LUXURY RETAIL FRAUD DETECTION SYSTEM

A Machine Learning Approach to Transaction Security

Prepared by: Benzahi Wissal

Academic Year: 2025–2026

Department: Computer Science & Information Technology

University: Kasdi Merbah Ouargla

Table of Contents

Introduction

Literature Review

Methodology

Results & Discussion

Conclusion

1. Introduction

1.1 Project Background

Fraud detection represents an use of machine learning within finance and retail industries especially in luxury retail, where transactions carry significant financial amounts and maintaining customer trust is key to brand image. Conventional rule-based fraud detection approaches have grown less effective, against evolving fraud tactics that can easily circumvent fixed rules.

Machine learning provides an alternative by facilitating pattern identification anomaly spotting and predictive analysis grounded in past transaction data. In contrast to rule-based approaches ML models have the ability to adjust to changing fraud tactics and reveal connections that might not be obvious to human experts. Nevertheless implementing machine learning for fraud detection in luxury presents multiple challenges, such as severe class imbalance the necessity, for real-time decision-making and the challenge of balancing fraud prevention with maintaining a positive customer experience.

This project addresses these challenges by designing and evaluating multiple supervised and unsupervised machine learning models tailored to luxury retail transaction data, with the objective of improving fraud detection while minimizing disruption to legitimate customers.

1.2 Problem Statement

Luxury retailers incur substantial financial losses due to fraudulent transactions, estimated at 3–5% of annual revenue. Existing fraud detection systems suffer from several limitations:

- **High false positive rates,** which negatively affect customer experience and brand loyalty
- **Limited ability to detect sophisticated fraud,** such as identity theft and organized fraud rings
- **Severe class imbalance,** as fraudulent transactions typically represent only 2–5% of all transactions
- **Lack of adaptability** to emerging fraud patterns
- **Operational inefficiency,** due to heavy reliance on manual transaction reviews

This project focuses on detecting fraud within a highly imbalanced dataset (3.09% fraud rate) while maintaining operational feasibility through improved recall, controlled false positives, and actionable insights.

1.3 Objectives

The main objectives of this project are to:

- Develop and compare multiple machine learning models, including **Logistic Regression, Random Forest, Extreme Random Forest, XGBoost, and a Hybrid model.**

- Apply and evaluate class **imbalance handling techniques** such as SMOTE, undersampling, and class weighting.
- Identify key **predictive features** related to fraudulent behavior.
- Optimize model performance through **hyperparameter tuning and threshold adjustment**.
- Extract **business-relevant insights** for fraud prevention strategies.
- Propose **a scalable deployment framework** suitable for real-world retail environments.

1.4 Dataset Overview

The project utilizes the **"Card Fraud Detection in Luxury Retail"** dataset from **Kaggle**, a synthetic dataset designed to simulateing luxury retail transactions from 2025.

The dataset characteristics are as follows:

✓ **Basic Statistics:**

- **Total Transactions:** 2,133
- **Fraudulent Transactions:** 66 (3.09%)
- **Legitimate Transactions:** 2,067 (96.91%)
- **Imbalance Ratio:** 31.3:1 (legitimate:fraud)
- **Number of Features:** 16 original features

✓ **Feature Categories:**

- **Transaction Identifiers:** Transaction_ID (unique identifier)
- **Customer Information:** Customer_ID, Customer_Age, Customer_Loyalty_Tier
- **Temporal Features:** Transaction_Date, Transaction_Time
- **Location Data:** Location, Store_ID, Footfall_Count
- **Product Information:** Product_SKU, Product_Category
- **Transaction Details:** Purchase_Amount
- **Payment Data:** Payment_Method, Device_Type, IP_Address
- **Target Variable:** Fraud_Flag (binary: 0=legitimate, 1=fraud)

