Group 5

Fraud Detection for Bank Payments

Modern Data Architectures for Big Data II

Christelle El-Haddad
Flavio Valerio
Luis Wilhelmi
Paula Caceres
Radwan Fenaer
Regina Ortega

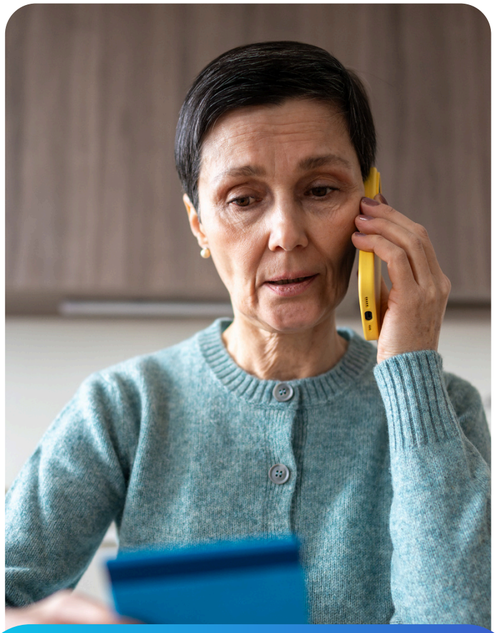
Table of content

The Current Landscape – Why Fraud Detection Matters

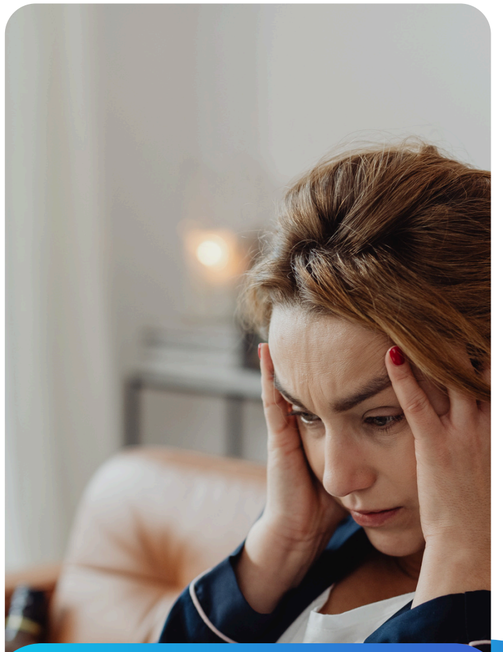
Fraud may be rare, but its impact is massive. As digital payments grow, institutions face new forms of risk that are harder to detect and costlier to manage.

Rising digital payments
→ more opportunities for fraud

Manual reviews are slow and inconsistent



Missed fraud = direct financial loss + chargebacks + damaged trust



Over-flagging frustrates good customers

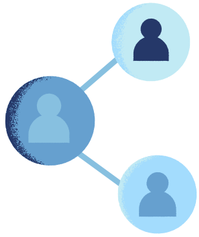
The Business Problem

WHAT …

Need we are trying to solve for

With the rise of digital payments, banks face an increasing number of fraudulent transactions that are difficult and costly to detect manually.

Traditional rule-based systems often miss sophisticated patterns or create too many false alerts, showing the need for intelligent, scalable detection systems.



WHY …

Is this important for end users and financial institutions

Fraud is a billion-dollar issue for financial institutions, leading to direct losses and damaged customer trust.
Every missed fraud increases risk, while false positives frustrate legitimate users.

Detecting fraud effectively is essential to protect both banks and customers in a digital-first world.

HOW …

Are we going to deliver real value to users and institutions

Using the BankSim dataset, our model applies Spark Machine Learning to learn behavioral patterns from synthetic bank payments.

By combining statistical features — such as amount, merchant, and transaction timing — the model assigns a risk score before authorization, allowing proactive and efficient fraud prevention.

Data Landscape

Transaction-level data reveals behavioral patterns of fraud

We analyzed thousands of transactions, each representing real customer behavior across merchants and categories.

