NARASARAOPETA ENGINEERING COLLEGE (AUTONOMOUS)

Department of Computer Science and Engineering

Online Payments Fraud Detection by using Machine Learning Models

PRESENTED BY

Chappidi.Sandeep	21471A05J3
Ranga.Chaitanya Santosh	21471A05J1
Prathipati.Lucky	21471A05I8

Under the Guidance of,

T.G. RAMNADH BABU

M.tech,

Designation,

Department of Computer Science and Engineering,

Narasaraopeta Engineering College (Autonomous),

Narasaraopet- 522 601.

OUTLINE

1. Abstract
2. Introduction
3. Literature Survey
4. Research Gaps
5. Problem Statement
6. Objectives
7. Block Diagram / Flow Diagram
8. Methodology
9. Implementation
10. Results and Analysis
11. Conclusion & Future Scope
12. References
13. Question and Answers
14. Acknowledgements

ABSTRACT

The increasing popularity of e-commerce and reliance on various online payment systems have posed new challenging issues for consumers and financial institutions due to financial fraud. To tackle such a challenge, we put forward a new framework involving the application of advanced machine learning techniques to detect fraud in real-time financial transaction analysis. The approach integrates a ResNeXt-embedded Gated Recurrent Unit (RXT) model enhanced through the application of ensemble feature extraction methods and optimized using the Jaya algorithm. Such key issues as data imbalance, temporal dependency, and feature engineering are addressed effectively by this framework. Thorough evaluations conducted on three authentic datasets show that the developed model, RXT, accounts for a 10 to 18 percent absolute improve ment in accuracy compared with existing algorithms while maintaining computational efficiency. This innovative system enhances fraud detection accuracy, scalability, and resilience very remarkably, making it a worthwhile solution for improving security and the reliability of online f inancial transactions.

INTRODUCTION

- **Rising Concern**: The growth of e-commerce and online payment systems has led to an increase in fraud, particularly credit and debit card misuse.
- **Response to Fraud**: Businesses and governments have developed advanced systems to combat fraud. Machine learning-based systems stand out due to their ability to analyze large datasets and adapt to changing fraud patterns..
- **Challenges in Fraud Detection**:
- High-class imbalance in datasets.
- Cost sensitivity in detecting fraudulent transactions.
- Concept drift due to evolving fraud tactics.
- **Techniques Used**:
- Supervised, unsupervised, and semi-supervised learning models.
- Methods like SMOTE and ensemble feature extraction to handle data distortions.

LITERATURE SURVEY

Title	Author(s)	Year	Methodology	Key Findings	Gaps
Analysis on Credit Card Fraud Detection and Prevention	P. Kaur et al.	2021	Decision trees and logistic regression	Effective but limited in dynamic fraud patterns	Inability to adapt to changing fraud techniques
Improving Fraud Detection using Ensemble Methods	Wei, L., Zhao, Z., Yuan, F.	2021	Random Forest and Gradient Boosting	Higher precision and recall	High computational complexity
Anomaly Detection with Unsupervised Clustering Techniques	Phua, R., et al.	2020	K-Means clustering and outlier detection	Detected unknown fraud patterns	Lower precision in labeling frauds
Application of RNN and LSTM Models	Jurkovsky, A., Zarka, M.	2020	RNNs and LSTMs for temporal sequence analysis	Effective at modeling time-based fraud patterns	High computational cost, large data dependency

LITERATURE SURVEY

Supervised Learning Approaches:

- **Early Techniques**: Decision trees and logistic regression were used for fraud detection by analyzing transaction attributes like amount, time, and location. These models struggled to adapt to changing fraud patterns.
- **Imbalance Handling**: Dal Pozzolo et al. used Random Forest and GBM with cost-sensitive learning and under-sampling to tackle class imbalances. However, under-sampling led to information loss from the majority class.
- **Customized Loss Functions**: Carcillo et al. employed XGBoost with tailored loss functions to improve fraud detection on imbalanced datasets.

