Predicting Disney Stock Price

David Berberena

Bellevue University

DSC 680 Applied Data Science

Amirfarrokh Iranitalab

October 20, 2024

Predicting Disney Stock Price

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.

Background/History

The large corporation the world knows as the Walt Disney Company, which holds many entertainment offerings such as Pixar, ABC, ESPN, Marvel Entertainment, Lucasfilm, FOX, Hulu, and more, began in 1923 as Disney Brothers Cartoon Studio, founded by Walt and Roy Disney (Tenebruso, 2020). In 1928, *Steamboat Willie* came to be, introducing one of the most beloved animated characters of all time, and in 1937, the cartoon studio's *Snow White* film was named

the highest-grossing movie of that time. In 1955, Walt and Roy Disney opened their first theme park in Anaheim, CA, and called it Disneyland. The year 1957 was when Walt Disney Productions (as the Company was known at that time) went public with an IPO of \$13.88 on the stock market after its successes in both the world of media and theme park entertainment. Since then, the Company's stock has split eight times and has reached an all-time high of over two hundred dollars in 2021 before the COVID-19 pandemic. With the Walt Disney Company known across the globe and its performance remaining consistently profitable, many current and potential investors are interested in either growing or maintaining their wealth within the Company's stock.

Data Explanation

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 (Patel, 2024). The variables within the dataset include the date of each market day (Monday through Friday of each week from January 1, 1962 to August 20, 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.

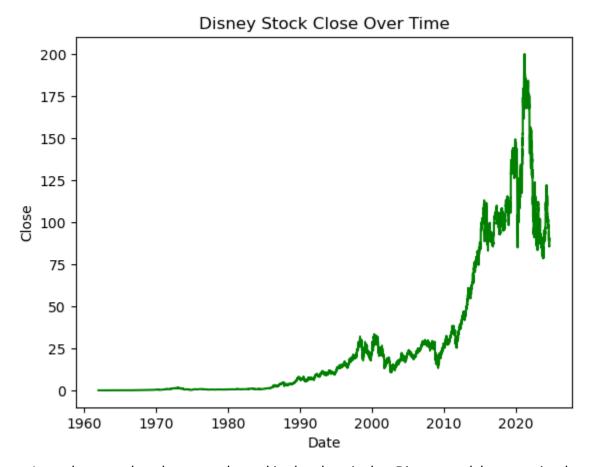
Questions that I believe end users would be interested in knowing about the prediction of Disney's stock price are vast, but there are a few key ideas that come to mind that the general public may inquire as to how this study will yield viable future results:

- 1. How can you confidently predict Disney's stock price?
- 2. Will a recommendation be provided once future stock prices are predicted, such as the Walt Disney Company being a Strong Buy, Hold, or Sell stock?
- 3. What external factors will be taken under consideration that are not within the data for the stock price prediction?
- 4. Is this research illegal, as it should be considered insider information?
- 5. How do I know the Walt Disney Company is a safe stock to invest in with my money?
- 6. What stock price within the data are you using to base predicted stock prices since the data has four price categories to choose from?
- 7. Can you use the model from this study to show future stock prices so I can buy/sell Disney shares depending on the results?
- 8. Does the volume of shares traded impact the stock price?
- 9. What trends appear in Disney's stock price from the 1960s to now in 2024?
- 10. Will the Company be profitable in the coming week? Month? Year?

Methods

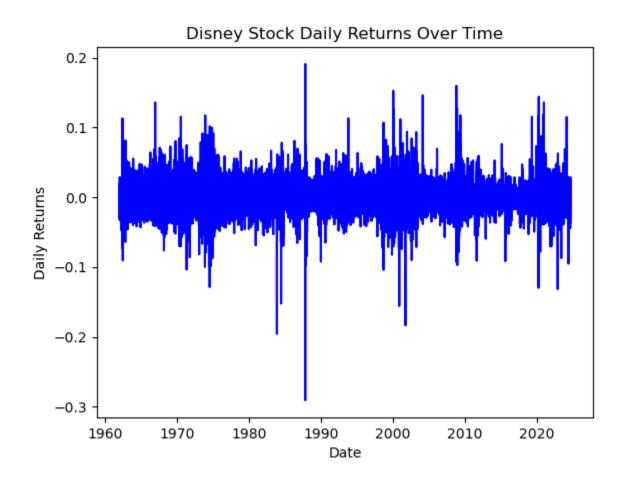
After preparing the data by changing the Date column within the dataset to the datetime data type so Python can recognize the data as time series data, the exploratory data analysis of the Disney stock price data could begin. The first order of business was to plot each

