

# THE ILLUSION OF SECURITY

*Presented by:*  
*Rayan HAMADEH*  
*Yusra AL AISSAMI*  
*Luca RAGAGLIA*  
*Michele PAGANO MARIANO*  
*Mehdi MAMLOUK*  
*Rachid SEFRIOUI*



# BLUE VS RED PILL



BLUE PILL

- **trust** legacy, rules-only systems and **hope** yesterday's signatures still catch tomorrow's fraud.



RED PILL

- **admit** the attacker adapts faster than rules and measure reality with data



We chose the **red pill** and built a transparent ML pipeline that lets banks see where risk truly lives, not where we wish it lived



# THE REAL PROBLEM COLD START

- Fraud is rare, evolving and asymmetric
- IBM constraint: Fraudulent clients in training are different from those in evaluation → we must generalize to new clients.
- Additional realities: no eval leakage, temporal drift, severe class imbalance.
- Translation: we're detecting first-time, never-seen behavior—not replaying past attacks.

# OBJECTIVES AND SUCCESS CRITERIA

- Predict fraud and rank transactions by risk for analyst review
- Operational focus: high precision at small  $k$  (top of the queue) and strong PR-AUC
- Success = generalization to unseen clients with auditable decisions



# DATA OVERVIEW

- Transactions (train):  $\approx$  210k (2016–2018) + labels.
- Evaluation: unlabeled transactions for scoring only.
- Side tables: cards, users, MCC codes (merchant categories).
- Signals used: amounts, time, channel/type (online/chip/swipe), card/profile, geo/state, MCC.



# DESIGN PRINCIPLES

- No leakage: evaluation features never used for training
- Temporal honesty: past  $\rightarrow$  future split to mimic production
- Group awareness: GroupKFold by client/card to avoid “same customer in train & val”
- Imbalance-aware: optimize PR-AUC/Precision@k; class\_weight=“balanced”



# SOLUTION OVERVIEW

1. Ingest & Clean (CSV/JSON → pandas/parquet)
2. Feature Layer (temporal, merchant/MCC, channel/type, monetary, card/profile, geo)
3. Modeling (Logistic Regression pipeline)
4. Thresholding (picked on validation for F1 / capacity)
5. Ranked Queue (top-risk transactions)
6. Dashboard (monitoring & action)

