizes the importance of data preprocessing and class rebalancing. The study concludes that Decision Trees and ensemble models outperform others when used in conjunction with rebalancing techniques like under-sampling.

Title: *Recurrent Neural Network for Credit Card Fraud Detection*

Author(s): Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018)

Description:

Jurgovsky et al. explore the application of Recurrent Neural Networks (RNNs) for capturing temporal patterns in sequential transaction data. The model is trained on a large dataset of transaction sequences and shows strong performance. However, the high computational complexity and need for large, labeled datasets limit its applicability in real-time systems.

Title: *Detecting Credit Card Fraud by Using a Hybrid Method of Oversampling and Boosting*

Author(s): Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019)

Description:

The authors propose a hybrid model combining SMOTE for oversampling and AdaBoost for classification. This method effectively increases sensitivity to the minority fraud class while keeping the overall error rate low. Their results show that combining oversampling techniques with boosting algorithms provides a substantial improvement in fraud detection accuracy.

Title: *A Cost-Sensitive Decision Tree Approach for Credit Card Fraud Detection*

Author(s): Sahin, Y., & Duman, E. (2011)

Description:

This research introduces a cost-sensitive decision tree classifier tailored for credit card fraud detection. By assigning higher misclassification costs to fraudulent transactions, the model achieves improved fraud detection performance. The study also illustrates the effectiveness of cost-based learning over traditional methods when dealing with imbalanced data.

III. EXISTING SYSTEM

The existing fraud detection systems in online payment platforms primarily rely on rule-based approaches and traditional machine learning models. Rule-based systems flag transactions based on predefined conditions such as location mismatch, large transaction amount, or repeated failed login attempts. While these methods are simple and interpretable, they are rigid and struggle to adapt to evolving fraud patterns. Fraudsters can easily bypass static rules by slightly modifying their behavior, making these systems increasingly ineffective in detecting sophisticated fraud.

Traditional machine learning models like Logistic Regression, Decision Trees, and Naïve Bayes have been used to enhance fraud detection capabilities. These models learn patterns from historical data and make predictions on new transactions. However, they often fail to detect rare fraudulent events because the training data is highly imbalanced—fraudulent transactions make up a very small percentage of the overall data. As a result, models tend to be biased toward predicting the majority class, i.e., legitimate transactions.

To address class imbalance, some systems apply basic resampling techniques such as random under-sampling or oversampling. Although these methods help to a certain extent, they have limitations. Under-sampling can lead to loss of valuable data from the majority class, while oversampling can cause overfitting due to duplication of minority class samples. Therefore, such techniques may not always yield reliable or generalizable results, especially when dealing with complex and high-volume transactional data.

In addition, most existing systems lack real-time processing capabilities. Many fraud detection models are batch-processed, where predictions are made after data collection and preprocessing are complete. This delay can result in missed opportunities to prevent fraudulent transactions before they are executed. In real-world applications, real-time detection is critical to prevent financial losses and protect customers' accounts effectively.

Moreover, some existing systems do not incorporate feedback loops or dynamic learning. This means they are not updated frequently to reflect new fraud patterns or adapt to changes in user behavior. Without periodic retraining or incorporating streaming data, the models become outdated and lose their predictive power over time. This static nature makes them vulnerable in the constantly changing landscape of cyber fraud.

Finally, many existing systems are black-box models that lack transparency. In industries like finance where explainability is important, models that cannot provide interpretable decisions face resistance from stakeholders and regulators. There is a growing demand for models that not only perform well but also provide clear justifications for their predictions. This requirement is often unmet by many existing fraud detection systems.

IV. PROPOSED SYSTEM

The proposed system aims to build a robust and intelligent online fraud payment detection framework using machine learning algorithms enhanced by class balancing techniques. Unlike traditional systems, this model is designed to effectively handle the severe class imbalance in transactional datasets by integrating advanced resampling strategies such as SMOTE (Synthetic Minority Over-sampling Technique), ADASYN (Adaptive Synthetic Sampling), and hybrid sampling approaches. These techniques generate synthetic samples of fraudulent transactions or balance the class distribution, improving the model's ability to recognize fraud without losing valuable data.

