

TEAM SMARTBLOCK

D.A.T.A HACKATHON

AML WALLET RISK SCORING VIA ORACLE
(BTC/ETH DATA)

INTRODUCTION:

TEAM SMARTBLOCK :

KESHAV BHOTIKA

NIRVIK AGARWAL

APARNA AMARNANI

ASTHA DAD

PROBLEM STATEMENT:

Non-custodial wallets lack institutional oversight and on-chain risk ratings, with no real-time risk signals in smart contracts to assess AML compliance dynamically.

Oracle-driven dynamic risk checks integrated with smart contracts enable immediate AML validation, producing compliance status and risk scores to reduce AML exposure in FI and DeFi platforms.

Solution: SmartBlock

- Detection approach combines unsupervised ML (**Isolation Forest**) with **rule-based scoring** to improve accuracy.
- Inputs analyzed include the last **10–20 BTC/ETH** transactions along with optional ETH contract approvals.
- Risk scoring system rates wallets on a scale of **1–10**, with scores of ≥ 7 automatically flagged for review.
- Data pipeline leverages **BigQuery** for blockchain data and **Postgres** for wallet-level feature storage.

Technologies we aim to use:

ML Backend

Lightweight, scalable APIs with Python Flask/FastAPI powering real-time inference.

Oracle Service

Decentralized data bridge built in TypeScript, securely fetching and delivering on-chain/off-chain data to the ML engine.

ML Model

Isolation Forest for anomaly detection, combined with rule-based logic for fraud prevention.

Data Source

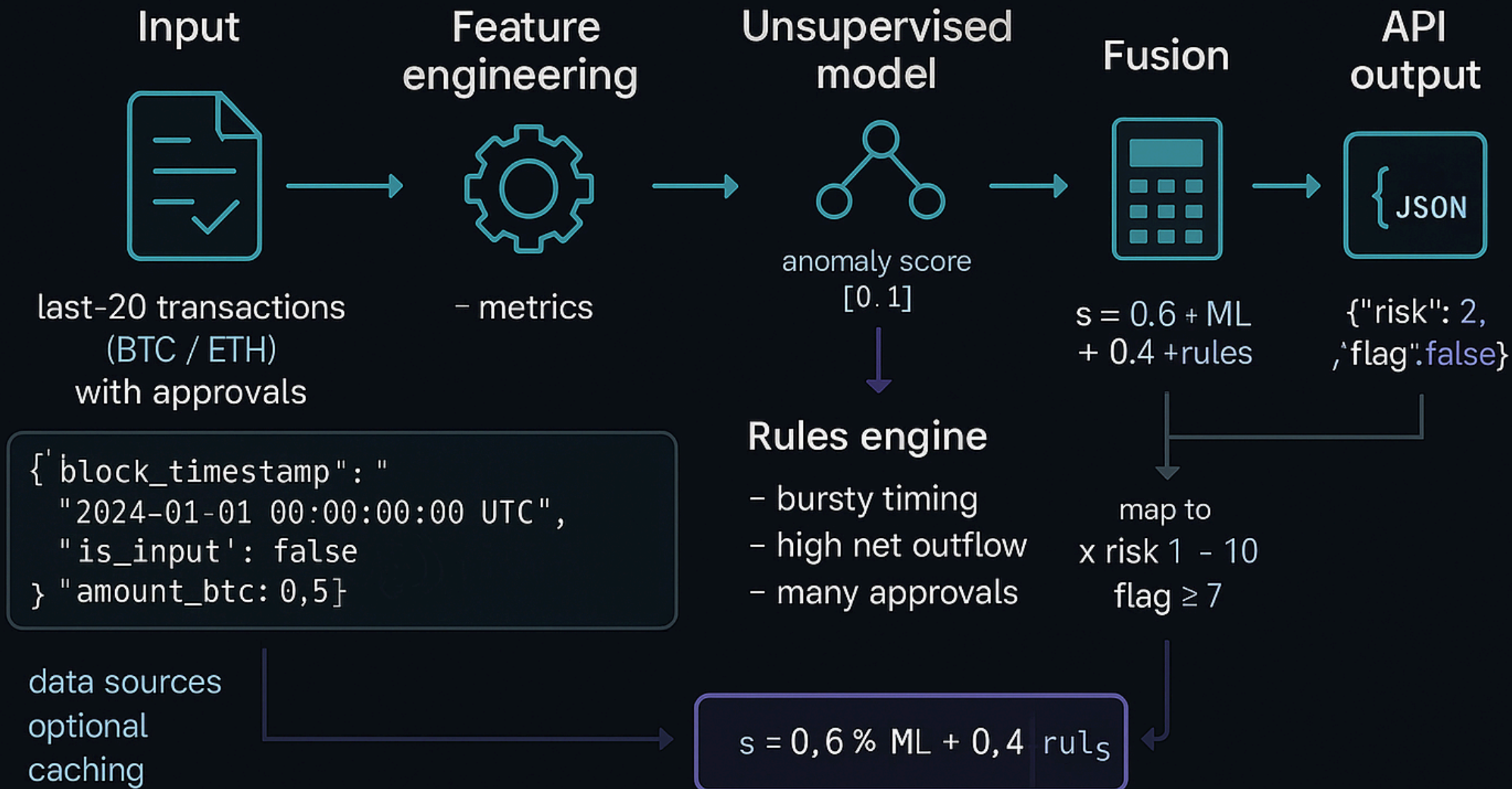
Google BigQuery APIs for public BTC/ETH blockchain transactions.

Architecture Diagram:



Scoring Methodology

AWL Scoring Pipeline



Challenges & Mitigations

01

Challenges Faced

Key challenges included lack of labeled data requiring unsupervised learning, massive blockchain data volume demanding sampling and materialized views, diverse log formats needing normalization, and cold-start wallets risking inaccurate scoring.

02

Strategies Implemented

Mitigation involved combining unsupervised ML with explicit rules, using materialized views for efficient feature extraction, robust log parsing for data consistency, and fallback heuristics for wallets with limited data.

FUTURE SCOPE

We can integrate with other chains and data sources to enrich wallet history—more addresses, longer timelines. This bigger dataset will make our unsupervised model steadier and our explanations more reliable.

We'll can improve the rule layer: tune thresholds with real distributions, add contextual rules (time of day, chain congestion), and weight rules by confidence. Combining calibrated rules with the model's anomaly score will reduce false flags while catching sharper risk patterns earlier.

Add explainability with SHAP/LIME to show which features drive each wallet's risk score.



THANK YOU