variable on a line chart, as the time series data called for such visualizations that best matched with the nature of the data. As the closing stock price was the outcome variable, I wished to learn of the Close variable's trends over time. Here is the line chart visualization generated from the data shown below.



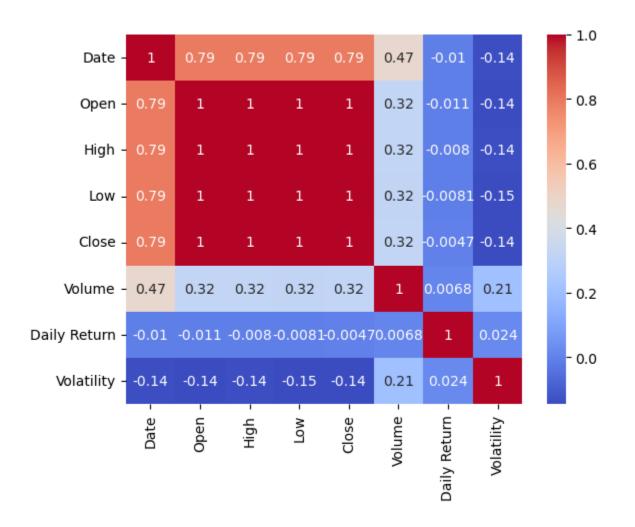
It can be seen that the general trend in the chart is that Disney stock has remained at a constant price from the beginning date of the dataset (January 1, 1962) to around the mid-1980s, then steadily increases until the 2010s where the stock takes off until it reaches an all-time high in the early 2020s with a significant drop following the spike. While the data does not tell us why, the data does illustrate this pattern and how over the tremendous amount of time within the data, the Walt Disney Company's stock behaves in different ways.

After looking at the few features that were present in the data, it was determined that feature engineering could be performed to add new features that were derivatives of the current variables. These features were known as Daily Return, or the percent change from the previous trading day's closing price, and Volatility, defined in the data's context as the standard deviation of the ten-day rolling average of the Daily Return variable. These were constructed to give the selected model choices more data to work with that mathematically stems from the original features within the data. I have also plotted these variables to see their behavior relative to the change in time. The chart for the newly added Daily Return variable is shown below.



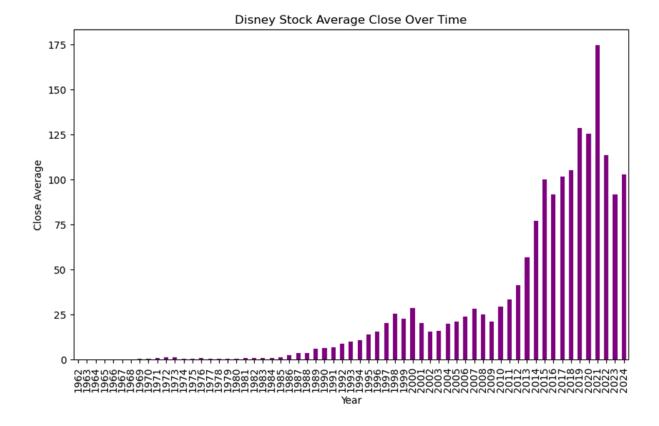
The plot shows that Disney's stock is relatively stable over time, as the only major concern is an almost twenty percent increase in closing price followed closely by an approximate thirty percent decrease in the late 1980s. The trend looks to be a change of less than twenty percent either way, which is fairly consistent.

