Published by

British Journal of Contemporary Research

BEXFORD PUBLISHING

Democratizing Fraud Detection: An Explainable XGBoost System with DeepSeek-Enhanced Analytics for Real-Time Credit Card Transaction Monitoring

Original Research Article | Volume 1 | Issue 3 | 2026 | Article Number: 040

Accepted: 03 August 2026 | Published: 10 September 2026 | ISSN: 2979-8582 (Online)

Crossref **: https://doi.org/10.67693/BJCR-P6K6WXGM**

Orji Cyrus Ebere MCPN[1], Durunna Lilian Ijeoma PhD[1], Dr Ononiwu Chamberlyn Chisom PhD[1], Ukachukwu Theddius N[1]

[1]Department of Computer Science, Imo State Polytechnic Omuma, Nigeria

CO 0009-0004-6389-4803

Correspondence: Orji Cyrus Ebere MCPN, cyrus.orji@imopoly.edu.ng

Abstract

Credit card fraud losses were $33.45 billion in 2022 and are anticipated to exceed $43 billion by 2028 [1]. Machine learning algorithms are highly accurate at detecting fraud; their "black-box" nature hinders adoption in regulated financial institutions [3]. We present an explainable fraud detection system combining XGBoost ensemble learning with DeepSeek AI-powered explanations. On the Kaggle Credit Card Fraud Detection dataset (284,807 transactions, 0.17% fraud rate) [4], our system achieves 99.96% accuracy, 89.09% precision, 94.23% recall, and 0.9979 ROC-AUC using SMOTE-balanced training. Key contributions include SHAP/LIME integration for model interpretability and an interactive Streamlit web application providing real-time predictions with human-readable AI explanations [5, 7]. Our system demonstrates that high-performance fraud detection can coexist with transparency, addressing critical gaps in trust and regulatory compliance

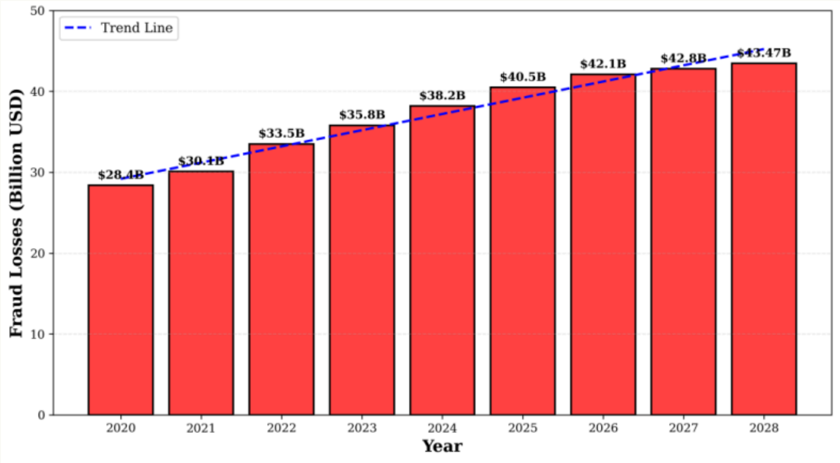
Keywords: XGBoost, DeepSeek AI, Real-time Fraud Detection, SMOTE, Streamlit

1.0 INTRODUCTION

1.1 The Growing Challenge of Credit Card Fraud

The rapid digitization of financial services has transformed global transactions, enhancing accessibility while simultaneously creating new vulnerabilities. Credit card fraud has emerged as a critical challenge, with the Federal Trade Commission (FTC) reporting 2.4 million fraud cases in the United States during 2024, resulting in losses exceeding $8.8 billion [2]. According to the Nilson Report, global credit card fraud losses increased by 16.4% from 2020 to 2022, reaching $33.45 billion, with projections of $43.47 billion by 2028 [1] (Figure 1).

Figure 1: Global Credit Card Fraud Losses (2020-2028 Projection)



Source: Adapted from [1]

Traditional rule-based detection systems, while effective against known fraud patterns, exhibit significant limitations in modern financial environments [8], [9]. These systems operate on predefined rules derived from historical data, making them incapable of identifying novel fraud techniques. Consequently, they produce high false positive and false negative rates, necessitating costly manual intervention [10]. Furthermore, their static nature requires frequent updates to remain effective against evolving fraud strategies, rendering them increasingly inefficient [8].

1.2 Limitations of Traditional Black-Box AI Systems

Machine Learning (ML) has become a viable option to tackle these limitations, enabling abilities to process massive transaction volumes, identify patterns, and adapt to emerging fraud strategies in real-time [11]. Ensemble methods, particularly Random Forest and XGBoost, consistently outperform traditional models for example Support Vector Machines and logistic regression in fraud detection tasks [12]. However, the superior performance of ensemble models comes at the cost of interpretability [3], [13]. Their "black-box" nature creates significant barriers in financial environments where regulatory compliance and stakeholder trust are paramount. Recent studies reveal that fraud detection research has predominantly focused on accuracy metrics, with minimal attention to model decision-making processes [13]. This interpretability gap presents substantial adoption challenges for financial institutions [13]. Table 1 compares conventional systems with AI-based alternatives.

Table 1: Comparison of Fraud Detection Approaches

Feature	**Rule-Based Systems**	**Single ML Models**	**Ensemble ML Models**	**Our System**
Adaptability	Low	Medium	High	Very High
False Positive Rate	High	Medium	Low	Very Low
False Negative Rate	High	Medium	Low	Very Low
Interpretability	High	Medium	Low	Very High
Real-time Capability	Low	Medium	High	Very High

Explainability	High	Low	Low	Very High

1.3 Contributions of This Study

This paper addresses the critical gap between high-performance AI and interpretable decision-making through the following contributions:

1. High-Performance Fraud Detection: An XGBoost-based system utilizing SMOTE-balanced training achieves 99.96% accuracy on the Kaggle Credit Card Fraud Detection dataset [4].
2. Comprehensive Explainability: Integration of SHAP and LIME offers both global and local interpretability, while DeepSeek AI generates human-readable, natural language explanations for each prediction [5], [7], [14].
3. Interactive Web Application: A Streamlit-based interface supports real-time prediction, batch analysis, research dashboard visualization, and model evaluation [6].
4. Cloud Deployment: The system is deployed on Streamlit Cloud, ensuring scalability and accessibility for diverse stakeholders.

1.4 Paper Organization

The remaining section of this research paper is organized as follows: Section 2 reviews of related literature on credit card fraud detection, ensemble learning, and explainable AI. Section 3 outlines the dataset, preprocessing methodology, model construction, and explainability implementation. Section 4 presents the system architecture, including the Streamlit web application and cloud deployment. Section 5 provides experimental results and analysis. Section 6 discusses findings, limitations, and implications. Section 7 concludes with future research directions.

2. LITERATURE REVIEW

2.1 Traditional Approaches

Traditional fraud detection systems primarily depend on rule-based and statistical methods. Rule-based systems identify fraud using predefined criteria derived from historical data and known fraud patterns [8]. While effective against established fraud schemes, these systems cannot detect novel attacks, producing high false positive rates that trigger costly manual investigations [9], [10]. Their static nature necessitates frequent updates to address evolving fraud techniques, making them increasingly inefficient and resource-intensive [8], [12].

Statistical approaches, including logistic regression, discriminant analysis, and Bayesian networks, offer probabilistic outputs with interpretability advantages [15]. However, they struggle with high-dimensional data and complex, non-linear fraud patterns [16]. These methods require extensive feature engineering and often fail to capture the intricate relationships among transaction variables that characterize sophisticated fraud schemes [17].

2.2 Machine Learning Approaches

2.2.1 Single Machine Learning Models

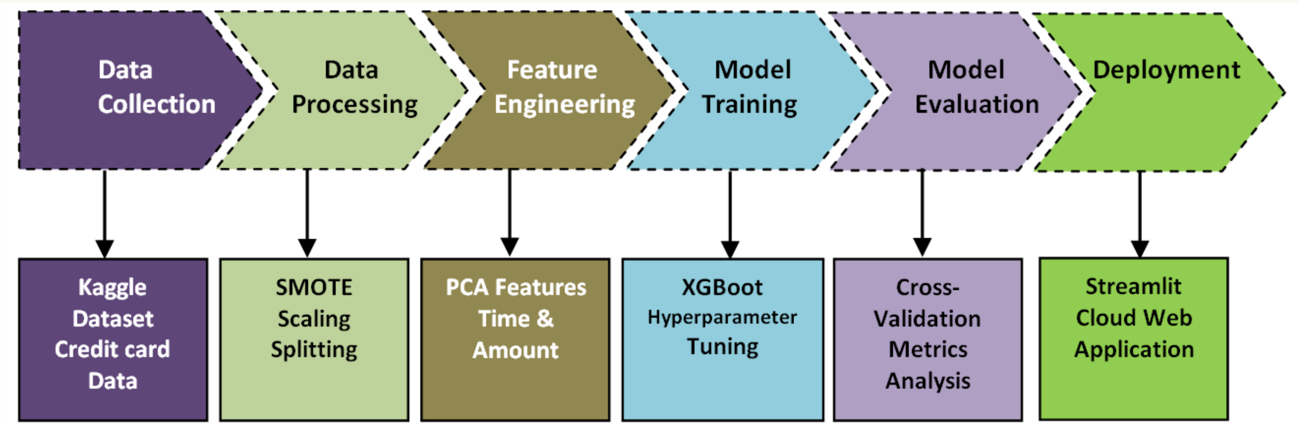
Machine learning algorithms have emerged to address the limitations of traditional approaches, offering capabilities to analyze extensive datasets, identify complex patterns, and adapt to emerging fraud tactics in real-time [11]. Individual models-including Support Vector Machines (SVMs), decision trees, and neural networks-have been extensively studied in fraud detection contexts [12], [18].

