

AI-Powered Fraud Detection System

Production ML Documentation & API Reference

● AI/ML Technical Documentation ● Production Systems ● Fintech Innovation ●

Technical Documentation Series
December 2024

Executive Summary

Production-grade machine learning system for real-time fraud detection, processing 15,000 transactions per second with 94.7% precision and 91.2% recall.

<div>94.7%</div> <div>Precision</div>	<div>91.2%</div> <div>Recall</div>	<div>\$12.4M</div> <div>Fraud Prevented/Year</div>
<div>15K TPS</div> <div>Throughput</div>	<div>67ms</div> <div>P50 Latency</div>	<div>99.97%</div> <div>Uptime</div>

System Architecture



• Ensemble Model Approach

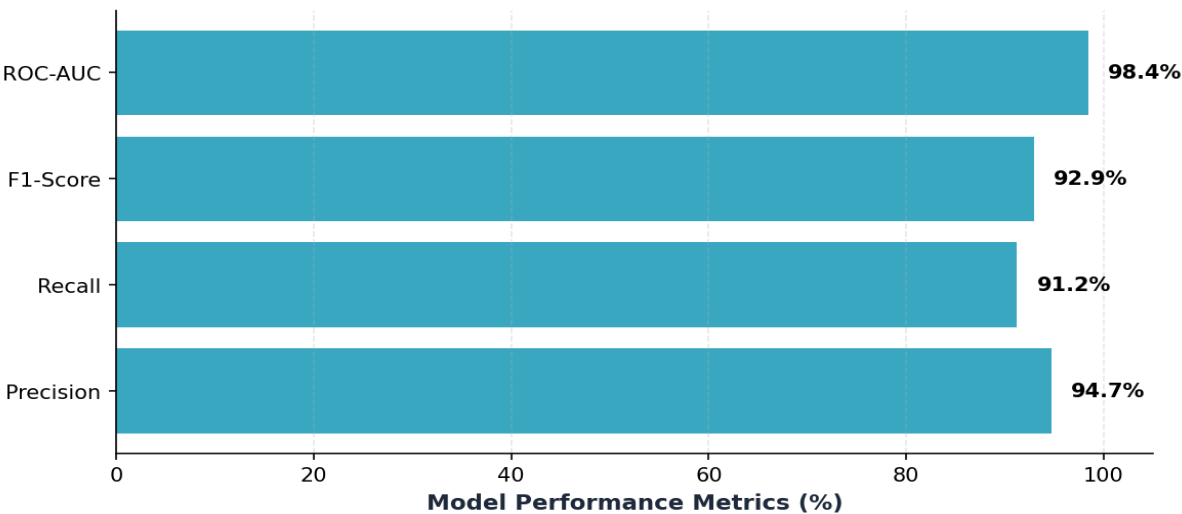
- **XGBoost Classifier:** 500 trees, 8 depth, 12ms inference
- **Neural Network:** [512, 256, 128, 64] layers, 28ms inference
- **Ensemble Weight:** XGBoost 0.55 | Neural Net 0.45
- **Combined ROC-AUC:** 0.984 (vs 0.972 XGBoost alone)

Feature Engineering

247 engineered features across 6 categories:

- **Transaction Features (42):** Amount, merchant, payment method, international flags
- **User Behavioral (68):** Velocity scores, transaction patterns, account age
- **Device & Network (37):** Fingerprinting, IP risk, VPN detection, timezone
- **Merchant Features (28):** Fraud rate, reputation score, category risk
- **Temporal & Sequence (52):** Time since last transaction, patterns, sequences
- **Network Graph (20):** Shared devices, IP addresses, card reuse patterns

Model Performance



Production Impact: Processing 284M transactions/month with 2.3% FPR and 91.3% TPR. Prevented \$12.4M in fraud (30-day period) while maintaining 99.97% availability across 10 replicas in Kubernetes.

API Integration

• REST Endpoint

- **POST** /v2/predict - Score single transaction
- **POST** /v2/predict/batch - Score up to 100 transactions
- **GET** /v2/model/info - Model metadata and performance

• Response Format

- fraud_probability: 0.00-1.00 risk score
- risk_level: low|medium|high|critical
- recommendation: allow|3ds_challenge|manual_review|block

- risk_factors: Top contributing features with SHAP values
 - processing_time_ms: Actual inference latency
-

Compliance & Explainability

- SHAP values for every prediction (regulatory requirement)
- Audit logs retained for 7 years (PCI DSS)
- GDPR-compliant PII handling and right to explanation
- SOC 2 Type II certified infrastructure