Predicting Disney Stock Price Using Time Series Data

David Berberena

Bellevue University

DSC 680 Applied Data Science

Amirfarrokh Iranitalab

September 29, 2024

<center>**Predicting Disney Stock Price Using Time Series Data**</center>

Topic

   To better align my study of data science with my company of choice to work for and

showcase the skillset I have learned using data involving the company, I will be looking at time

series data concerning Disney stock to predict future stock price. With Disney being a market

leader in the entertainment industry and many shareholders invested in its continued growth,

the stock price is one of the quickest indicators of overall business health, development, and

stability. Analyzing the change over time also may highlight market volatility, which can be

factored into the stock price in the future.

Business Problem

   The Walt Disney Company has grown exponentially from its humble start as an

animation studio to what may very well be the most common household name in the United

States regarding the entertainment industry. The Company comprises various parts that create

its economic fingerprint: Parks and Resorts, Media Networks, Studio Entertainment, and

Consumer Products. Within each sector of the Company, major developments are occurring

that could positively or negatively affect the brand, and shareholders are most concerned with

the longevity of their return on investment within the Company as indicated by the stock price,

a direct designation of a business's overall market value. Stock price prediction will allow

investors to make more informed decisions on how they will allocate their money to the Walt

Disney Company, or if they should even continue to allocate their funds within the Company.

Buying, selling, or holding stocks is mainly determined by stock price and volatility, hence why

the prediction of Disney stock would help manage the risk that shareholders take when funneling their assets towards the Company.

Datasets

Kaggle contains a dataset that shows the stock prices for the Walt Disney Company from the early 1960s to the current year of 2024, outlining more than 60 years of price changes. The variables within the dataset include the date of each market day (Monday through Friday of each week from January 1, 1962 to August 19, 2024), the opening market price, the high price during the day, the low price of the day, the stock price at market close, and the volume of shares exchanged each day. With this dataset being so well maintained and updated, it is possible to accurately predict future Disney stock prices provided a predictive model trains well on the data and can properly handle the trends and patterns within the observations.

Methods

Initial analysis of the business problem and the data prompts exploratory data analysis of the dataset. I plan to look at how the other variables correlate with my outcome variable of closing stock price by using a heatmap to display the correlation coefficients and their relationship to the closing variable (Close) variable. I also would like to plot other charts such as line charts showcasing the movement over time for each variable. Another area of interest I hold in the data is adding potentially useful features such as volatility and daily return variables. After the exploratory data analysis and any other transformations made to the data are complete, I will be testing the data for stationarity with the Augmented Dickey-Fuller (ADF) test.

If differencing does not need to be performed, then I will split the dataset into training and test sets at an 80/20 ratio in preparation for my selected model.

As the data lends itself to time series analysis, the predictive model I have chosen is the Autoregressive Integrated Moving Average (ARIMA) model. If the ARIMA model performs poorly, I will employ a random forest regressor to predict future stock prices provided lagged features are created. The metric that will be used to calculate the ARIMA model's predictive performance is the RMSE (Root Mean Squared Error). Should I need to craft the random forest algorithm, the RMSE and the R^2 value (coefficient of determination) will be used to measure its performance. The RMSE value will provide an answer to the difference in model predictions and actual values, while the R^2 value indicates the amount of variance in the data explained by the model. I will create a grid search to identify which parameters for the ARIMA model yield the best-performing model, and plot the predictions of the random forest regressor against the actual observations should this model be needed.

Ethical Considerations

Ethics are always a concern when dealing with data; this research is no different. An apparent ethical concern is that I currently work for the Company, thereby asserting some form of bias to be recognized and set aside to be an impartial viewer of the data to report results that are not obscured by my direct affiliation with the Company. Adding onto this is the privacy of all employees (myself included) and shareholders of the Walt Disney Company. While the stock prices are public data to be obtained by anyone interested, if how the stock price fluctuates is a direct or indirect result of anyone linked to the Company, their identity is to be

protected and/or anonymized for data protection purposes. As a Cast Member, I have signed a confidentiality agreement stating that all private and secure data concerning the Company is to be kept as such to avoid such disciplinary action including, but not limited to, termination and legal action.

Another ethical debate I must clarify before the data analysis is my financial stake within the Company. To maintain transparency throughout this research, I must confess that I no longer possess any financial stake within the Walt Disney Company aside from being employed by the Company, as I owned stock within the Company over a year ago. I wish to conduct this study solely for the ability to report on the Company's stock price data using time series forecasting through predictive modeling techniques. There shall be no conflict of interest within the confines of this educational research.

A lasting ethical concern is the impact of the results on shareholders, consumers, and Cast Members. Should the resulting stock price predictions prove to be less than the current stock price, shareholders may react negatively to this information, further driving revenue away from the Company when this study is meant to showcase the predictive power of time series forecasting models on Disney stock prices and is in no way meant to persuade or dissuade current or potential investors to buy, hold, or sell shares of the Company's stock. Employees should not see the results of this study as a reason to lose morale or perform any other negative action towards the Company for this research is not intended for any ill or malicious purpose. No insider intelligence is at play within this study either, as my role as an employee has no bearing on the outcome of this research.

Challenges/Issues

What I believe may be considered a challenge is showing the audience of this study's results that there is no skewing of statistics and figures output by the Company because of my affiliation with the Company. All information provided within the data obtained is accessible to the public, and any feature engineering performed to add new variables to the original will be based purely on mathematical transformations. The resulting stock price predictions are also not to be taken as financial truth since this study is being performed under educational constraints and is not directly associated with real-world research. No financial advice should be derived from the data showcased within the study, and the communication of this recommendation and the previous statements is a key issue that if not handled carefully could lead to backlash that could adversely affect the Company's financial position and business reputation as well as my person as a direct associate of the Company.

References

In addition to the Kaggle dataset found which holds the data I am to work with in this project, I have also located an article providing insight as to how stock prices are defined and how they differ from stock value. The article links are below:

1. https://www.kaggle.com/datasets/krupalpatel07/disney-stock-data

2. https://www.investopedia.com/articles/stocks/08/stock-prices-fool.asp