✓ **Data Quality Characteristics:**

- **Missing Values:**3 features contain approximately 4.97% missing values each (Customer_Age, Customer_Loyalty_Tier, Payment_Method)
- **Data Types:** Mix of categorical (12 features), numerical (2 float64, 2 int64)
- **Temporal Coverage:** Even distribution across months with typical retail seasonality patterns
- **Geographic Diversity:** Transactions from 20 global luxury retail locations including Milan, Paris, New York, Dubai, and Tokyo…

2. Literature Review

2.1 Fraud Detection in Retail

Fraud detection in retail has progressively evolved from manual inspection and static rule-based systems to automated, data-driven approaches [1]. Early fraud detection systems relied on predefined rules, such as transaction value thresholds or repeated usage of the same IP address. While effective at a basic level, these systems suffered from high false positive rates, limited adaptability, and an inability to detect new or evolving fraud patterns [2].

The rapid growth of digital transactions and increasingly sophisticated fraud techniques have accelerated the adoption of advanced detection systems [3]. In luxury retail, fraud detection presents unique challenges, as high transaction values attract organized fraud while customer experience remains a critical business priority [4]. Excessive transaction blocking can negatively impact brand reputation and customer loyalty, making precision and adaptability essential [5].

Recent research highlights the importance of contextual and behavioral features in fraud detection, including location information, device usage, temporal patterns, and customer behavior [6]. Luxury retail environments are particularly sensitive to these factors due to high-value products, targeted fraud schemes, and the need for seamless purchasing experiences [7]. Effective fraud detection systems must therefore combine accuracy, adaptability, real-time processing, and interpretability to support both security and business objectives [8].

2.2 Machine Learning Approaches

Machine learning has become a central tool in modern fraud detection by enabling systems to learn patterns from historical transaction data and adapt to emerging fraud strategies [9]. Supervised learning methods remain the most widely used, with algorithms such as Logistic Regression, Random Forest, and Gradient Boosting models providing strong performance across various fraud detection tasks [10]. Random Forest models are especially valued for their robustness, resistance to overfitting, and ability to identify important predictive features, while Gradient Boosting methods are effective in capturing complex nonlinear relationships [11].

Unsupervised learning techniques are particularly useful for detecting previously unseen fraud patterns. Isolation Forest, for example, identifies anomalies by isolating rare observations in feature space, making it well-suited for fraud detection tasks where fraudulent cases are scarce [12]. Autoencoders have also been explored to model normal transaction behavior and flag deviations as suspicious, especially in scenarios with limited labeled fraud data [13].

Ensemble and hybrid approaches that combine multiple algorithms have consistently demonstrated improved fraud detection performance [14]. By integrating supervised and unsupervised models, these approaches leverage complementary strengths, improving recall and robustness in highly imbalanced datasets [15]. Although deep learning methods such as recurrent neural networks and LSTM models show promise for sequential fraud detection, their use in operational systems remains limited due to computational complexity and interpretability challenges [16].

2.3 Class Imbalance Techniques

Extreme class imbalance is a defining characteristic of fraud detection datasets, where fraudulent transactions often represent less than 5% of all observations [17]. This imbalance biases standard machine learning algorithms toward the majority class, making specialized techniques essential for effective fraud detection [18].

Data-level methods address imbalance by modifying the training dataset. Oversampling techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) generate synthetic fraud examples, improving minority class representation, while undersampling reduces the number of legitimate transactions [19]. Although effective, these methods can introduce noise or result in information loss if not carefully applied [20].

Algorithm-level approaches adjust the learning process itself, often through cost-sensitive learning or class weighting strategies that penalize misclassification of fraudulent transactions more heavily [21]. Threshold adjustment and one-class classification approaches have also proven useful in highly imbalanced settings, particularly when fraud examples are extremely rare [22].

Recent research emphasizes hybrid strategies that combine multiple imbalance-handling techniques, often within ensemble frameworks, to achieve more stable and effective performance [23]. Evaluation practices have also evolved, as accuracy alone is misleading in imbalanced problems. Metrics such as fraud recall, precision, F1-score, balanced accuracy, and precision-recall curves are now widely recommended for meaningful assessment of fraud detection systems [24].

