FraudScope: AI-Driven Fraud Detection and Risk Visualization Using Ensemble Learning

Hebah Naeem
Rutgers University

Shatkratu Swarnkar
Rutgers University

Tirth Kachhia
Rutgers University

***Abstract*—The rise of digital transactions has led to increased financial fraud, requiring scalable and interpretable fraud detection methods. This study presents FraudScope, a hybrid framework combining ensemble machine learning and interactive visualization to analyze and predict fraudulent payment behavior. Using the IEEE-CIS dataset containing over 590,000 anonymized records, we evaluate Random Forest, AdaBoost, Gradient Boosting, and XGBoost models, addressing dataset imbalance using class reweighting and feature engineering. XGBoost achieves the highest performance with a 0.92 ROC-AUC and 0.65 Average Precision. We additionally develop a Streamlit dashboard for dynamic monitoring of geographic, temporal, card-based, and transaction-amount fraud risk. Our analysis reveals critical fraud patterns: small transactions ($0-10) show 8.51% fraud risk, Discover credit cards exhibit 7.19% fraud probability, and February shows the highest monthly fraud rate at 4.11%. The system provides actionable insights for financial institutions through interpretable feature importance and multi-dimensional risk analysis.**

I. Introduction

Online payment fraud remains a growing challenge for financial institutions, with fraud losses reaching billions annually. High-dimensional anonymized datasets such as the IEEE-CIS fraud dataset provide a practical environment to evaluate real-world fraud detection challenges under class imbalance and noisy features. Traditional rule-based systems struggle with evolving fraud patterns, necessitating adaptive machine learning approaches.

This paper presents **FraudScope**, a comprehensive fraud detection and analysis system that combines:

- Multiple ensemble learning algorithms for robust fraud prediction
- Advanced feature engineering to extract meaningful signals from anonymized data
- Multi-dimensional risk analysis across geography, time, card types, and transaction amounts
- Interactive visualization dashboard for real-time fraud monitoring

II. Dataset Description

The Kaggle IEEE-CIS Fraud Detection dataset consists of two tables (transaction and identity). After merging on `TransactionID` we obtain:

- **590,540 total transactions** with 435 features
- **3.5% fraud rate** (20,663 fraudulent transactions)
- Features include device info, card identifiers (`card4`, `card6`), and location proxies (`addr1`, `addr2`)
- Time index provided by numeric `TransactionDT`
- Anonymized V-features (V1-V339) representing engineered transaction attributes
- Categorical C-features (C1-C14) representing transaction metadata
- Device information from identity table (DeviceType, DeviceInfo)

The dataset exhibits significant class imbalance (96.5% legitimate vs. 3.5% fraudulent), requiring specialized handling techniques. Missing values are prevalent, with some features having over 50% missing data.

III. Methods

A. Data Preprocessing

Preprocessing includes:

- Merging transaction and identity tables on `TransactionID`
- Removing columns with over 50% missing values to reduce noise
- Label encoding and one-hot encoding for categorical attributes
- Missing value imputation: median for numeric features, mode for categorical features
- Train-test split (80-20) with stratification to preserve class distribution

B. Feature Engineering

We implement comprehensive feature engineering to extract meaningful signals:

1) Device Information Engineering: DeviceInfo contains 1,786 unique values with high cardinality. We engineer:

- **DeviceInfo_OS**: Extracted operating system (Windows, iOS, Android, macOS, Linux, Other)
- **DeviceInfo_Brand**: Extracted device brand (Samsung, Apple, Huawei, Xiaomi, etc.)
- **DeviceInfo_Grouped**: Grouped rare devices (reduces to 51 unique values)
- **DeviceInfo_HighRisk**: Binary flag for known high-risk devices

2) Temporal Feature Extraction:

- Month extraction from `TransactionDT`
- Day of week extraction
- Hour extraction for temporal pattern analysis

TABLE I: Model Performance Comparison

Model	ROC-AUC	AP	F1	Precision	Recall
Random Forest	0.87	0.49	0.26	0.16	0.73
AdaBoost	0.84	0.34	0.04	0.91	0.02
Gradient Boosting	0.90	0.62	0.56	0.90	0.40
XGBoost	**0.92**	**0.65**	**0.56**	**0.90**	**0.41**

