Finance and Analytics Club

SUMMER PROJECT
SCIENCE AND TECHNOLOGY COUNCIL
IIT KANPUR

Documentation on

Initial Public Offering Analysis

Through Predictive Modelling

Under the Mentorship of

Dakshita Mittal

Gaurav Raj

Sujal Harkut

Contents

ACKNOWLEDGEMENT

We are immensely grateful to Finance and Analytics club, IIT Kanpur and Science and Technology council, IIT Kanpur for providing us with resources and motivation for this summer project. We thank our mentors, the coordinators of Finance and analytics club Dakshita Mittal, Gaurav Raj and Sujal Harkut for their guidance, constant support and invaluable inputs without which we would not be able to do this project in a timely manner.

Introduction to IPO

1. What exactly is an IPO?

An initial public offering (IPO) is the first time a company issues shares to the public. This is when a private company decides to go 'public'. In other words, a company that was privately-owned until then becomes a publicly-traded company. Before the IPO, a company had very few shareholders. This includes the founders, angel inves tors and venture capitalists. But during an IPO, the company opens its shares for sale to the public. The transition of a company from private to public is achieved by a considerable infusion of cash by institutional investors and the general public. Firms go public for different reasons, the often-cited reason is the firm needs additional cash for growth and raises it by divesting a part of the company through the sale of common stock.
The IPO process involves a lot of Institutional players. This includes Investment banks and major Stock brokers. In our country, the latter dominates this list due to the absence of a substantial Investment firm. The IPO process begins with negotiations with investment banks, continues with road shows, and ends with either the SEC/CIBIL approving the company's stock to be listed on the specific stock market or disapproved.

2. Basic Terminologies

1.Risk Factors

Application structures for applying/offering for shares are accessible with all organisation individuals, assortment focuses, the specialists to the issue and the brokers to the issue. IPO (Initial Public Offering) is a sort of public offering where shares are offered to establishments, who reciprocally, offer their shares to the overall population. There can be various purposes behind beginning an IPO, notwithstanding, the essential explanation is to build the liquidity on the off

chance that you plan to apply through a new process presented by SEBI for example APPLICATIONS SUPPORTED BY BLOCKED AMOUNT (ASBA), you may get the ASBA application structures from the Self Certified Syndicate Banks. Any financial specialist who needs to put resources into an issue ought to have a PAN which is needed to be referenced in the application structure. It is to be particularly perceived that the copy of the PAN isn't needed to be connected alongside the application structure at the hour of making an application. However, when it comes to applying for an IPO, there are a variety of risks that are involved and must be duly considered. These risk factors can deter any meaningful investment and can hamper the chances of your fund growth. The various risks that are involved in applying for an IPO are as under:

1.*Risk related to Equity Shares and the Issue*

Fluctuations in price of Equity shares due to no assurance of liquid market for the shares and active trading.

2.*Legal Risks*

Changing laws, rules, regulations and legal uncertainties including adverse application of tax laws

3. *Regulatory Risks*

Non compliance with certain statutory provisions, applicable laws, rules or regulations

4.**Financial Risks**

Equity shares issued during previous year at a price below offer price

5.**Managerial Risks**

Influence of promoter group on outcome of matters submitted to the shareholders for approval. Such a concentration of ownership may also have the effect of delaying, preventing or deterring a change in control

2. *Jargons*

- **Underwriting-** IPO underwriters are financial specialists who work closely with the issuing body to determine the initial offering price of the securities, buy the securities from the issuer, and sell the securities to investors.

- **Lead manager-** Lead Managers are independent financial institutions appointed by the company going public to initiate the IPO processing.
- **Syndicate**- An underwriting syndicate is a band of investment banks and broker dealers who come together to sell new offerings of equity or debt securities of a firm to investors when the issue is too large for a single firm to handle.
- **Underwriting spread-** The difference between the price paid to the issuer, and the price received from investors.
- **Due Diligence -**Due diligence is the process of gathering important information about a business or assets
- **Primary Market**- IPO market
- **DRHP**- Draft Red Herring Prospectus.
- **Price band**- Price band is a price range between which the stock gets listed.
- **SEBI**- Securities and Exchange Board of India
- **Book Building**-The process of collecting all the price points along with the respective quantities is called Book Building.
- **Closure**- After the book building window is closed (generally open for a few days) then the price point at which the issue gets listed is decided.
- **Listing Day-** This is the day when the company actually gets listed on the stock exchange.
- **Green Shoe Option**- Part of the underwriting agreement which allows the issuer to authorise additional shares (typically 15 percent) to be distributed in the event of oversubscription.
- **Under and Over Subscription-** If more bids are received than the shares offered it is over subscription and when less bids are received than the shares offered it is under subscription.
- **Stabilisation-** The underwriter will typically go in and buy to support the price of the underwrite.