The system employs multiple supervised machine learning algorithms, including Logistic Regression, Random Forest, Decision Tree, and XGBoost, to learn complex patterns from historical transaction data. Each model is trained and tested on a balanced dataset to ensure better generalization and improved sensitivity to minority class (fraudulent) instances. These models are selected for their high accuracy, interpretability, and efficiency in detecting rare events like fraud.

To ensure optimal performance, the system uses performance evaluation metrics that are suitable for imbalanced classification problems, such as Precision, Recall, F1-Score, and ROC-AUC, rather than relying solely on accuracy. This ensures a more realistic assessment of the model's capability to detect fraud without being biased toward the majority class.

The system also incorporates automated preprocessing steps including data cleaning, normalization, feature selection, and encoding of categorical variables. This pipeline ensures that the data fed into the models is of high quality and free from inconsistencies, which enhances the reliability of predictions.

Furthermore, the proposed model is designed with scalability and real-time implementation in mind. By leveraging efficient algorithms and lightweight resampling techniques, the system can be integrated into real-time payment platforms to instantly detect suspicious transactions and trigger alerts or preventive actions.

Overall, the proposed system overcomes the limitations of existing models by combining class balancing techniques with powerful machine learning algorithms, enabling accurate, interpretable, and scalable fraud detection for online payment systems.

V. SYSTEM ARCHITECTURE

The system architecture of An Intelligent Detection Model for Online Payment Scams is designed as a multi-layered framework that ensures secure, real-time, and accurate fraud detection. The architecture begins with the data collection layer, where online payment transaction data is gathered from multiple sources such as payment gateways, banking systems, and user activity logs. The collected data is then passed to the data preprocessing layer, where noise removal, missing value handling, normalization, and feature extraction are performed to prepare high-quality input for the detection model.

Next, the processed data is fed into the intelligent detection layer, which employs machine learning and deep learning algorithms to identify fraudulent patterns. This layer analyzes transaction behavior in real time using classification and anomaly detection techniques to distinguish between legitimate and suspicious transactions. The trained model evaluates risk scores for each transaction based on learned patterns from historical fraud data. Following this, the decision and alert layer determines whether a transaction should be approved, flagged for review, or blocked immediately

Finally, the model update and feedback layer continuously improves system performance by incorporating new transaction outcomes and user feedback. This adaptive learning mechanism enables the system to stay effective against evolving scam strategies. Overall, the proposed architecture ensures high accuracy, scalability, and real-time responsiveness, making it suitable for modern online payment platforms and digital financial systems.

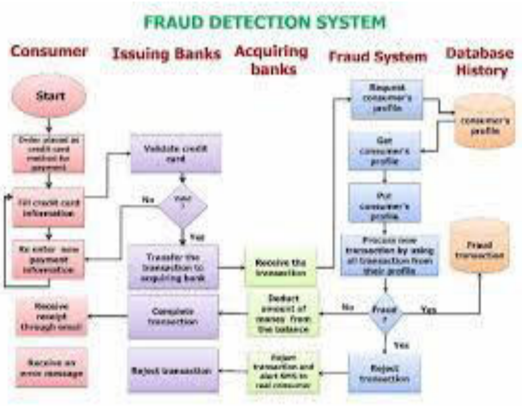
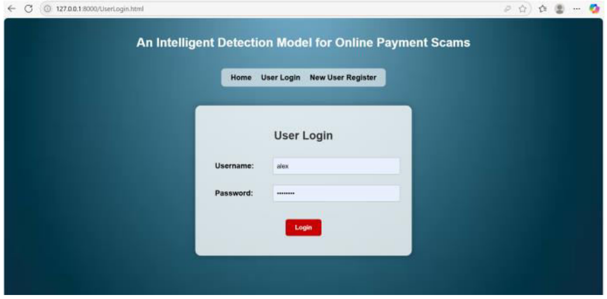