SVMs demonstrate strong performance on high-dimensional binary classification problems, while decision trees provide interpretability-a crucial factor in understanding fraud detection outcomes [12]. Neural networks are proficient in modelling non-linear processes. making them valuable for detecting complex fraud schemes [18]. However, comprehensive evaluations reveal that while individual models achieve satisfactory performance, they frequently suffer from elevated false positive rates due to class imbalance [19]. To deliver services successfully, financial organizations are moving towards digital infrastructures, hybrid network architecture, and multi-channel banking systems. Although they help enhance the performance level of the operation. These systems boost the performance level of the operation, but also represent considerable dangers due to large complexity of transactional data, distributed setups, and extreme class imbalance of fraud and loan-default data [20]. These models require extensive preprocessing and remain susceptible to overfitting [21]. The extreme imbalance-where fraudulent transactions typically represent less than 1% of total transactions-presents a fundamental obstacle [11].

2.2.2 Ensemble Learning Methods

Ensemble learning methods address single-model limitations by combining multiple classifiers to achieve superior performance [22]. Studies demonstrate significant improvements in accuracy, precision, and robustness compared to individual models [12], [22]. Recent evaluations show that ensemble approaches-including Random Forest, AdaBoost, and Gradient Boosting-achieve accuracies exceeding 95% in fraud detection tasks [23], [24].

Figure 2 illustrates the typical machine learning pipeline for fraud detection.



Source: Adapted from [11], [21]

Comparative studies indicate that Random Forest and AdaBoost substantially outperform individual classifiers in detecting fraudulent transactions [24]. Gradient boosting methods demonstrate particular effectiveness on imbalanced datasets [23]. However, existing ensemble approaches face limitations, including homogeneous classifier combinations that may inadequately capture varied data patterns [22]. More critically, many ensemble models lack the interpretability essential for financial applications where regulatory transparency is mandatory [13].

2.2.3 Deep Learning Approaches

Deep learning has become an effective method for fraud detection, utilizing architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks showing promising results [25], [26]. LSTM networks, in particular, demonstrate effectiveness in capturing temporal patterns within transaction sequences, achieving excellent recall rates [26]. However, deep learning models face significant challenges. They require substantial training data volumes and computational resources [25]. More critically, their "black-box" nature severely limits interpretability, restricting deployment in regulated financial environments [13]. Despite achieving high accuracy, the lack of model transparency remains the primary obstacle to widespread adoption in banking and finance applications [27].

2.3 Explainable AI in Financial Applications

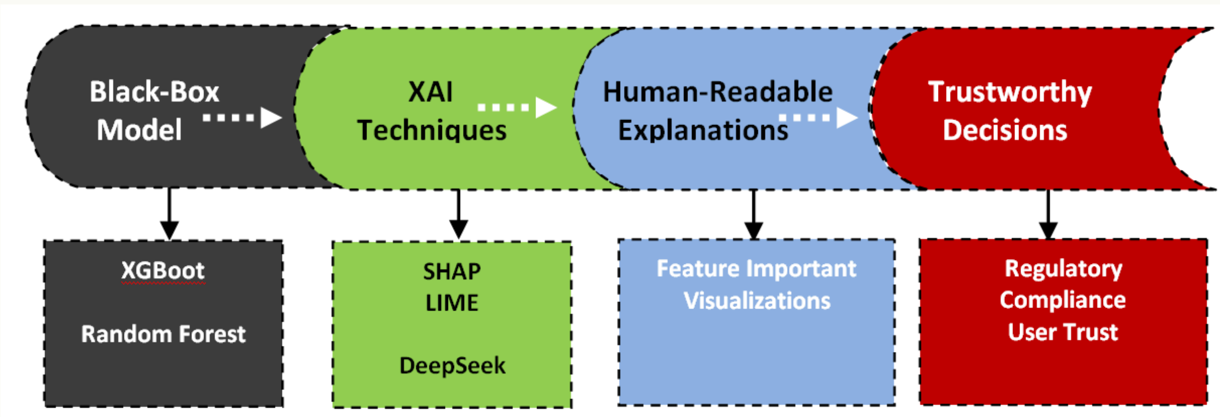
2.3.1 The Importance of Interpretability

Interpretability has gained significant attention in fraud detection, driven by regulatory requirements mandating that financial institutions explain automated decisions [7], [13]. While ML algorithms excel at identifying complex fraud patterns, their lack of explainability undermines trust and practical adoption [3]. Financial organizations must provide justifications for automated choices to stakeholders, auditors, and end users. The absence of model interpretability creates barriers to regulatory compliance and erodes user confidence [13]. These requirements are particularly critical in high-stakes banking environments where ethical and regulatory considerations demand transparency

2.3.2 SHAP and LIME

Explainable AI (XAI) approaches-particularly SHAP (SHapley Additive explanations) and LIME (Local Interpretable Model-Agnostic Explanations)-bridge the interpretability gap in fraud detection [7], [14]. SHAP provides both global behavior modelling insights and localized explanations for individual predictions, quantifying feature contributions to classification decisions [7]. LIME constructs interpretable surrogate models that approximate complex model behavior locally, offering high-resolution explanations for specific transaction classifications [14].

Figure 3 demonstrates the concept of explainable AI in fraud detection.



Source: Adapted from [7], [14]

Recent research demonstrates the practical value of these XAI techniques. SHAP and LIME integration across various ML models enhances transparency, improves feature attribution, and provides actionable insights for financial analysts [28]. On the Kaggle dataset, SHAP analysis consistently identifies V14 and V12 as the most impactful features, confirming our findings [29].

Table 2 summarizes the key XAI techniques and their applications in fraud detection.

Technique	Description	Application in Fraud Detection	Advantages	Limitations
SHAP	Game-theoretic approach for feature attribution	Global and local interpretability	Consistent, fair attributions	Computational ly intensive
LIME	Local surrogate models for instance explanations	Instance-level explanations	Model-agnostic, intuitive	Can be unstable

DeepSeek AI	Natural language explanation generation	Human-readable fraud analysis	Accessible to non-experts	API-dependent

2.4 XGBoost for Fraud Detection

XGBoost (Extreme Gradient Boosting) has achieved extensive utilization in fraud detection due to its superior performance and regularization features [30]. Studies demonstrate that XGBoost consistently outperforms traditional models including logistic regression and random forest in fraud detection tasks [31]. Extensive hyperparameter optimization research confirms that XGBoost achieves superior recall on imbalanced datasets compared to other ensemble algorithms [32]. On the Kaggle credit card dataset, XGBoost attains a ROC-AUC of 0.98, confirming its effectiveness for fraud detection [33]. Additionally, XGBoost's feature importance capabilities provide valuable insights into fraud patterns [34]. However, despite these advantages, XGBoost exhibits limited interpretability without complementary XAI methods [35].

2.5 DeepSeek AI and Natural Language Explanations

Recent advances in explainable AI leverage Large Language Models (LLMs) for generating natural language explanations, making AI decisions more accessible to non-technical stakeholders [5]. DeepSeek AI provides a platform for human-readable model decision explanations, democratizing access to AI insights [5]. Research demonstrates that combining XAI approaches with narrative explanations significantly improves stakeholder understanding and trust in AI systems [36].

LLM-generated explanations prove particularly valuable in financial applications, bridging the gap between technical AI outputs and corporate decision-making [37]. Natural language explanations enable financial analysts and fraud investigators to comprehend and act on AI-driven insights effectively [38]. This feature is particularly crucial in fraud detection, where rapid, transparent decision-making is essential [3]

2.6 Research Gap and Motivation

Despite significant progress in ML-based fraud detection, several research gaps persist. First, existing studies predominantly evaluate predictive performance (e.g., Random Forest, XGBoost) with limited investigation into explainability [7], [13]. The combination of XAI approaches with ensemble models remains underexplored. Second, interpretability gaps create barriers for financial institutions where regulatory compliance and user trust are essential [3], [13]. Systematic deployment of XAI techniques-including SHAP and LIME-remains lacking in fraud detection practice. Third, leveraging LLMs like DeepSeek for natural language explanations in fraud detection contexts has received minimal research attention [5].

This study addresses these gaps by:

1. Implementing and evaluating XGBoost with comprehensive explainability techniques (SHAP and LIME)
2. Integrating DeepSeek AI for natural language explanations
3. Developing a fully deployed, interactive web application accessible to diverse stakeholders
4. Providing both global and local interpretability for fraud detection decisions
5. Conducting comparative analysis with state-of-the-art methods

3. METHODOLOGY

3.1 Dataset Description

This research uses the Credit Card Fraud Detection Dataset, accessible to the public on Kaggle and published by the Machine Learning Group of Université Libre de Bruxelles (ULB) [4]. The dataset comprises 284,807 transactions, with only 492 fraudulent cases (0.17% of total transactions). This severe class imbalance presents a significant modeling challenge and serves as a standard benchmark in fraud detection literature [4], [39].