- **Common Stock-** A class of stock, usually having voting power, conferring residual ownership of assets of a corporation after all liabilities have been satisfied. Known as ordinary share capital in the UK.

- **Flotation-** Issuing new shares or stocks.

- **Issue-** Financial securities that are made available for sale.

- **Anchor investors**- Investors that have to hold their shares for min. of a month after the listing day before trading them.

INTRO TO PYTHON AND ITS APPLICATIONS

1. *What is Python?*

Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis.It is simple and easy to use, portable, extensible, built-in structure, and also, it is open-source. In addition, its language constructs and object-oriented approaching to aim to help programmers write clear, logical codes for small and large-scale projects.

2. How is Python used in this project?

These are the various Python libraries learned-

1.BeautifulSoup:

Beautiful Soup is a Python library that is used for web scraping purposes to pull the data out of HTML and XML files. It creates a parse tree from page source code that can be used to extract data in a hierarchical and more readable manner.

2.Openpyxl:

Openpyxl is a Python library that is used to read from an Excel file or write to an Excel file. A spreadsheet is represented as a workbook in openpyxl.

```
#!/usr/bin/env python
# coding: utf-8

# In[6]:


from bs4 import BeautifulSoup
import requests, openpyxl

excel = openpyxl.Workbook()
print(excel.sheetnames)
sheet= excel.active
sheet.title = 'IPOSubscriptionVsListingDayGain'
print(excel.sheetnames)
sheet.append(['Issuercompany','Issueprice','Issueprice(cr)','Total','Open
Price','Low Price','High price','Close price','%change'])

source=
requests.get('https://www.chittorgarh.com/report/ipo_report_listing_day_g
ain/98/')
code=BeautifulSoup(source.text,'html.parser')
companies = code.find('tbody').find_all('tr')

for company in companies:
     issuercompany = company.find_all('td')
     issues=[a.text for a in issuercompany]
     del(issues[4:8])
     del(issues[1:2])

     sheet.append(issues)
     print(issues)
     excel.save("IPO Analysis.csv")


# In[ ]:
```

FUNDAMENTALS OF RATIO ANALYSIS

It is not always possible to value a company and compare it with others only qualitatively and through sentiment analysis. Ratios often let investors compare companies as they indicate profitability, debt and asset management, as well as help avoid misleading signs.

1.Types of ratios-

- Valuation ratios
- Leverage ratios
- Liquidity Ratios

Valuation Ratios

1. P/E ratio:

●Price to earning ratio - price of stocks /earning per share = profit after tax/ total no. of shares

●Basically shows how much the shares are overvalued or undervalued compared to the company's earnings.

●A high P/E ratio shows the company is overvalued, while a lower ratio shows its undervalued, and its cheaper.

2. P/S ratio:

- Price to sales ratio - price of stocks / Sales per share = (Total Sales(or) Revenue/shares)
- Use for small companies that hardly have any earnings, but do have revenue. Same goes for newly established companies as well.

3. P/BV ratio:

- Price to sales ratio - price of stocks (market price per share)/ Book Value per share = (Book Value/Shares).
- Base metric, Book Value, tends to be a stabler metric than the other metrics we talked about.

Leverage Ratios

1. Debt Ratio:

- Shows how much of the company's assets are funded by equity and how much by debt.
- Is calculated as = Total debt / Total Assets
- A higher Debt ratio shows that most of the company's assets are met by debt, hence more liabilities, and the company may be at a risk of defaulting if the interest rates increase.