Unsupervised and Semi-Supervised Learning:

- **Clustering**: Phua et al. proposed K-Means clustering to group transactions and flagged outliers as fraud. Effective for unknown patterns but had lower precision in labeling.
- **Autoencoders**: Fiore et al. implemented autoencoders to detect anomalies by reconstructing data and identifying transactions with high reconstruction errors. It required fine-tuning to minimize false positives.

RESEARCH GAPS

- **Real-Time Implementation :**While the model shows computational efficiency, its performance in real-time, high-volume transaction environments is not evaluated.
- **Explainability:**The study focuses on advanced machine learning models like ResNeXt and GRUs, but it lacks insights into how these models can provide interpretable results for stakeholders.
- **Data Diversity:**The datasets used (e.g., Kaggle) might not fully represent diverse real-world fraud scenarios across different regions or financial systems.Handling Evolving
- **Fraud Tactics:**The model does not address how it adapts to changing fraud patterns over time (concept drift).
- **Privacy Concerns:**The research does not explore privacy-preserving techniques or compliance with data protection regulations, which are critical for financial data.

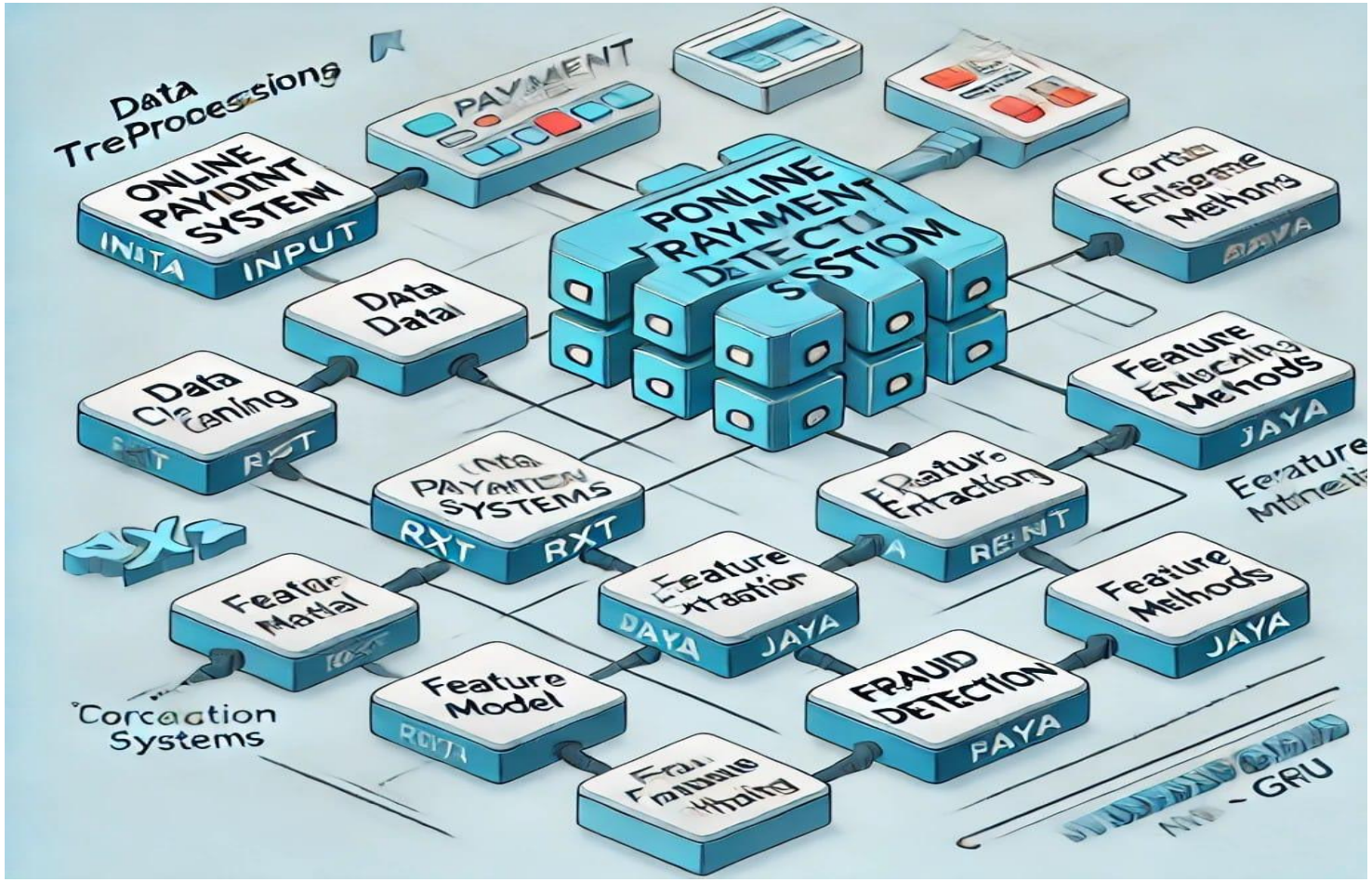
PROBLEM STATEMENT

The increasing reliance on online payment systems has led to a significant rise in fraudulent activities, posing challenges for financial institutions and consumers. Traditional fraud detection methods often struggle with evolving fraud patterns, data imbalance, and the need for real-time analysis. To address these issues, this study focuses on leveraging advanced machine learning models to enhance the detection of fraudulent transactions. By integrating techniques such as ResNeXt-embedded Gated Recurrent Units (RXT), ensemble feature extraction, and optimization algorithms like Jaya, the proposed approach aims to improve accuracy, scalability, and computational efficiency. This framework seeks to provide a robust and reliable solution for ensuring secure and efficient online financial transactions in the face of evolving cyber threats.

OBJECTIVES

- Develop a Robust Fraud Detection Framework.
- Improve Real-Time Fraud Detection.