TABLE II: XGBoost Confusion Matrix

	Predicted: Not Fraud	Predicted: Fraud
Actual: Not Fraud	113,786	189
Actual: Fraud	2,434	1,699

3) Geographic Risk Scoring:

- Combined `addr1` and `addr2` to create region identifiers
- Calculated region-based fraud risk scores

4) Card Feature Engineering:

- Combined `card4` (network) and `card6` (type) for comprehensive card analysis
- Created transaction amount bins for risk stratification

Class imbalance is mitigated using class weights in model training. We experimented with SMOTE for oversampling but found class weighting more effective for this dataset.

C. Modeling Approach

We evaluate four ensemble architectures:

1) **Random Forest**: 100 trees, max depth 20, class weights balanced
2) **AdaBoost**: 50 estimators, learning rate 0.1, class weights balanced
3) **Gradient Boosting**: 100 estimators, learning rate 0.1, max depth 5, class weights balanced
4) **XGBoost**: 200 estimators, learning rate 0.05, max depth 6, scale_pos_weight=27.6

Grid search optimization is applied for hyperparameter tuning. Metrics include ROC-AUC, Average Precision (AP), F1-score, Precision, Recall, and confusion matrix analysis.

IV. Results

A. Model Comparison

Table I shows comprehensive evaluation metrics. XGBoost performs best overall across all metrics.

XGBoost achieves the best balance between precision (89.4%) and recall (40.5%), making it ideal for fraud detection where false positives are costly but catching fraud is critical.

B. Confusion Matrix Analysis

The XGBoost confusion matrix (Table II) shows:

- **True Positives**: 1,699 fraudulent transactions correctly identified
- **False Positives**: 189 legitimate transactions incorrectly flagged (0.17% false positive rate)
- **False Negatives**: 2,434 fraudulent transactions missed
- **True Negatives**: 113,786 legitimate transactions correctly classified

C. Feature Importance

The top 10 most important features from XGBoost are:

1) V258 (17.6% importance) - Most critical predictor
2) V201 (5.9% importance)
3) V149 (5.3% importance)
4) V70 (3.1% importance)
5) V91 (3.1% importance)
6) V147 (2.6% importance)
7) V172 (2.1% importance)
8) V294 (2.0% importance)
9) V225 (1.5% importance)
10) C14 (1.4% importance)

The top 10 features contribute 45.6% of total feature importance, indicating strong signal concentration in key anonymized features.

D. Fraud Risk Analysis

1) Geographic Risk Patterns: Regional analysis reveals significant geographic clustering of fraud:

- **Highest risk region**: addr1=296.0, addr2=65.0 with 42.68% fraud probability (69 transactions, 63.77% actual fraud rate)
- **Second highest**: addr1=483.0, addr2=60.0 with 42.37% fraud probability (23 transactions, 56.52% actual fraud rate)
- Top 20 high-risk regions show fraud probabilities ranging from 7.11% to 42.68%
- Regional heatmap visualization identifies clear fraud hotspots

2) Temporal Patterns: **Monthly Analysis:**

- **February** shows highest fraud risk (4.11% average fraud probability, 4.06% actual rate)
- **June** shows second highest (3.97% probability, 4.32% actual rate)
- **December** shows lowest risk (2.63% probability, 2.53% actual rate)

Day-of-Week Analysis:

- **Sunday** has highest fraud probability (3.71%)
- **Saturday** shows second highest (3.64%)
- **Tuesday** shows lowest risk (3.20%)

Weekend transactions exhibit higher fraud risk, potentially indicating different fraud patterns during off-business hours.