2. Debt to equity Ratio:

- Shows how much of the company's assets are funded by equity and how much by debt.
- Is calculated as = Total debt / Shareholder's equity

Liquidity Ratios

1. Current Ratio:

- Is calculated as the ratio of Current Assets and current Liabilities of a company
- Indicates the ability of a company to pay off its short term obligations, due within a year.
- A current ratio greater than 1 is normally desirable, but a very high current ratio indicates an inefficient management of assets.

2. Quick Ratio:

- Is calculated as (Current Asset - Inventory)/ Current Liability.
- Indicates the ability of a company to pay off its short term obligations, due within a year, however, takes into account only the most liquid assets, or those assets which can be converted into cash most quickly.
- An acid test to determine a company's financial health.

2.Return on equity (ROE)-

- Calculated as the ratio of Net Income to the Average SHareholders' equity.
- Since the Average Shareholders' equity can be equated to the Assets minus debts of a company, it can be thought of return on Net Assets, or how good is the company able to use the equities to generate profit.
- A good RoE shows the management is efficient in converting investors' money into profit.

ASSIGNMENT 2 BASED ON RATIO ANALYSIS

Problem Statement

"You will have to find financial statements from DRHPs of these companies listed below and perform ratio analysis on any 3 of your choice (1 from each category)
Category-1
Paras defence and space technologies ltd
Exxaro Tiles ltd
Category-2
Sapphire foods India ltd
FSN E-commerce ventures ltd
Category-3
Fino payments bank ltd
Chemplast sanmar ltd
The submission could be in the form of a doc/spreadsheet."

FLS AND THEIR ANALYSIS WITH VARIOUS MODELS

1. *What are FLSs?*

A Forward-looking statement is an indicative document that gives a brief idea about the plans and proceeding a management is eyeing in near and far future. FLSs have informational value in applications such as predicting stock prices. Management Discussion & Analysis (MD&A) sections in initial public offering (IPO) prospectuses contain FLSs that provide prospective information about the company's future growth and performance.

During the initial phase of the release of an IPO, A common practice in the U.S. IPO is a book building process, where underwriters collect information (i.e., market perception of the issuing company and stock) from informed investors. The book building process helps the underwriter set the preliminary price range of the offer price, and then underwriters and issuers start a "road show" to market the company to the prospective investors.The road show helps the underwriter gauge the potential market demand of the stock. If the demand is high, the underwriter sets a higher offer price defined as P_{ipo} . However, in order to mitigate potential risks, underwriters more often set an offer price lower than the market expectation hence, the *underpricing* phenomenon. The pre-IPO price adjustment ΔP can be defined as:

$$\Delta P = (P_{ipo} - P_{mid})/P_{mid}$$

If the P_{ipo} is lower than P_{mid} , the offer price is said to be below market perception and the stock is underpriced. Finance literature also argues that the underpricing phenomenon usually affects short-run performance of IPOs and hence studies have been using the first-day closing price P_{1day} as the metric to capture the post-IPO initial returns. The post-IPO initial returns IR can be defined as:

