Project Plan

Explorations in Data Science

Project Objective: In today's world where information is readily available and at your fingertips, we think and believe that the paradigm of how stocks rise and fall based on information has changed in a way where we can use data science to use available data to decide if the outcome will cause a positive or negative impact.

Project Approach: The Project is twitter sentiment analysis. Sentiment Analysis is the automated process of analyzing text data and sorting it into sentiments positive, negative, or neutral. Performing Sentiment Analysis on data from Twitter using machine learning can help companies understand how people are talking about their brand. Monitoring Twitter allows companies to understand their audience, keep on top of what's being said about their brand and their competitors, and discover new trends in the industry. Are users talking positively or negatively about a product? Well, that's exactly what sentiment analysis determines.

Our first step towards this project is to fetch the tweets from Kaggle which are pre-processed twitter data and then perform sentiment analysis on the same. Similarly, get the stock prices for the respective company for the dates we fetched the tweets. We considered the percentage change in stock values.

Team Structure

- Shashank Shekhar
- Likhitha vanga

We started this project by initially dividing our work into 2 halves, so that we can contribute equally to the project. Below are the individual duties:

Shashank:

· Prepared data for processing by deleting unwanted columns and rows.

· Selecting the algorithms and analyzing why the chosen is better than other.

· Implementing an algorithm.

· Documentation.

Likhitha:

· Fetched data from Kaggle.

· Understand the algorithms we are using for this project.

· Implementing an algorithm.

· Documentation.

Project Milestones

- Started the project on 06/27 and tentative end date is 08/10.
- The key deliverables will be being able to predict next year stock values.
- Training the model by 07/21.
- Testing on scalability by 08/01.
- Documentation by 08/10.