Determining Parameters of Success on Initial Coin Offerings

The dataset *ICO_Data.csv* contains information for 2318 initial coin offerings (ICOs) from April 2015 to January 2020.

Part 1: Descriptive analytics

i) Examine the data to identify trends and relationships. Then, provide a summary including relevant visualizations that help the reader of report to understand the data.

ii) Report your data preparation (handling missing/incorrect entries, outliers, etc.)

Split the dataset into training and test set. Determine the partition percentages by yourself.

Part 2: Clustering to determine the success of ICOs

An ICO is considered to be successful if it hits its softcap (if any) *or* the amount of money raised is more than $0.5 million in the absence of a softcap, and unsuccessful otherwise.

i) Report which variables you include in your analysis, explain if you manipulate the data (i.e., categorical variables, new features, etc.)

ii) Use any clustering method (except Logistic regression) to determine ICO success.

iii) Report the model performance using confusion matrix in training and test set.

iv) Discuss your results.

Part 3 : Linear regression to estimate the return of ICOs *ICO_Return.csv.*

Estimate 1-week, 1-month, and 6-month returns using the ICOs given in the dataset *ICO_Return.csv.*

i) Report which variables you include in your analysis, explain if you manipulate the data (i.e., categorical variables, new features, etc.)

ii) Explain the impact of independent variables on the ICO return. Then, build a linear regression model using significant variables.

iii) Report the model performance.

iv) Discuss the impact of factors on ICO return.

Part 4: Identify additional factors that could have an impact on ICO success and/or return. Collect new ICOs after January 2020 and determine the ICO success and/or return. Compare the new model you build with the models in the previous parts. e

We draw on Fisch et al. (2021) in defining an ICO is an event where a startup sells tokens to a crowd using blockchain technology.

look at post-ICO performance as a measure of success, such as trading volume, liquidity, first-day underpricing and long-run returns, with the recent study

Software: You are allowed to use only Python for this project. Microsoft Excel can be used for transforming or formatting the data. If you use Excel or a macro in Excel for data preparation, you do not need to submit your Excel files, but do not forget to explain your steps in Part 1.

Submission: You can work in a group of up to 2 members. In Moodle, one of the group members needs to upload a zip folder including your report and codes in Python.

Dataset Description

Below are the descriptions of independent variables that could be used in the analysis:

- *ICO* – Name of the firm.
- *Token* – Name of token.
- *Softcap* – The venture may set a softcap, which is the minimum amount to be raised. If this amount is not hit, the project is not launched, and funds are usually returned to investors.
- *Hardcap* – The venture usually sets a hardcap, which is the maximum amount to be raised by the project.
- *Start* – Start date of ICO.
- *End* – End date of ICO.
- *Quarterstart* – Quarter when the project started.
- *Duration of offering* – The length of the campaign, and defined in days between the ICO start and end date.
- *Country* – Country of registration of the token.
- *Category* – Companies choose among 29 industry categories determined by icobench.com which shows the future sphere of activity. The number of categories the ICO falls into is a measure for diversification. A high number of industries is an indicator of the broader areas of future usage of a company's products.
- *ERC20* – Ethereum is the prominent platform for conducting ICOs. The Ethereum standard (ERC20) provides a set of rules for transfer tokens, and investors may recognize ICOs using Ethereum's infrastructure safer than other token exchange platform. *ERC20* is 1 if ICOs build on ERC20 Blockchain and 0 if the project uses its own or other technology.
- *Rating* – Experts' rating for business ventures from zero to five, zero being the lowest and five being the highest quality of ICO.
- *Number of experts* – The number of experts providing ratings.
- *Bonus* – 1 if a bonus is offered to early investors, 0 otherwise.
- *Policy* – *kyc* if the project implemented a Know Your Customer (KYC) policy only, *whitelist* if the project implemented a whitelist policy only, *kyc&whitelist* if both KYC and whitelist policies implemented, 0 otherwise. KYC policy requires the potential investors provide information to verify their identity. Whitelist policy requires registration for participation.

- *Presale* – 1 if a presale is carried out before the main crowdsale, 0 otherwise.
- *Number of team members* – A large team creates trust in the investors, believing that there will be enough people to carry out the project. As the number of contacts increases, there will be more contributions to the project's future endeavors.
- *Distributed in ICO* – The ratio of the tokens offered in sales to the total supply of tokens. This ratio is indicative of how successful the ICO will be in terms of the trust it creates in the market. The more tokens a venture retains ownership share, the more quality it signals.
- *US restriction* – 1 if ICOs do not accept US-based investors and 0 otherwise.
- *Restricted countries* – Countries that are not admitted to take part in ICO.
- *Accepted fiat* – 1 if investors could buy tokens with fiat currencies (Dollar, Euro, Yuan, and other government-issued currencies) and 0 if otherwise.
- *Major cryptocurrencies* - 1 if investors could buy tokens only with four major cryptocurrencies, i.e., Bitcoin, Ethereum, Litecoin, and Ripple.
- *Amount raised* – The amount raised in ICO.