Next, I chose to craft a correlation matrix through the construction of a heatmap to see the correlation coefficients of each variable as they relate to one another. This visualization was considered mainly as a reinforcement of the selected model choice and why time series data should not be handled by linear-assuming models such as ordinary least squares regression models. The correlation matrix heatmap can be seen here:



Looking at our outcome variable of Disney's closing stock price, we can see that Open, High, and Low are perfectly positively correlated with the Close variable, showcasing the existence of multicollinearity within the features, solidifying the determination that regression algorithms are poor model choices for this time series data. It is also intriguing to note that the heatmap shows almost no correlation between the added Daily Return feature and Close, and a slightly negatively correlation between Volatility and Close, which theoretically makes sense since the more volatile a stock is, the more the closing stock price reliant on such volatility. However, since the Walt Disney Company's stock has shown its stability over time, volatility correlates very little with the closing stock price.

Continuing the exploratory data analysis, I wished to see the yearly averages of every variable, which was showcased on a bar chart. This allowed me to see which years the stock price surged and dipped to better understand which time periods have affected the most recent closing stock prices. Here is the visualization highlighting the yearly average closing stock price for the Walt Disney Company.



As illustrated by the above chart, it can be seen that from the mid-1980s to the year 2000, a sizable bump in the average closing stock price occurred. From 2010 to 2021, Disney stock skyrocketed in average closing price, then took a sharp dive in the years after leading to the current time. This particular chart may be useful when deciding how to split the data when training the model selections to predict Disney's closing stock price. With the EDA stage complete, we can now launch into the model choices that have been selected: the time series-oriented Autoregressive Integrated Moving Average (ARIMA) model and the random forest regressor algorithm. The ARIMA model is my first model choice, as it lends itself to time series data and is designed to handle fluctuations in the data caused by trends such as seasonality. The random forest regressor model choice is also a strong choice yet will be treated as a contingency plan should the ARIMA model perform poorly.

Model Creation and Analysis

After careful consideration of the insights gained from exploratory data analysis, I have decided to subset the data to only take into account the last ten years of Disney stock price data. With the large variation in Disney's stock price throughout the entirety of the dataset, many models may not be able to handle these observations as they look like outlier data points. Also, this study's goal is to predict Disney's closing stock price for the near future, meaning the most recent figures are more impactful in determining future numbers in the short term future. With that addressed, time series data used to train an ARIMA model needs to be tested for stationarity before being split into training and test sets. Stationarity refers to the presence of trends in the data and the test to measure the stationarity of our data is the Augmented Dickey-Fuller (ADF) test. Upon performing the ADF test on the data, the p-value, being the major indicator of stationarity, showed that the data contained periodic trends and needed to be handled by way of differencing. Differencing is necessary to help the ARIMA model better generalize the data to make more accurate predictions, as the mean, variance, and autocorrelation factors are not constant at any given point within the data, which differencing assists in normalizing.

Now seeing that the data needs to be differenced, I split the data into training and test sets using the 80/20 ratio typical of data science projects. I then crafted the ARIMA mode using a function that automatically found the parameters needed for the model to perform its best, which also allowed the model to difference the data automatically. To provide more insight, the ARIMA model has three impactful parameters: the autoregressive factor which is represented by the variable p, the differencing factor which is notated as d, and the moving average factor

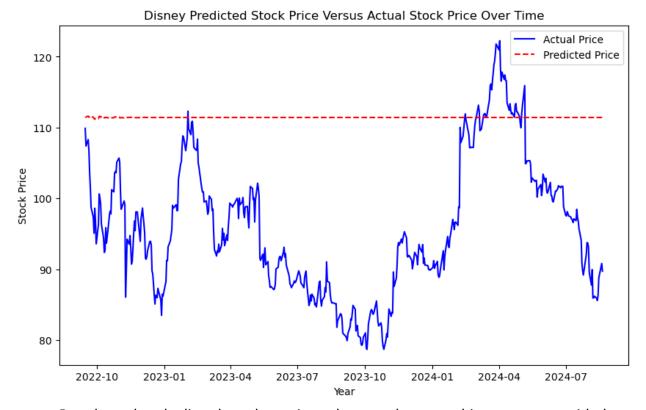
best known as q. The function used to create the time series model, called auto_arima(), automatically runs through the most common combinations of these variables with certain parameter values and outputs the best-performing model based on a certain metric. The metric I used as the indicator of the best-performing ARIMA model was the Akaike Information Criterion (AIC). The AIC is a statistic that, while holding very little meaning to a predictive model on its own, showcases a model's performance comparable to other predictive models just like it. The model with the lowest AIC is determined to be the best-performing time series model, so once the auto_arima() function ran, the results were as shown below.

