ISTANBUL TECHNICAL UNIVERSITY

Agentic Fraud Detection System

A Machine Learning Approach to Financial Security

Business Analytics for Managers – Term Project
Oğuzhan Kır • Fırat Ölçüm
December 2025

Agenda

Problem Statement: Financial fraud landscape & challenges

Exploratory Data Analysis: Pattern discovery & anomaly identification

Feature Engineering: Temporal, geospatial & behavioral features

Model Development: Training strategy & optimization

Results & Performance: Model evaluation & comparison

System Architecture: Agentic framework & explainability

Business Impact: ROI analysis & cost-benefit

Conclusion & Future Work: Key takeaways & next steps

Introduction & Problem Statement

The Challenge

Financial fraud costs billions annually

Traditional systems lack adaptability and have high false positives

ML models need explainability for stakeholder trust

Extreme class imbalance: 0.58% fraud rate (172:1 ratio)

Our Solution

XGBoost Model: 99.76% AUC on temporal test data

Agentic Framework: ReAct-based reasoning for transparency

Real-time: <50ms end-to-end inference latency

Dataset Overview

- 1,296,675 synthetic credit card transactions
- 7,506 fraudulent cases (0.58% prevalence)
- Temporal coverage: January 2019 – June 2020 (18 months)
- Features: Temporal, geospatial, transactional, demographic
- Designed to mirror real-world transaction distributions

EDA: Distributional Analysis

Key Findings:

- Fraud shows bimodal distribution (low & high amounts)
- Log transformation improves class separation
- Similar median distances but fraud shows higher variance

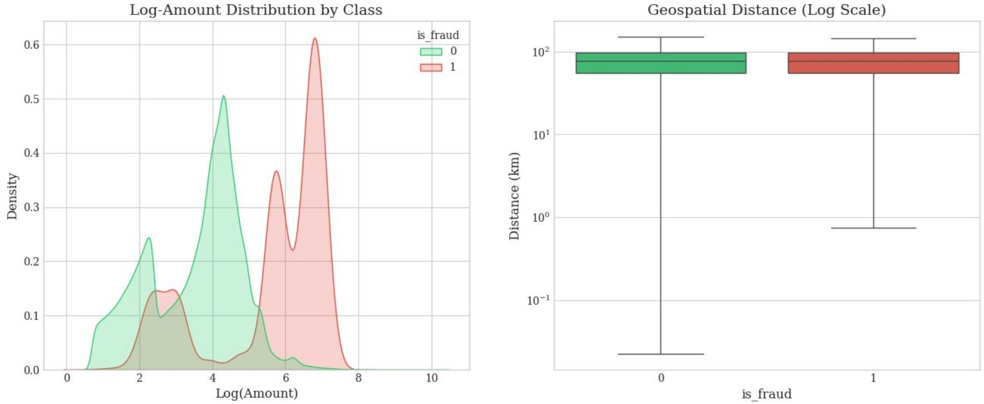


Figure 1: Distributional differences in Amount and Location

Log-Amount Distribution by Class

Geospatial Distance (Log Scale)

EDA: Temporal Pattern Recognition

Key Findings:

- 25%+ of fraud occurs in 21:00-23:00 window
- Legitimate transactions uniform (3-5%) across day
- Near-zero fraud during business hours (05:00-20:00)
- Created hour_risk_score feature with elevated weights

