Methodology Document: Wallet Credit Scoring

Objective

To develop a robust and explainable scoring system (0-100) for wallets interacting with a lending protocol, capturing behavioral patterns such as borrowing discipline, repayment reliability, and transaction engagement.

Initial Observations & Iteration Journey

- At first, we noticed a major flaw: after running the initial clustering-based score model on the three largest JSON files, only 2 wallets scored above 90, and just 3 wallets above 50 out of 1,000 - clearly an overly compressed distribution.

- This behavior prompted suspicion about score clumping and lack of resolution across wallet tiers. Upon further debugging, we found that the KMeans cluster-label-based adjustments were overly dominant, hiding genuine feature differences.

- We then pivoted to a more feature-driven, linear, and explainable scoring pipeline, allowing score differentiation without black-box clustering artifacts.

Data Loading & Preprocessing

- The input data consisted of JSON files containing user-level interactions categorized under: deposits, withdraws, borrows, repays, and liquidations.
- Each action was flattened into a uniform transaction structure containing: wallet_address, action, amount, and timestamp.
- Malformed or missing entries were gracefully skipped, improving robustness.

Feature Engineering: Extracting Wallet Behavior

For each wallet, we computed the following behavioral features:

- Total Deposits
- Total Borrows
- Total Repayments
- Repayment Ratio
- Borrow-to-Deposit Ratio
- Transaction Frequency
- Avg. Transaction Interval

Each feature was carefully chosen to reflect financial responsibility, activity level, and risk-taking behavior.

Score Computation Logic

Rather than relying on KMeans clusters, we computed a weighted linear score from normalized features:

score = (+0.20 * norm(total_deposits)

+0.25 * norm(repayment_ratio)

-0.30 * norm(num_liquidations)

-0.15 * norm(borrow_to_deposit_ratio)

+0.10 * norm(txn_frequency)

-0.05 * norm(avg_txn_interval)

+0.05 * norm(num_unique_actions))

Weight tuning was based on behavioral intuition and experimentation. Repayment and liquidation dominate to encourage responsible borrowing. Diversity and frequency give credit to genuinely active wallets.

Normalization and Rescaling

Each feature was min-max normalized with epsilon smoothing to avoid zero-division. The final raw score was linearly scaled to a 0-100 range for intuitive understanding.

Result Insights

The new logic produced a full score spread from 0 to 100. The top 12 wallets scored in the 99.95-100 range, indicating healthy differentiation. Even marginal behavioral differences yielded distinguishable scores - ideal for ranking or airdrops.

Robustness & Consistency

The model handles missing keys, empty actions, and inconsistent behaviors. It avoids domination by any single feature or clustering artifact. Works reliably across diverse dataset sizes.