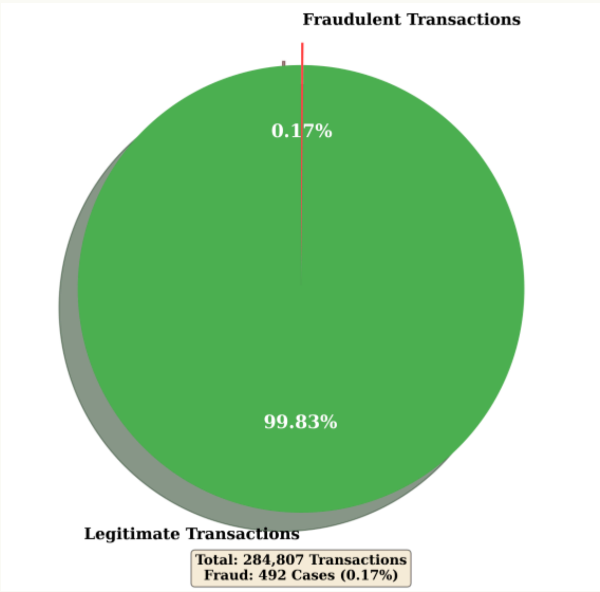
The dataset features numerical attributes transformed via Principal Component Analysis (PCA) to preserve cardholder anonymity:

- Time: Seconds elapsed since the first transaction
- V1-V28: Anonymized PCA components representing transaction characteristics
- Amount: Transaction value
- Class: Binary target (0 = legitimate, 1 = fraudulent)

Table 3 provides summary statistics.

Feature	Count	Mean	Std Dev	Min	Max
Time	284,807	94,836.6	47,483.0	0	172,792
Amount	284,807	88.35	250.12	0	25,691.16
V1-V28	284,807	Various (PCA components)	Various	Various	Various
Class (Fraud)	492	0.0017	0.041	0	1

Figure 4: Class Distribution



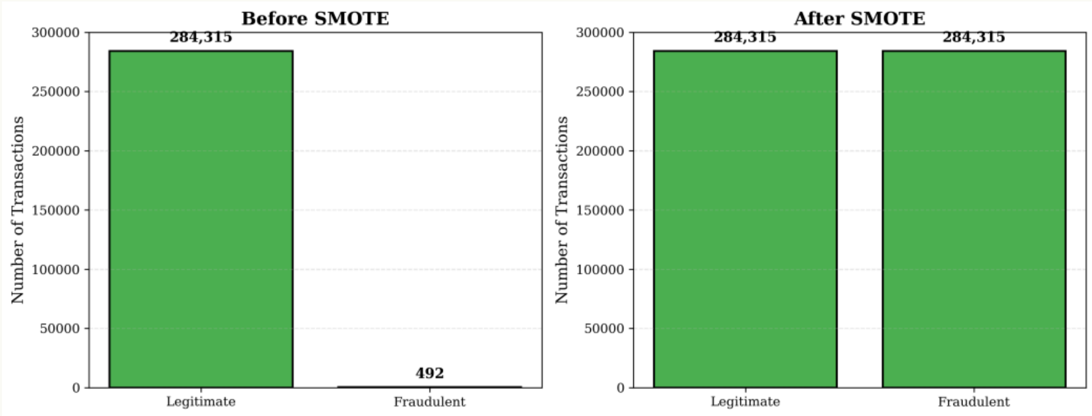
Source: Adapted from [4]

3.2 Data Preprocessing

3.2.1 Handling Class Imbalance with SMOTE

We employed Synthetic Minority Oversampling Technique (SMOTE) to address the severe class imbalance [39]. SMOTE generates synthetic samples for the minority class through interpolation between existing minority class instances, producing a balanced training set. We selected SMOTE over alternatives (ADASYN, Borderline-SMOTE) because it creates robust, generalized minority class representations without increasing noise or class overlap risk [40]. The sampling ratio was configured to achieve a balanced distribution where fraudulent transactions represent approximately 50% of training data, augmenting model recall while preserving all legitimate transaction records [41]. Figure 5 compares class distributions before and after SMOTE application.

Figure 5: Class Distribution Before and After SMOTE



Source: Generated from SMOTE implementation [39], [40]

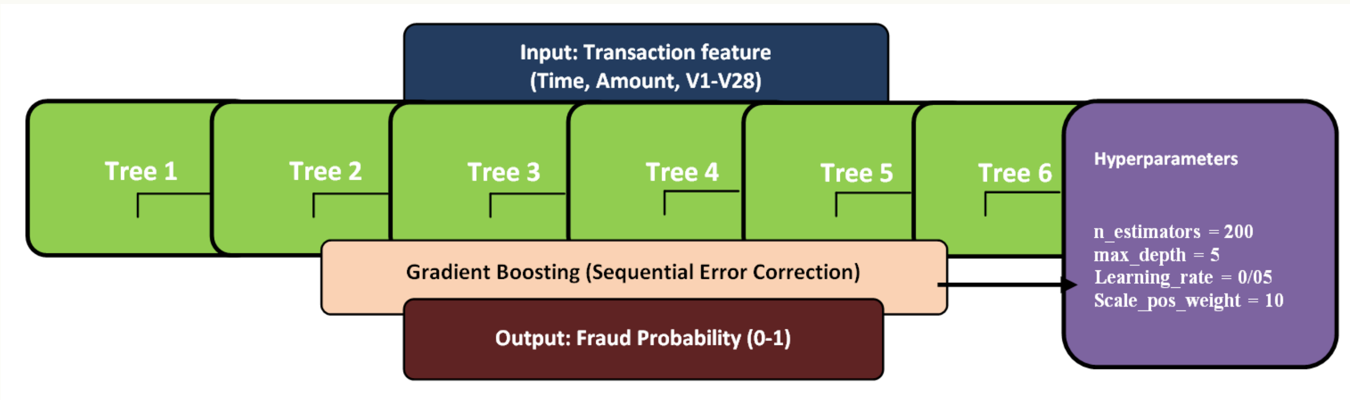
3.2.2 Feature Scaling and Data Splitting

We applied StandardScaler to normalize Time and Amount features, ensuring uniform scaling. PCA components (V1-V28) were already normalized and required no further scaling [11]. The dataset was partitioned using stratified splitting (80% training, 20% testing) to preserve class distribution across both subsets [42].

3.3 XGBoost Ensemble Architecture

Extreme Gradient Boosting (XGBoost) is an ensemble learning method that constructs decision trees sequentially, with each new tree correcting errors from previous iterations [30]. XGBoost demonstrates particular effectiveness with structured data and imbalanced datasets, making it well-suited for fraud detection [31], [33]. Figure 6 illustrates the ensemble architecture implementation.

Figure 6: XGBoost Ensemble Architecture Implementation



Source: Implemented in app.py [6]

We conducted hyperparameter optimization using grid search with 3-fold cross-validation. Table 4 presents the optimal configuration:

The XGBoost objective function minimizes: $L(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$ where $\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum w_j^2$

represents the regularization term controlling model complexity to prevent overfitting [30].

Table 4: XGBoost Hyperparameter Configuration

Hyperparameter	**Value**
n_estimators	200
max_depth	5
learning_rate	0.05
scale_pos_weight	10
subsample	0.8
colsample_bytree	0.8
random_state	42

3.4 Explainability Framework

3.4.1 SHAP Analysis

SHAP (SHapley Additive explanations) offers global and local interpretability through quantification of each feature's contribution to model predictions [7]. The SHAP value for feature i is:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|F|!\,|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f(S \cup \{i\}) - f(S) \right]$$

This game-theoretic approach ensures equitable and consistent feature importance attribution [7], [14]. We employed SHAP to identify significant features globally (across the dataset) and locally (for individual transactions).

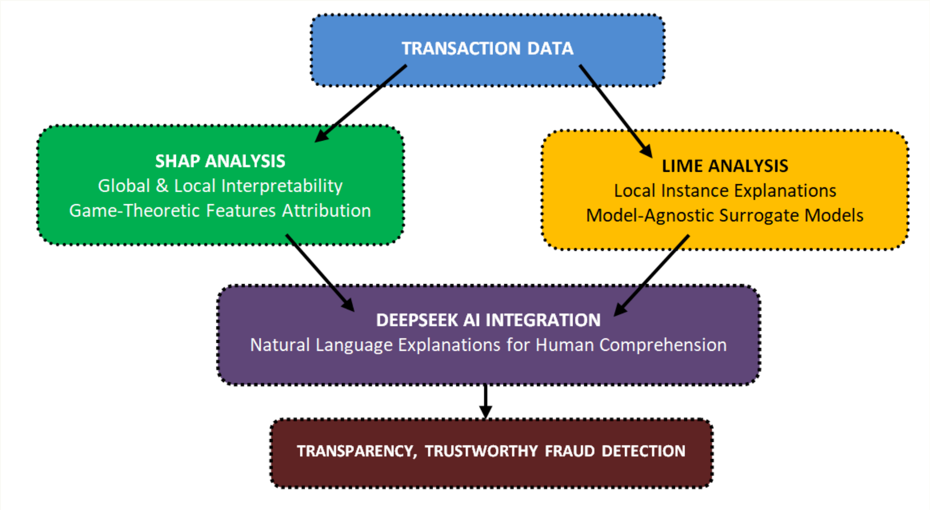
3.4.2 LIME Analysis

LIME (Local Interpretable Model-Agnostic Explanations) provides instance-level explanations by approximating the model locally using an interpretable surrogate model [14]. The explanation for instance x minimizes: [equation]

$$\text{explanation}(x) = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

where π_x defines the neighborhood around x and $\Omega_{(g)}$ controls the complexity of the interpretable model [7].