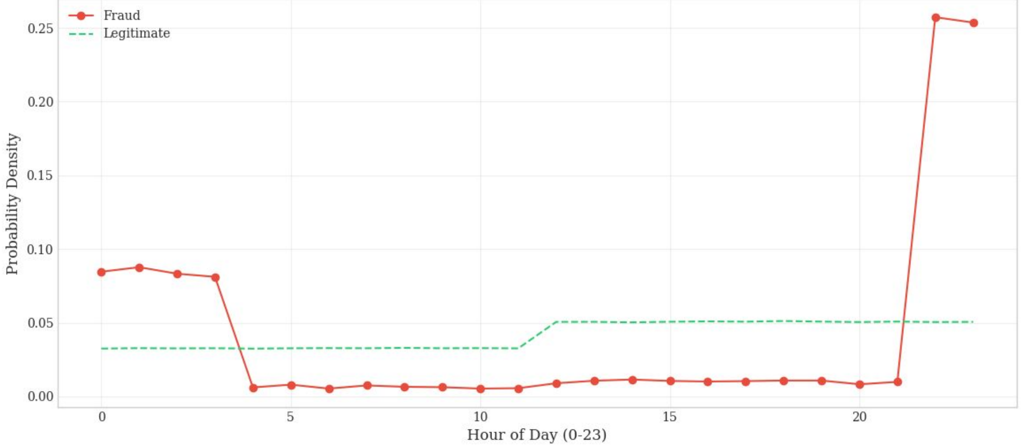


Figure 2: Diurnal Transaction Patterns

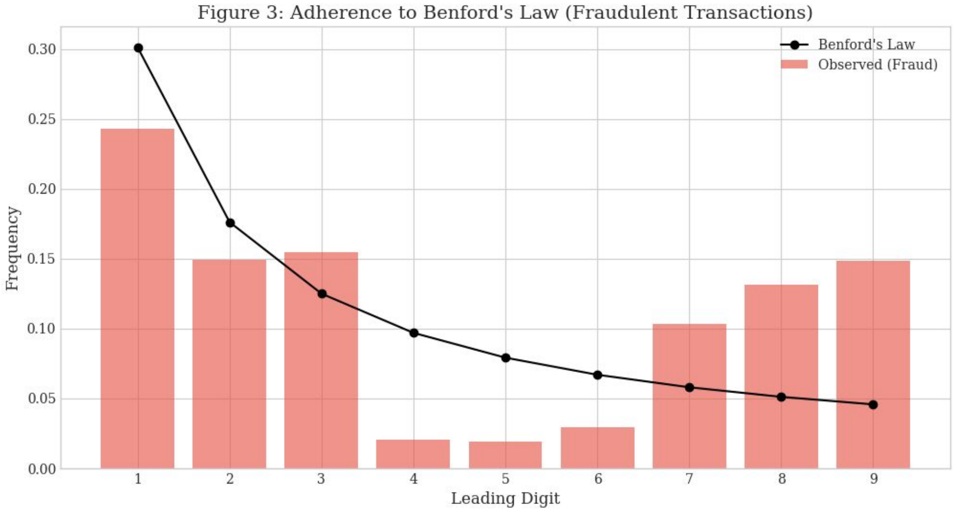
EDA: Benford's Law Analysis

What is Benford's Law?

In natural datasets, leading digits follow logarithmic distribution. Deviation indicates manipulation/fraud.

Key Findings:

- Fraud shows strong deviation from expected
- Over-representation of digits '1' (24%) and '9' (15%)
- Engineered benford_log_prob feature



Feature Engineering Strategy

Temporal Features

Cyclical encoding: hour_sin, hour_cos, day_sin, day_cos

Hour risk score based on EDA findings

Geospatial Analytics

Haversine formula for geodesic distance

Financial Profiling

log_amt, relative amounts, Benford's Law probability

Behavioral Aggregates

Rolling windows (24h, 7d, 30d): transaction counts, amounts, frequencies

Model Training Strategy

Temporal Split (Forward Validation)

Training Set (80%): Jan 2019 – Mar 2020 (1,037,340 txs)

Test Set (20%): Apr 2020 – Jun 2020 (259,335 txs)

