Stock Market Analysis

Zaki Kidane (4502) & Kristian Montoya (4502)

Introduction

Understanding and predicting stock market movements is useful for individual stock market traders as well as investment firms and banks. So far, the task of successfully predicting stock prices has been done by people. We ask the question, “Can ML algorithms effectively outperform the average trader?” and seek to answer that question. By introducing the performance history of different stocks, we can visualize the data and look for correlations between stock attributes/features to help determine the current/future performance of any given stock. We intend to store the collected data in a data warehouse for efficient retrieval to be used by a machine learning model (possibly a linear regression).

Related Work

- https://www-sciencedirect-com.colorado.idm.oclc.org/science/article/pii/S0957417419304142
- https://www-sciencedirect-com.colorado.idm.oclc.org/science/article/pii/S0957417416305115
- https://www-sciencedirect-com.colorado.idm.oclc.org/science/article/pii/S0957417415005126

Proposed Work

- Taking uncleaned datasets from Kaggle to extract, prepare, and store the data in a dataframe.
  - https://www.kaggle.com/dgawlik/nyse
- Visualize and analyze trends and correlations between features to find if any such relationships exist.
- Reduce unnecessary features and normalizing the data in preparation for a linear regression model.
- Use the cleaned data to predict whether or not a stock should be bought or not.

Evaluation

Based off of the features we drop/reduce, we want to look at the performance of the of the model we use to see what features correspond to the overall price of a stock (what feature drives a stock the most, or, what feature(s) are most important to a stock's price). With that in mind, the evaluation of the project can be determined by rate of successful predictions of our selected model.

Milestones

1. Download and prepare data for storage
2. Create visual reports and interpretations of data
    a. Report correlations and trends found
3. Create a revised dataset from the dataframe that will be used for a Linear Regression model.
4. Test the model and report accuracy based off of features selected for use.