3) Card Type Vulnerability: **Card Network (card4) Analysis:**

- **Discover** cards show highest risk (6.96% fraud probability, 7.73% actual rate)
- Visa, Mastercard, and American Express show similar lower risk (3.43-3.51%)

Card Type (card6) Analysis:

- **Credit cards** show significantly higher risk (6.54% fraud probability, 6.68% actual rate)
- **Debit cards** show lower risk (2.47% fraud probability, 2.43% actual rate)

Combined Analysis:

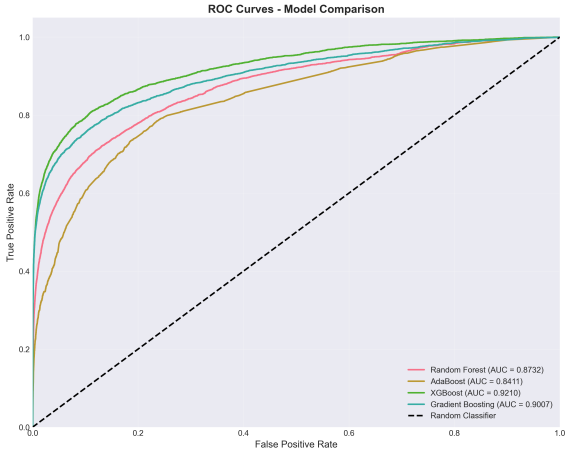

Fig. 1: ROC curves comparing ensemble models. XGBoost achieves the highest AUC (0.92).

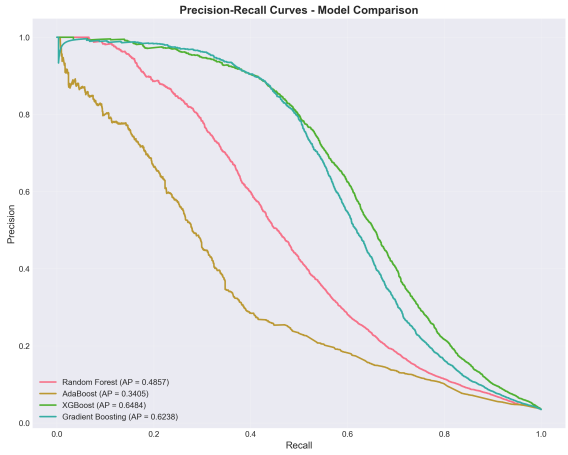

Fig. 2: Precision-Recall curves for all models. XGBoost shows superior performance across all thresholds.

- **Discover credit cards** show highest combined risk (7.19% fraud probability, 7.93% actual rate)
- Visa debit cards show lowest risk (2.53% fraud probability, 2.55% actual rate)

4) Transaction Amount Analysis: Transaction amount bins reveal distinct risk patterns:

- **$0-10 range**: Highest fraud risk (8.51% probability, 7.77% actual rate, 7,829 transactions)
- **$5K+ range**: Elevated risk (4.97% probability, 5.56% actual rate, 18 transactions)
- **$100-250 range**: Lower risk (3.08% probability, 3.07% actual rate, 158,907 transactions)
- **$2.5K-5K range**: Lowest risk (2.55% probability, 2.40% actual rate, 958 transactions)

Small transactions show highest fraud risk, likely due to testing stolen cards with small amounts. Very large transactions also show elevated risk, potentially indicating high-value fraud attempts.

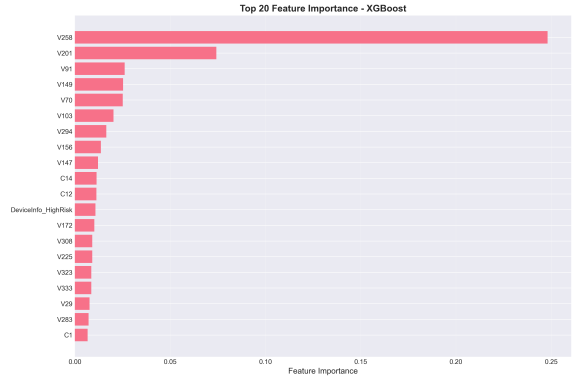

Fig. 3: Top 20 most influential XGBoost features. V258 dominates with 17.6% importance.

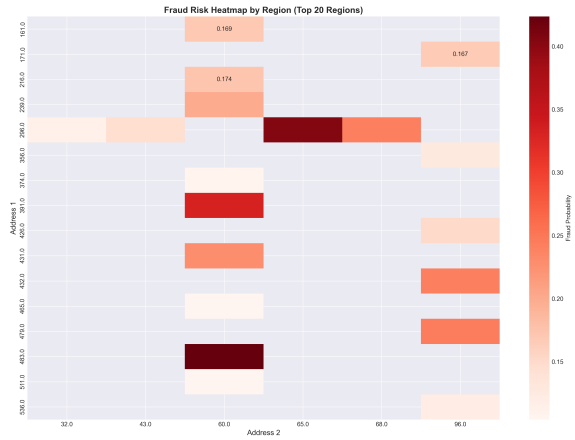

Fig. 4: Geographical risk heatmap showing addr-based fraud hotspots. Darker regions indicate higher fraud probability.