Fields

Amount, time step, merchant, category, customer traits

Labeled Target

Whether a transaction is fraudulent

Patterns

Enough variety to learn patterns across merchants and time

SPARK MACHINE-LEARNING PIPELINE

MLlib provides a uniform set of high-level APIs built around the concept of Pipelines, which include data preprocessing, feature extraction, model fitting, and evaluation.

Data Source

BankSim Dataset

Transaction data (amount, time, merchant, category, fraud label)

Cleaning & Feature Engineering

Handle missing values, encode categoricals, cast numerics

Data Preparation

Model Pipeline (MLlib)

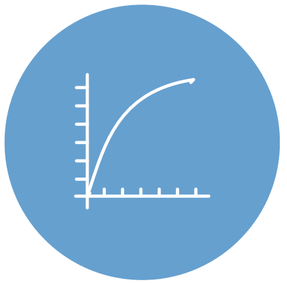
Spark ML Pipeline → Random Forest

(**Transformers**: StringIndexer, VectorAssembler , **Estimator**: RandomForestClassifier)

Model Performance (AUC = 0.9976)

Confusion matrix + threshold tuning

Evaluation



Insights & Reporting

Business Impact / Risk Scoring

Identify high-risk transactions → reduce losses

Challenges Adressed

Class Imbalance

Fraud = only 1.2% of all transactions → required metrics beyond accuracy.

Fraud Clusters

Some merchants & categories had extreme fraud concentration, creating risk of overfitting.

Precision vs Recall

High amounts correlate with fraud → tuned thresholds to avoid false positives.

Fairness & Governance

Different base rates by gender → kept explainability & audit logs for transparency.

Data & Engineering Quality

Ensured correct model fitting, consistent data types, and stable transformations.

Business Approach – From data to decisions

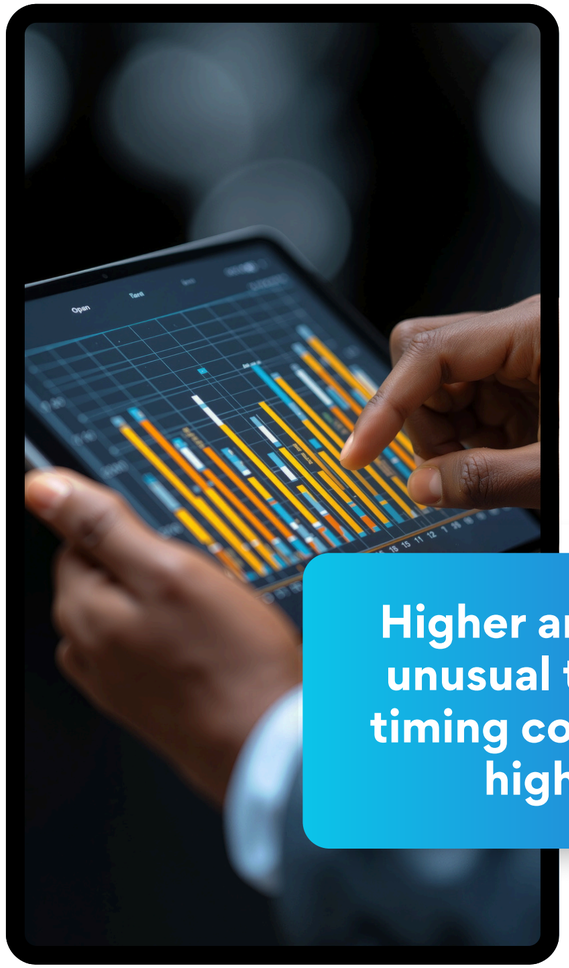
Enough variety to learn patterns across merchants and time