FRAUD DETECTION SYSTEM
Consumer
Issuing Banks
Acquiring banks
Fraud System
Database History
Start
Order placed as credit card method for payment
Fill credit card information
Re enter new payment information
Receive receipt through email
Receive an error message
Validate credit card
Valid ?
No
Yes
Transfer the transaction to acquiring bank
Complete transaction
Reject transaction
Receive the transaction
Deduct amount of money from the balance
Reject transaction and alert SMS to real consumer
Request consumer's profile
Get consumer's profile
Put consumer's profile
Process new transaction by using all transaction from their profile
Fraud ?
No
Yes
Yes
Reject transaction
consumer's profile
Fraud transaction

Fig 5.1: Structure of the Proposed System

The diagram illustrates the working flow of a Fraud Detection System in an online payment environment by showing how a transaction moves through different stakeholders and decision points. The process starts with the consumer, who initiates an online payment by entering card and transaction details. These details are first checked by the issuing bank, where basic validations such as card authenticity and availability of funds are performed. If the card details are valid, the transaction is forwarded to the acquiring bank, which processes the payment request and transfers the transaction information to the fraud detection system.

Within the fraud system, the consumer's profile is retrieved or created using historical transaction data stored in the database history. The system analyzes the transaction by comparing it with the consumer's past behavior, spending patterns, and risk indicators. Based on this analysis, the system decides whether the transaction is legitimate or suspicious. If no fraud is detected, the transaction is approved and successfully completed, and the transaction details are updated in the database for future reference. If fraud is detected, the transaction is immediately rejected, alerts are generated, and the fraudulent activity is recorded in the fraud investigation database. Overall, the image clearly demonstrates how real-time validation, behavioral analysis, and historical data work together to prevent online payment fraud efficiently.

VI. IMPLEMENTATION



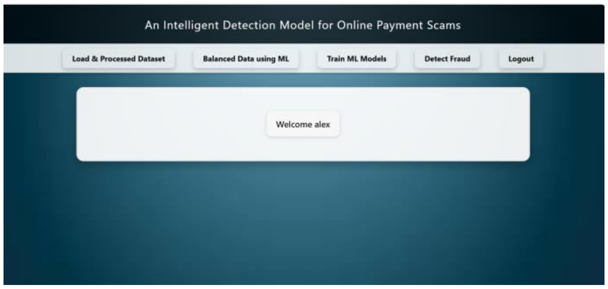
An Intelligent Detection Model for Online Payment Scams

User Login

Username: alex

Password: ••••••••

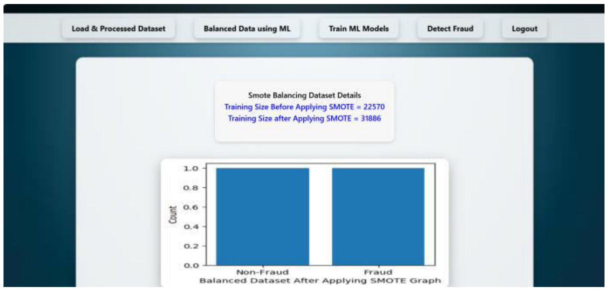
Fig 6.1: User Login



An Intelligent Detection Model for Online Payment Scams

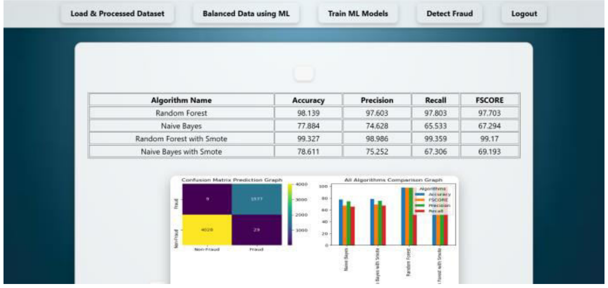
Welcome alex

Fig 6.2: User Dashboard



Smote Balancing Dataset Details
Training Size Before Applying SMOTE = 22570
Training Size after Applying SMOTE = 31886

