Stock Price Predictive Analysis: Capturing Idiosyncratic Risks Based on News headlines and Modelings

Amber Lee, Cindy Zhang, Wafer Hsu, Hejia Zhang & Oretha Domfeh (TA)

Highlights

- Sentiment Analysis of Finance News headlines for possibly investment decision making
- Predictive model of stock prices for investors

Background

Over the years, investment firms have increasingly embraced technology and data science to forecast financial market trends. Even with the expansion of accessing abundant data availability, the event-driven stock price changes are still an unpredictable part to all investors.

Our project aims to enhance stock price prediction by employing a comprehensive quantitative and qualitative approach create a predictive model which allows us to harness the power of ubiquitous data to empower investors of all scales in making more accurate and effective investment decisions.

Data

The time-series ranged from 01-2018 to 06-2023 within top 5 U.S. companies: Amazon, Apple, Google, Microsoft, and Nvidia.

Stock Prices: The adjusted closing stock prices and volume of stocks sold for each day. The data is captured from Quandl via the NASDAQ.

S&P 500 Tickers: The Standard and Poor's 500 is a stock market index tracking the stock performance of the largest companies listed on stock exchanges in the U.S.. This dataset contains company name, symbol, and weight of the companies.

Fama-French 5 Factor Model:
This data captures five factors where can be used to assess the exposure of each stock to the Fama-French risk premium.

News Headlines: Web scraping Google News headlines through GNews library and will be used to conduct sentiment analysis through the NLTK library.

Models

We leverage diverse data sources and analytical techniques aiming to provide comprehensive insights into stock price prediction.

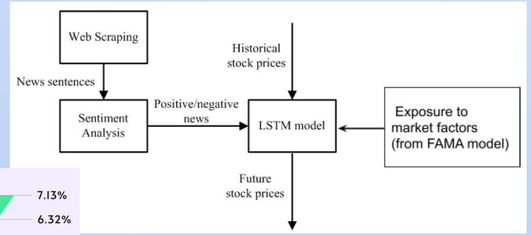
1. Time-Series Data Prediction: The Long Short-Term Memory (LSTM) Network is applied to forecast the future stock price based on the previous prices. This methodology will also be used to predict the following days' time-series trends.
2. The Fama-French 5 Factor model: this approach is used to estimate excess return of an investment asset. The model adjusts for this outperforming tendency, which is thought to make it a better tool for evaluating performance. The dependent variable $R_{ft} - R_{Ft}$ is the difference between the return in a period t and risk free rate which alter the factors described above. The model is given by:

$$R_{ft} - R_{Ft} = \beta_0 + \beta_1 MKT + \beta_2 SMB + \beta_3 HML + \beta_4 RMW + \beta_5 CMA + \varepsilon$$

(*Market*) (*Size*) (*Value*) (*Profitability*) (*Investment*)

3. Sentiment Analysis: Conducted Natural Language Processing (NLP) on Financial News headlines and determine whether the sentiment is positive, negative or neutral (scale 0-1).

Our approach that integrates Financial News sentiment analysis, historical stock returns, and idiosyncratic exposure to Fama-French risk premiums. We strive to discover a comprehensive assessment of the total risks associated with each equity. This holistic approach ensures that both systematic and idiosyncratic factors are adequately incorporated, allowing us to offer a well-rounded evaluation of equity risks and facilitate more informed investment decisions. This could presumably empower investors of all scales in making more accurate and effective investment decisions.



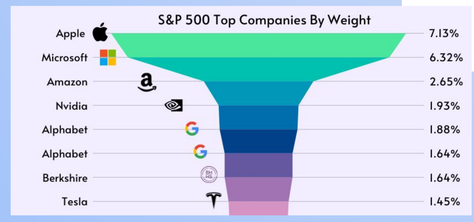


Fig 1. S&P Companies (2023); *created by Abheey from Finasko*

(convert models into charts/ EDA results)
(results/images/flowchart)

Results

Our results will be included here.

Including some visualizations

Future Work

Based on our conclusions, do our approaches work? Can we rely on our assumptions?
If do/don't, what are our next steps?