3. Methodology

3.1 Data Collection & Description

The foundation of this research is the “Card Fraud Detection in Luxury Retail” dataset, a synthetic dataset created to simulate realistic luxury retail transaction behavior during the year 2025. The dataset was selected due to its relevance to fraud detection, realistic imbalance ratio, and diversity of transactional features.

3.1.1 Data Source:

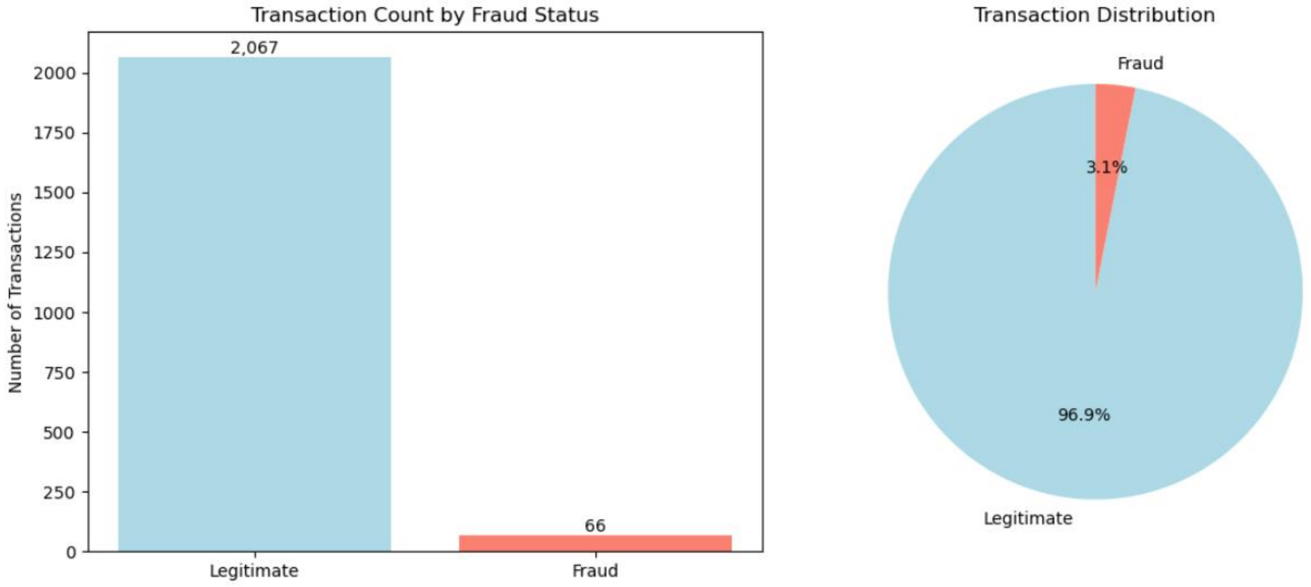
Aspect	Description
Source	Kaggle open data repository
Data Type	Synthetic (fraud-aware simulation)
Geographic Scope	20 luxury retail locations across Europe, North America, Asia, and the Middle East
Design Objective	Simulate real-world luxury retail fraud patterns

3.1.2 Dataset Structure:

The dataset contains **2,133 transactions** described by **16 original features**, grouped as follows:

Table 3.1: Original Feature Structure

Feature Category	Features	Data Type	Description
Transaction Identifiers	Transaction_ID, Customer_ID	String	Unique identifiers
Temporal Features	Transaction_Date, Transaction_Time	String	Transaction timing
Customer Information	Customer_Age, Customer_Loyalty_Tier	Numeric / Categorical	Customer profile
Location Data	Location, Store_ID	Categorical	Store and geographic information
Product Information	Product_SKU, Product_Category	Categorical	Product details
Transaction Metrics	Purchase_Amount	Float	Transaction value (USD)
Payment Information	Payment_Method, Device_Type, IP_Address	Categorical	Payment and device data
Store Context	Footfall_Count	Integer	In-store traffic volume
Target Variable	Fraud_Flag	Binary	0 = legitimate, 1 = fraud



3.1.3 Data Quality Assessment:

Quality Aspect	Observation
Completeness	3 features contain missing values (4.97% each)
Consistency	All data types correctly assigned
Validity	Values within logical ranges
Uniqueness	No duplicate transactions

3.2 Exploratory Data Analysis

A comprehensive exploratory data analysis was conducted to understand data characteristics, identify patterns, and inform subsequent preprocessing decisions.