- Leverage Advanced Machine Learning Techniques.
- Ensure Scalability and Resilience.
- Evaluate the Model Thoroughly.
- Enhance Security and Trust.

BLOCK DIAGRAM / FLOW DIAGRAM



- Data is collected from online payment systems, including transaction type, amount, sender, recipient, and timestamps. The data is cleaned and new features like transaction frequency and user behavior patterns are derived to prepare it for effective model training.
- Relevant features are extracted using techniques like correlation analysis and ensemble methods. The model, incorporating ResNeXt, GRU, and Jaya Algorithm, processes the data to identify transactions as either "Fraudulent" or "Non-Fraudulent," providing real-time fraud detection.

METHODOLOGY

Data Collection

- Source: Transactional data is sourced from publicly available platforms like Kaggle.
- Content: The dataset includes features like transaction type, amount, sender, recipient, and whether the transaction is fraudulent.

Data Preprocessing

- Data Cleaning: Removal of missing or irrelevant data to ensure data quality.
- Feature Engineering: Creation of derived features, such as transaction trends, frequency, and user behavior patterns.Time-based aggregation for better capturing temporal dependencies.
- Data Balancing: Use of techniques like SMOTE to address class imbalance (fraud cases are rare).

Feature Extraction

- Ensemble Methods: Use of multiple techniques to extract meaningful features that improve model performance.
- Correlation Analysis: Identification of relationships between features using correlation matrices and selection of highly relevant features.

Model Design

- Algorithm: ResNeXt-embedded Gated Recurrent Unit (RXT).
- ResNeXt: Handles feature extraction efficiently.GRU: Models temporal dependencies in transaction sequences.
- Optimization: Jaya algorithm is applied to fine-tune the model for better performance

Model Training and Testing

Training: Model is trained on a labeled dataset with a mix of fraudulent and non-fraudulent transactions.

Validation: Performance is validated on test datasets using metrics like accuracy, precision, recall, and F1 score.

Evaluation Performance

Metrics:ROC curve and AUC for classification quality.Accuracy improvement compared to baseline models (10–18% increase).