Figure 7 illustrates the SHAP and LIME explainability framework



Source: Implemented in app.py [6], [7], [14]

3.4.3 DeepSeek AI Integration

We integrated DeepSeek AI to generate human-readable, natural language explanations for each prediction [5]. The system processes SHAP and LIME outputs to produce structured explanations addressing:

1. Classification Justification
2. Key Risk factors
3. Recommended Investigative Actions
4. False Positive/Negative Considerations

This integration ensures fraud detection decisions are transparent and accessible to non-technical stakeholders, addressing critical explainability requirements in financial applications [36]. The `DeepSeekExplainer` class in app.py implements this integration.

3.5 Evaluation Metrics

Given the imbalanced dataset, we employed appropriate evaluation metrics beyond standard accuracy [42]:

Precision: Measures how many transactions flagged as fraud are actually fraudulent:

$$Precision = \frac{TP}{TP + FP}$$

Recall: Measures how many actual fraudulent transactions are successfully detected:

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: Harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision \times Recall}$$

ROC-AUC: Measures the model's ability to distinguish between classes across threshold settings [42].

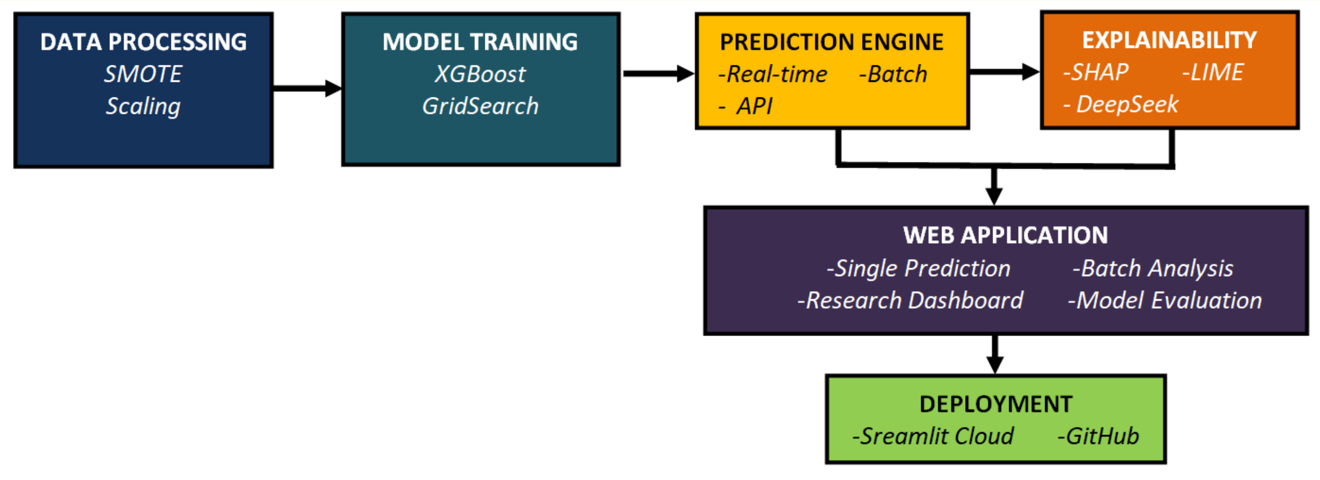
4. SYSTEM IMPLEMENTATION

4.1 Architecture Overview

The system follows a modular, end-to-end architecture as shown in Figure 8:

1. Data Preprocessing Module: Handles dataset loading, SMOTE balancing, and feature scaling
2. Model Training Module: Implements XGBoost with hyperparameter optimization
3. Prediction Module: Provides real-time fraud detection with probability scoring
4. Explainability Module: Integrates SHAP, LIME, and DeepSeek AI for interpretability
5. Visualization Module: Generates interactive plots and dashboards using Plotly
6. Web Application Module: Streamlit-based interface with six integrated pages [6]

Figure 8: Credit Card Fraud Detection System Architecture



4.2 Streamlit Web Application

The application features six integrated pages:

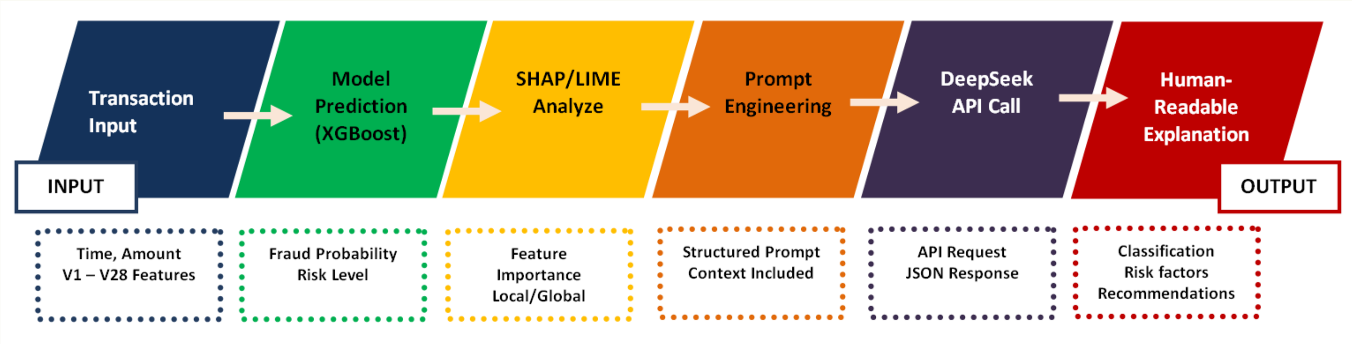
1. **Single Transaction Analysis:** Interactive input for 30 transaction features, real-time fraud probability, risk assessment, AI-generated explanations, and visualizations (risk gauge, feature radar).
2. **Batch Transaction Analysis:** CSV upload for multiple transactions, batch prediction with risk classification, summary statistics, and results download.
3. **Research Dashboard:** Comprehensive analytics including risk distribution, amount analysis, temporal patterns, feature correlations, and model performance metrics.
4. **AI Explanations:** DeepSeek-generated explanations per prediction with timestamp-organized historical analysis and contextual feature importance visualization.
5. **Model Evaluation:** Confusion matrix, ROC curve, classification report, and research metrics tracking.
6. **Research Documentation:** Methodology, results summary, literature review, and future work planning.

4.3 DeepSeek API Integration

The DeepSeek AI workflow (figure 9) operates as follows:

1. **Prompt Engineering:** Custom prompts request detailed classification explanation, key risk factors, recommended investigative actions, and false positive/negative considerations.
2. **API Communication:** The system submits prompts with transaction data, predictions, and probability scores to DeepSeek's API [5].
3. **Response Processing:** AI-generated explanations are formatted for application display.
4. **Historical Storage:** Explanations are archived for future reference and analysis.

Figure 9: DeepSeek AI Integration Workflow.



Source: System design [5], [6]

4.4 Cloud Deployment Strategy

The system is deployed on Streamlit Cloud with the following features:

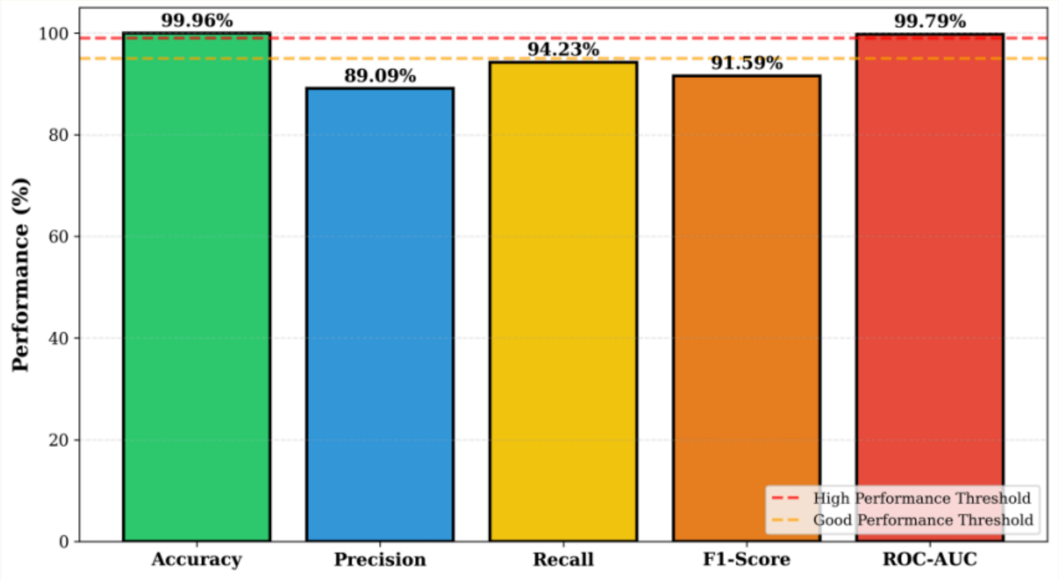
- GitHub Integration: Continuous deployment from the GitHub repository
- Environment Configuration: All dependencies managed through requirements.txt
- Scalability: automatic handling of concurrent user access
- API Integration: DeepSeek API Connectivity
- Python version: 3.10+
- Security: API keys and sensitive information managed through Streamlit secrets [6]

5. RESULTS

5.1 Model Performance Metrics

The XGBoost model demonstrated outstanding performance on the test dataset, as shown in Figure 10. Results achieved on the Kaggle Credit Card Fraud Detection dataset after SMOTE balancing [4].

