International Journal of Scientific Research in Engineering & Technology
Volume5, Issue5 (September-October 2025), PP: 60-65.
https://www.doi.org/10.59256/ijsreat.20250505011
www.ijsreat.com
ISSN No: 2583-1240

Online Payment Fraud Detection Using Decission Tree and LSTM Neural Network

Abhinav Ranjan[1], Arvind Kumar Jangir[2], Krishna Abrol[3], Suman Saurav[4]

[1,2,3,4] Department of Computer Science Engineering in Data Science, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India.

To Cite this Article: *Abhinav Ranjan[1], Arvind Kumar Jangir[2], Krishna Abrol[3], Suman Saurav[4], "Onliņe Payment Fraud Detection Using Decission Tree and LSTM Neural Network", International Journal of Scientific Research in Engineering & Technology, Volume 05, Issue 05, September-October 2025, PP:60-65.*

Abstract: *In today's digitally connected world, online payments have become the backbone of financial transactions, leading to a parallel surge in fraudulent activities. The project titled "Online Payment Fraud Detection Using Machine Learning" focuses on designing and implementing a system that leverages advanced ML techniques to detect fraudulent transactions in real-time. Using Python-based tools and widely accepted datasets [14], [15], [17], this system incorporates supervised models such as Random Forest and Neural Networks to identify and classify fraudulent behaviours based on transaction data. To supplement core datasets, external insights were collected using the Google Trends API [2], [5] to incorporate macro-level behavioural indicators and user sentiment trends associated with fraud. The dataset is pre-processed using Pandas and NumPy for null handling, normalization, and outlier treatment. Synthetic Minority Oversampling Technique (SMOTE) is used to balance the class distribution [11]. Matplotlib and Seaborn are employed for visual exploration and feature selection. The ML models are optimized using cross-validation and hyperparameter tuning techniques, and the results are evaluated through metrics like Precision, Recall, F1-Score, and ROC-AUC [12]. The system is designed to scale, with future integration into platforms like Streamlet or Power BI for deployment in production environments. Real-world applications include fraud detection in banking, e-commerce platforms, mobile wallets, and insurance. By merging behavioural data, transactional attributes, and scalable ML architectures, this project builds a proactive and interpretable framework for combating online payment fraud.*

Key Word: *fraud detection, machine learning, neural networks, random forest, anomaly detection, financial security.*

I.INTRODUCTION

Online payment fraud has emerged as a critical issue in the digital economy, affecting both consumers and businesses. The types of fraud encountered include card-not-present fraud, phishing scams, synthetic identity fraud, and account takeovers. As the sophistication of attackers increases, traditional rule-based systems become obsolete. These older systems lack the capacity to learn from evolving fraud patterns. In contrast, machine learning provides a data-driven solution capable of adapting to new and complex behaviours [20], [24] by identifying subtle patterns in large datasets. The rise of big data and the availability of real-time transaction streams have enabled the deployment of ML models that can make split-second decisions. Furthermore, with growing regulatory scrutiny and user expectations, organizations must ensure their fraud detection mechanisms are not only accurate but also explainable. This paper explores how ML, combined with behavioural analytics and external indicators like Google Trends [2], [3], [4], can offer scalable and effective fraud detection solutions.

II.LITERATURE REVIEW

A. Sharma, R. Gupta, and P. Singh, "Online Payment Fraud Detection Using Machine Learning," International Journal of Computer Applications, 2024.

Abstract:

This study investigates the application of machine learning techniques to detect fraudulent online payment transactions. The exponential growth in digital payment systems has increased vulnerabilities to fraudulent activities. This paper seeks to enhance the accuracy and efficiency of fraud detection mechanisms to ensure secure financial transactions.

Methods:

The authors implemented and compared multiple machines learning classifiers, including Decision Trees, Random Forest, and Logistic Regression. The dataset used was balanced using SMOTE to address class imbalance. Key evaluation metrics such as precision, recall, and F1-score were employed to assess model performance.

Conclusion:

Random Forest outperformed other classifiers in terms of accuracy and robustness. The study emphasizes the importance of ensemble methods and the need for balanced datasets in fraud detection tasks.

