Research Question

A. The research question for my performance assessment is; “To what extent can the closing price (Y) of Apple stock be predicted with 90% accuracy using MAPE/MSE?

Justification: I am looking at this from the perspective of an employee for a company who is interested in purchasing shares of Apple. Moreover, I chose to use Apple as the company whose stock price I wanted to analyze and forecast due to Apple being a widely followed company with a significant impact on the stock market.

Context: The context of this research question lies in the field of financial analysis and investment decision-making. By studying Apple's stock price data, I can gain insight into market trends, investor sentiment, and the financial stability of the company due to them being publicly traded.

Hypothesis: The hypothesis is that there are identifiable patterns and trends in Apple's stock price data that can be used to make informed investment decisions. These patterns are influenced by factors such as company performance and market conditions. Below are my null and alternate hypotheses.

Null hypothesis: The closing price of Apple stock cannot be predicted with 90% accuracy using MAPE/MSE.

Alternate Hypothesis: The closing price of Apple stock can be predicted with 90% accuracy using MAPE/MSE.

Data Collection

B. **Data Collection Process**: The data I used was pulled from a publicly available dataset on Kaggle. The csv file contains stock price data such as; the date, high, low, close, adjusted close, and the volume of shares from the initial IPO until March 2022. I chose to use a past date rather than present day due in order to ensure I am not recommending investment advice to others.

Advantage: One advantage of using this dataset is the comprehensive historical record it provides, allowing for a detailed analysis of long-term trends.

Disadvantage: A disadvantage is that the dataset does not include the most up-to-date stock price data, which could limit the accuracy of the analysis.

Challenges Overcome: To overcome the limitation of this analysis being used for investment advice. I was able to overcome this challenge by downloading historical data from Kaggle of dates that have already occurred as opposed to scraping data from Bloomberg or Yahoo Finance.

Data Extraction and Preparation

C. **Data Extraction and Preparation Process**: The data was extracted from the Kaggle dataset using Python and the pandas library. The data was then cleaned and prepared for analysis, including handling checking for null values and formatting the data for analysis.

Tools and Techniques: Python and pandas were used for data extraction and preparation due to their efficiency in handling large datasets and their flexibility in data manipulation.

> **Advantage**: One advantage of using pandas is the ease of use and wide range of functionality for data manipulation.
>
> **Disadvantage**: A disadvantage is that pandas tends to struggle with larger datasets which can result in slower loading time and they are more intensive on the machine.

Analysis

D. **Data Analysis Process**: The data was analyzed using various statistical techniques, including calculating summary statistics, visualizing trends, and then I performed three separate forecasting techniques using ARIMA, Prophet, and LSTM.

Forecasting Techniques

ARIMA: ARIMA is typically used for time series data, in order to prepare the data for analysis using this method I had to difference the data.

Prophet: Prophet is an open-source forecasting tool that was created by Facebook. I chose to use this method due to the level of automation that can be created using this library, it also works well with handling daily observations.

LSTM: This was the final technique I employed for this forecasting analysis. LSTM is typically used to learn, and classify sequential data.

Justification of Selection

ARIMA: Because this is stock market data, and ARIMA excels in demand forecasting it was an easy choice to use for this assignment.

Prophet: I chose this method due to the way it can handle daily data as the values in our time series are daily throughout the week.

LSTM: I chose this method due to the way it is able to capture complex patterns in the data. Since our data begins in 1980 this was extremely helpful the first few times I performed the analysis on the full dataset.

Advantage

ARIMA: Can capture patterns and behaviors in the data easily.

Prophet: Works extremely well with daily, weekly, and hourly data.

LSTM: Able to process longer sequences of data.

Disadvantage

ARIMA: The parameters of the model (p,d,q) must be manually defined and finding the most accurate fit can take more time.

Prophet: It does not look for casual relationships between the data, it essentially just finds the best curve to fit the data.

LSTM: they are more intensive compared to RNN and run slower.

Data Summary and Implications

E. **Data Analysis Implications**: The use of ARIMA, Prophet, and LSTM techniques to the data analysis process provided informative results about the behavior of the time series data. Prophet successfully captured daily patterns, ARIMA offered a reliable forecasting model, and LSTM proved it was able to capture complex trends and long-term dependencies.

Limitation: One limitation of the analysis with the ARIMA model is that the model assumes the data is linear. Due to this it may not capture the nonlinear patterns of the data.

Recommendation: Based on the analysis, even though we have to reject the null hypothesis, I would recommend that the forecast for the stock looks strong and that we should try to consider it as a buy if it aligns with our companies stated investment objectives. The forecast was unable to accurately predict Apple's stock price over the long term, with the lower and upper closing forecast values being significantly lower than what the actual stock price was.Moreover, I would want to select one of the three techniques I used to present and roll that up into an executive summary for an ELT management team.

Future Study Directions: Future studies could focus on using more recent data to improve the accuracy of predictions. Moreover, we could also explore feature engineering in order to enhance the models performance. For example, we could incorporate more external factors such as recent news and create a relationship between the stock price and customer sentiment. This could assist the LSTM model in its accuracy by using sentiment to give it more context for the data.

Sources:

- APPLE Stock Data. (n.d.). Www.kaggle.com. https://www.kaggle.com/datasets/varpit94/apple-stockdata-updated-till-22jun2021
- *Prophet: forecasting at scale - Meta Research*. (n.d.). Meta Research. https://research.facebook.com/blog/2017/2/prophet-forecasting-at-scale/
- *Time Series Forecasts using Facebook's Prophet*. (2018, May 10). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/05/generate-accurate-forecasts-facebook-prophet-python-r/
- Filho, M. (2023, February 20). *Multiple Time Series Forecasting with Prophet in Python*. Forecastegy.com. https://forecastegy.com/posts/multiple-time-series-forecasting-with-prophet-in-python/