**Wallet Risk Scoring Model: Methodology Report**

**1. Objective**

The objective of this project was to develop a risk scoring model for a given list of Ethereum wallets based on their on-chain transaction history with the Compound V2 and V3 protocols. The final score for each wallet ranges from 0 (highest risk) to 1000 (lowest risk).

**2. Data Collection Method**

Transaction data was retrieved using the Etherscan API. To ensure a comprehensive dataset, the collection process was enhanced to include:

* **Normal Transactions:** Standard wallet-initiated transactions.
* **Internal Transactions:** Contract-to-contract calls, which are critical for capturing the full scope of DeFi interactions like receiving cTokens or claiming rewards.
* **Expanded Contract List:** The list of monitored addresses was expanded to include a wide range of Compound cTokens and core governance contracts, ensuring all relevant protocol interactions were captured.

This comprehensive approach provides a complete picture of each wallet's activity, capturing both actions initiated by the user and those received from the protocol.

**3. Advanced Feature Engineering**

The core of the model lies in its feature engineering, which moves beyond simple metrics to analyze the *intent* and *financial behavior* of each wallet.

* **Behavioral Classification (Transaction Type Analysis)**: Transactions were classified into supply, withdraw, and borrow\_or\_repay categories by analyzing the direction of value transfer. This allows the model to differentiate between risk-reducing behaviors (supplying liquidity, repaying loans) and risk-increasing behaviors (borrowing).
* **Leverage Proxy (borrow\_to\_supply\_ratio)**: This crucial feature (borrow\_count / supply\_count) serves as a direct proxy for a user's leverage. A high ratio indicates heavy borrowing against collateral, a primary risk factor for liquidation in volatile market conditions.
* **Financial Health Proxy (health\_factor\_proxy)**: Calculated as repay\_count / borrow\_count, this feature measures responsible loan management. A value greater than 1 signifies that a user is repaying more than they borrow, indicating a healthy, lower-risk position.
* **Standard Health Indicators**: Foundational features were also included:
  + successful\_tx\_rate: Measures technical competence.
  + wallet\_age\_days & is\_active\_recently: Indicate stability and current engagement.
  + unique\_contracts\_interacted: Shows the breadth of a user's protocol knowledge.

**4. Normalization Method**

All engineered features were scaled to a uniform range of 0 to 1 using **Min-Max Scaling**. This is a critical step to prevent features with large absolute values (like wallet age) from disproportionately influencing the score, ensuring that each feature's contribution is based purely on its assigned importance (weight).

**5. Scoring Method and Justification**

A **weighted linear model** was used to calculate the final score, with specific enhancements to handle different types of risk indicators.

* **Bi-directional Weighting**: Features were assigned positive (risk-reducing) or negative (risk-increasing) weights.
  + **Positive Weights**: Applied to desirable behaviors like a high successful\_tx\_rate and health\_factor\_proxy.
  + **Negative Weights**: Applied to risky behaviors, most notably a high borrow\_to\_supply\_ratio.
* **Impact-Driven Weights**: The weights were calibrated to give the highest impact to the most direct risk indicators. The borrow\_to\_supply\_ratio received the largest negative weight (-0.20), as high leverage is the most significant and immediate risk in a lending protocol.
* **Final Score Calculation**: The final raw\_score combines the weighted influence of all features. For negative indicators, the logic is inverted so that a higher value (e.g., more borrowing) correctly *decreases* the final score. This raw score is then scaled to the required 0-1000 range.

This advanced model provides a nuanced and realistic assessment of wallet risk by focusing on the actual financial behaviors exhibited within the Compound protocol.