Learn from historical outcomes to estimate today’s risk

Score each transaction before authorization or settlement

Route the riskiest cases to automated blocks or fast human review

Learn ⇨ **Score** ⇨ **Act**



Key Insights from the Data

Fraud concentrates in specific conditions

Higher amounts and unusual transaction timing correlate with higher risk

Certain merchant categories show elevated rates

Legitimate customers show steadier patterns across time and merchants

Performance Snapshot

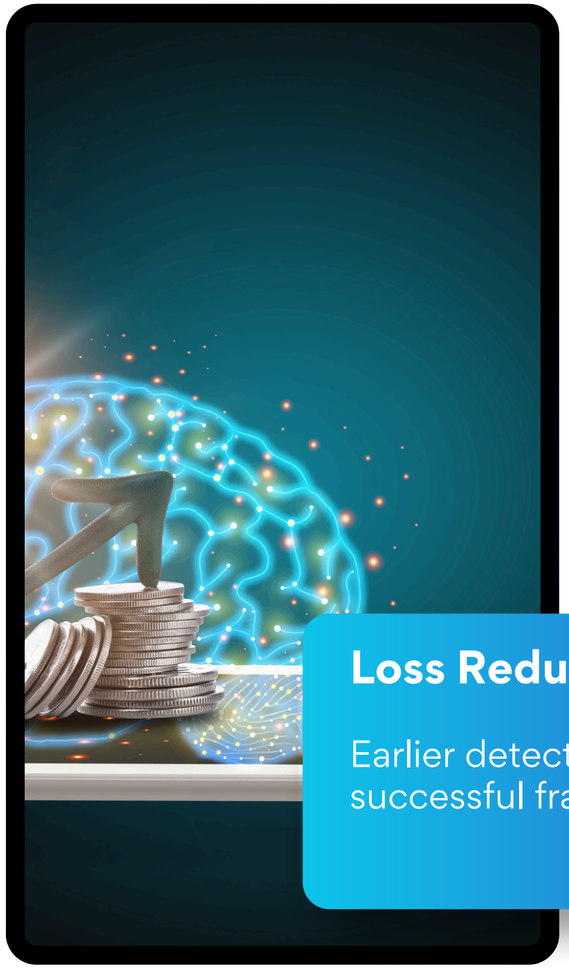
Strong separation between fraud and legitimate transactions

Model discrimination (AUC): 0.99

The risk score ranks truly fraudulent transactions higher most of the time

Confusion matrix shows low false-positive rate with strong fraud capture

Thresholds can be tuned for the desired balance of customer friction vs fraud catch



Impact on Operations

Reduce losses, accelerate reviews, protect experience

Loss Reduction

Earlier detection → fewer successful frauds

Efficiency

Prioritize top-risk cases → analysts focus where it matters

Customer Trust

Fewer false declines with targeted controls

Responsible AI

Governance and Trust

Explainable AI Builds Trust and Compliance Confidence

Explainability	Fairness checks	Auditability
• High transaction amount raised the risk score • Unusual merchant or time of day increased risk • Regular customer patterns lowered risk	Regularly audit for bias across gender, age, or geography to ensure the system treats customers fairly	Maintain decision logs and explanations for every scored transaction so investigators and compliance officers can review the rationale behind each flag

Risks and Mitigation

Manage operational and customer risks proactively

False Positives

Start with soft blocks / step-up verification

Data drift

Weekly monitoring and retraining

Regulatory

Maintain explanations and audit logs

Business Questions Answered

How can we identify fraudulent transactions before authorization?	The model assigns a fraud risk score per transaction → enables real-time prevention.The model assigns a fraud risk score per transaction → enables real-time prevention.
Which behaviors are most associated with fraud?	Fraud often involves high amounts, specific merchants, and unusual timing.
How can we balance fraud detection vs customer experience?	Threshold tuning allows minimizing false positives while keeping strong recall.
How effective is the model?	AUC = 0.9976 → accurately separates fraud from legitimate transactions.
What business impact can it create?	Reduces losses, increases analyst efficiency, and strengthens customer trust

Conclusions

PATTERNS ARE LEARNABLE.

Combining amount, timing (step), and merchant/category signals delivers consistent ranking of risky payments.

EFFECTIVE DISCRIMINATION.

Our Spark-ML Random Forest shows strong separation between fraud and legitimate transactions (AUC ≈ 0.998).

OPERATIONAL VALUE

Scoring before authorization can reduce losses and focus analysts on the highest-risk cases while preserving good customer experience.

RESPONSIBLE AI MINDSET.

We structured the pipeline for explainability and threshold control, so risk can be tuned to business needs.

Next Steps – Scale & Resilience

Make it real-time, adaptive, and enterprise-grade.