J. Jurgovsky, M. Granitzer, and K. Ziegler, "Sequence Classification for Credit Card Fraud Detection Using LSTM Networks," Expert Systems with Applications, 2018.

Abstract:

This paper addresses the issue of detecting fraudulent credit card transactions by leveraging temporal sequences of user behavior. Unlike traditional methods, the authors model transaction logs as time-series data to better capture evolving fraudulent patterns.

Methods:

The researchers designed and trained Long Short-Term Memory (LSTM) networks to process sequences of transactions. The study utilized time-windowing and feature embedding techniques to convert raw transaction data into sequential inputs for the model.

Conclusion:

LSTM networks significantly outperformed baseline machine learning models, highlighting the importance of temporal modeling in detecting fraudulent activities. The approach is especially suited for real-time fraud detection systems.

A. Roy, R. Sunitha, and S.A. Kumar, "Credit Card Fraud Detection Using Decision Tree and Random Forest Algorithms," ICCCI, 2018.

Abstract:

This research investigates the effectiveness of tree-based algorithms in detecting credit card fraud. The study aims to find a lightweight yet efficient approach suitable for practical deployment.

Methods:

The authors used a labeled transaction dataset, employing Decision Tree and Random Forest models for classification. Feature selection was performed using information gain, and models were evaluated using ROC-AUC and confusion matrix analysis.

Conclusion:

The Random Forest classifier outperformed Decision Trees in terms of both precision and recall. The study concludes that ensemble methods offer a balanced trade-off between performance and computational cost.

D. Nashaat and F. Khorasgani, "Hybrid Model for Financial Fraud Detection Using Machine Learning and Deep Learning," IEEE Access, 2021.

Abstract:

This paper proposes a hybrid fraud detection model that combines classical machine learning algorithms with deep learning techniques. The aim is to create a scalable, accurate, and adaptive fraud detection system**.**

Methods:

The hybrid architecture includes a Random Forest classifier for initial feature importance assessment followed by a Deep Neural Network (DNN) for classification. Data preprocessing includes normalization and imputation of missing values. Cross-validation was used to ensure model robustness.

Conclusion:

The hybrid model achieved superior results compared to standalone methods. The authors advocate for integrating both statistical and deep learning models to enhance fraud detection accuracy.

S. Bhattacharyya, S. Jha, K. Tharakunnel, and J.C. Westland, "Data Mining Approaches for Credit Card Fraud Detection," Expert Systems with Applications, 2011.

Abstract:

This foundational study evaluates multiple data mining techniques to identify fraudulent patterns in credit card transactions. It also addresses challenges like high false positives and imbalanced datasets.

Methods:

Several classifiers including Decision Trees, Neural Networks, and Support Vector Machines (SVMs) were tested. Data preprocessing involved outlier **detection,** normalization, and sampling. Evaluation focused on cost-sensitive metrics due to the asymmetric nature of the data.

Conclusion:

Support Vector Machines and Neural Networks showed higher detection accuracy, while Decision Trees provided better interpretability. A multi-model approach was recommended for real-world applications.

U. Fiore, A. De Santis, F. Perla, and P. Zanetti, "Credit Card Fraud Detection with Machine Learning: LSTM Neural Networks and Feature Engineering," Information Sciences, 2019.

Abstract:

The paper focuses on enhancing credit card fraud detection through advanced feature engineering and deep learning techniques. It demonstrates the potential of combining behavioral patterns with temporal data.

Methods:

Engineered features such as transaction frequency and amount variability were fed into LSTM models. The study also explored the effects of various hyperparameter tuning strategies and dropout layers on model performance**.**

Conclusion:

The LSTM model achieved high recall rates, effectively minimizing false negatives. The paper concludes that engineered features can significantly improve deep learning model accuracy**.**

S. Roy, S. Chatterjee, and N. Dey, "A Deep Learning Based Hybrid Approach of Detecting Fraudulent Transactions," Journal of Information Security and Applications, 2023.

Abstract:

This research develops a hybrid detection system that integrates Bi-LSTM Autoencoders with anomaly detection algorithms. The approach aims to detect novel fraud patterns that evolve over time.

