STUDY OF TEXT-BASED DATA AS DRIVERS OF THE STOCK PRICE FOR TOP PHARMACEUTICAL COMPANIES

Introduction

Financial market analysis focuses primarily on understanding the value of a stock to project its short & long-term values. While technical indicators focus on capturing changepoints in the trend of a stock, fundamental indicators focus on the accounting of the company (P&L, earnings per share, revenue, gross profit, among other indicators) to assess the potential growth of the company that will also positively impact its stock price.

Recent enhancements in Natural Language Processing (NLP) for analyzing & interpreting text-data have led to a numerous among of possibilities for understanding the financial market behavior. Below we will explore two use-cases: earnings calls & press releases.

Earnings calls are periodic (often quarterly) statements produced by high-level executives of a company (tends to involve the CEO & CFO) who reflect on the evolution of the company during the last period with the aim to influence the investors' expectations of the past and future performance. Earnings calls are oral communications that are then transcribed for public review.

Press releases are news covered by several media that reflect an important event for the company (can be positive or negative) that will likely impact on the organization or financial statement.

Scope

The scope of this project includes the Top 10 Pharmaceutical Industries by revenue. Those include: *AABV (AbbVie), AZN (AstraZeneca), BMY (Bristol-Myers Squibb Co), JNJ (, Johnson & Johnson), LLY (Eli Lilly And Co) , MRK (Merck & Co Inc), NVO (Novo Nordisk A/S), NVS (Novartis AG), PFE (Pfizer Inc.), ROG (Rogers Corp)* - sorted alphabetically -.

Objectives

· Perform Exploratory Data Analysis (EDA) to evaluate the price evolution of each stock at different temporalities: daily, weekly & yearly level. Also check its correlation with the volume or other price-derivative indicators (eg. moving averages, relative strength index, etc.), that are commonly used as technical indicators.

· Assess the volatility of a pharmaceutical stock along the years & its impact versus both micro & macroeconomic events.

· Use NLP for topic modeling and word extraction from the earnings calls & press-releases with the aim to characterize the sentiment of the text and find a predictive relationship with the stock price fluctuation. This should be the main focus of the project and given high priority.

Data Sources

Data provided is public & available through different APIs obtained from Financial Modeling Prep (https://site.financialmodelingprep.com/) & StockNews API (https://stocknewsapi.com/). Therefore, no data anonymization has been performed in any of the provided datasets.

Datasets Provided are:
- Press Release: Obtained from Financial Modeling Prep API
- Earnings Call Transcripts: Obtained from Financial Modeling Prep API
- General Market Sentiment: Obtained from StockNews API
- Daily Stock Price Data: Obtained from Financial Modeling Prep API

Link to Data & Metadata:
- https://drive.google.com/drive/folders/1ifIzZEeSDnUoo1Vyzi3I4lXQnmzNInUp?usp=share_link

The students are not restricted to the above datasets and can leverage any external information they may find suitable for the achievement of the objectives.