**Methodology Document: Scoring Compound V2 Wallets**

**Objective**

The objective is to assign a **credit score (ranging from 0 to 100)** to each wallet that has interacted with the Compound V2 protocol. This score reflects the wallet's historical behavior and reliability. Higher scores indicate responsible, long-term, protocol-aligned usage, while lower scores are associated with risky, short-term, or exploitative behavior.

**Data Overview**

We utilized raw transaction-level logs from the Compound V2 protocol. The transactions include key actions such as deposit, withdraw, borrow, repay, and liquidation. To ensure representative coverage, the three largest files by size were selected from the provided dataset, maximizing wallet diversity and volume.

**Behavioral Assumptions**

**Positive (Good) Behavior:**

* Regular, consistent interaction with the protocol
* Higher repay ratios (i.e., repayment of borrowed funds)
* Minimal or no liquidation events
* Balanced use of deposits and borrows
* High overall activity volume and active days

**Negative (Risky) Behavior:**

* High liquidation rates
* Little or no repayments
* One-off or extremely short-lived interaction
* Unbalanced behavior (e.g., borrowing without collateral deposits)

**Feature Engineering**

The raw logs were aggregated into wallet-level features capturing behavioral signals:

| **Feature** | **Description** |
| --- | --- |
| net\_deposit\_usd | Total deposits minus total withdrawals |
| net\_borrow\_usd | Total borrows minus total repays |
| active\_days | Number of distinct days the wallet interacted with the protocol |
| total\_volume\_usd | Sum of all transaction volumes in USD |
| deposit\_to\_borrow\_ratio | Ratio of deposits to borrows (proxy for overcollateralization) |
| repay\_ratio | Total amount repaid divided by total amount borrowed (capped at 1) |
| liquidation\_rate | Number of liquidations divided by borrow events |

Missing values (e.g., due to zero activity in certain categories) were imputed with neutral values such as zero.

**Scoring Approach**

All features were normalized using MinMaxScaler to bring them into a common 0–1 range. Each feature was assigned a specific weight reflecting its contribution to trustworthiness:

| **Feature** | **Weight** |
| --- | --- |
| net\_deposit\_usd | 0.20 |
| net\_borrow\_usd | 0.10 |
| active\_days | 0.15 |
| total\_volume\_usd | 0.15 |
| deposit\_to\_borrow\_ratio | 0.15 |
| repay\_ratio | 0.15 |
| liquidation\_rate | -0.10 |

The final credit score was calculated using a weighted sum of the scaled features:

ini

Copy code

score = (scaled\_features \* weights).sum(axis=1) \* 100

This ensures the resulting score lies between 0 and 100.

**Labeling Logic**

To support downstream categorization, scores were labeled as follows:

* **Highly Trustable (Label 2):** Score ≥ 80
* **Medium Trust (Label 1):** 50 ≤ Score < 80
* **High Risk (Label 0):** Score < 50

**Output Files**

The following CSV files were generated:

* top\_1000\_wallets.csv: Top 1,000 wallets by score (descending)
* bottom\_1000\_wallets.csv: Lowest 1,000 wallets by score (ascending)

These files can be used for further wallet behavior analysis and validation.

**Justification**

This approach is:

* **Unsupervised and explainable:** All features are derived and weighted based on logical protocol usage assumptions.
* **Domain-aligned:** Feature engineering directly maps to mechanisms in Compound V2 (e.g., repayments, collateral, liquidations).
* **Generalizable:** The methodology can be applied to other lending protocols with similar transaction types.