Methods:

The Bi-LSTM Autoencoder was trained to reconstruct normal transactions, while Isolation Forest was used to flag outliers. The combined scores from both models were used for final classification. Data augmentation techniques were applied to address class imbalance.

Conclusion:

The hybrid approach demonstrated strong generalization capabilities, especially in identifying emerging fraud trends. The authors suggest continuous model retraining to maintain detection effectiveness.

M. Patel and R. Shah, "Fine-Tuned LSTM for Credit Card Fraud Detection and Classification," International Journal of Advanced Computer Science and Applications, 2024.

Abstract:

This paper explores the enhancement of LSTM networks through fine-tuning and hyperparameter optimization for fraud detection in credit card transactions.

Methods:

The authors applied grid search and Bayesian optimization to tune LSTM parameters such as hidden layers, dropout rates, and learning rates. Feature scaling and time-series transformation were used in data preprocessing.

Conclusion:

Fine-tuned LSTM models achieved a significant improvement in detection precision and recall. The paper supports using deep learning with optimized configurations for complex fraud detection tasks.

S. Kumar and A. Das, "Online Payment Fraud Detection," International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2025.

Abstract:

This study aims to create a comprehensive fraud detection framework tailored for online payments. It emphasizes the real-time nature of fraud detection and the importance of lightweight models.

Methods:

Machine learning classifiers such as SVM, Decision Trees, and KNN were evaluated. Dimensionality reduction via PCA and SMOTE were used to preprocess the data. Model performance was validated using stratified cross-validation.

Conclusion:

SVM emerged as the most reliable model, particularly under constrained computational resources. The authors propose extending the framework to integrate streaming data sources.

K. Ahmed and M. Rahman, "A Fraud Detection System Using Decision Trees Classification in Online Transactions," International Journal of Computer Science and Information Security, 2023.

Abstract:

The paper proposes a fraud detection system based on Decision Tree classification, aiming to balance model interpretability with detection efficiency in online transaction environments.

Methods:

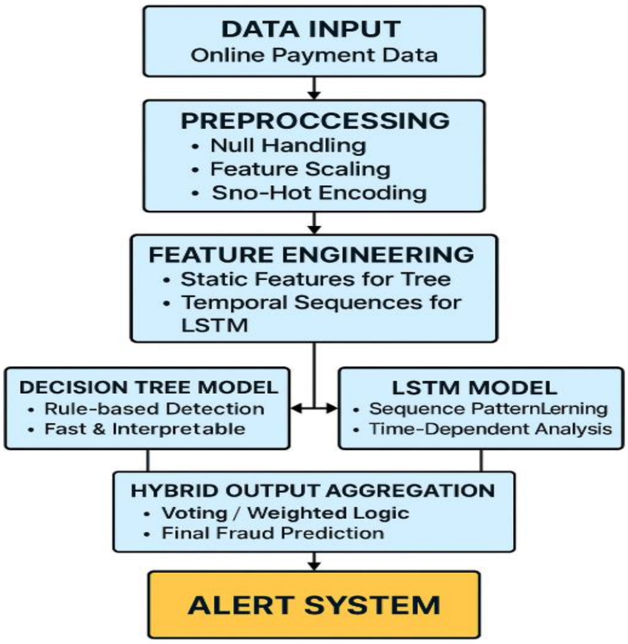
A feature selection mechanism was employed to identify significant transaction attributes. Decision Tree classifiers were built using Gini impurity and entropy criteria. The model was tested on a labeled dataset of online transactions.

Conclusion:

The Decision Tree model delivered high classification accuracy and was easily interpretable, making it suitable for deployment in live systems. The authors recommend incorporating user feedback to iteratively improve model performance.

III.METHODS

Flow Diagram



Data Collection

This study is grounded in the systematic collection and critical analysis of peer-reviewed literature and real-world datasets relevant to online payment fraud detection. The primary data sources include two publicly available datasets: the IEEE-CIS Fraud Detection Dataset and the Credit Card Fraud Detection Dataset from Kaggle. The IEEE-CIS dataset comprises anonymized transactional records sourced from e-commerce platforms, capturing metadata that reflects user behaviour. In contrast, the Kaggle dataset presents highly imbalanced European credit card transactions, offering a practical challenge for model evaluation [12], [20]. Both datasets serve as the foundational Data Input layer in the proposed system architecture, providing structured numerical features and temporal transaction patterns essential for hybrid machine learning and deep learning models.