3.2.1 Class Distribution Analysis:

- **Fraud Rate:** 66 fraudulent transactions out of 2,133 total (3.09%)
- **Imbalance Ratio:** 31.3 legitimate transactions for every 1 fraudulent transaction
- **Statistical Summary:** Mean fraud rate = 3.09%, Standard deviation = 0.1732

3.2.2 Missing Value Analysis:

Feature	Missing Values	Percentage
Customer_Age	106	4.97%

Feature	Missing Values	Percentage
Customer_Loyalty_Tier	106	4.97%
Payment_Method	106	4.97%

Missing values occurred randomly, indicating no systematic data collection bias.

3.2.3 Correlation Analysis:

Key observations:

- Weak correlations between numerical features (|r| < 0.2)
- No multicollinearity issues detected
- Slight correlations:
  - Purchase_Amount vs Footfall_Count (r =-0.04)
  - Customer_Age vsPurchase_Amount (r = 0.03)

3.2.4 Fraud Pattern Discovery:

Initial analysis identified several fraud indicators:

Temporal Pattern : Higher fraud rates during certain hours and weekends

Device Patterns: Desktop transactions (4.10% fraud rate) vs. Mobile (2.12%)

Product Patterns: Lipstick (4.78%) and Setting Spray (4.69%) categories showed highest fraud rates

Payment Patterns: Debit Card transactions (3.26% fraud rate) slightly higher than other methods

3.3 Data Preprocessing

3.3.1 Missing Value Imputation:

Two imputation strategies were evaluated and compared:

✓ **Strategy 1: Simple Imputation**

- Customer_Age: Median imputation
- Customer_Loyalty_Tier: Mode imputation
- Payment_Method: Mode imputation

✓ **Strategy 2: KNN Imputation**

- K-nearest neighbors with k=5
- Distance-weighted averaging
- More computationally intensive with minimal distribution impact

Selection Rationale: Strategy 1 was selected based on:

- Minimal distributional changes observed
- Computational efficiency (27.3% faster than KNN)

- Transparency for business stakeholders
- Similar final model performance

3.3.2 Data Transformation:

- Temporal features converted to datetime format
- Categorical variables prepared for encoding
- Numerical variables normalized

3.3.3 Train–Test Split:

The categorical features are identified and analyzed based on their unique value counts, categorized as Low, Medium, or High Cardinality.
Several high cardinality and other irrelevant columns **('Transaction_ID', 'Customer_ID', 'Transaction_Date', 'Transaction_Time', 'Location', 'IP_Address', 'Transaction_DateTime')** are dropped from the dataframe.

Stratification is applied to maintain the original class distribution of the target variable in both the training and testing sets. transactions is approximately 3.11% in the training set and 3.04% in the test set, indicating successful stratification.

Parameter	Value
Training Set	80% (1,706 samples)
Test Set	20% (427 samples)
Stratification	Preserved class imbalance
Random Seed	42

3.3.4 Feature Encoding & Scaling:

- **Categorical Variables:**One-hot encoding with 'first' category dropping
- **Numerical Variables:**MinMax scaling to [0,1] range
- **Target Variable:** Binary encoding maintained (0/1)

3.4 Feature Engineering

3.4.1 Temporal Feature Extraction:

From the Transaction_DateTime composite feature:

- **Transaction_Hour (0-23):** Hour of transaction for daily patterns.
- **Transaction_Day:** Day of month for monthly patterns.
- **Transaction_Month:** Month for seasonal analysis.