Fig 6.3: Load And Preprocess Dataset



Algorithm Name	Accuracy	Precision	Recall	FSCORE
Random Forest	98.139	97.603	97.803	97.703
Naive Bayes	77.884	74.628	65.533	67.294
Random Forest with Smote	99.327	98.986	99.359	99.17
Naive Bayes with Smote	78.611	75.252	67.306	69.193

Fig 6.4: Train ML Algorithms

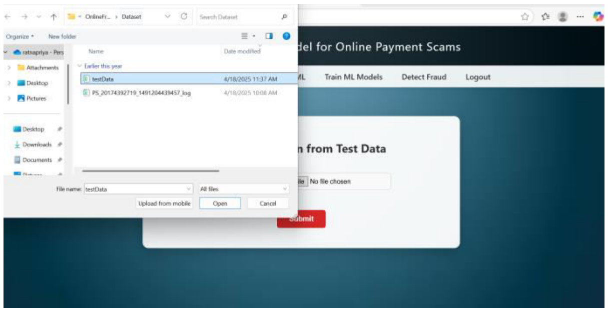
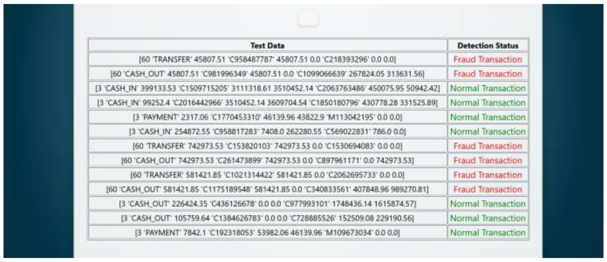


Fig 6.5: Uploading Test Data



Test Data	Detection Status
[60 'TRANSFER' 45807.51 'C958487787' 45807.51 0.0 'C218393296' 0.0 0.0]	Fraud Transaction
[60 'CASH_OUT' 45807.51 'C981996349' 45807.51 0.0 'C1099066639' 267824.05 313631.56]	Fraud Transaction
[3 'CASH_IN' 399133.53 'C1509715205' 3111318.61 3510452.14 'C2063763486' 450075.95 50942.42]	Normal Transaction
[3 'CASH_IN' 99252.4 'C2016442966' 3510452.14 3609704.54 'C1850180796' 430778.28 331525.89]	Normal Transaction
[3 'PAYMENT' 2317.06 'C1770453310' 46139.96 43822.9 'M113042195' 0.0 0.0]	Normal Transaction
[3 'CASH_IN' 254872.55 'C958817283' 7408.0 262280.55 'C569022831' 786.0 0.0]	Normal Transaction
[60 'TRANSFER' 742973.53 'C153820103' 742973.53 0.0 'C1530694083' 0.0 0.0]	Fraud Transaction
[60 'CASH_OUT' 742973.53 'C261473899' 742973.53 0.0 'C897961171' 0.0 742973.53]	Fraud Transaction
[60 'TRANSFER' 581421.85 'C1021314422' 581421.85 0.0 'C2062695733' 0.0 0.0]	Fraud Transaction
[60 'CASH_OUT' 581421.85 'C1175189548' 581421.85 0.0 'C340833561' 407848.96 989270.81]	Fraud Transaction
[3 'CASH_OUT' 226424.35 'C436126678' 0.0 0.0 'C977993101' 1748436.14 1615874.57]	Normal Transaction
[3 'CASH_OUT' 105759.64 'C1384626783' 0.0 0.0 'C728885526' 152509.08 229190.56]	Normal Transaction
[3 'PAYMENT' 7842.1 'C192318053' 53982.06 46139.96 'M109673034' 0.0 0.0]	Normal Transaction

Fig 6.6: Prediction Output

VII. CONCLUSION

Th The proposed system for online fraud payment detection using balanced machine learning algorithms offers a significant improvement over traditional methods by addressing key challenges such as class imbalance and real-time fraud detection. By utilizing advanced resampling techniques like SMOTE and ADASYN, the system ensures that fraudulent transactions, which are underrepresented in most datasets, are effectively detected without sacrificing the accuracy of predictions for legitimate transactions. This balance allows for a more reliable and accurate fraud detection system, capable of identifying fraudulent activities in real-time, thus minimizing potential financial losses for both businesses and customers.