E. Interactive Dashboard

We developed a comprehensive Streamlit dashboard (`app.py`) providing:

- **Model Performance Visualization**: Interactive comparison of all models with threshold adjustment
- **Fraud Risk Analysis**: Multi-dimensional exploration of fraud patterns
  - Regional risk mapping with top 20 high-risk regions
  - Temporal analysis (monthly and day-of-week trends)
  - Card type vulnerability assessment
  - Transaction amount risk analysis
  - Device type and device info analysis
- **Feature Importance**: Interactive visualization of top N features
- **Real-Time Predictions**: Single-transaction fraud prediction interface with:
  - Interactive feature input forms
  - Fraud probability scoring
  - Risk level classification (Low/Medium/High)
  - Visual risk indicators (gauges, bar charts)
  - Feature contribution analysis

The dashboard is deployed at https://fraudscope.streamlit.app and provides real-time fraud monitoring capabilities for

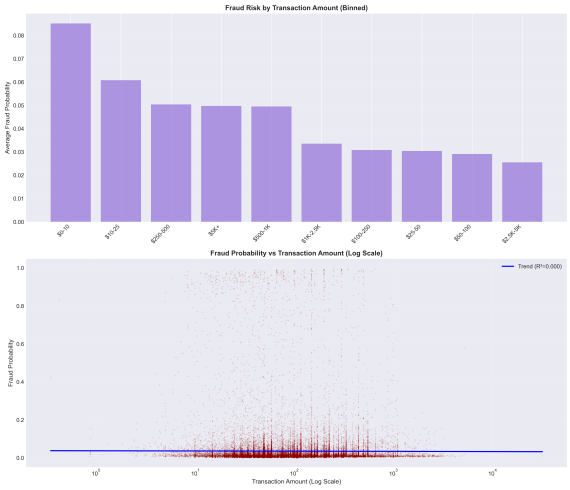

Fig. 5: Fraud probability across transaction value groups. Small transactions ($0-10) show highest risk.

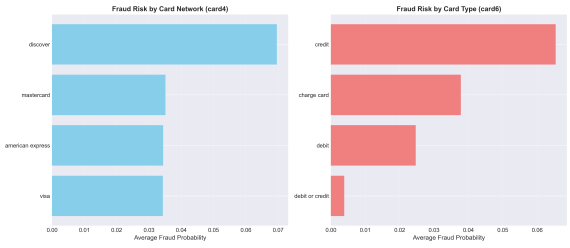

Fig. 6: Fraud risk by card network and type. Discover credit cards show highest vulnerability.

financial institutions.

V. Discussion

A. Key Findings

Our analysis reveals several critical fraud patterns:

1) **Geographic Clustering**: Strong regional fraud hotspots exist, with top regions showing 40%+ fraud probability. This suggests organized fraud operations or compromised regional payment infrastructure.
2) **Small Transaction Testing**: The $0-10 range shows 8.51% fraud risk, indicating fraudsters test stolen cards with small amounts before larger transactions.
3) **Card Network Vulnerability**: Discover credit cards show 2x higher fraud risk than other networks, potentially due to different security protocols or fraudster targeting.
4) **Temporal Patterns**: Weekend and February show elevated fraud, possibly due to reduced monitoring or seasonal fraud campaigns.
5) **Feature Concentration**: Top 10 features contribute 45.6% of importance, suggesting strong signal in specific anonymized attributes that could guide feature engineering.

