

BU3001 Project Synopsis

Predicting IPO Performance Using Quantitative & Qualitative Factors

13 October 2024

Team Members

Devansh Upadhyaya, Hemant Gupta, Ona Dubey, Ronit Kadakia, Yashvi Maheshwari

Team Leader

Yashvi Maheshwari

Roles & Responsibilities

- **Data Analyst:** Ona Dubey
- **Data Cleaner:** Ronit Kadakia
- **Model Developer:** Hemant Gupta
- **Presenter:** Everyone
- **Documentation & Reviewer:** Devansh Upadhyaya

Project Scope

This project aims to predict the performance of Initial Public Offerings (IPOs) before they list on the market by analyzing pre-market data from the Draft Red Herring Prospectus (DRHP), competitor and industry performance data, and social media sentiment. We will develop ML models to provide early predictions of IPO performance, enabling more informed investment decisions.

Objective

The objective is to create a predictive model to evaluate the potential performance of an IPO, with performance defined by metrics such as first-day returns, long-term stock performance, subscription rates, among others. By analyzing a company's financial health, market positioning within the industry, and public sentiment, the model will predict the likely performance of an IPO.

Data Exploration

- **DRHP Data:** We will collect financial metrics like revenue growth, net income, debt ratios, and risk factors from publicly available DRHP filings.
- **Competitor and Industry Data:** We will gather financial performance metrics, stock trends, and valuation ratios of industry competitors from financial platforms like Yahoo Finance to compare the IPO company's market position.
- **Sentiment Analysis:** We will use social media APIs (e.g Twitter API) and scraping tools to assess public sentiment about the IPO, applying sentiment analysis techniques to quantify overall market perception leading up to the IPO.

Potential Models

The project will explore a variety of machine learning models to predict IPO performance based on the available data. Some of the models under consideration include:

- **Logistic Regression**
- **Random Forest**
- **Gradient Boosting Models (XGBoost, LightGBM)**
- **Support Vector Machines (SVM)**
- **Neural Networks**
- **Sentiment Models (VADER, BERT)**

Each model will be evaluated based on accuracy, precision, recall, and other relevant performance metrics to determine the best approach for predicting IPO performance.

Timeline

- **Week 1:** Collect DRHP, competitor, industry, and sentiment data using APIs and financial platforms. Begin data cleaning.
- **Week 2:** Complete data cleaning and pre-processing, handling missing values and formatting data for analysis.
- **Week 3:** Perform exploratory data analysis (EDA), identifying key correlations and trends in the data. Engineer necessary features.
- **Week 4-5:** Train and validate the machine learning models (e.g., logistic regression, random forest) using historical IPO data. Tune the models based on evaluation metrics.
- **Week 6:** Evaluate the models' accuracy, interpret results, and identify the most important predictors of IPO performance.
- **Week 7:** Finalize the report, documentation, and prepare the presentation for the project's findings.