Best model: ARIMA(1,1,0)(2,0,2)[5] Total fit time: 65.622 seconds							
SARIMAX Results							
Dep. Variable Model: Date: Time: Sample:	SARI	MAX(1, 1, 0)		1, 2], 5) Oct 2024	No. Observati Log Likelihod AIC BIC HQIC		1939 -4177.590 8367.181 8400.597 8379.470
=======	coef	std err	Z	P> z	[0.025	0.975]	
ar.S.L5 ar.S.L10 ma.S.L5 ma.S.L10	-1.0476 -0.6049 1.0464 0.6498		-5.529 -3.840 5.696 4.401	0.000 0.000		-0.676 -0.296	
Ljung-Box (L1) (Q): Prob(Q): Heteroskedasticity (H): Prob(H) (two-sided):			0.00 0.98 7.11 0.00	Prob(JB): Skew:	a (JB):	12864.26 0.00 0.67 15.55	

The model results shown above state that the model that was output is a SARIMA model and not an ARIMA model. The main reason why an ARIMA model was not crafted is because of the presence of seasonality in the data, which ARIMA is not built to handle while the SARIMA and SARIMAX models are. Even though the model results say this is a SARIMAX model, a

SARIMA model was created here since there are no other external factors placed in the model, such factors being shareholder sentiment, unemployment and/or inflation, company restructuring, or any other that could potentially impact Disney's closing stock price but are not derived from the dataset. Looking further into the results, we can see that the p, d, and q factors were calculated along with the seasonal versions of these factors (P, D, and Q) as SARIMA models need the correct seasonal factors as well, and the number of days within a seasonal cycle (the model results show five as the number of days within the cycle which can be seen as a trading week).

For time series models, the performance metric I have chosen to evaluate the model is the root mean squared error (RMSE), and after calculating the RMSE when computing model predictions, the output came out to 18.24. Since the RMSE statistic is expressed in the same scale as the outcome variable, the margin of error between the SARIMA model's predictions and the actual observations is rather high, as an \$18 difference in actual versus predicted observation could translate to between a ten to twenty percent difference depending on the day observed. After plotting the predictions and actual data points, here is the resulting chart:

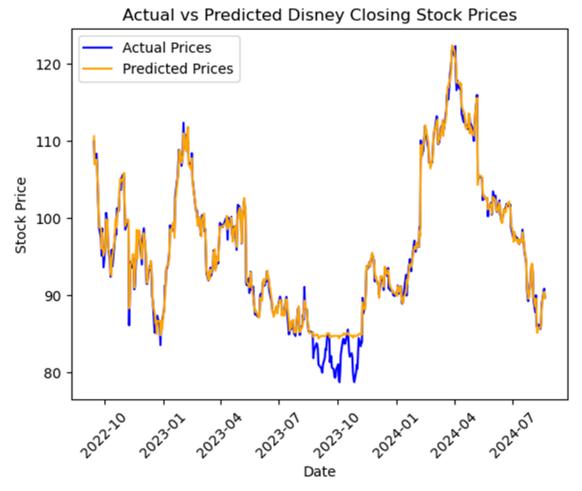


Based on what the line chart shows, it can be seen that something went awry with the construction of the SARIMA model as the predictions have resulted in an almost linearly constant stock price for the duration of the test set's closing price observations. The model does not seem to grasp the fluctuations within the stock data, therefore a deeper dive into this model's parameters would be necessary for the SARIMA model to yield better results, for as it stands this model does not accurately predict Disney's closing stock price.