Prevents temporal leakage - model trained on historical data only

Nested Cross-Validation

Outer Loop (5-fold): Performance estimation

Inner Loop (3-fold): Hyperparameter optimization with Optuna

100 trials per model using TPE sampler

Candidate Models

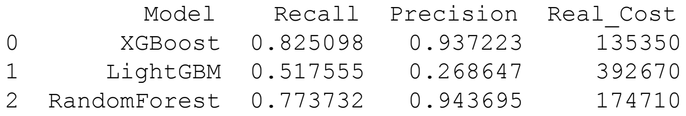
XGBoost, LightGBM, Random Forest

Results: Model Performance Comparison

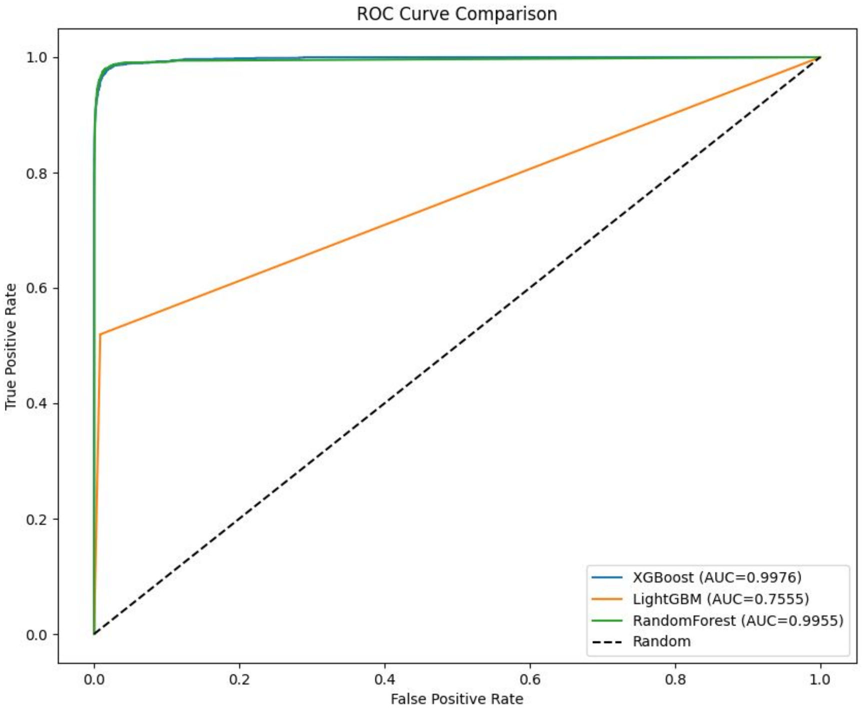
Out-of-Time Test Performance:

XGBoost: AUC 0.9976, Precision 93.7%, Recall 82.5%
Random Forest: AUC 0.9955, Precision 94.4%, Recall 77.4%
LightGBM: AUC 0.7555 (underperformed)

Winner: XGBoost with near-perfect class separation



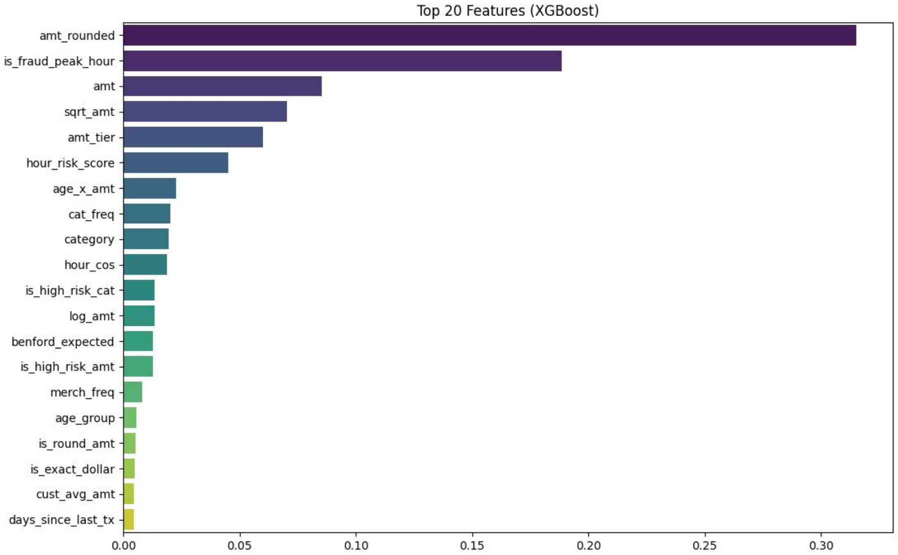
```
         Model    Recall  Precision  Real_Cost
0      XGBoost  0.825098   0.937223     135350
1     LightGBM  0.517555   0.268647     392670
2  RandomForest  0.773732   0.943695     174710
```