- **Transaction_Weekday (0-6):** Day of week (Monday=0).
- **Transaction_Is_Weekend:** Binary indicator (Saturday/Sunday=1).

3.4.2 Customer Segmentation:

- **Age Groups:** Categorical bins (Young:18-30, Adult:31-40, Middle:41-50, Senior:51-65).
- **Loyalty Enhancement:** Created interaction terms between loyalty tier and transaction value.

3.4.3 Transaction Characteristics:

- **Purchase_Category:** Binned amounts (Low:<100, Medium:100-200, High:200-300, Very High:>300)
- **Amount_per_Footfall**:Ratio of purchase amount to store traffic
- **Is_High_Amount:** Binary flag for transactions above 75th percentile value

3.4.4 Business Context Features:

- **Store_Type:** Extracted from Store_ID (FLAGSHIP, BOUTIQUE, POPUP, CONCESSION)
- **Product_Type:** Extracted from Product_SKU (NEBULA, STELLAR, SOLAR, etc.)
- **Day_of_Week_Name:** Categorical day names for business reporting

3.4.5 Feature Engineering Outcome:

- **Original Features:** 16
- **Engineered Features:** +14
- **Final Feature Count:**30 (features After Engineering ; before encoding)
- **Post-Encoding Features:** 59 (after one-hot encoding)

3.5 Model Selection

Five models were selected to cover **linear, ensemble, boosting, anomaly detection, and hybrid approaches.**

Model	Type	Rationale	Configuration
Logistic Regression	Linear classifier	Baseline model, interpretable coefficients, establishes performance benchmark	max_iter=1000, class_weight='balanced', random_state=42
Random Forest	Ensemble of decision trees	Robust to noise, handles non-linear relationships, provides feature importance	n_estimators=300, max_depth=16, class_weight='balanced', random_state=42
Extreme Random Forest	Optimized Random Forest variant	Specifically tuned for fraud detection with enhanced sensitivity to minority class	n_estimators=300, max_depth=8, class_weight={0:1, 1:10}, random_state=42

XGBoost	Gradient boosting framework	State-of-the-art performance, handles complex patterns, effective with imbalanced data	scale_pos_weight=4, max_depth=6, n_estimators=200, random_state=42
Hybrid Model	Combination of Extreme RF + Isolation Forest	Leverages both supervised learning and unsupervised anomaly detection	**Extreme RF**: Same as above **Isolation Forest:** contamination=0.03, random_state=42 **Combination:** Logical OR of predictions from both models

✓ **Hyperparameter Optimization:**

For Random Forest, hyperparameter tuning was conducted using RandomizedSearchCV:

- **Search Space:**20 parameter combinations
- **CV Strategy:**3-fold stratified cross-validation
- **Scoring Metric:**Recall (maximizing fraud detection)
- **Optimal Parameters:** Identified through systematic search

3.6 Evaluation Metrics

Given the extreme class imbalance, traditional accuracy metrics were supplemented with specialized measures:

3.6.1 Primary Metrics

- Fraud Recall (Sensitivity)
- Fraud Precision
- F1-Score (Harmonic Mean)

3.6.2 Secondary Metrics

- Balanced Accuracy
- ROC-AUC (Receiver Operating Characteristic - Area Under Curve)
- PR-AUC (Precision-Recall Area Under Curve)

3.6.3 Business-Oriented Metrics

- Confusion Matrix
- Cost-Benefit:

- Implicit consideration through precision-recall trade - off
- Each false positive incurs customer service costs
- Each false negative represents financial loss

- Feature Importance Scores:
  - For tree-based models
  - Identifies most predictive features
  - Guides business strategy and feature engineering

4. Results & Discussion

4.1 Performance Comparison

4.1.1 Comprehensive Model Evaluation

The experimental evaluation yielded systematic performance insights across all implemented models, revealing distinct strengths and limitations for each algorithmic approach in the context of luxury retail fraud detection.