We have now reached the need to craft our second model choice, which is the random forest regressor model. For this model to properly handle time series data like our current dataset, I have engineered more features into the data, specifically lagged values which alert the model that it is dealing with time series data. To mirror the number of trading days in one week, I have created five lagged features in preparation for the random forest regressor to train on the data using these lagged values as indicators of past values being used to predict future

values. To provide the random forest model with more information to help it perform better, I also engineered three-day, five-day, and ten-day rolling averages of the Close and Volume variables into the data. Before splitting the data into training and test sets, I subset the data just as in the SARIMA to include the stock price history starting January 1, 2015 to August 20, 2024.

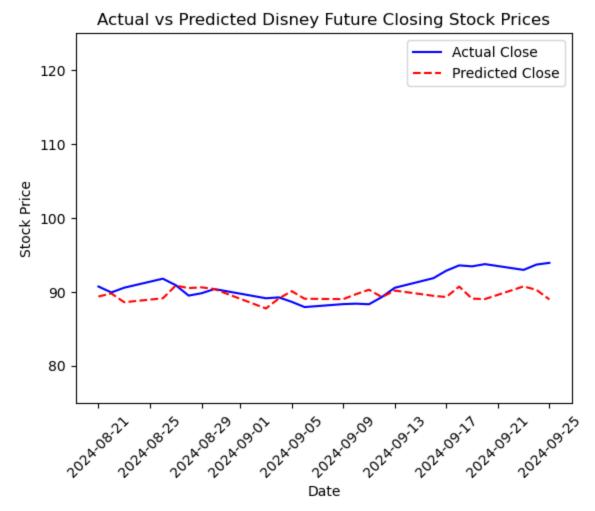
Once the data had been split using train_test_split() at an 80/20 ratio making sure to keep the oldest dates in the training set and the most recent dates in the test set, I crafted the model and realized its performance metrics, which were the RMSE once again and the R-squared value. In conjunction with the RMSE's ability to account for the difference between the predictions and actual values, the R-squared value is a good measure of model performance here since it allows one to see what percentage of the data's variance is explained by the model. The random forest regressor's RMSE value was 1.24 while the R-squared value was 0.98. These metrics demonstrate a much more accurate model than the SARIMA model as the RMSE value is very close to zero, indicating a small margin of error between predictions and actual observations, and the R-squared value is close to one, highlighting a high percentage of data variance explained by the random forest algorithm. The last performance indicator I wished to see was the line chart comparing the random forest model's predicted closing Disney stock prices against the Company's true closing stock prices. The resulting plot is shown here:



Here we can see clearly that the predictions created by the random forest regressor model closely mimic the Walt Disney Company's actual closing stock prices for the timeframe showcased in the visualization, which means that the random forest regressor is a good fit for the data and can predict Disney's closing stock price with a high degree of accuracy (98% judging by the model's R-squared value). I then extracted feature importances from the model and ordered them by decreasing importance to show the top five most important features in predicting Disney's closing stock price. At this stage of the model's development, I wish to see how the random forest regressor would perform when generating Disney closing stock price predictions for stock prices unseen by the data, mimicking what it would look like for the model to be deployed.

With the model tasked to predict closing stock prices on data not provided in the original dataset, some creativity is necessary. A new DataFrame object was created to hold future dates beginning from August 21, 2024 extending twenty-five business days forward, just over a month after the original dataset ends (I also made sure to skip September 2, 2024 for Labor Day since the stock market was closed on that day). The new DataFrame was then populated with the same variables as the original dataset along with the features engineered right before it was trained using the random forest algorithm. The values for these variables were derived from the original dataset's most recent values with a small numerical variation added to prevent stagnant observations and to mimic the minute changes stock prices undergo from day to day. With the future Disney stock price DataFrame complete, the trained random forest regressor was used to create future predictions using this future DataFrame and added to it through the creation of a new variable named Predicted Close. Now that we have predicted future values, I accessed the Internet to look for the Walt Disney Company's actual closing stock prices for the timeframe outlined in my future DataFrame (ending on September 25, 2024). Upon finding the true closing stock prices, I added them into the DataFrame under the guise of Actual Close (Yahoo!, 2020).