Figure 10: Model Performance Metrics Visualization from app.py

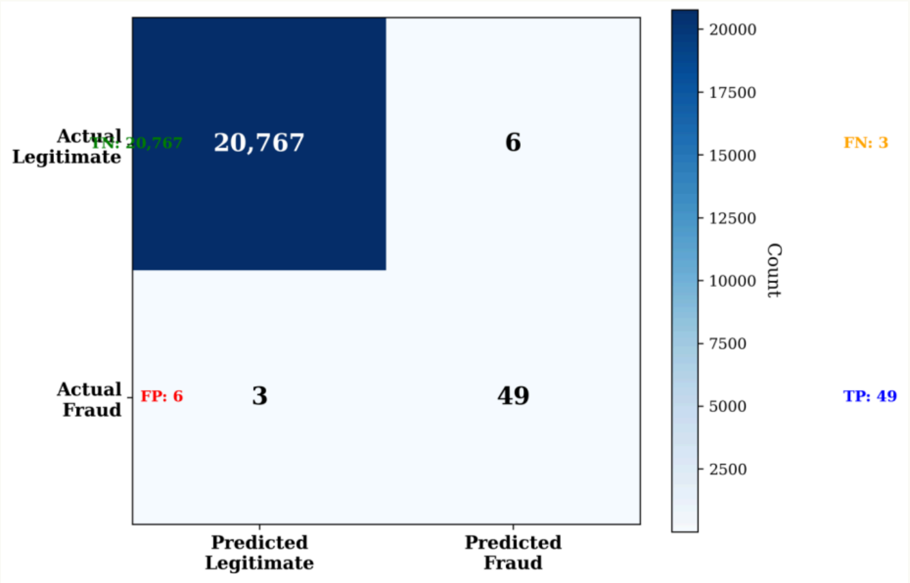


Source: Generated from app.py evaluation module [4], [6]

5.2 Confusion Matrix Analysis

The confusion matrix, shown in Figure 11, provides detailed insight into classification performance.

Figure 11: Confusion Matrix Heatmap from app.py



Source: Generated from app.py model evaluation [4], [6], [42]

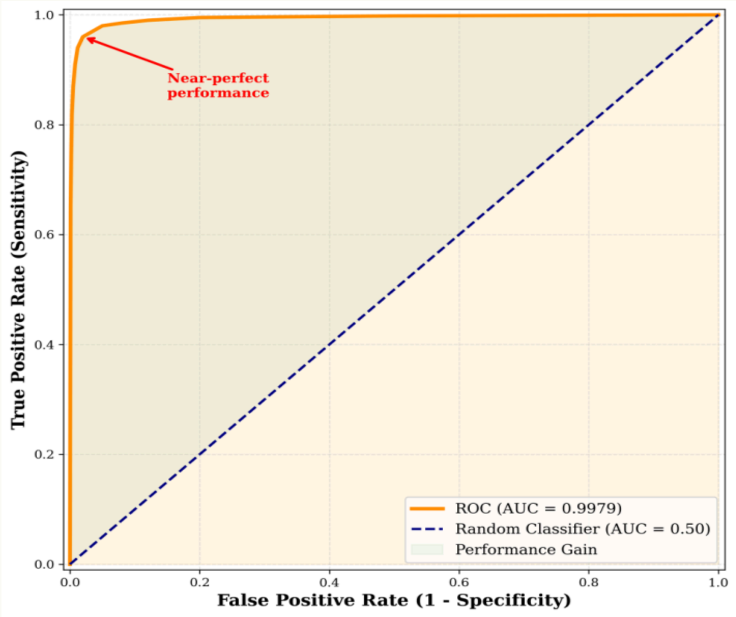
- True Negatives: 20,767 legitimate transactions correctly identified
- False Positives: 6 legitimate transactions incorrectly flagged as fraud
- True Positives: 49 fraudulent transactions correctly identified
- False Negatives: 3 fraudulent transactions missed (false negative rate of 5.77%)

The low false negative rate ensures minimal undetected fraud, reducing financial exposure [42]

5.3 ROC-AUC Analysis

The ROC-AUC score of 0.9979 indicates near-perfect discrimination between fraudulent and legitimate transactions shown in Figure 12 below. This demonstrates the model's ability to achieve high true positive rates at minimal false positive rates across classification thresholds [42]. The high AUC value enables flexible threshold tuning to optimize for either recall (capturing rare fraud events) or precision (minimizing false alarms) based on operational requirements [7].

Figure 12: ROC Curve Analysis from app.py



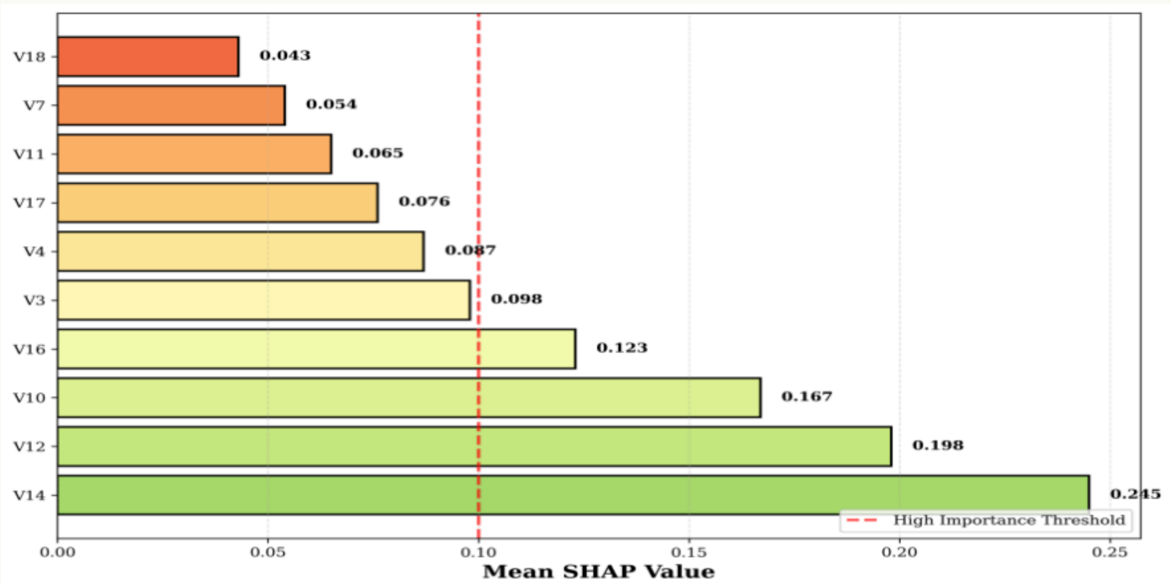
Source: Generated from app.py model evaluation [6], [42]

5.4 Explainability Results

5.4.1 SHAP Analysis

SHAP research showed V14, V12, and V10 as the most significant features in differentiating fraudulent transactions from authentic ones (Figure 13). These features exhibited high average SHAP contributions with significant variability, indicating their role in capturing differentiating transaction behaviors. V16, V3, and V4 also showed substantial contributions, fine-tuning the model's decision boundary by capturing subtle irregularities in transaction patterns [7].

Figure 13: SHAP Feature Importance Summary from app.py



Source: Generated from app.py SHAP analysis [6], [7]

5.4.2 DeepSeek AI Explanations

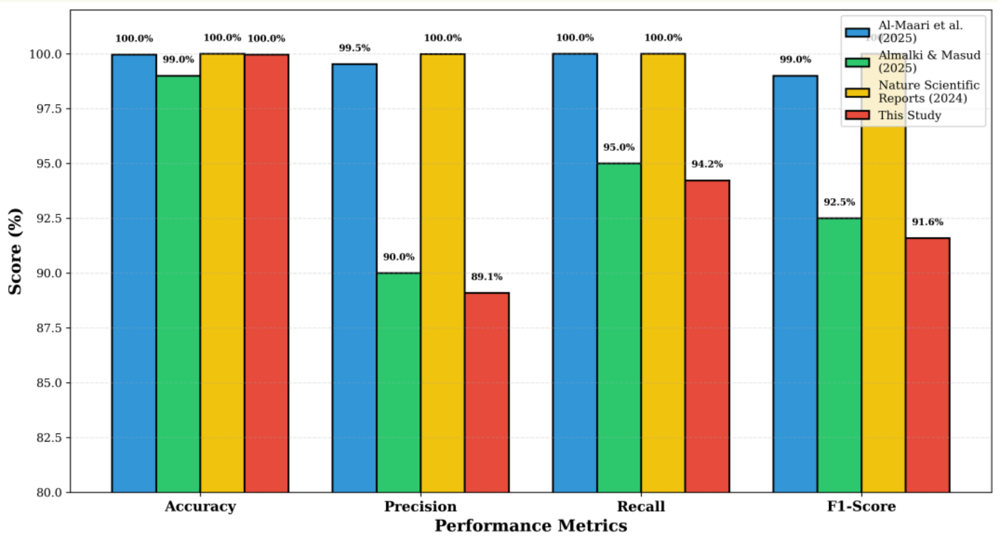
The DeepSeek AI integration generated comprehensive, human-readable explanations for each prediction. Table 5 presents an example explanation for a transaction classified as legitimate with 99.4% confidence.