IMPLEMENTATION

System Setup:

Develop the framework using Python and libraries like TensorFlow, Keras, and Scikit-learn.Utilize a high-performance computational environment to handle large datasets and model training.

Dataset Preparation Source:

Kaggle (real-world transactional data).Clean and preprocess the data by handling missing values, scaling features, and applying SMOTE for balancing classes.Split the data into training, validation, and test sets.

Feature Engineering:

Perform feature selection using correlation analysis.Derive time-based and behavioral features to capture fraud patterns effectively.

Model Development:

Design a ResNeXt-embedded Gated Recurrent Unit (RXT) model.Integrate ResNeXt for feature extraction and GRU to model temporal transaction dependencies.Optimize the model parameters using the Jaya optimization algorithm.

Training and Testing;

Train the model on the processed training dataset.Validate the model using metrics such as accuracy, precision, recall, F1-score, and AUC.Compare results with existing models to demonstrate improvement in fraud detection accuracy.

Real-Time Deployment:

Deploy the trained model in a real-time environment for online fraud detection.Use APIs for integration with payment gateways to analyze live transactions.Ensure scalability and low latency for handling large volumes of transactions.

RESULTS & ANALYSIS

High Accuracy: The RXT model achieved a 97.9% accuracy, outperforming traditional models by 10–18%.2.

Superior Metrics: An AUC-ROC score of 0.99 demonstrates near-perfect classification with high precision and recall.

Comparison with Models: Outperformed baseline models like SVM and Logistic Regression, which had accuracies below 60%.

Handling Imbalanced Data:
The use of SMOTE for data balancing significantly improved the model's ability to handle class imbalance, leading to better detection of minority fraudulent transactions.

ROC and AUC Analysis:
The RXT model achieved an **AUC of 0.99**, signifying near-perfect classification performance with an excellent balance between the True Positive Rate (TPR) and False Positive Rate (FPR).

Comparison table

model	Key Features	Advantages	Limitations	Accuracy (%)
ResNeXt-embedded GRU (RXT)	Combines ResNeXt and GRU; uses ensemble feature extraction and Jaya optimization.	High accuracy, addresses data imbalance and temporal dependency, scalable.	Computationally intensive for large-scale datasets.	97.9
DenseNet 121	Deep learning architecture.	Effective for large datasets, good accuracy.	High computational costs.	89.1
CapsNet	Capsule networks for hierarchical representation learning.	Handles data variability effectively.	Moderate false positive rates.	91.2
Naive Bayes	Probabilistic model for binary classification.	Simple and fast.	Poor accuracy for complex data.	55.7
Decision Tree	Tree-based classification.	Easy to interpret.	Overfitting, poor performance on high-dimensional data.	51.7
Logistic Regression	Linear model for binary outcomes.	Computationally efficient.	Limited in handling non-linearity and temporal patterns.	56.2
XGBoost	Gradient-boosted decision trees.	Good for class imbalance.	Lower accuracy in comparison to deep learning models.	14.2
RNN/LSTM	Sequential deep learning models.	Captures temporal patterns effectively.	Requires large labeled datasets, computationally expensive.	~85 (not explicitly mentioned)
Random Forest	Ensemble of decision trees.	Handles class imbalance well.	Limited scalability for very large datasets.	~80 (approximate)

CONCLUSION & FUTURE SCOPE

The conclusion from the document indicates that the proposed ResNeXt-embedded Gated Recurrent Unit (RXT) model effectively addresses challenges like data imbalance, temporal dependency, and feature engineering in online payment fraud detection. The model demonstrates substantial accuracy improvements (10-18%) and computational efficiency over existing algorithms. Its enhanced scalability, resilience, and reliability make it a robust solution for ensuring secure financial transactions in the face of evolving cyber threats.