$$IR = (P_{1day-}P_{ipo})/P_{ipo}$$

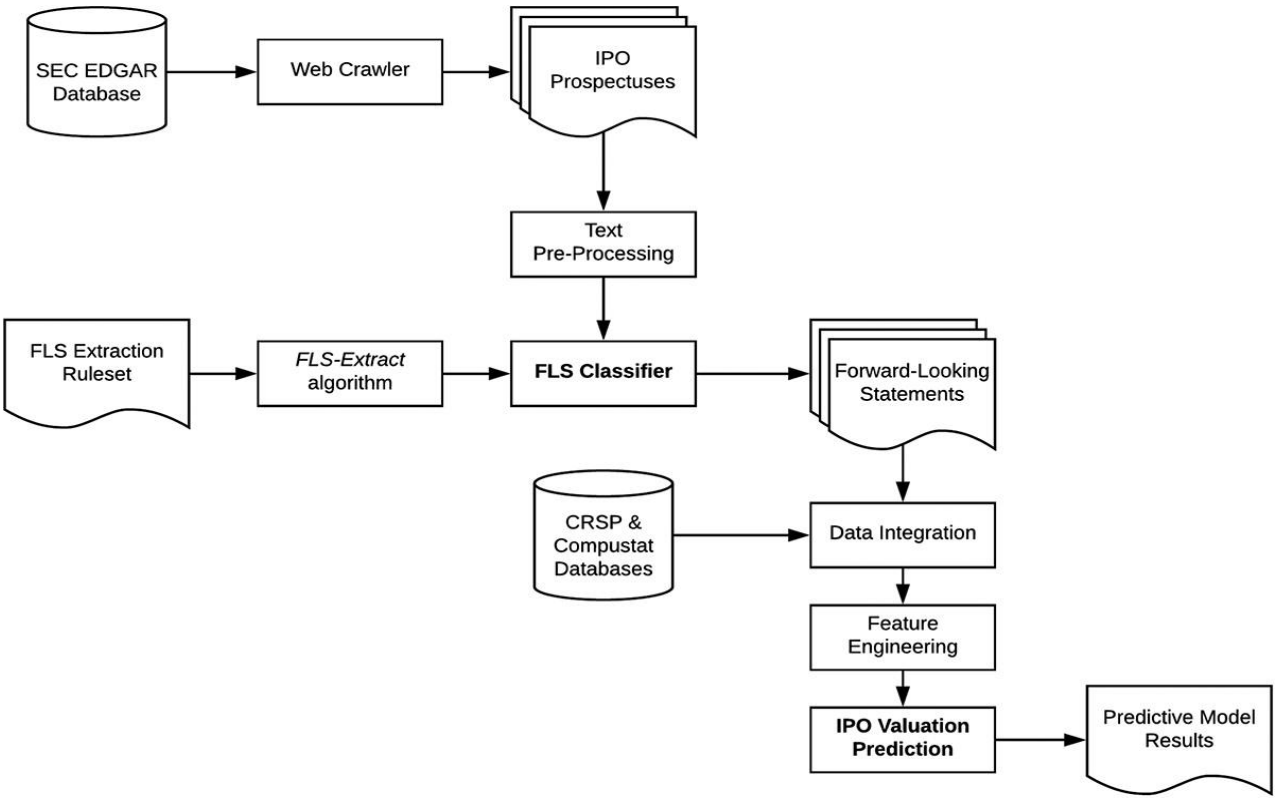
2. Model to predict future with FLSs

Text Mining

Text analytics, also known as text mining, is a cluster of techniques utilising machine learning for the analysis of unstructured textual data, as in a variety of documents including the IPO prospectuses, to extract valuable information from them, and to construct structured variables/features for further analysis. Based on the underlying machine learning techniques, text analytics can be categorised as supervised or unsupervised. The basic idea behind this technique is to search specific keywords that can be used to directly or indirectly relate to the future of the firm. The list for a text miner:

- Can contain "will" or "future", both of which indicate future time;
- Can contain combinations of adjectives such as "next", "subsequent", "following", "upcoming", "incoming", "coming", "succeeding", and "carryforward" and time indicators such as "month", "quarter", "year", "fiscal", "taxable", and "period".
- Have a verb list that includes: "aim", "anticipate", "assume", "commit", "estimate", "expect", "forecast", "foresee", "hope", "intend", "plan", "project", "seek", "predict", and "target".

A popular but outdated model based on this approach was the **FLS classifier** that used round-wise searches and faced false errors elaborated ut infra.
The first round searches for key phrases such as "will" or "next year" in a sentence; the second round searches for conjugations of certain verbs like "expect" or "foresee" in a sentence; the third round searches for the mentions of a future year in a sentence. Even though these approaches have led to successful subsequent analyses, it is acknowledged that both type I (false positive) and type II (false negative) errors may exist in the FLS identification stage, and that type II errors may be more severe.



Naive Bayes Classifier

There are two general approaches for conducting content analysis: a rule-based approach (i.e., dictionary approach) and a statistical approach. The first approach uses a "mapping" algorithm in which a computer program reads the text and classifies words (or phrases) into different categories based on predefined rules (i.e., dictionary) as described in the previous section.
The second approach, which was pioneered by computer scientists and mathematicians, relies on statistical techniques to infer the content of text and classify documents based on statistical inference. For instance, the algorithm may calculate the statistical correlation between the frequency of some keywords and the document type to draw inference.