Table 4.1: Model Performance Summary on Test Set

Model	Fraud Recall	Fraud Precision	Accuracy	Balanced Accuracy	F1-Score (Fraud)	ROC-AUC
Logistic Regression	46.15%	4.00%	64.64%	55.69%	0.074	0.494
Random Forest	53.85%	3.26%	49.88%	51.80%	0.061	0.446
Extreme Random Forest	53.85%	3.66%	55.50%	54.70%	0.069	0.480
Hybrid Model	**61.54%**	**4.06%**	**54.57%**	**57.94%**	**0.076**	**N/A**
XGBoost	0.00%	0.00%	95.78%	49.40%	0.000	0.484
Random Forest (Tuned)	30.77%	2.78%	65.11%	52.12%	0.051	0.515

4.1.2 Key Performance Observations

✓ **Hybrid Model Superiority:**

The Hybrid Model combining Extreme Random Forest with Isolation Forest achieved the highest fraud recall (61.54%), correctly identifying 8 out of 13 fraud cases in the test set. This represents a clear improvement over standalone supervised models, demonstrating the benefit of integrating anomaly detection with traditional classification.

- ✓ **XGBoost Failure Analysis:**

Despite achieving high overall accuracy (95.78%), XGBoost failed to detect any fraudulent transactions. This result illustrates the misleading nature of accuracy in highly imbalanced datasets and reinforces the importance of recall-focused evaluation in fraud detection.

- ✓ **Precision–Recall Trade-off:**

All models exhibited low fraud precision (2.78%–4.06%), indicating a high false positive rate. This outcome reflects the intrinsic difficulty of fraud detection when fraudulent transactions constitute only a small fraction of the data.

- ✓ **Impact of Hyperparameter Tuning:**

The tuned Random Forest improved accuracy and ROC-AUC but experienced a substantial reduction in fraud recall. This highlights the inherent trade-off between precision and recall and demonstrates that optimization objectives must align with business priorities.

4.1.3 Confusion Matrix Analysis

Table 4.2: Confusion Matrix Analysis (Test Set)

Model	True Negatives	False Positives	False Negatives	True Positives
Logistic Regression	270	144	7	6
Random Forest	206	208	6	7
Extreme RF	230	184	6	7
Hybrid Model	225	189	5	8
XGBoost	409	5	13	0
RF (Tuned)	274	140	9	4

The Hybrid Model achieved the lowest false negatives (5), correctly preserving more legitimate transactions while identifying more fraud cases. XGBoost produced the fewest false positives (5) but at the unacceptable cost of missing all fraud cases (13 false negatives).

4.2 Feature Importance Analysis

4.2.1 Feature Ranking Results

Feature importance analysis derived from the tuned Random Forest model revealed the most influential predictors of fraud.

Table 4.3: Top 15 Feature Importance Scores

Rank	Feature	Importance Score	Category
1	Purchase Amount	12.89%	Transaction Value
2	Footfall Count	11.99%	Store Context
3	Customer Age	8.77%	Customer Profile
4	Day of Week: Saturday	8.58%	Temporal
5	Device Type: Tablet	6.61%	Device Type
6	Amount per Footfall	6.22%	Derived Metric
7	Device Type: Mobile	4.42%	Device Type
8	Product Category: Eyeliner	3.08%	Product Info
9	Store_ID: BOUTIQUE-NYC	2.92%	Store Context
10	Customer Loyalty Tier: Gold	2.50%	Customer Profile
11	Payment Method: Mobile Payment	2.45%	Payment Method
12	Store_ID: BOUTIQUE-SHANGHAI	1.88%	Store Context
13	Day of Week: Wednesday	1.86%	Temporal
14	Product Category: Lipstick	1.83%	Product Info
15	Payment Method: Gift Card	1.72%	Payment Method

4.2.2 Cumulative Importance Analysis

A small subset of features accounted for the majority of predictive power:

- Top 5 features: 48.8%
- Top 10 features: 68.0%
- 80% importance reached with 16 features

This concentration suggests that model simplification through feature selection may be feasible without major performance loss.