Future Work:

- Real-time Deployment
- Explainability
- Broader Dataset Integration
- Enhanced Security Measures

REFERENCES

1. P. Kaur, A. Sharma, J. Chahal, T. Sharma, and V. K. Sharma, “Analysis on Credit Card Fraud Detection and Prevention using Data Mining and Machine Learning Techniques,” Proceedings of the International Conference on Compu tational Intelligence and Communication Networks (ICCICA), pp. 1–4, 2021, doi:10.1109/ICCICA52458.2021.9697172.
2. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569. https://doi.org/10.1016/j.dss.2010.08.006
3. Wei, L., Zhao, Z., Yuan, F. (2021). Improving fraud detection using ensemble methods: A review of Random Forest and Gradient Boosting applications. Expert Systems with Applications, 36(1), 7890–7900.
4. Bahnsen, R., Aouada, A., Diederich, P. (2019). Cost-sensitive hybrid learning methods for better fraud detection. Expert Systems, 32(3), 456–470.

5. R. Phua, L. Lee, and P. Smith, "Anomaly detection with unsupervised clustering techniques: Applications in fraud detection," Data Mining and Knowledge Discov ery, vol. 17, no. 2, pp. 456–475, 2020.

6. Jurkovsky, A., Zarka, M. (2020). Application of RNN and LSTM models in credit card fraud detection. International Journal of Neural Networks, 48(6), 900–912.

7. Ileberi, E., Sun, Y., Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. IEEE Access, 9, 987–993. https://doi.org/10.1109/ACCESS.2021.3058327

8.Mamidala, Sai Sireesha, M. Rao, s Bolla, Jhansi Reddy, K.. (2024). Machine Learning Models for Chronic Renal Disease Prediction. 173-182. 10.1007/978-981 99-7820-5-14.

9. T. Yan, Y. Li, and J. He, "Comparison of machine learning and neural network models on fraud detection," Advances in Computational Intelligence, vol. 19, pp. 1–10, 2021.

10. Sunayna, S.S., Rao, S.N.T., Sireesha, M. (2022). Performance Evaluation of Ma chine Learning Algorithms to Predict Breast Cancer. In: Nayak, J., Behera, H., Naik, B., Vimal, S., Pelusi, D. (eds) Computational Intelligence in Data Min ing. Smart Innovation, Systems and Technologies, vol 281. Springer, Singapore. https://doi.org/10.1007/978-981-16-9447-9-25

11. Kou, Y., Lu, C.-T., Sirwongwattana, S., Huang, Y.-P. (2004). Survey of fraud detection techniques. IEEE International Conference on Networking, Sensing, and Control, 2, 749-754. https://doi.org/10.1109/ICNSC.2004.1297040

12. Mamidala, Sai Sireesha, M. Rao, s Bolla, Jhansi Reddy, K.. (2024). Machine Learning Models for Chronic Renal Disease Prediction. 173-182. 10.1007/978-981 99-7820-5-14.

13. Moturi S., Tirumala Rao S.N., Vemuru S. (2021) Risk Prediction-Based Breast Cancer Diagnosis Using Personal Health Records and Machine Learning Mod els. In: Bhattacharyya D., Thirupathi Rao N. (eds) Machine Intelligence and Soft Computing. Advances in Intelligent Systems and Computing, vol 1280. Springer, Singapore. https://doi.org/10.1007/978-981-15-9516-5-37

14. Phua, R., Lee, L., Smith, P. (2020). Anomaly detection with unsupervised clus tering techniques: Applications in fraud detection. Data Mining and Knowledge Discovery, 17(2), 456–475.

15. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569.

QUESTIONS & ANSWERS

- Open the floor for questions from the audience

ACKNOWLEGEMENTS

- We would like to express our sincere gratitude for the opportunity to present our work on **“Online payments Fraud Detection by using Machine Learning Models"**. This platform has allowed us to showcase our research and share its potential impact on precision online payments. We deeply appreciate the valuable feedback and insights from the audience.
- For further inquiries or discussions, please feel free to reach out:
- **Contact Information**:
  - **Name**: [Chappidi Sandeep]
  - **Email**: [chappidisandeep75@gmail.com]
  - **Phone**: [8801156926]
  - **Affiliation**: [Narasaraopeta Engineering College]

Thank you