The Naïve Bayes algorithm is a supervised learning algorithm based on text mining and Bayes theorem is used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. The Naïve Bayes Classifier is one of the simplest and most effective Classification algorithms which helps in building fast machine learning models that can make quick predictions.

A Model presented by Feng Li (2010) based on this approach used the Bayesian learning algorithm to categorise the tone and content of about 13 million FLS from more than 140,000 10-Q and 10-K filings between 1994 and 2007. The results indicate that firms with better current performance, lower accruals, smaller size, lower market-to-book (MTB) ratio, less return volatility, lower MD&A Fog index, and longer history tend to have more positive FLS in their MD&As.. Under this method, a given sentence is first reduced to a list of words (words) with each word weighted in some fashion (e.g., by frequency in the sentence). The goal is to classify the sentence into a specific category (cat) from a set of all possible categories (cats). In this paper, there are four possible tone categories: positive, negative, neutral, and uncertain. The Naïve Bayesian algorithm chooses the best category by solving the following problem.

$$cat^* = \underset{cat \in cats}{argmax} \frac{P(words \mid cat)P(cat)}{P(words)}.$$

Since P_{words} does not change over the range of categories, it can be eliminated. The problem thus become

$$cat^* = \underset{cat \in cats}{argmax} P(words \mid cat)P(cat).$$

Finally, if $w_1, w_2, \ldots, w_n$ are the words in the document and their probability of appearing in a sentence is assumed to be independent expression is equivalent to:

$$cat^* = \underset{cat \in cats}{argmax} P(w_1 \mid cat) * P(w_2 \mid cat) * \cdots * P(w_n \mid cat) * P(cat),$$

The last step is the only non rigorous one in the derivation and thus the "naïve" part of the Naïve Bayesian technique. It assumes that the probability of each word appearing in a document is unaffected by the presence or absence of each other word in the document. The independence assumption simplifies the computation and avoids the "curse of dimensionality" problem . Independence is assumed even though it is not true. For example, in the financial statement setting, the words ``adverse effect" are more likely to appear together with the word "material." However, empirical results from other fields suggest that making this assumption even if it is not true may have little effect on the result.

TONE ANALYSIS

After collecting all the Forward looking statements, the future of the company is predicted with the help of the tone of these statements. To judge the nature of the statement, it is noted that statements which reflect the growth of the company tend to be positive FLS. On a similar basis, FLS are classified into ones with negative, neutral and undefined tones. Following a set of rules listed below, helps one to decide the nature of statements with more clarity.

The Rules

1. Current firm performance

   While there is substantial theoretical and empirical literature on the amount of disclosure, little work exists on the tone. Many arguments made in the disclosure level literature can apply to the tone as well. For example, litigation concern may encourage firms with good current performance to be more cautious in discussing future events in their MD&A. Momentum in firm performance also suggests that the MD&A FLS may be more positive for firms with good current performance. However, earnings are mean-reverting, which implies a more negative tone in FLS for firms with better current performance. Therefore, ex ante the relation between tone and current performance is unclear.

2. Accruals

   It is well documented that accruals are negatively associated with future firm performance and investors underreact to this information. If this is true, it implies that managers understand the implications of the accruals for future performance. Alternatively, accruals may simply proxy for a firm's economic conditions (e.g., distress) and managers are likely to understand (at least partially) the implications of accruals for future earnings. In either case, a negative relation between accruals and the MD&A tone is expected. How- ever, if managers are overconfident or fixated on current earnings, they may not understand the implications of accruals for future performance and hence no relation between accruals and MD&A tone is expected. Also, if managers understand the implications of accruals but have in to mislead investors, a positive relation between accruals and MD&A tone might result.

3. Firm Size

Size captures many aspects of a firm's operational and business environment. The accounting literature has used firm size as a proxy for a firm's political cost. Larger firms may have more cautious FLS because of the higher political and legal cost due to their vis

4. Market-to-book ratio

High MTB firms are different from low MTB firms in many aspects, including the investment opportunity set and growth poten- tial. To the extent that growth firms face more uncertain future economic conditions, a negative relation between the MTB ratio and MD&A tone is expected.