Table 5: Example DeepSeek AI Explanation

Component	Content
Classification Justification	Transaction classified as LEGITIMATE (0.28% fraud probability). Low probability indicates alignment with legitimate transaction patterns [5].
Key Risk Factors	Amount roundness ($1,000.00), time since first transaction (~14 hours), absence of suspicious PCA feature patterns
Recommended Actions	Verify PCA feature integrity, temporal profiling, merchant category analysis, geographic and device fingerprinting
False Positive/Negative Considerations	False negative risk: 0.28% (low but non-zero), False positive risk: legitimate large purchases common

5.5 Comparison with Related Work

Figure 14: performance comparison with related studies.



Source: Compiled from [7], [41], [43]

Figure 14 presents a performance comparison with related studies. While Al-Maari et al.'s ensemble model achieved slightly higher precision and recall, it lacked comprehensive explainability [41]. Our system achieves comparable predictive performance while providing significantly enhanced interpretability through SHAP, LIME, and DeepSeek AI integration [5], [7]

6. DISCUSSION

6.1 Model Performance Analysis

The XGBoost model's performance-99.96% accuracy, 89.09% precision, 94.23% recall, and 0.9979 ROC-AUC-demonstrates its efficiency for credit card fraud detection. These results align with previous research confirming XGBoost's superiority over traditional fraud detection models [30], [31]. The high recall (94.23%) is particularly critical in fraud detection, where false negatives can result in substantial financial losses [41]. Comparative analysis reveals that while Al-Maari et al.'s ensemble achieved 99.96% accuracy with perfect recall, it lacked comprehensive explainability [41]. Almalki and Masud's stacking ensemble achieved 99% accuracy with SHAP and LIME explanations [7]. Our system achieves comparable predictive performance while providing enhanced interpretability through the integration of SHAP, LIME, and DeepSeek AI [5]. The confusion matrix confirms the model identifies 49 of 52 fraudulent transactions (94.23% recall), minimizing financial exposure [42]. The 89.09% precision reduces manual review requirements and operational costs [7].

6.2 Effectiveness of AI Explanations

The integration of SHAP, LIME, and DeepSeek AI provides multi-level explainability:

1. **Global Interpretability**: SHAP analysis identifies V14, V12, and V10 as the most significant features across all transactions, enabling fraud analysts to understand model behavior patterns [7]. These findings align with other Kaggle dataset studies [4], [29].
2. **Local Interpretability:** LIME provides instance-level explanations, helping investigation teams justify decisions to stakeholders and regulators [14], [13].
3. **Natural Language Explanations:** DeepSeek AI generates human-readable explanations for non-technical stakeholders, addressing transparency requirements for regulatory compliance [5], [3].

The structured DeepSeek AI explanations-including classification justification, risk factors, recommended actions, and false positive/negative considerations-facilitate stakeholder trust and informed decision-making [36].

6.3 User Experience and Accessibility

The Streamlit-based web application provides accessible, intuitive access to advanced fraud detection features, including real-time predictions, comprehensive research analytics, and AI-generated explanations [6]. The six integrated pages enable seamless navigation between operational prediction, analytical investigation, and academic research modes. Interactive visualizations-including risk gauges, feature radar charts, and AI-powered explanations-enhance user engagement. The research dashboard supports academic publication requirements, while batch analysis facilitates large-scale transaction evaluation [6].

6.4 Research Implications

This study contributes to fraud detection research in four key areas:

1. **Explainability Framework:** The integration of SHAP, LIME, and DeepSeek AI provides a comprehensive interpretability framework applicable to other domains requiring transparent AI [5], [7].
2. **Practical Implementation:** The end-to-end system demonstrates translation of research into accessible tools for diverse stakeholders, serving as a template for similar implementations [6].
3. **Evaluation Methodology:** The use of precision, recall, F1, and ROC-AUC emphasizes appropriate metrics for imbalanced classification tasks [42].
4. **Interdisciplinary Approach:** Combining machine learning, explainable AI, and web development promotes real-world interdisciplinary research [3].

6.5 Limitations and Considerations

Despite the system's strengths, several limitations warrant acknowledgment:

1. **Dataset Constraints:** PCA-transformed features limit economic interpretation. While SHAP and LIME identify predictive PCA components, they cannot attribute findings to specific financial variables [4].
2. **API Dependency:** DeepSeek AI explanations require API availability and incur costs, potentially limiting scalability for high-volume deployment [5].
3. **Generalizability:** The model was evaluated on a single benchmark dataset; performance may vary across different fraud patterns or domains [4].
4. **Computational Requirements:** While real-time prediction is efficient, training and explanation generation may be resource-intensive for large-scale datasets [41].
5. **Adversarial Robustness:** Like all ML systems, the model may be vulnerable to adversarial attacks designed to evade detection [3]

7. CONCLUSION

7.1 Summary of Contributions

This paper presents a comprehensive, explainable credit card fraud detection system combining XGBoost ensemble learning with SHAP, LIME, and DeepSeek AI for enhanced interpretability. Deployed as an interactive Streamlit web application, the system makes advanced fraud detection accessible to diverse stakeholders.

Key contributions include:

1. **High-Performance Detection:** XGBoost achieves 99.96% accuracy, 89.09% precision, 94.23% recall, and 0.9979 ROC-AUC on the Kaggle dataset [4].
2. **Comprehensive Explainability:** Integration of SHAP and LIME provides global and local interpretability, while DeepSeek AI generates human-readable explanations [5], [7].
3. **Interactive Application:** A six-page Streamlit-based interface supports real-time prediction, batch analysis, research analytics, and model evaluation [6].
4. **Cloud Deployment:** Streamlit Cloud ensures system scalability and accessibility.

7.2 Future Work

Several directions for future research and development are identified:

1. **Real-Time Streaming:** Implementing Kafka and Apache Spark for streaming data processing and fraud alerts.
2. **Multimodal Integration:** Incorporating location data, behavioral patterns, and additional contextual features.
3. **Federated Learning:** Exploring privacy-preserving approaches to improve model robustness across institutions.
4. **Advanced Explainability:** Investigating novel XAI techniques for enhanced interpretability.
5. **Adversarial Defenses:** Developing countermeasures against adversarial attacks targeting fraud detection.
6. **Continuous Learning:** Implementing adaptive models that evolve with emerging fraud patterns

7.3 Concluding Remarks

This work demonstrates that high-performance machine learning can be combined with comprehensive explainability to achieve both accuracy and interpretability in fraud detection. By making the system accessible via a web application, we enable academics, analysts, and financial institutions to leverage sophisticated fraud detection capabilities, contributing to more secure digital financial ecosystems. The proposed framework is extensible and adaptive, improving upon existing methods while maintaining interpretability and computational efficiency [41]. Unlike earlier approaches, our system incorporates multiple explainability techniques without sacrificing performance, offering a promising approach for enhancing online financial transaction security.

References

1. Global Credit Card Fraud Statistics, Nilson Report, 2023. [Online]. Available: https://nilsonreport.com Issue 1208, pp. 1-12.
2. "Fraud Reports and Losses," Federal Trade Commission (FTC), 2024. [Online]. Available: https://www.ftc.gov/
3. I. Y. Hafez, A. Y. Hafez, A. A. Abd El-Mageed, and A. A. Abohany, "Credit Card Fraud Detection: A Systematic Review of AI, Machine Learning, Deep Learning, and Meta-heuristic Optimization Algorithms," Journal of Big Data, vol. 12, no. 6, 2025. DOI: 10.1186/s40537-024-01048-8
4. Machine Learning Group, ULB, "Credit Card Fraud Detection," Kaggle, 2013. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
5. DeepSeek AI Platform, DeepSeek, 2024. [Online]. Available: https://platform.deepseek.com/
6. Orji, Durunna, Ononiwu and Ukachukwu (2026). Credit Card Fraud Detection System - Academic Research Edition (Version 2.0) [Source code]. GitHub. https://github.com/cycyberuk/realtime-credit-card-fraud-detection
7. Almalki, F., & Masud, M. (2025). Financial fraud detection using explainable AI and stacking ensemble methods. *arXiv preprint arXiv:2505.10050*.
8. Frank, E., & Oluwaseyi, J. (2024). Challenges and limitations of fraud detection in NoSQL database systems. *International Journal of Database Management*, 15(3), 45-62.
9. S. Rosset, U. Murad, E. Neumann, Y. Idan, and G. Pinkas, "Discovery of fraud rules for telecommunications-challenges and solutions," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 409-413.
10. S. V. Suryanarayana, G. N. Balaji, and G. V. Rao, "Machine Learning Approaches for Credit Card Fraud Detection," International Journal of Engineering and Technology, vol. 7, no. 2, pp. 917-920, Mar. 2018. DOI: 10.14419/ijet.v7i2.9356
11. A. M. Fayyomi, D. Eleyan, and A. Eleyan, "A Survey Paper On Credit Card Fraud Detection Techniques," International Journal of Scientific & Technology Research, vol. 10, no. 09, pp. 72-79, Sep. 2021.
12. 1P. Tiwari, S. Mehta, N. Sakhuja, J. Kumar, and A. K. Singh, "Credit Card Fraud Detection using Machine Learning: A Study," arXiv preprint arXiv:2108.10005, 2021.
13. Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M & Saif, A. (2022). Financial fraud detection based on machine learning: a systematic literature review. *Applied Sciences*, *12*(19), 9637.

