Project Name:

Market Moods

Team Name : StatSquad
Team Members:
> Rushikesh Joshi
> Lakshay Sharma

Problem

Big Idea: Assessing the Impact of Press Coverage on Stock Prices: Is All Press Really Good Press?

- **Problem to Solve:** The problem we aim to solve is on how different types of press coverage affect the stock prices of companies whether it is positive, negative, or neutral news.
- **Question to Answer:** How does press coverage influence stock prices? Our goal is to determine:
    - If stock prices increase after positive news and decrease after negative news.
    - How long the stock price is directly affected in response to the news article.
    - Whether the magnitude of news impact varies by different industries.
- **Importance of This Problem:**
    - Investors and financial analysts can use this information to refine investment strategies.
    - Can lead to corporate PR teams to manage communications effectively.
    - Give academics and economists an outlet to better understand market behavior dynamics.
- **Reason for Choosing This Problem:** The importance of digital news and its real-time access influences financial markets significantly. With the increase in algorithmic trading where decisions are partly based on news parsing, it's critical to quantify the impact of media to determine the direct affect on its stock price.
- **Specific Hypotheses:**
    - The market recently is very volatile in regards to both good news and bad news.
    - Negative news or hikes in rates give investors a good reason to perform big moves and trades.
    - Recently, technology stock prices have been greatly impacted by AI, one of the examples is $NVDA.

Data

Data Collection:

- **Stock Price Data Collection:**
    - We will be using historical stock price data from a Nasdaq financial API.
    - We have selected three potential APIs for accessing historical stock price data: Polygon.io, Yahoo Finance, and Alpha Vantage.
- **News Data Collection:**
    - Along with financial data, we will also require data regarding articles written about companies.
    - For this reason, we will also be using NewsAPI.

With extensive documentation available for each, data collection should be straightforward.

Types and size of data:

- **Stock Data:**
    - Since there are 3432 companies listed in the Nasdaq stock market index, our financial data will contain 3432 rows of data.
    - The types of data we will require is the company name, stock symbol, stock price and date.
- **News Data:**
    - Our news data will be very large as it will likely contain multiples articles about various different various companies.
    - From this data, we will require the fields: company name, negative or positive article, date of article

As we will be checking for a correlation between the News data and the Stock data, it will be essential to retrieve the publication date of the article, along with the stock price before and after it's release, to verify the impact of the article on the company's stock price. Furthermore, flagging the articles as either positive, negative or neutral will also be crucial as it will indicate whether the article influenced the stock price in the expected direction.

Solution

Scope of project/next steps:
The project will involve gathering stock market data and news articles, using machine learning to predict stock movements, and building a simple user interface. Next steps are cleaning the data, starting the model, and designing the user interface.

End result and Techniques to analyze data:
The final tool will give personalized investment advice with helpful visuals, like stock predictions and news impact charts. We'll also be using machine learning to analyze how news affects stock prices.

Interactive or static:
The tool will be interactive, letting users input preferences and receive personalized, real-time advice.

What do you hope to have achieved for the Progress Report?:
By the progress report, we want to have cleaned the data, explored it, and built a basic model to start predicting stock trends based on news.

Tentative Timeline

Finding Data Sources: By Oct 10th

Data Cleaning: By Oct 20th

Statistical Analysis: By Oct 30th

Data Visualization: By Nov 10th

Code Optimization: By Nov 20th

Professor Check-In Date : Oct 11th, 2024