Credit Scoring Methodology for Zeru Finance Protocol

Problem Definition

The objective was to create a credit scoring system to assess the creditworthiness of wallets on the Zeru Finance platform. This score, ranging from 0 to 100, reflects how well a wallet is managing its financial activities, such as deposits, borrows, repayments, and engagement with the ecosystem. It helps categorize wallets as low or high risk, informing decisions related to lending, borrowing, and other financial services within the platform.

Data Exploration and Feature Selection

We began by analyzing the available wallet data. Key features were selected to provide insights into wallet behavior:

- Total Deposits (USD): The total funds deposited by the wallet.
- Total Borrows (USD): The total funds borrowed by the wallet.
- Total Repays (USD): The total funds repaid by the wallet.
- Number of Liquidations: The count of times a wallet was liquidated due to insufficient collateral.
- Active Duration (Days): The period for which the wallet has been active on the platform.
- Borrow-to-Repay Ratio: This ratio compares the total borrowed funds to the repaid funds, helping assess financial responsibility.
- Days Since Last Activity: The number of days since the wallet last interacted with the platform.
- Account Age (Days): The length of time since the wallet's creation.

These features were selected as they provide a comprehensive picture of a wallet's activity, engagement, and financial health.

Data Preprocessing

To prepare the data for modeling, several preprocessing steps were carried out:

- Handling Missing Values: Missing values in features such as the number of liquidations were filled with zeros, reflecting that the wallet hadn't experienced that event yet.

- Outlier Detection and Removal: Extreme values in transaction amounts (e.g., very high deposits or borrows) were identified using the Interquartile Range (IQR) method and removed to ensure the data remained within reasonable bounds.

Feature Transformation

Some features needed transformation to ensure consistency across the dataset:

- Borrow-to-Repay Ratio: This ratio was computed as Total Borrows / Total Repays. A high ratio indicates that a wallet is borrowing more than it is repaying, which suggests potential financial risk. We scaled this feature to ensure comparability between wallets with differing activity levels.

- Outlier Handling: After identifying outliers using the IQR method, extreme values were removed to ensure the analysis was based on wallets within a reasonable transaction range.

Modeling the Credit Score

We implemented both rule-based scoring and regression modeling to calculate the credit score.

- Rule-based Scoring Model: A scoring system was devised, where wallets were assessed based on their engagement, activity, and risk factors. These categories (Engagement, Activity, and Risk) were combined to form an overall credit score between 0 and 100.

- Regression Model: To further refine the score prediction, a regression model was built. The model utilized the features listed above to predict the credit score directly. Using advanced regression techniques (such as linear regression or other suitable models), the model achieved an impressive R^2 score of 0.98, indicating that it could explain 98% of the variance in wallet behavior. This exceptional performance demonstrates the model's high accuracy and predictive power in assessing creditworthiness.

The final credit score was calculated by combining the outputs from both the rule-based and regression approaches, creating a robust evaluation system.

Thresholds for Classification

Once the credit scores were computed, we defined thresholds to categorize the wallets into three risk groups based on percentiles:

- High Risk: Wallets scoring below the 50th percentile.

- Medium Risk: Wallets scoring between the 50th and 75th percentiles.

- Low Risk: Wallets scoring above the 75th percentile.

This classification helps identify wallet behavior and potential risk, providing clear insights into wallet reliability.

Analysis of High and Low Scoring Wallets

The analysis revealed distinct patterns between high and low-scoring wallets:

- High-Scoring Wallets: These wallets exhibited responsible behavior, characterized by steady deposits and repayments, low borrow-to-repay ratios, longer account and active durations, and fewer liquidations. These wallets are financially healthy and considered reliable for future interactions.
- Low-Scoring Wallets: These wallets often showed signs of risk, including high borrow-to-repay ratios, periods of inactivity, and frequent liquidations. These wallets are considered high-risk, and further actions, like lending or financial product engagement, may need to be restricted or closely monitored.

Visualization of Results

Visualizations such as bar charts and distribution plots were created to show the distribution of credit scores across wallets. These visualizations helped us understand the range of wallet behaviors and the frequency of different score groups, enabling more informed decisions regarding wallet engagement and risk assessment.

Conclusion

By integrating various financial metrics, engagement indicators, and a high-performing regression model, we developed a robust credit scoring system. The model, which achieved an R^2 score of 0.98, ensures high accuracy in assessing wallet behavior and creditworthiness. This dynamic and transparent system provides valuable insights into wallet health and can be used for decision-making in lending, monitoring risk, and optimizing wallet engagement within the Zeru Finance protocol.