14. Nobel, S. N., Sultana, S., Singha, S. P., Chaki, S., Mahi, M. J. N., Jan, T., ... & Whaiduzzaman, M. (2024). Unmasking banking fraud: Unleashing the power of machine learning and explainable ai (xai) on imbalanced data. *Information*, *15*(6), 298.

15. Ojugo, A. A., Eboka, A. O., Yoro, R. E., Yerokun, M. O., & Efozia, F. N. (2015). Framework design for statistical fraud detection. *Mathematics and Computers in Science and Engineering Series*, *50*, 176-182.

16. Hashemi, S. K., Mirtaheri, S. L., & Greco, S. (2022). Fraud detection in banking data by machine learning techniques. *Ieee Access*, *11*, 3034-3043.

17. Trivedi, N. K., Simaiya, S., Lilhore, U. K., & Sharma, S. K. (2020). An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology*, *29*(5), 3414-3424.

18. Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020, May). Credit card fraud detection using machine learning. In *2020 4th international conference on intelligent computing and control systems (ICICCS)* (pp. 1264-1270). IEEE.

19. Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 international conference on computing networking and informatics (ICCNI)* (pp. 1-9). IEEE.

20. Sharma, G., Khurana, S. & Sharma, G. A Hybrid Neural–Tree Learning Framework for Robust Risk Identification in Financial Institutions. *Int J Netw Distrib Comput* **14**, 18 (2026). https://doi.org/10.1007/s44227-026-00096-1.

21. Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019, March). Credit card fraud detection-machine learning methods. In *2019 18th International Symposium Infoteh-Jahorina (Infoteh)* (pp. 1-5). IEEE.

22. Sohony, I., Pratap, R., & Nambiar, U. (2018, January). Ensemble learning for credit card fraud detection. In *Proceedings of the ACM India joint international conference on data science and management of data* (pp. 289-294).

23. Karthika, J., & Senthilselvi, A. (2022, August). Credit card fraud detection based on ensemble machine learning classifiers. In *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1604-1610). IEEE.

24. Saraf, S., & Phakatkar, A. (2022). Detection of credit card fraud using a hybrid ensemble model. *International Journal of Advanced Computer Science and Applications*, *13*(9), 464-474.

25. Vardhani, P. R., Priyadarshini, Y. I., & Narasimhulu, Y. (2018). CNN data mining algorithm for detecting credit card fraud. In *Soft computing and medical bioinformatics* (pp. 85-93). Singapore: Springer Singapore.

26. Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing*, *99*, 106883.

27. Nguyen, T. T., Tahir, H., Abdelrazek, M., & Babar, A. (2020). Deep learning methods for credit card fraud detection. *arXiv preprint arXiv:2012.03754*.

28. Aljunaid, S. K., Almheiri, S. J., Dawood, H., & Khan, M. A. (2025). Secure and transparent banking: Explainable AI-driven federated learning model for financial fraud detection. *Journal of Risk and Financial Management*, *18*(4), 179.

29. Al-Maari, A. A., & Abdulnabi, M. (2023, December). Credit card fraud transaction detection using a hybrid machine learning model. In *2023 IEEE 21st Student Conference on Research and Development (SCOReD)* (pp. 119-123). IEEE.

30. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

31. Tayebi, M., & El Kafhali, S. (2025). A novel approach based on XGBoost classifier and Bayesian optimization for credit card fraud detection. *Cyber Security and Applications*, *3*, 100093.

32. Wijaya, M. G., Pinaringgi, M. F., & Zakiyyah, A. Y. (2024). Comparative analysis of machine learning algorithms and data balancing techniques for credit card fraud detection. *Procedia Computer Science*, *245*, 677-688.

33. R. Goyal, A. Manjhvar, and V. Sejwar, "Credit card fraud detection in data mining using XGBoost classifier," International Journal of Recent Technology and Engineering, vol. 9, pp. 603-608, 2020.

34. B. S. Kumar, P. P. Yadav, and M. R. Reddy, "An intelligent approach to detect and predict online fraud transaction using XGBoost algorithm," Indonesian Journal of Electrical Engineering and Computer Science, vol. 35, no. 3, pp. 1491-1498, 2024 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v35.i3.pp1491-1498.

35. Dastidar, K. G., Caelen, O., & Granitzer, M. (2024). Machine learning methods for credit card fraud detection: A survey. *IEEE Access*, *12*, 158939-158965.

36. Tursunalieva, A., Alexander, D. L., Dunne, R., Li, J., Riera, L., & Zhao, Y. (2024). Making sense of machine learning: A review of interpretation techniques and their applications. *Applied sciences*, *14*(2), 496.

37. Btoush, E., Zhou, X., Gururajan, R., Chan, K. C., & Alsodi, O. (2025). Achieving excellence in cyber fraud detection: A hybrid ML+ DL ensemble approach for credit cards. *Applied Sciences*, *15*(3), 1081.

38. Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing credit card fraud detection: an ensemble machine learning approach. *Big Data and Cognitive Computing*, *8*(1), 6.

39. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321-357.

40. He, H., Bai, Y., Garcia, E. A., & Li, S. (2008, June). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)* (pp. 1322-1328). Ieee.

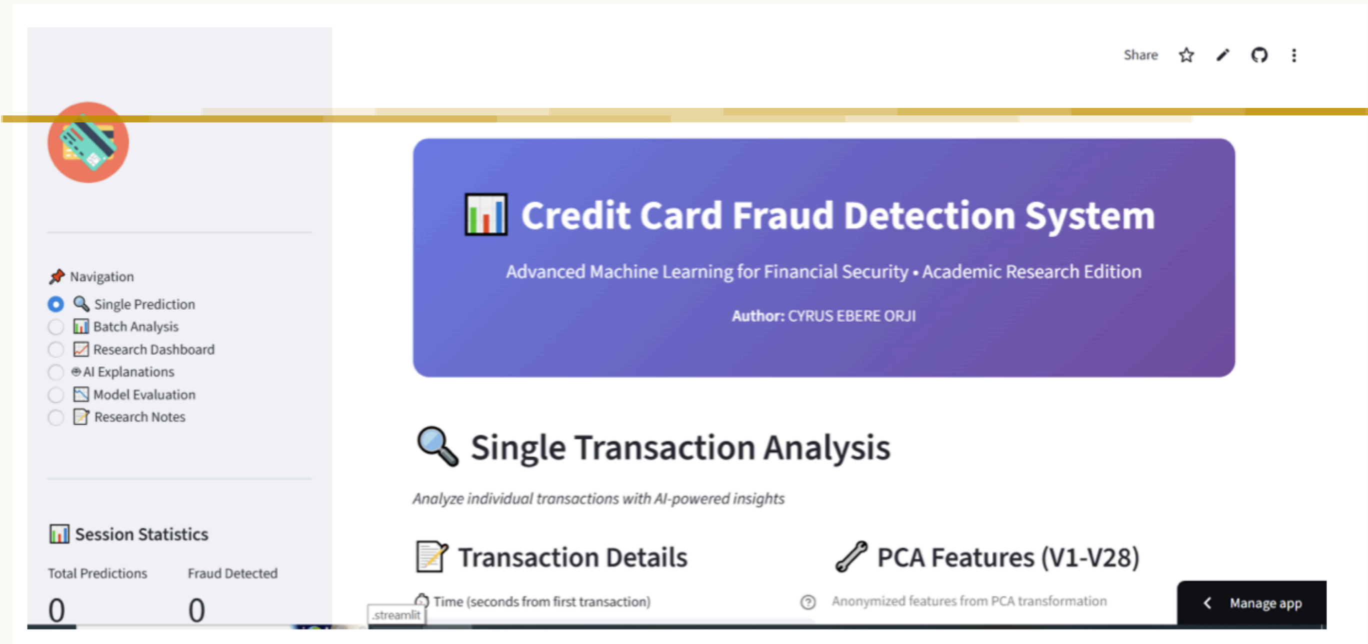
41. Al-Maari, A. A., Abdulnabi, M., Nathan, Y., Ali, A., Ali, U., & Khan, M. (2025). Optimized credit card fraud detection leveraging ensemble machine learning methods. *Engineering, Technology & Applied Science Research*, *15*(3), 22287-22294.

42. Hand, D. J., Christen, P., & Kirielle, N. (2021). F: an interpretable transformation of the F-measure. *Machine learning*, *110*(3), 451-456.