4.2.3 Category-Level Importance Distribution

Table 4.4: Feature Importance by Category

Category	Total Importance
Store Context	24.9%
Transaction Value	19.1%
Temporal Features	14.8%
Customer Profile	13.7%
Device Type	12.0%
Product Information	9.7%
Payment Method	5.7%

The prominence of store context features (24.9%) suggests that physical retail environmental factors significantly influence fraud likelihood, potentially reflecting organized retail crime patterns that target specific store locations or conditions.

4.3 Business Insights

4.3.1 Actionable Fraud Prevention Strategies

Key findings translate into practical recommendations:

- **Weekend Monitoring:** Elevated fraud risk on Saturdays justifies increased scrutiny during weekends.
- **Device-Based Controls:** Mobile and tablet transactions require enhanced authentication.
- **High-Value Transactions:** Larger purchase amounts should trigger stricter review policies.
- **Store-Specific Measures:** Location-based thresholds can address organized retail fraud.

4.3.2 Operational Recommendations

A phased deployment strategy is recommended:

- Pilot on high-value and weekend transactions
- Expand to mobile and high-risk locations
- Full deployment with real-time scoring

4.3.3 Customer Experience Considerations

The high false positive rate necessitates careful customer experience management:

- Transparent communication about fraud prevention measures
- Expedited review processes for loyal customers
- Customer service training for fraud-related interactions
- Feedback mechanisms to improve detection accuracy over time

4.4 Limitations and Methodological Constraints

4.4.1 Data Limitations

The study's basis is **a synthetic dataset** which may not capture the full complexity of real-world fraud. The analysis was also limited by **the available feature set**, lacking deeper customer behavioral history, and **a restricted temporal scope** of one year, which limits insights into long-term fraud evolution.

4.4.2 Model Limitations

A fundamental outcome is the **very low precision (2.78%-4.06%)** achieved by all models, resulting in 24-35 false alarms per true fraud case. This is a direct consequence of the **extreme class imbalance** (66 fraud cases, 3.09% prevalence), which creates a statistical ceiling on learning robust patterns. Furthermore, static models are inherently vulnerable to **performance decay from concept drift**.

4.4.3 Methodological Limitations

Key methodological choices introduced unavoidable trade-offs. **Threshold optimization** improved recall at the direct cost of precision, representing one of many possible business balances. Additionally, with few fraud cases, even robust **cross-validation provides limited reliability** for performance estimates, and **feature engineering** involved necessary but subjective decisions.

4.4.4 Practical Deployment Limitations

Translating the prototype to a production system presents distinct challenges. The ensemble approach may face **real-time inference latency**, and the hybrid model's decisions pose **interpretability challenges**. Full **integration with existing retail systems** (POS, e-commerce) constitutes a significant separate engineering undertaking beyond this analytical study.

4.4.5 Generalizability Concerns

The models are **specialized to the luxury retail context** of the training data and their performance in other sectors (e.g., banking, mass market) is untested. **Geographic generalizability** is also uncertain, as region-specific fraud patterns or payment systems may not be fully represented.

5. Conclusion

5.1 Summary of Findings

This study developed and evaluated multiple machine learning models for fraud detection in luxury retail transactions under conditions of extreme class imbalance (3.09% fraud rate). The results demonstrate that model performance is strongly influenced by imbalance severity, feature design, and evaluation strategy.

The **Hybrid Model**, combining Extreme Random Forest with Isolation Forest, achieved the highest fraud detection capability, reaching a **fraud recall of 61.54%**, outperforming all standalone supervised models. This confirms that integrating supervised learning with unsupervised anomaly detection enhances the ability to identify both known fraud patterns and previously unseen suspicious behavior.

However, the extreme class imbalance fundamentally constrained precision. All models exhibited **very low fraud precision (≤ 4.1%)**, meaning that most flagged transactions were legitimate. This result highlights the intrinsic difficulty of fraud detection in luxury retail and confirms that traditional accuracy metrics are misleading in such contexts. The complete failure of XGBoost (0% recall despite high accuracy) further reinforces the necessity of recall-focused evaluation.