5. Volatility of operations

Firms with more volatile business environments may be more cautious in discussing future events because of information uncertainty with regard to future performance. These firms are also more likely to have severe information asymmetry between managers and in- vestors. Finally, performance variability may be related to MD&A tone because of its effect on a firm's vulnerability to legal action. These factors all posit a negative relation between volatility and MD&A tone.

6. Complexity of operations

More complex operations are more likely to lead to complex disclosures. I use the numbers of segments and the number of nonmissing financial items in Compustat to capture the complexity of operations and examine how they are associated with MD&A tone.

7. Firm age

Young firms face more uncertainties. Managers of young companies are likely to be more cautious when discussing future outlook. A negative relation between MTB ratio and MD&A tone is expected.

8. Firm Event

I include two firm-event dummies as potential determinants of the MD&A tone - seasoned equity offering and merger and acquisition activities. Firms that make seasoned equity offerings or acquisitions might have incentives to discuss future outlook more positively. Also, firms with more positive outlook are more likely to issue new equity or make acquisitions. Therefore, a positive association between a firm's equity offering and merger transactions and its MD&A tone is expected.

9. Special item

Firms with a significant amount of special items are more likely to experience unusual performance. I expect firms with more negative special items to have more negative MD&A tone.

10. Incorporation state

Due to the different corporate laws and investor protections, companies are more likely to receive takeover bids and be acquired. Such firms are also valued higher than similar firms incorporated elsewhere. I include a dummy variable for Delaware- incorporated firms to examine whether such firms tend to have different MD&A tone.

11. Disclosure transparency

The firms' annual report readability captures management's obfuscation behaviour. When performance is poor, annual reports tend to be less readable and have a high Fog index. I include the Fog index of the MD&A section as a determinant of the MD&A tone and expect Fog to be negatively correlated with the MD&A tone.

CHARACTERISTIC PARAMETERS

We now discuss the data and research methodology used in this study. The analytical pipeline is depicted earlier. In brief, the IPO prospectuses collected from the data sources are pre-processed in the text pre-processing module, the core of which is the word embedding based on a continuous-space language model. Following this, the FLS classifier is trained. With the FLSs classified, feature engineering is done where new predictive features such as topics, sentiments, readability, and semantic similarity are constructed for the application at hand, i.e., IPO valuation prediction. We adopt the following key IPO characteristics based on financial/ accounting data:

1. Pre-IPO price revisions (Y1)

   If the IPO offering price is lower than the midpoint of the price range disclosed in the initial prospectus (Form S-1), then 1; otherwise, 0.

2. First-day returns (Y2)

   If the IPO offering price is lower than the closing price of the first trading day, then 1; otherwise, 0.

3. Trailing EPS (EPS)

   The trailing Earnings Per Share (EPS) in the 12 months prior to IPO;

4. Prior NSE15-day Return (PRIOR_RTRN)

   The buy-and-hold returns of the Nasdaq value weighted index over the 15 days before the IPO date.

5. Share Overhang (SHR_OVERHNG)

   Number of shares retained divided by the number of shares offered in the IPO

6. Trailing Sales in Million $ (SALES)

   Trailing sales in millions of dollars, disclosed in the 424B filings.

7. Up Revision (UP_REV)

Upward revision (in percentage) from the mid-point of the price range disclosed in the initial prospectus (Form S-1) to the offering price, if the latter is greater than the former, otherwise zero.

ASSIGNMENT-3 BASED ON CHARACTERISTICS PARAMETERS

Problem Statement

"You have done the fundamental analysis for 3 companies as a part of assignment 2; taking that into account and consequently, identifying and classifying FLSs according to their tone you will be able to get the characteristics parameter talked about in class. After gathering all this information, you need to grade these 3 IPOs according to your sentiment regarding buying or refraining from buying these. All this analysis has to be done for all the three companies from assignment 2 and one of these two companies listed below. 1) Adani Wilmar Ltd 2) Aether Industries Ltd For one of the above 2 companies, you will have to prepare a presentation of a min of 3 slides detailing the fundamental analysis, company operations, FLSs and their segregation into different tones, characteristic ratios and valuation metrics and your inference from these analyses as well as would you as an investor invest in it or not on the last slide with reasoning. (You will have to extract all the information for your analysis from the DRHPs only)"

is