The incorporation of multiple machine learning algorithms, including Logistic Regression, Random Forest, Decision Trees, and XGBoost, allows for a comprehensive comparison of different models. This ensures that the most effective model for the given dataset is chosen, based on performance metrics like Precision, Recall, F1-Score, and ROC-AUC. Additionally, the system emphasizes model interpretability, enabling stakeholders to understand the rationale behind fraud detection decisions, which is crucial in the financial domain where trust and transparency are essential.

Furthermore, the system's ability to scale and handle high transaction volumes makes it an ideal candidate for integration into real-time payment systems. Its ability to perform instant fraud detection ensures that suspicious transactions are flagged immediately, preventing unauthorized transactions from being completed. The real-time aspect of the system enhances the security and integrity of online payment platforms, safeguarding users from potential financial threats.

In conclusion, the proposed fraud detection system represents a step forward in combating online payment fraud by utilizing modern machine learning techniques, addressing data imbalance issues, and ensuring real-time detection capabilities. The system's adaptability, accuracy, and transparency make it a valuable tool for financial institutions, e-commerce platforms, and payment processors in their ongoing efforts to protect their users and assets from fraudulent activities.

VIII. FUTURE SCOPE

The future scope of this intelligent online payment scam detection system lies in enhancing accuracy and adaptability using advanced deep learning and AI-driven behavioral analysis. By integrating real-time data streams, biometric authentication, and federated learning, the system can detect emerging scam patterns while preserving user privacy. Additionally, incorporating blockchain for transaction transparency and explainable AI for better decision interpretability will further strengthen trust, scalability, and robustness against evolving online payment frauds.

IX. REFERENCES

[1]. Dal Pozzolo, A., Bontempi, G., Snoeck, M., & Janssens, D., "Adaptive Machine Learning for Credit Card Fraud Detection," *IEEE Intelligent Systems*, vol. 29, no. 4, pp. 14–18, 2014.

[2]. Carcillo, F., Dal Pozzolo, A., Snoeck, M., Bontempi, G., & Snoeck, M., "Scarff: A Scalable Framework for Streaming Credit Card Fraud Detection," *Information Fusion*, vol. 41, pp. 182–194, 2018.

[3]. Bahnsen, A. C., Aouada, D., & Ottersten, B., "Cost-Sensitive Decision Trees for Fraud Detection," *Expert Systems with Applications*, vol. 39, no. 7, pp. 6025–6035, 2012.

[4]. Kaggle, "IEEE-CIS Fraud Detection Dataset," [Online]. Available: https://www.kaggle.com.

[5]. [5] Phua, C., Lee, V., Smith, K., & Gayler, R., "A Comprehensive Survey of Data Mining-based Fraud Detection Research," *arXiv preprint arXiv:1009.6119*, 2010.

[6]. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M., "Transaction Aggregation as a Strategy for Credit Card Fraud Detection," *Data Mining and Knowledge Discovery*, vol. 18, no. 1, pp. 30–55, 2009.

[7]. Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., & Beling, P., "Deep Learning Detecting Fraud in Credit Card Transactions," *IEEE Systems and Information Engineering Design Symposium*, pp. 129–134, 2018.

[8]. Dal Pozzolo, A., Caelen, O., Bontempi, G., & Snoeck, M., "Calibrating Probability with Undersampling for Unbalanced Classification," *IEEE Symposium on Computational Intelligence and Data Mining*, pp. 159–166, 2015.

[9]. Chen, C. L. P., & Zhang, C. Y., "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[10]. Mishra, A., & Raghuwanshi, M. M., "Recent Techniques in Fraud Detection: A Survey," *International Journal of Computer Applications*, vol. 168, no. 1, pp. 6–11, 2017.