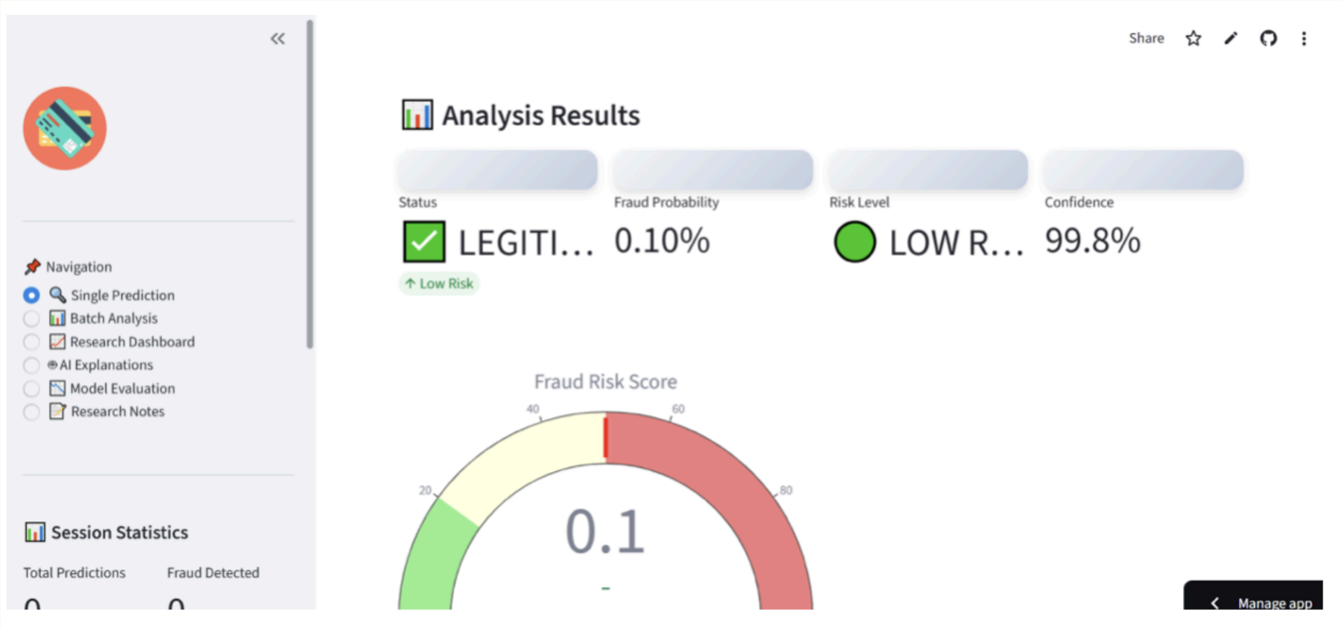
43. Stacking Ensemble for Credit Card Fraud Detection," Nature Scientific Reports, vol. 14, no. 1, pp. 1-12, 2024.

Appendix A

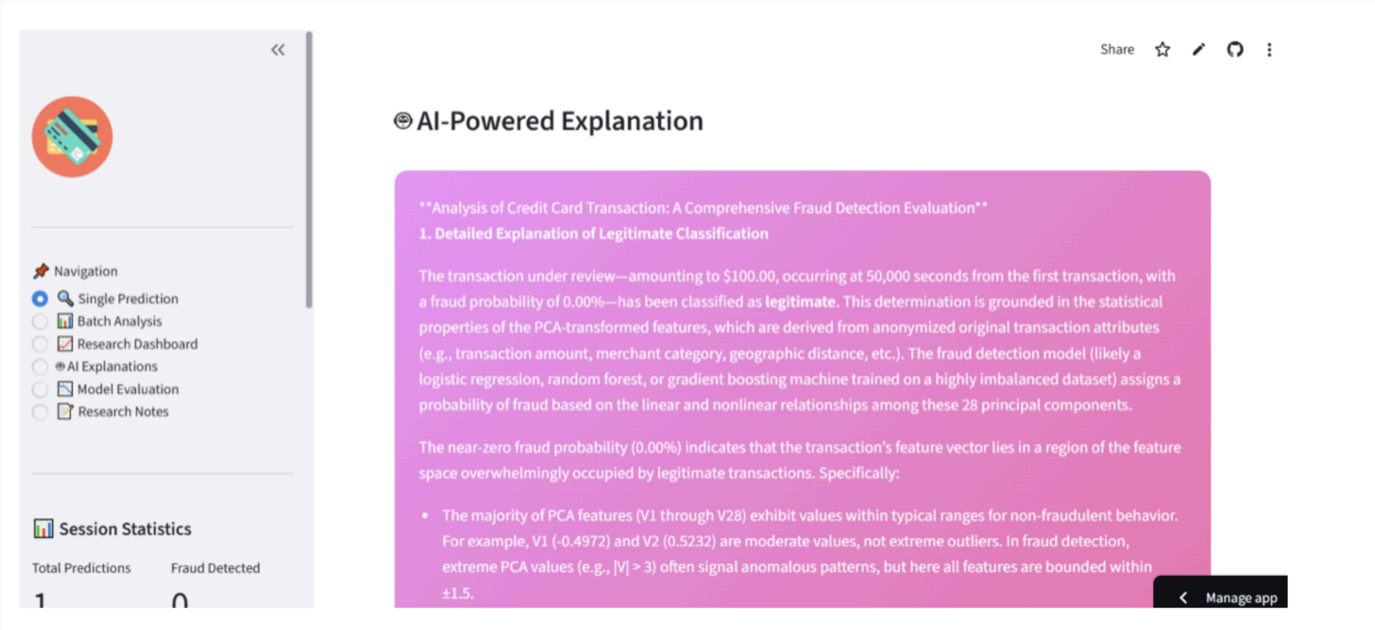
Main page



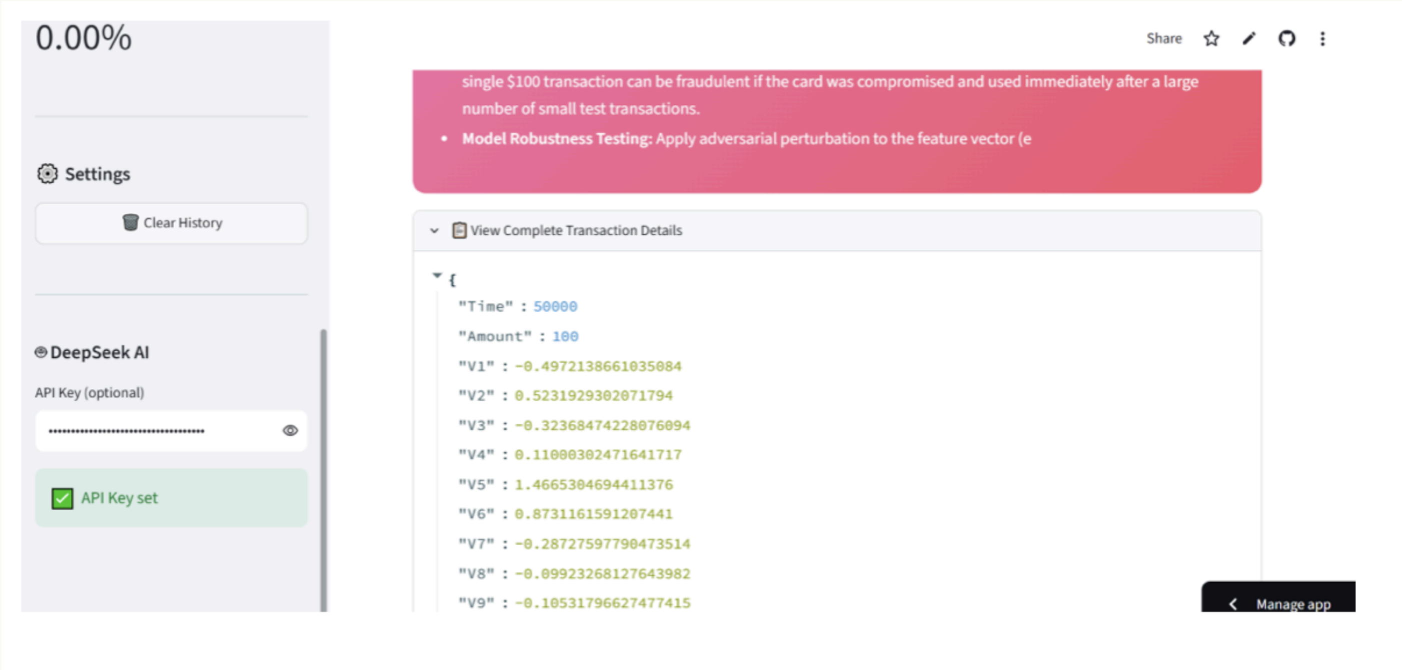
Results



AI Explanation



API key



Appendix B: Source Code Availability

The complete source code for Credit Card Fraud Detection System v2.0 is publicly available on GitHub to facilitate reproducibility, further research, and community contributions. The repository includes all source files, documentation, and example datasets used in this study.

Repository Details	**Information**
Repository Name	Credit Card Fraud Detection System
GitHub URL	https://github.com/cycyberuk/realtime-credit-card-fraud-detection
Streamlit App URL	https://realtime-credit-card-fraud-detection-j2qqy5ijxas3jiuonmykxz.streamlit.app/
Version	v2.0
License	MIT License
Language	Python 3.9
Last Updated	13th July 2026