**Team 25**

**Team Leader:** Anastasia Gkousioudi

**Members:** Wanying Ye, Cindy Chen, Tanya Gonzalez, Anastasia Gkousioudi, and Swadha Rai

**Overview of an industry, business or problem**

Stock markets are vital components of a free-market economy because they enable indiscriminate access to trading and exchange of capital for all kinds of investors. Prediction of the movement of a stock market is essential to the whole financial business since it supports investors to make smart trading decisions and develop strategies for optimal timing of stock trading. Several factors influence a stock market trend, one of which is news headlines.

**Why this is interesting and important**

Stock market prediction is one of the most relevant but highly challenging tasks in financial research. The abundance of financial data has motivated the use of machine learning and computational intelligence methods to develop models for stock market prediction.

Most researchers utilize either technical analysis or fundamental analysis to forecast the market. Specifically, technical analysis focuses on examining previous price trends to forecast stock prices, such as using time series modeling techniques. On the other hand, fundamental analysis relies on the exploration of unstructured textual data, such as financial news and company reports, to study their impact on stock prices fluctuations.

**Define the specific problem that should be solved**

In this project, we will compare these two approaches in terms of prediction accuracy. Specifically, we want to answer questions such as, is daily world news a good predictor for the stock market trend? As an extension, we plan to incorporate both approaches to develop a more comprehensive modeling framework to predict the stock market performance that investors can utilize for their stock trading decisions.

For this project, we are interested in exploring various machine learning (ML) models and Natural Language Processing (NLP) techniques to predict the stock market movement (whether the prices increase or decrease compared to the previous day) based on the daily news headlines. If time permits, we will also utilize the time series feature of the data and develop a modeling framework to predict the stock market performance based on both daily news headlines and previous market performance data.

**Data:**

The original data can be downloaded from [Kaggle](https://www.kaggle.com/datasets/aaron7sun/stocknews?resource=download) and there are two subsets of the data.

|  | Description |
| --- | --- |
| News data  “RedditNews.csv” | The historical news headlines data is crawled from Reddit WorldNews Channel. They are ranked by reddit users’ votes, and only the top 25 headlines are considered for a single date. |
| Stock data  “DJIA\_table.csv” | It contains the Dow Jones Industrial Average (DJIA) of open, close, high, low, and adj close for each day.  (Range: 2008-08-08 to 2016-07-01).  "1" when DJIA Adj Close value rose or stayed the same; "0" when DJIA Adj Close value decreased. |

**Reference:**

1. Alzazah et al., “Recent Advances in Stock Market Prediction Using Text Mining: A Survey”, IntechOpen, 2020
2. Jaitly et al., “Forecasting Stock Market Trends Using News Headline Analysis”, IRJET, 2021
3. Kumbure et al., “Machine learning techniques and data for stock market forecasting: A literature review”, ESWA, 2022
4. [What is the Stock Market?](https://www.investopedia.com/terms/s/stockmarket.asp#:~:text=Stock%20markets%20are%20vital%20components,price%20discovery%20and%20efficient%20dealing)