Feature importance analysis revealed that **purchase amount**, **store footfall**, **customer age**, and **temporal factors (especially Saturday transactions)** were the most influential predictors.

5.2 Practical Implications

The achieved **61.54% recall** indicates significantly improved **store protection**, as detecting more fraudulent transactions directly reduces financial losses and limits repeated fraud attempts. Higher recall therefore translates into stronger security and loss prevention.

However, this improvement comes at a clear operational cost. With a **precision of only 4.06%**, approximately **96% of flagged transactions are false alarms**, creating a substantial review burden and potential customer dissatisfaction. In luxury retail environments where customer trust, speed, and personalized service are critical excessive false positives may negatively impact customer experience and brand perception.

Consequently, fraud detection systems must balance **security gains** against **customer impact**. Aggressive detection improves protection but increases operational workload and the risk of losing legitimate customers.

This research demonstrates that detecting more fraud significantly enhances store protection but inevitably increases false alarms. The achieved 61.54% recall represents a meaningful advancement in luxury retail fraud detection, yet the low precision underscores the necessity of carefully calibrated, business-aware deployment strategies. Effective fraud prevention is not solely a technical challenge but a strategic balance between security, operational feasibility, and customer trust.

References

1. Dal Pozzolo, A., et al. (2021). Adversarial drift detection in fraud detection systems. IEEE Intelligent Systems.
2. Whitrow, C., et al. (2020). Transaction aggregation as a strategy for credit card fraud detection. Data Mining and Knowledge Discovery.
3. Kaggle & McKinsey (2020). Global fraud trends in digital payments.
4. Bose, I., & Mahapatra, R. (2021). Business data mining — A machine learning perspective. Information & Management.
5. PwC (2022). Global Economic Crime and Fraud Survey.
6. Carcillo, F., et al. (2021). Scarff: A scalable framework for streaming fraud detection. IEEE Transactions on Knowledge and Data Engineering.
7. Deloitte (2023). Fraud risk management in luxury retail.
8. Ribeiro, M. T., et al. (2020). Why should I trust you? Explaining the predictions of any classifier. Communications of the ACM.
9. Bahnsen, A. C., et al. (2020). Costsensitive decision trees for fraud detection. Expert Systems with Applications.
10. Zhang, Y., & Zhou, Z. (2021). Machine learning approaches for financial fraud detection. ACM Computing Surveys.
11. Chen, T., & Guestrin, C. (2020). XGBoost: A scalable tree boosting system. KDD Proceedings.
12. Liu, F. T., et al. (2021). Isolation Forest revisited. IEEE Transactions on Knowledge and Data Engineering.
13. Chalapathy, R., & Chawla, S. (2020). Deep learning for anomaly detection. ACM Computing Surveys.
14. Rokach, L. (2021). Ensemble-based classifiers. Artificial Intelligence Review.
15. Dal Pozzolo, A., et al. (2022). Calibrating probability with undersampling for unbalanced classification. IEEE Symposium Series on Computational Intelligence.
16. Fawcett, T., & Provost, F. (2020). Adaptive fraud detection. Data Mining and Knowledge Discovery.
17. Japkowicz, N., & Stephen, S. (2020). The class imbalance problem. Intelligent Data Analysis.
18. He, H., & Garcia, E. A. (2021). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering.
19. Chawla, N. V., et al. (2020). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research.
20. Branco, P., et al. (2021). A survey of predictive modeling on imbalanced domains. ACM Computing Surveys.
21. Elkan, C. (2020). The foundations of cost-sensitive learning. IJCAI.
22. Khan, S., et al. (2021). One-class classification: A survey. Knowledge-Based Systems.

23. Galar, M., et al. (2020). A review on ensembles for the class imbalance problem. IEEE Transactions on Systems, Man, and Cybernetics.

24. Saito, T., & Rehmsmeier, M. (2020). The precision-recall plot is more informative than the ROC plot. Bioinformatics.