

Taking Stock of the Situation

If chosen we would both like to be part of this project, 10/10.

Description:

We want to use historical data and statistics for a specific stock, Pfizer, to predict the Adjusted Closing Price of the next trading day. With the dataset that we plan to generate, we want to compare results between different time series models (RNNs, LSTMs, Transformers) and simpler models that aren't optimized for time series, such as the Perceptron. In the case of models not optimized for time series, we plan on simply using the feature inputs of the previous trading day to predict the next Adjusted Closing Price.

Daily Opening Price	Daily Closing Price	Daily High	Daily Low	Daily Adjusted Close	Daily Trading Volume	Earnings per Share	Price to Earnings Ratio	Debt Ratio	Return on Equity	Shares Short Ratio	Market Cap
32.72	32.98	32.36	32.96	28.9	25,698,648	3.76	8.62	66%	23.07%	2.1	182.929B

Features:

We will start with the 13 features displayed, some will be given in the Yahoo Finance dataset we download and others are going to be based on calculation. The data can be broken into 3 main groups of features:

1. Price data: This data, including highs, lows, and closings, is going to be helpful for tracking changes and giving a general gauge of investor sentiment.
2. Financial Ratios: Ratios like the Return On Equity, Debt Ratio, and Earnings Per Share will give an indication of the company's growth and internal financials.
3. Broader Market Data: Individual companies are greatly influenced by overall market trends. The market being in recession for example can greatly influence the performance, how other sectors are performing, etc.

Data Gathering and Labeling:

Browsing online proves that it is hard to find any substantial datasets that are publicly available. Many different sites will provide free access to APIs with registration of an account, although these APIs have rate limits and don't extend more than a couple of decades into the past. We found that Yahoo Finance will allow us to download the Opening Price, High Price, Low Price, Close Price, Adjusted Close Price, and Volume for Pfizer daily since June 2nd, 1972. We've thought through and selected a few specific features to start our model with, as listed above. To get access to some of these features historically, however, we plan on signing up for a Yahoo Finance Plus Essential free 14-day trial to fully flesh out our dataset. Once we have all the data downloaded, we foresee some additional issues in dataset preparation, such as only keeping data entries from trading days (and not the weekend), making sure we only use features that have values for all trading days, and finding the right methods of normalization to keep our inputs in the same space.