With both predicted and actual future stock price observations, I could now evaluate the random forest model's performance on unseen data using the RMSE and the plot of predictions versus actual observations. After calculating the RMSE of the model using the future stock price predictions the random forest regressor created, the metric's value came to be 2.32, which indicates another low margin of error between predictions and real values. Plotting the predictions against the actual future stock price observations yielded this visualization:



The final chart highlights the high accuracy with which the random forest regressor model has predicted closing stock prices as there is little difference between actual values and the predictions. This last evaluation of the performance of the random forest regressor on unseen data proves that the model is ready for deployment since this exercise was a form of deployment in itself.

Recommendations/Implementation Plan

With the model performance results in, we can now identify actionable insights for end users to act on should they wish to do so. Financial risk management regarding investments in

the stock market is important to take note of multiple times throughout a shareholder's investment in the companies that make up their stock portfolio. The Walt Disney Company's stock is no different, as the prediction of the Company's closing stock price will allow shareholders to better assess their financial risk in buying, selling, or holding Disney stock. Now that we have seen the accuracy demonstrated by the random forest regressor model, the recommendations to be outlined follow four financial choices relating to the acquisition of Disney stock. What can be recommended as shown by the model's results are as follows:

- 1. If you are someone who wishes to invest in Disney stock and become a shareholder, the data suggests that the Company is a good option to purchase provided its vast history of increasing stock price, low levels of volatility, consistent volume of traded shares, relatively low fluctuations of daily returns, and propensity to overcome price dips with higher surges given time. Looking at the initial predictions versus actual values for the random forest algorithm, higher lows can be seen at each dip in the past couple of years, indicating an increasing trend over time.
- 2. If you are a current shareholder and are looking to purchase more stock to increase your position, an evaluation of your current position is necessary before purchasing. If the goal of purchasing is to buy down your stock average, the addition of more shares may not be advantageous unless you are willing to play the long game as Disney's stock has not exploded significantly since 2021 and has fallen significantly since then to a more sustainable level with little volatility. Yet if the goal is to acquire more shares to have a larger presence in the Company, this may prove beneficial, which once again suggests a long game.

- 3. If you are a current shareholder looking to hold your position with the Walt Disney Company, this strategy may be lucrative in the long run. With Disney having established itself as a safe and stable stock, there is no immediate reason why a shareholder choosing to hold their position would lead to a stark decrease in their portfolio valuation. For extra precaution, however, the shareholder may wish to consider placing stop limit orders to mitigate any unforeseen risk should they not be attentive to the Company's swift movements.
- 4. If you are a shareholder and are looking to sell part or all of your position, an evaluation of your current position is necessary before initiating a sale. The price at which the initial stock was purchased greatly impacts the desire to sell as the monetary gain from the sale is unique for each individual. If the money gained from the sale of some or all of a position in the Company is satisfactory enough to relinquish partial ownership of the Company, then even the results of this study would not be enough to dissuade someone so intent on turning over their shares.

The implementation of the results of this research will most likely fall on professional stockbrokers who have access to predictive modeling techniques as they are the ones to can present clients with detailed advice on how best to manage their portfolio provided the Walt Disney Company is a part of it. Being able to back up claims of short-term or long-term stock success with data-backed evidence would establish a better client-broker relationship and increased confidence in the stock, potentially leading to increased consumer investment and more market expenditure to fuel the economy. Professional financial advisors and stockbrokers would be able to access and use this research wisely as any notion of collusion or insider

trading is heavily punishable by the Securities Exchange Commission and other large financial corporations and organizations.

Assumptions

An assumption made in the research was the absence of seasonality within the original dataset being queried, as the selection of the ARIMA model automatically rules out the question of the data being seasonal in nature. Granted this assumption was addressed in the construction of the SARIMA model using the auto_arima() function, yet an ARIMA model was expected without studying the data more thoroughly before choosing my initial model choice. The SARIMA model also did not generalize the data well, suggesting that there was an underlying trend or pattern within the pattern that was not identified and noted for the construction of the SARIMA model. The other assumption I made involving the erroneous ARIMA model was that I only needed to check for stationarity when there was more going on in the data than was originally considered. This underestimation of the data's complexity invalidated the ARIMA model choice and output an underwhelming SARIMA model.

