Predicting IPO excess returns: A Sentiment Analysis Approach

The central issue we are addressing involves predicting IPO excess returns for the years 2019 to 2021 across different time intervals using various predictor variables.

We are particularly interested in exploring whether a broad sentiment score, derived from Twitter sentiment analysis concerning tech giants Amazon, Apple, Google, Microsoft, and Tesla, can help predict IPO excess returns.

Context

The rise of social media platforms, such as Twitter, has fundamentally changed financial market dynamics. Investor sentiment is increasingly shaped and influenced by discussions in these virtual spaces. Our project aims to quantify this effect alongside other variables that could predict IPO returns.

Criteria for Success

The project's success will be gauged by the prediction models' accuracy and reliability in forecasting IPO excess returns. Our objective is to create a model that accurately predicts excess returns, with a particular emphasis on understanding the impact of a broad Twitter sentiment indicator on these returns.

Scope of Solution Space

Our approach will involve constructing machine learning models that include traditional IPO factors, such as company age, IPO price, and venture capital involvement, along with novel predictors like aggregated sentiment scores derived from Twitter data.

Constraints

The project's primary constraints revolve around data availability and quality. Twitter data, in particular, can be noisy and challenging to process and may not fully represent broader market sentiment.

We are also relying heavily on third-party databases, the accuracy and completeness of which cannot be fully verified.

Stakeholders

The main stakeholders are the investors and financial analysts interested in IPO performance. The findings from this project may help them better understand the market dynamics and enhance their decision-making process.

Data Sources

Our primary data sources are:

- University of Florida's IPO Database: This provides us with historical IPO data for the predictor variables, specifically variables such as: whether the IPO was a Rollup, whether it involved multiple share classes, whether it was an internet company, and the year of founding.
- Kaggle - Company IPOs (2019 - 2021): We used this dataset to obtain the IPO offer price.
- Kaggle - Tweets about the Top NASDAQ Companies (2015 - 2020): This dataset provides us with the raw tweets for sentiment analysis.
- Yahoo Finance API: This API provides us with stock price data necessary for our analysis.

The work will be organized in a GitHub repo containing all the steps of the project, a slide deck, and a project report.