## Why Logistic Regression?

- Speed
- Sustainability
- Explainability for risk & compliance



# FEATURE ENGINEERING

**01** Temporal: hour, day of week, weekend/night flags

—

**02** Channel/Type: online vs chip vs swipe (encodes attack surface)

—

**03** Merchant: MCC categories + frequency/rarity effects

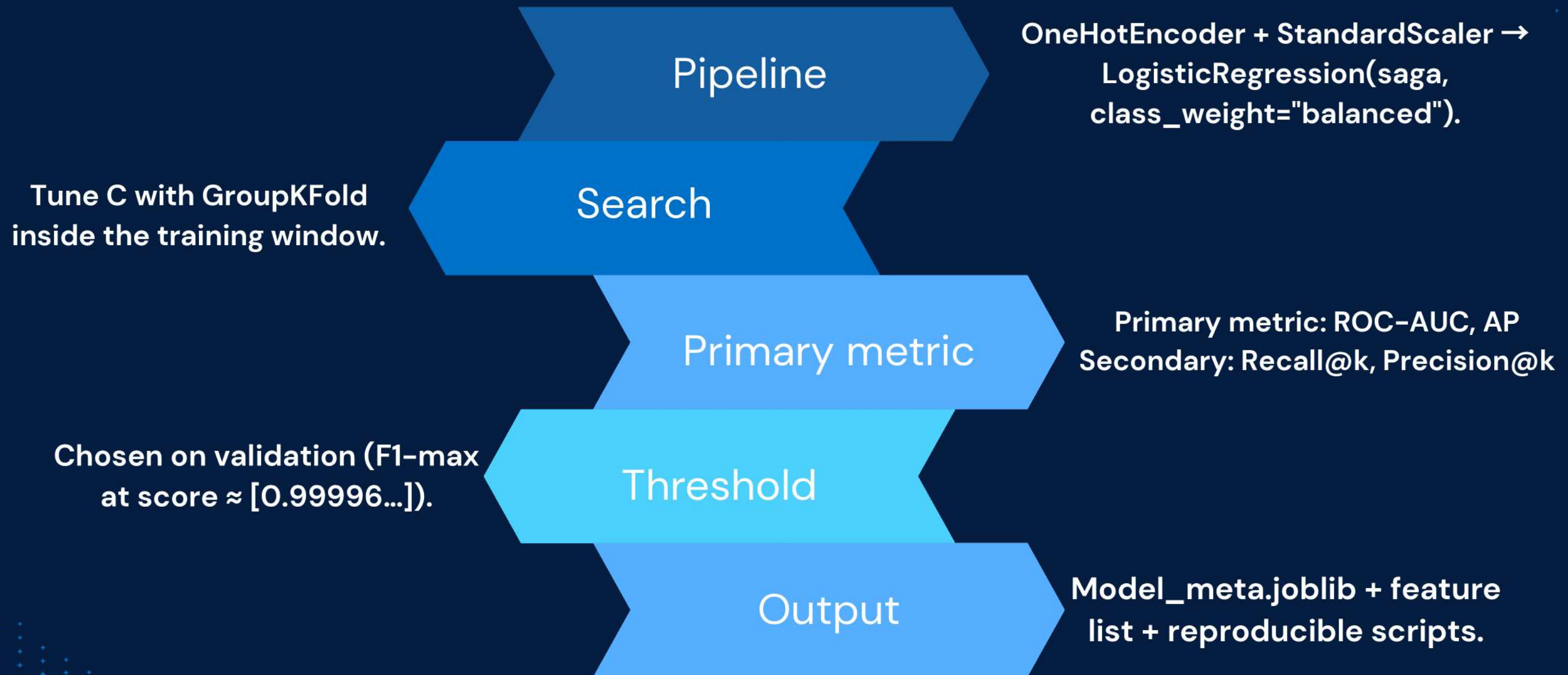
—

**04** Monetary: raw, log-scaled, clipped amounts, overlimit flags

—

**05** Profile/Geo: card brand/type, state/region signals

# MODEL & VALIDATION



# RESULTS

**0.47551**

AP

**0.97783**

ROC-AUC

**0.33962**

P@0.5

**0.69231**

R@0.5

**0.99996**

Threshold





# DASHBOARD

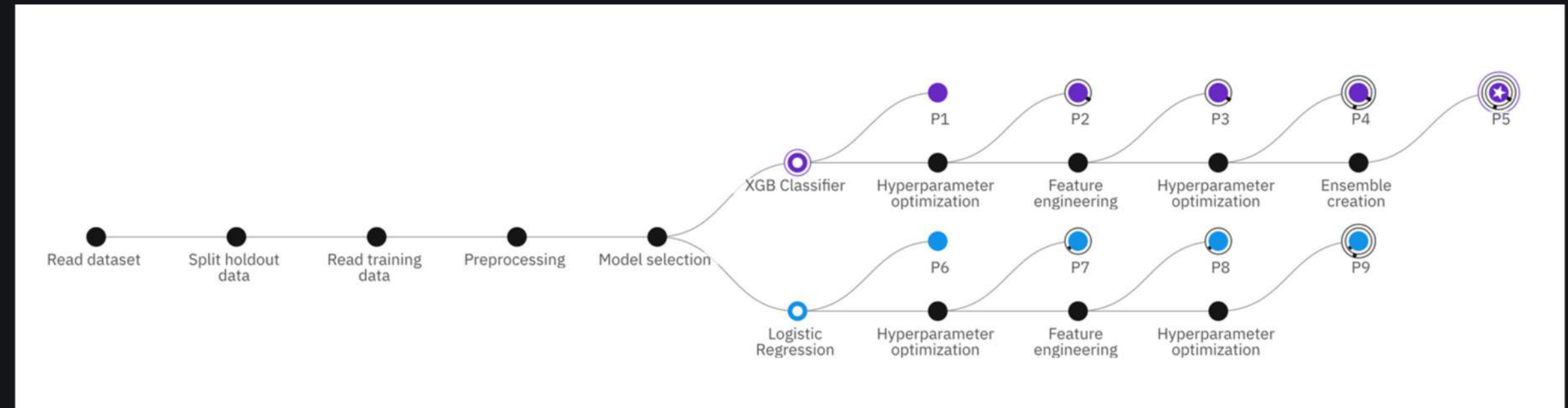
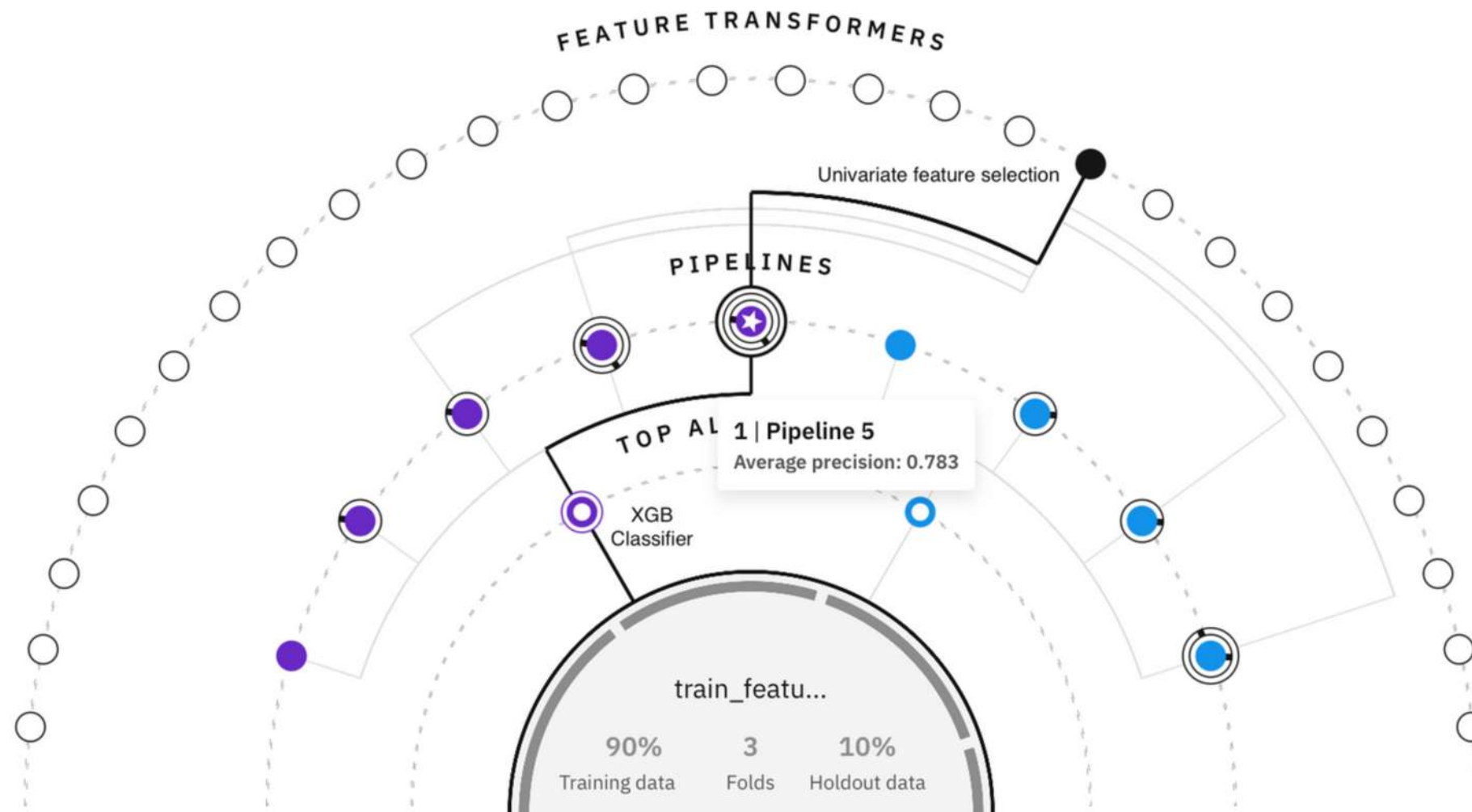




# IBM'S AUTOAI VS TRADITIONAL PROGRAMMER

- Accelerated Development & Efficiency: AutoAI automates the most time-consuming tasks in model building, including data preparation, feature engineering, algorithm selection, and hyperparameter optimization
- Optimized Model Performance: The platform systematically explores thousands of potential pipeline combinations
- Streamlined MLOps & Deployment
- Enhanced Trust & Explainability
- Increased Productivity & Accessibility

# SOLUTION OVERVIEW



# RESULTS

Measures	Holdout score	Cross validation score
Accuracy	0.999	0.999
Area under ROC	0.997	0.995
Precision	0.889	0.894
Recall	0.516	0.606
F1	0.653	0.720
Average precision	0.728	0.783
Log loss	0.004	0.003

IBM

# SEEING THE CODE FROM DASHBOARD TO ACTION

## When



fraud spikes mid-afternoon (~14:00–16:00) → smart shift staffing / rules tightening

## How



Online channel shows higher fraud rate than chip/swipe → prioritize 3DS, device fingerprinting, velocity checks

## Who



Top-20 highest-risk transactions → immediate review/blocking before settlement

## Why



Score distribution + threshold line make the decision transparent

# RISK & VALUE

## Value now

- Loss containment: investigators start with the highest-impact 1–2%.
- Lower friction: calibrated threshold reduces false positives on good customers.
- Audit-ready: explainable LR coefficients for risk & compliance

## Roadmap

- calibration
- comparison with even more models
- cost-based thresholding
- continuous drift monitoring.

## The red pill

- move from reactive defense to proactive detection
- see the code → act before the loss



# Thank's For Listening

Connect with us....  
For an internship maybe?