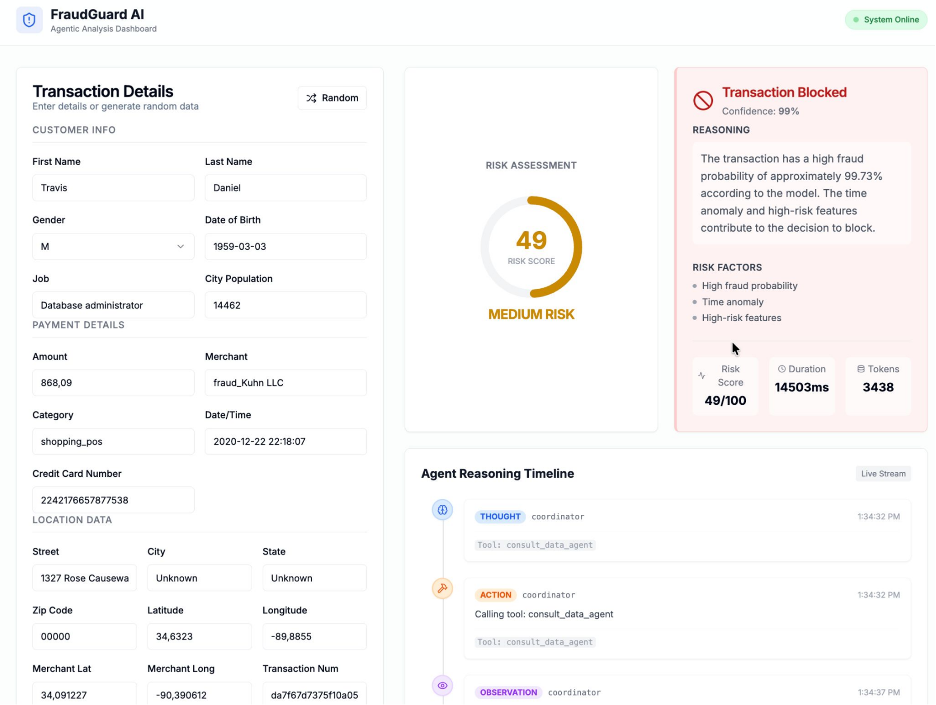
Feature Importance Analysis

Top Features:

- amt_rounded (0.30): Round amounts
- is_fraud_peak_hour (0.20): 21:00-23:00 flag
- amt, sqrt_amt, amt_tier: Amount variations
- hour_risk_score (0.05): Temporal risk
- Amount features dominate top 6 positions



System Architecture: Agentic Framework



FraudGuard AI
Agentic Analysis Dashboard

● System Online

Transaction Details

Enter details or generate random data

Random

CUSTOMER INFO

First Name	Last Name
Travis	Daniel
Gender	Date of Birth
M	1959-03-03
Job	City Population
Database administrator	14462

PAYMENT DETAILS

Amount	Merchant
868,09	fraud_Kuhn LLC
Category	Date/Time
shopping_pos	2020-12-22 22:18:07
Credit Card Number	
2242176657877538	

LOCATION DATA

Street	City	State
1327 Rose Causewa	Unknown	Unknown
Zip Code	Latitude	Longitude
00000	34,6323	-89,8855
Merchant Lat	Merchant Long	Transaction Num
34,091227	-90,390612	da7f67d7375f10a05

RISK ASSESSMENT

49
RISK SCORE

MEDIUM RISK

Transaction Blocked
Confidence: **99%**

REASONING

The transaction has a high fraud probability of approximately 99.73% according to the model. The time anomaly and high-risk features contribute to the decision to block.

