ICO Success Prediction

Jaewon Lee | jaewon.lee@columbia.edu

PROBLEM

- There has been an enormous craze over Initial Coin Offerings (ICO)

- It is often regarded as a get-rich-quick scheme;
    - *Scams and ponzi-schemes are prevalent*
    - *Investors are blinded by greed and false hope, and fail to see the true value*
    - *Legitimate projects do not get the healthy support*
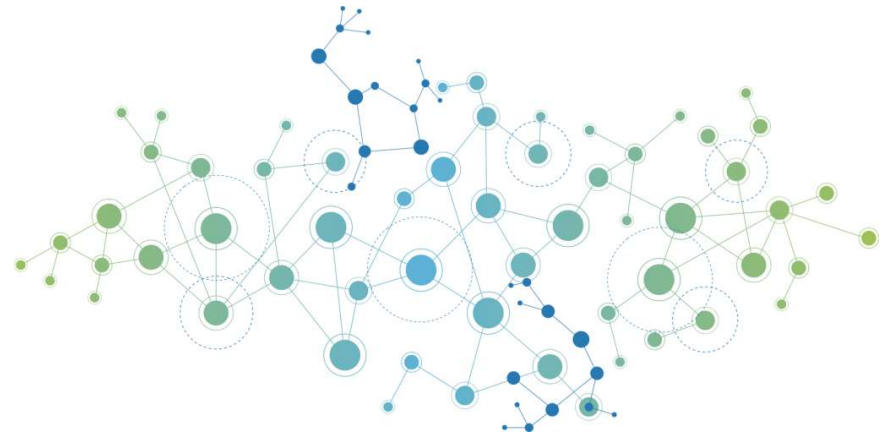   - *Many projects have failed, lost value, or even ceased to exist*

SOLUTION

Prediction model available to public that predicts project success based on data

Key Elements in ICO/Project Success

- Soft/Hard Cap
  - *Total Amount of Money to be Raised, Set by Team*
- Token Distribution Percentage
  - *How tokens are distributed among participants*
- Team Composition
  - *Development team, advisory board, etc*
- Length of Project
  - *How long the project has been in development*
- Whitepaper Quality
- Development Activity
  - *Amount of work actually done on code repository*
- Origin of Project
  - *Where the project was originated*
- Project Type
  - *Finance/Entertainment/Digital Assets/Legal/etc*
- Payment Form Received
  - *Common payments include BTC/ETH/NEO/USD*

How to Define "SUCCESS"

Although it is not the sole indicator that represents "success", the current market capitalization and Return of Investment of cryptocurrencies will be used to define the success of a project

DATA

Data will be collected from various sources that contain historical ICO data on projects

https://icostats.com/

https://www.coindesk.com/ico-tracker/

https://tokentops.com/ico/past/

https://www.ico-list.com/

https://icorating.com/

https://icotimeline.com/

https://coinmarketcap.com/

https://www.trackico.io/ico/

Example of Data

NEO – Formerly Known as Antshares

Token: ANS

Token ICO Price: 0.03 USD

Total Raised: 28,056,944 USD / 6,119.53 BTC

Token for Sale: 20,000,000 ANS

Payment Form Accepted: BTC

Available Supply: 10.0 million ANS
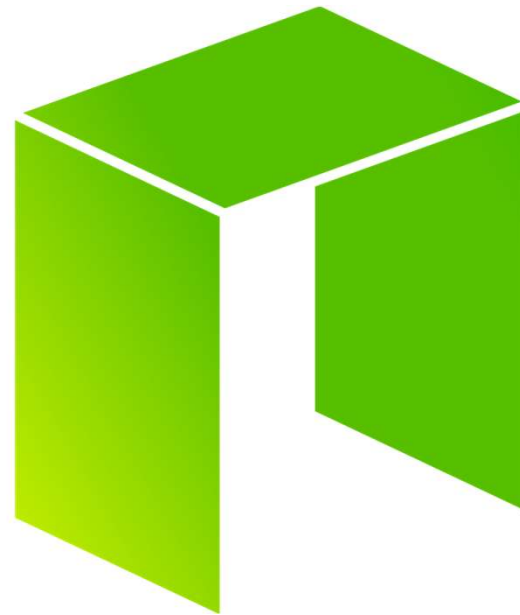
Total Supply: 16.4 million ANS

Current Price: 68.627460 USD

Country of Origin: China

ROI (in USD): 228,758.2%

ICO Date: July 9th 2016 To July 9th 2016

Original# Team Members: 4

APPROACH

1. Quantify qualitative data

    Sentiment Analysis applied on project reviews and comments; "convert" text into negative/neutral/positive

    Frequency of key words such as "scam", "ponzi"

2. Qualitative features used to build a regression model

3. Predict [Y = Return of Investment As of Today] from [X = features]

4. Use 80% of data as training set; 20% as test set

5. Extensive data available to validate the model

Sentiment Analysis

```
VADER is smart, handsome, and funny.
compound: 0.8316, neg: 0.0, neu: 0.254, pos: 0.746,
VADER is smart, handsome, and funny!
compound: 0.8439, neg: 0.0, neu: 0.248, pos: 0.752,
VADER is very smart, handsome, and funny.
compound: 0.8545, neg: 0.0, neu: 0.299, pos: 0.701,
VADER is VERY SMART, handsome, and FUNNY.
compound: 0.9227, neg: 0.0, neu: 0.246, pos: 0.754,
VADER is VERY SMART, handsome, and FUNNY!!!
compound: 0.9342, neg: 0.0, neu: 0.233, pos: 0.767,
VADER is VERY SMART, really handsome, and INCREDIBLY FUNNY!!!
compound: 0.9469, neg: 0.0, neu: 0.294, pos: 0.706,
The book was good.
compound: 0.4404, neg: 0.0, neu: 0.508, pos: 0.492,
The book was kind of good.
compound: 0.3832, neg: 0.0, neu: 0.657, pos: 0.343,
The plot was good, but the characters are uncompelling and the dialog is not great.
compound: -0.7042, neg: 0.327, neu: 0.579, pos: 0.094,
A really bad, horrible book.
compound: -0.8211, neg: 0.791, neu: 0.209, pos: 0.0,
At least it isn't a horrible book.
compound: 0.431, neg: 0.0, neu: 0.637, pos: 0.363,
:) and :D
compound: 0.7925, neg: 0.0, neu: 0.124, pos: 0.876,
```

Natural Language Toolkit will be used to perform Sentiment Analysis

Taken from http://www.nltk.org/howto/sentiment.html

Posts and reviews on projects are good indicators of investor sentiment – this example shows the frequency of the word "Scam"

Potential Difficulties

- Data collection is done semi-manually; time/effort will be put but it may not be enough

- There are a lot of other factors that may cause inconsistencies (i.e. general market sentiment, politics, international regulations)