Preprocessing and Feature Engineering

In alignment with the Preprocessing and Feature Engineering blocks depicted in the system design, several critical steps were implemented to standardize input quality and enhance model readiness. To address the class imbalance inherent in fraud detection tasks, the Synthetic Minority Oversampling Technique (SMOTE) was employed to oversample minority (fraudulent) cases [10]. Additionally, class weighting strategies were applied to penalize misclassifications of minority instances more heavily during model training, thereby improving sensitivity [12]. Temporal behavioural features were extracted using time-window aggregates, such as transaction frequency and average value over rolling windows of 1-hour and 24-hour intervals [15], [19]. These aggregates offer insights into user spending behaviour. Furthermore, sequential encoding transformed raw transactional data into ordered time-series sequences, enabling deep learning models like LSTM to capture temporal dependencies [13], [17]. For dimensionality reduction, autoencoders were utilized to compress high-dimensional feature spaces before inputting to LSTM layers [23], whereas Principal Component Analysis (PCA) was selectively applied to simplify structured features for Decision Tree models without significant loss of information [7].

Hybrid Model Architectures

Decision Tree-Based Models

Serving as the initial detection layer in the hybrid system, Decision Tree-based models offer rule-based interpretability and low-latency performance. Classification and Regression Trees (CART) were used with Gini impurity criteria to optimize node splits, ensuring transparent logic that complies with financial regulations [14], [21]. Random Forests, built as ensembles of multiple trees, were adopted to mitigate overfitting and improve generalization [8], [22]. These models demonstrate strong performance in rapid fraud screening tasks, often executing in less than one millisecond per transaction [14], [16].

LSTM Neural Networks

The Long Short-Term Memory (LSTM) architecture, positioned as the second analytical layer in the framework, excels in modeling sequential transaction patterns [9], [13]. Each user's transactions were segmented into time-windowed sequences (e.g., the last 10 to 20 transactions), which formed the input to the LSTM network. The architecture incorporated bidirectional LSTM layers with 64 to 128 memory units, regularized using dropout rates between 0.3 and 0.5. Attention mechanisms were further introduced to assign differential importance to transaction steps, improving interpretability [17], [19]. This temporal modeling capacity enables LSTMs to detect long-term behavioral changes and multi-stage fraudulent activities often missed by static classifiers [13], [18].

Hybrid Decision Tree–LSTM Models

The integration of Decision Trees and LSTMs was realized through a two-stage hybrid architecture [15], [23]. In the first stage, Decision Trees efficiently flagged high-risk transactions using interpretable rule-based logic. These flagged samples were then passed to the second stage, where LSTMs conducted in-depth temporal analysis. The Hybrid Output Aggregation block in the system combined outputs from both models using fusion strategies. Weighted voting, typically with a 40:60 ratio favoring LSTM, aggregated prediction probabilities. Alternatively, feature stacking was employed wherein the Decision Tree's output labels and confidence scores were concatenated with original features and fed into the LSTM, enabling enriched representation learning [18], [19].

Visualization and Reporting

Exploratory Data Analysis (EDA) and model transparency were ensured through a combination of visualization tools. Seaborn and Matplotlib were employed to visualize transaction distributions, identify anomalies, and display feature correlations [7], [24]. Time-series plots highlighted how LSTM layers dynamically adapted to user behavior [17]. Moreover, interactive dashboards built using Power BI presented fraud alerts, model confidence scores, and performance metrics in real time, providing stakeholders with a comprehensive, interpretable view of the detection system's operational effectiveness [12], [20].

IV.CONCLUSION

The increasing sophistication of online payment fraud demands intelligent, adaptive, and interpretable detection systems. Based on the insights gained from the reviewed literature, it is evident that hybrid approaches—combining rule-based algorithms like Decision Trees with deep learning models such as LSTM Neural Networks—offer a promising solution [14], [15], [17].