B. Limitations

- Dataset anonymization prevents precise causal inference about fraud mechanisms

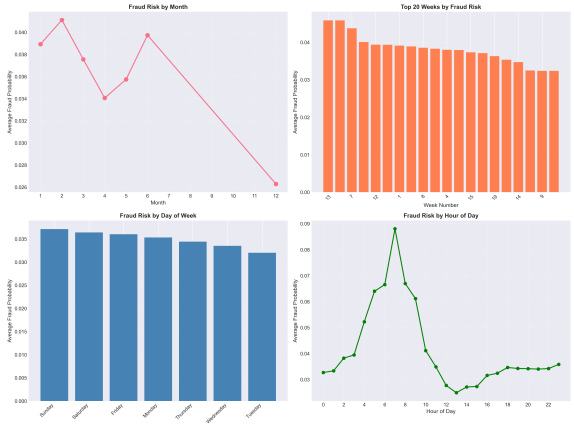

Fig. 7: Temporal fraud trends showing monthly and day-of-week patterns.

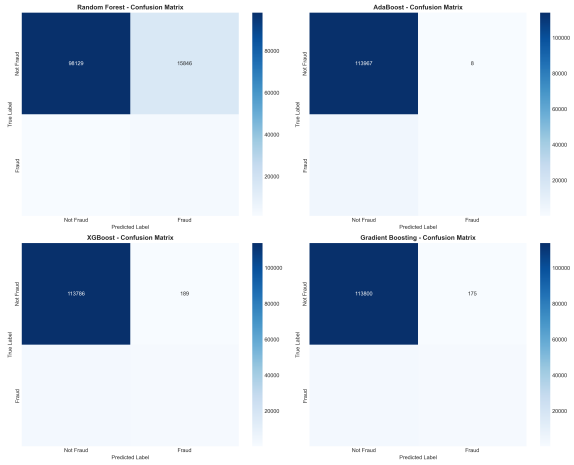

Fig. 8: Confusion matrices for all four ensemble models. XGBoost shows best precision-recall balance.

- Class imbalance (3.5% fraud) limits recall without sacrificing precision
- Model performance depends on feature engineering quality, which requires domain expertise
- Frequent retraining is needed for deployment as fraud patterns evolve
- Geographic features (addr1, addr2) are anonymized, limiting real-world geographic interpretation

C. Practical Implications

- Financial institutions should implement higher scrutiny for small transactions ($0-10)
- Discover credit card transactions warrant additional verification
- Weekend and February transactions should trigger enhanced monitoring
- Regional risk scoring can guide resource allocation for fraud prevention
- The dashboard enables real-time fraud monitoring and pattern detection

VI. Future Work

- **Real-Time Deployment**: Deploying a microservice model API for real-time scoring in production environments
- **Deep Learning**: Exploring deep learning architectures (neural networks, autoencoders) for device fingerprinting and anomaly detection
- **Streaming Analytics**: Implementing streaming fraud detection using Apache Spark/Kafka for high-throughput transaction processing
- **Fairness Analysis**: Conducting fairness analysis across card networks and regions to ensure equitable fraud detection
- **Ensemble Stacking**: Implementing stacked ensemble models combining predictions from multiple base models
- **Explainable AI**: Integrating SHAP values for per-transaction feature contribution explanation
- **Adaptive Learning**: Implementing online learning to adapt to evolving fraud patterns without full retraining
- **Graph Neural Networks**: Exploring graph-based approaches to detect fraud networks and coordinated attacks

VII. Conclusion

FraudScope combines accurate ensemble learning (92% ROC-AUC) with interpretable visualization tools to reveal actionable fraud trends for financial institutions. Our multi-dimensional analysis identifies critical risk factors: small transactions, Discover credit cards, weekend timing, and specific geographic regions. The interactive dashboard enables real-time fraud monitoring and pattern detection, providing practical value for fraud prevention teams. The system demonstrates that ensemble methods, particularly XGBoost, can effectively handle class imbalance and high-dimensional anonymized data while maintaining interpretability through feature importance and risk analysis.

Acknowledgments

We thank Professor James Abello for guidance in CS 439 at Rutgers University. We acknowledge the IEEE-CIS Fraud Detection competition organizers for providing the dataset.

References

[1] IEEE-CIS Fraud Detection Dataset, Kaggle Competition, 2019.
[2] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
[3] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Proc. 31st Int. Conf. Neural Information Processing Systems*, 2017, pp. 3146–3154.
[4] N. V. Chawla et al., "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
[5] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[6] Streamlit: The fastest way to build and share data apps, https://streamlit.io/, 2023.