Fundraise Predictor

Why fundraise prediction is needed

- For making better Investment Decisions.
- Making Investment Strategies.
- Helps investors (VCs, angels) prioritize startups with higher chances of success. are a part of overall market analysis.
- Replaces intuition-based decisions with data-driven models.
- Identifies potential red flags early, reducing the chances of bad investments.

Challenges

- Available data can be inconsistent, outdated, or exaggerated (e.g., pitch exaggerations).
- Elements like founder charisma, networking ability, and investor sentiment are hard to model.
- Successful fundraising events are fewer compared to unsuccessful ones, leading to class imbalance in training data.
- Trends and investor behavior shift rapidly; models can become outdated if not regularly retrained.

Data Description

- The study collected datasets from http://crunchbase.com/.
- The data has 36 features.

Feature Extraction

- The funding velocity were calculated based on funding_total_usd and age_last_funding_year features.
- The funding velocity provides velocity of funding and has 26 percent spearman correlation with target feature.

Handling missing values

- The collected data underwent rigorous preprocessing to address any inconsistencies or missing values
- We systematically identified and removed null values to ensure the integrity and reliability of the dataset.

Feature scaling

- To address the issue of widely varying data ranges, normalization was performed to scale the data values to a consistent range.
- Specifically, min-max normalization was employed, which rescales the data to a range of 0 to 1.
- This normalization technique ensures that all variables contribute equally to the analysis, regardless of their original scale.

Feature selection

- In the feature selection process, we calculated the correlation of each feature with the ”Status” variable.
- Features exhibiting a correlation higher than 10 percent were selected for further analysis.
- This threshold was chosen to focus on variables that demonstrate a relatively strong relationship with the target variable, ”Status”.

Feature selection

- Correlations used were linear correlation, spearman, kendall and mutual info correlation.
- Feature selection aim to prioritize those features that are likely to have a meaningful impact on predicting or understanding cryptocurrency market trends.

Evaluation Criteria

- The evaluation of the proposed scheme is done using mean absolute precision,recall ,f1-score , support and Accuracy (Acc).

Machine Learning and Deep Learning Model Used

- Machine Learning and Deep Learning Model used are GaussianNB,KNeighborsClassifier,LGBMClassifier,Support Vector Classifier , LogisticRegression and Artificial Neural Network.
- LogisticRegression and Artificial Neural Network gave best results with 75 percent accuracy.

COMPARISON OF RMSE, MAPE AND ACCURACY VALUES FOR ETH.

Model	Train Accuracy(%)	Test Accuracy(%)
GaussianNB	57.45	53.51
KNeighborsClassifier	80.35	72.97
LGBMClassifier	**100**	74.59
Support Vector Classifier	77.91	74.59
LogisticRegression	76.82	75.67
Artificial Neural Network	80.71	**76.75**

Table. 1. Accuracy loss comparison for each model

Conclusion

- Our comparative analysis of prediction models showed that artificial neural network performed best.
- Predictions models gave good generalized results.

Thank You