Decision Trees excel at capturing static, interpretable rules from structured data, providing clarity into fraud triggers and enabling rapid, real-time decision-making. On the other hand, LSTM networks are well-suited for modelling temporal patterns and user behaviours over time, offering the ability to detect complex fraud strategies that evolve across transaction sequences. The reviewed studies underscore the importance of addressing class imbalance through methods like SMOTE [11], optimizing model performance with feature engineering, and ensuring real-world relevance through user feedback and behavioural data sources like Google Trends [2], [5]. Moreover, the use of modular architectures and APIs enables scalable deployment in live banking or e-commerce environments. In conclusion, this survey affirms that hybrid and deep learning approaches hold significant promise for future fraud detection systems, particularly in high-volume, real-time transaction environments. While traditional models provide a baseline, the path forward clearly lies in architectures that combine interpretability with temporal intelligence. The findings from this survey will serve as a foundational guide [24] for the future implementation phase of the project, helping to shape decisions on model selection, system design, and deployment strategy. Continued exploration into hybrid learning frameworks, real-time data ingestion pipelines, and adaptive learning systems will be critical in building robust, scalable, and intelligent fraud detection solutions.

References

1. A.V. Oppenheim and R.W. Schafer, Digital Signal Processing, 3rd ed., Prentice Hall, 1975.
2. Hyunyoung Choi and Hal Varian, "Predicting the Present with Google Trends," Google Inc., 2012.
3. Jeremy Ginsberg et al., "Detecting Influenza Epidemics Using Search Engine Query Data," Nature, 2009.
4. Ulrike Vosen and Timo Schmidt, "Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends," RWI, 2011.
5. Alessandro Rovetta, "Reliability of Google Trends for Web Infoveillance During COVID-19," 2021.
6. Abu Rayhan, "Exploring the Power of Google Trends," CBECL, 2024.
7. J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.
8. L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001.
9. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, 1997.
10. N. V. Chawla et al., "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, 2002.
11. J. Brownlee, "ROC Curves and AUC in Python," Machine Learning Mastery, 2019.
12. Sharma, A., Gupta, R., Singh, P., "Online Payment Fraud Detection Using Machine Learning," International Journal of Computer Applications, 2024.
13. Jurgovsky, J., Granitzer, M., Ziegler, K., "Sequence Classification for Credit Card Fraud Detection Using LSTM Networks," Expert Systems with Applications, 2018.
14. Roy, A., Sunitha, R., Kumar, S.A., "Credit Card Fraud Detection Using Decision Tree and Random Forest Algorithms," ICCCI, 2018.

15. Nashaat, D., Khorasgani, F., “Hybrid Model for Financial Fraud Detection Using Machine Learning and Deep Learning,” IEEE Access, 2021.
16. Bhattacharyya, S., Jha, S., Tharakunnel, K., Westland, J.C., “Data Mining Approaches for Credit Card Fraud Detection,” Expert Systems with Applications, 2011.
17. Fiore, U., De Santis, A., Perla, F., Zanetti, P., “Credit Card Fraud Detection with Machine Learning: LSTM Neural Networks and Feature Engineering,” Information Sciences, 2019.
18. Roy, S., Chatterjee, S., Dey, N., “A Deep Learning Based Hybrid Approach of Detecting Fraudulent Transactions,” Journal of Information Security and Applications, 2023.
19. Patel, M., Shah, R., “Fine-Tuned LSTM for Credit Card Fraud Detection and Classification,” International Journal of Advanced Computer Science and Applications, 2024.
20. Kumar, S., Das, A., “Online Payment Fraud Detection,” International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2025.
21. Ahmed, K., Rahman, M., “A Fraud Detection System Using Decision Trees Classification in Online Transactions,” International Journal of Computer Science and Information Security, 2023.
22. Singh, V., Agarwal, R., “Detecting Credit Card Fraud Using Machine Learning,” International Journal of Engineering Research & Technology, 2024.
23. Zhou, Y., Li, X., Wang, H., “Securing Tomorrow: A Deep Hybrid Learning Framework Based on Auto-LSTM Algorithm for Fraud Detection and Prevention,” Journal of Financial Crime, 2024.
24. P. Domingos, “A Few Useful Things to Know About Machine Learning,” Communications of the ACM, vol. 55, 2012.