RISK FACTORS

- High fraud probability
- Time anomaly
- High-risk features

Risk Score	Duration	Tokens
49/100	14503ms	3438

Agent Reasoning Timeline

Live Stream

THOUGHT coordinator 1:34:32 PM

`Tool: consult_data_agent`

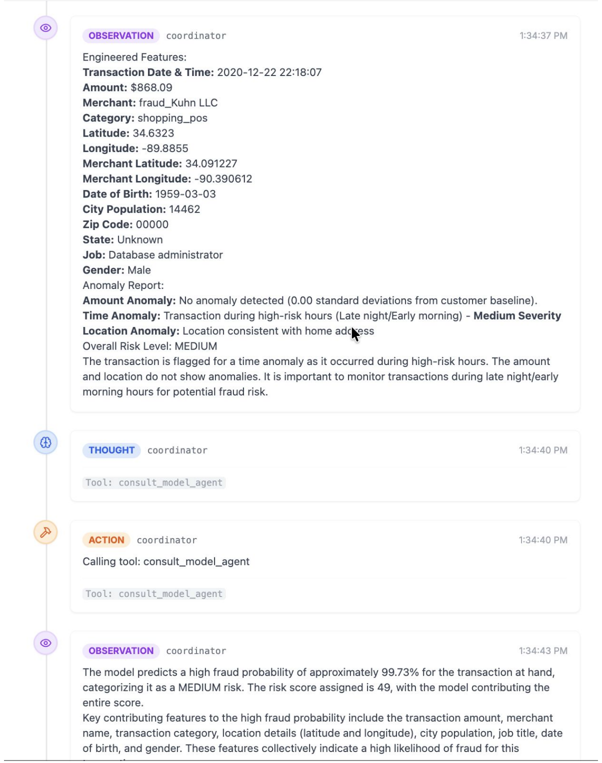
ACTION coordinator 1:34:32 PM

Calling tool: consult_data_agent

`Tool: consult_data_agent`

OBSERVATION coordinator 1:34:37 PM

FraudGuard AI Dashboard: Transaction details, risk assessment, and decision



OBSERVATION coordinator 1:34:37 PM

Engineered Features:
Transaction Date & Time: 2020-12-22 22:18:07
Amount: $868.09
Merchant: fraud_Kuhn LLC
Category: shopping_pos
Latitude: 34.6323
Longitude: -89.8855
Merchant Latitude: 34.091227
Merchant Longitude: -90.390612
Date of Birth: 1959-03-03
City Population: 14462
Zip Code: 00000
State: Unknown
Job: Database administrator
Gender: Male
Anomaly Report:
Amount Anomaly: No anomaly detected (0.00 standard deviations from customer baseline).
Time Anomaly: Transaction during high-risk hours (Late night/Early morning) - **Medium Severity**
Location Anomaly: Location consistent with home address
Overall Risk Level: MEDIUM
The transaction is flagged for a time anomaly as it occurred during high-risk hours. The amount and location do not show anomalies. It is important to monitor transactions during late night/early morning hours for potential fraud risk.

THOUGHT coordinator 1:34:40 PM

`Tool: consult_model_agent`

ACTION coordinator 1:34:40 PM

Calling tool: consult_model_agent

`Tool: consult_model_agent`

OBSERVATION coordinator 1:34:43 PM

The model predicts a high fraud probability of approximately 99.73% for the transaction at hand, categorizing it as a MEDIUM risk. The risk score assigned is 49, with the model contributing the entire score.
Key contributing features to the high fraud probability include the transaction amount, merchant name, transaction category, location details (latitude and longitude), city population, job title, date of birth, and gender. These features collectively indicate a high likelihood of fraud for this

THOUGHT ⟶ ACTION ⟶ OBSERVATION loop with tool calls and model predictions

Agent Architecture & ReAct Workflow

Multi-Agent Hierarchy

- Coordinator Agent: Decomposes tasks, delegates to sub-agents, synthesizes final decision
- Data Agent: Uses detect_anomalies tool (Z-scores, Benford's Law, temporal risk)
- Model Agent: Uses model_predictor and calculate_risk_score tools

ReAct Loop Implementation (THOUGHT → ACTION → OBSERVATION)

- THOUGHT: Agent analyzes current state and plans next action
- ACTION: Agent calls specific tool (consult_data_agent, consult_model_agent)
- OBSERVATION: Agent receives tool results and updates reasoning
- Loop continues until sufficient evidence gathered for decision

Tools & Capabilities

- detect_anomalies: Statistical analysis, Benford's Law, temporal patterns
- model_predictor: XGBoost inference via ONNX Runtime (5ms)
- calculate_risk_score: Combines probability + anomalies → 0-100 score
- get_customer_profile: Historical spending patterns and baselines

Real Example from Dashboard

- Transaction: $868.09 at 22:18:07 (high-risk late night hours)
- Model Prediction: 99.73% fraud probability
- Risk Score: 49/100 (MEDIUM RISK)
- Decision: BLOCK with 99% confidence based on time anomaly

Business Impact Analysis

Cost-Benefit Analysis

Baseline (no model): $1,501,200 in fraud losses

XGBoost Model: $908,220 net benefit

ROI: 60.5% reduction in fraud-related costs

Operational Metrics

6,194 fraud cases detected and prevented

Only 416 false positives (6.3% false alarm rate)

1,312 fraud cases missed (17.5%)

Additional Benefits

Real-time prevention at point-of-sale

Explainable decisions for regulatory compliance

Continuous improvement through analyst feedback

Deployment & Technical Stack

Backend Stack

FastAPI: Asynchronous Python web framework

LangChain: Agent orchestration & LLM integration

ONNX Runtime: Optimized model inference (5ms latency)

WebSocket: Real-time streaming to frontend

Frontend Stack

Next.js 14: React 18 + TypeScript framework

Tailwind CSS + shadcn/ui components

Real-time ReAct timeline visualization

Docker Deployment

Multi-container setup with Docker Compose

Backend: Python 3.11-slim (512MB memory limit)

Frontend: Node 20-alpine (256MB memory limit)

Production ready with health monitoring

Conclusion & Future Work

Key Achievements

State-of-the-art performance: ROC-AUC 0.9976

Production ready: <50ms latency, containerized deployment

Explainable AI: Multi-agent architecture bridges accuracy and trust

Business impact: $908K annual savings (60.5% ROI)

Future Enhancements

Federated Learning: Train on decentralized data while preserving privacy

Graph Neural Networks: Detect collusion rings and mule accounts

Reinforcement Learning: Optimize decision thresholds based on outcomes

Online Learning: Continuous adaptation to emerging fraud tactics

Recommendation

Deploy XGBoost with agentic explainability as primary system

Implement monthly retraining with sliding temporal window

LIVE DEMO