Challenges/Limitations

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 was 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, the previously stated recommendations based on the model's results, 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.

Another challenge that came with the data was its sheer volume, spanning over sixty years. After running into many roadblocks by trying to encapsulate all of the data within my time series model selections, I came to realize that over half of the data had to be considered outlier data due to the extremely low stock prices outlined, driving the average stock price down and muddling predictions. Only after I thought to remove over three-quarters of the data to focus on the ten most recent years did my models start to yield viable results. This research enlightened me to the realization that sometimes including too much data can be as fruitless as incorporating too little data.

The glaring limitation that would have been helpful to understand more about why Disney's stock prices fluctuated at certain times within the data was the absence of external factors within the data. These external factors could have been major events that happened during the timeframe the dataset covered, such as Walt Disney's untimely passing, the COVID-19 pandemic, the acquisition of certain brands like Marvel Entertainment, ESPN, Hulu, and others, the changing of CEOs, economic recession, consumer sentiment, and many more. Had these factors been represented in the data, they could have been placed into the auto_arima()

function to help realize a SARIMAX model that could have handled these external variables in addition to the Close variable and would most likely have yielded a more accurate time series model. However as none of these factors were present in the data to be identified as direct reasons for certain stock price movements, nothing can be confirmed definitively as to why the data behaves the way it does. Overall, while the study yielded viable results, more data could have been gathered relating to external variables in search of a more detailed explanation of the stock's underlying seasonal patterns and trends.

Ethical Assessment

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.

Future Uses/Additional Applications

This research is extremely beneficial to the stockbroker sector of the finance industry.

As predictive modeling can be advantageous in alerting shareholders to the need for preemptive decision-making regarding the buying, selling, or holding of the shares within their stock portfolio, these models can be extended to other areas of stock data analysis. Finance and trading apps like Webull and SoFi can incorporate predictive models into their guest user

interface to help investors better analyze the future performance of stock options they own and/or wish to own, guiding them to make more informed decisions with the capital they are looking to invest. The Walt Disney Company can harness this predictive modeling so the CEO, CFO, and the other members of the executive team can stay ahead of the market and plan their press releases accordingly, so positive news can be delivered when the stock price is down for a healthy rebound and negative news can be relayed during a future decrease in volume of shares traded to take a minimal potential loss. With the amount of resources at the Walt Disney Company's disposal, predictive modeling techniques like the future prediction of their stock price will help the Company to stay at the forefront of the entertainment industry.

Conclusion

After careful analysis, feature engineering, and training of the data using viable model choices, it has been found that the random forest regressor model performed the best, with a nearly one-dollar margin of error between current Disney current stock price predictions and actual closing prices and a two-dollar RMSE concerning future unseen predictions against actual future closing stock prices. With the random forest model explaining over ninety-eight percent of the data's variance as well, the model is readily equipped for the deployment stage with the addition of new data outlining external factors affecting the stock price. The Walt Disney Company is a stable company to invest in with a bright future ahead according to the data, and with predictive modeling techniques to aid the Company and its current and potential shareholders, that future should remain bright barring any typical minor volatility and/or major events.

Appendix/References

- Tenebruso, J. (2020, November 2). If You Invested \$1,000 in Disney's IPO, This is How
 Much Money You'd Have Now. Nasdaq. Retrieved October 4, 2024, from
 https://www.nasdaq.com/articles/if-you-invested-\$1000-in-disneys-ipo-this-is-how-much-money-youd-have-now-2020-11-22
- Patel, K. (2024, August 20). Disney Stock Data. Retrieved October 4, 2024, from https://www.kaggle.com/datasets/krupalpatel07/disney-stock-data
- 3. Yahoo! (2024, October 4). The Walt Disney Company (DIS). Yahoo!Finance. Retrieved October 5, 2024, from https://finance.yahoo.com/quote/DIS/history/