**STOCK MARKET FORECASTING OF LEADING COMPANIES**

**Resources**

GitHub repository: [<https://github.com/Rgmoogachiri/captsone_project/tree/main>]

**Members**

* Alice Nyamngo [alice.nyamngo@student.moringaschool.com](mailto:alice.nyamngo@student.moringaschool.com)
* Laura Mutheu [laura.mutheu@student.moringaschool.com](mailto:laura.mutheu@student.moringaschool.com)
* Emmanuellar Karisa [emmanuel.karisa@student.moringaschool](mailto:emmanuel.karisa@student.moringaschool)
* Richard Muriithi [richard.muriithi@student.moringaschool.com](mailto:richard.muriithi@student.moringaschool.com)
* Chepkemoi Ruto [chepkemoi.ruto@student.moringaschool.com](mailto:chepkemoi.ruto@student.moringaschool.com)
* Daniel Muriithi [daniel.muriithi@student.moringaschool.com](mailto:daniel.muriithi@student.moringaschool.com)

**Business Understanding**

**Business Overview**

In the fast-paced world of finance, timely access to real-time stock market data is essential for making well-informed decisions, crafting effective strategies, and managing risks. This data wealth offers a unique opportunity to develop predictive models capable of forecasting stock prices, spotting emerging trends, and providing actionable insights to empower investors and traders.

Our project introduces a comprehensive model designed to harness real-time stock market data for constructing predictive models. By leveraging sophisticated machine learning techniques and advanced data analysis methodologies, our project aims to bridge the gap between raw data and strategic decision-making. We believe that by providing stakeholders in the financial sector with robust predictive tools, we can help them navigate the intricate landscape of the stock market more effectively.

This project utilized real-time stock and historical market data, including data collection, preprocessing, feature engineering, model development, and rigorous evaluation. Through the collaborative efforts of multidisciplinary teams, combined with our commitment to innovation and excellence, we aspire to redefine the boundaries of predictive analytics in finance. Our goal is to create value, manage risks, and foster sustainable growth within the financial ecosystem.

**Business Objectives**

The **main objective** of this study is to develop predictive models leveraging real-time and historical stock market data to enhance decision-making processes for investors and traders.

Other objectives include:

1. **Bridge the Gap**: Bridge the gap between raw financial data and strategic decision-making by providing stakeholders with actionable insights derived from advanced data analysis techniques.

2. **Empower Stakeholders**: Empower stakeholders in the financial sector to navigate the complexities of the stock market more effectively through the use of robust predictive tools.

3. **Redefine Boundaries**: Redefine the boundaries of predictive analytics in finance by pioneering innovative approaches and methodologies that deliver tangible value and sustainable growth within the ecosystem.

**Success Criteria**

This project will be deemed successful when the following is achieved:

1. The project is to be completed on time using available resources.
2. Project achieving **functional requirements** such as

* Performing crucial data science skills such as cleaning and exploratory data analysis
* Building effective and accurate Machine Learning models using time series analysis
* Deployment of the model for easier interaction with the end user

1. Project achieving **non-functional requirements** such as:

* Achieving customer satisfaction
* Teamwork

1. The project meets all the stated objectives.

***Assessing the Situation***

1. ***Personnel***

Personnel: technical support and data scientists.

1. ***Datasets***

***S***tock Market dataset: - Yahoo Finance uses an API to collect the data and convert it to an xlsx file.

1. ***Assumptions***

There is an assumption that the data provided represents the actual interest of investors in Kenya.

1. ***Constraints***

Limited project period

Large data size

**Data Understanding**

**Overview**

The datasets used in this project are:

1. Stock Market dataset: - Yahoo Finance uses API to collect the data and convert it to an xlsx file.

**Data Description**

***Stock Market dataset***

The dataset consists of two xlsx files: “stock\_data.xlsx which contains the individual companies from Yahoo Finance and “update\_stock\_data.xlsx” which is a combination of the companies and contains over 11,000 stock prices.

It contains the following columns:

* close
* high
* low
* open
* private
* symbol
* volume
* id
* key
* subkey
* updated
* changeOverTime
* marketChangeOverTime
* uOpen
* uClose
* uHigh
* uLow
* uVolume
* fOpen
* fClose
* fHigh
* fLow
* fVolume
* label
* change
* changePercent

**Verifying Data Quality**

This is the unprocessed, raw data. But it has a high degree of integrity.  
 A few values are also absent. Data cleaning techniques are necessary to get the data ready for analysis.

**Data Preparation**

**1.** **Loading data:** the files were downloaded to our local machine. They were then loaded to our environment by reading them as xlsx files.

**2.** **Cleaning data:** The unnecessary columns were dropped then null values in the dataset were dropped. We then added a category column to the

**Preprocessing**

We performed feature engineering on our data to get the desired input to feed our

model. The following steps were taken:

1. We created new columns '7\_day\_ma' and '30\_day\_ma' which will highlight the long-term trends or cycles in the data.

2. New columns, ‘close\_lag\_1’ and ‘close\_lag\_7 were created to represent the past values of that variable over a specified period.

3. Volatility measures were done to help quantify the degree of variation in the price over time.

4. RSI was calculated because it helps identify potential overbought or oversold conditions in security, which traders can interpret as signals to buy or sell. This was stored in the column ‘rsi’.

5. Percentage change was calculated because it provides a normalized representation of price movements over time. Expressing price changes as percentages, it allows for better comparison across different securities, periods, and price levels. This normalization is crucial for analyzing stocks with different price ranges.

**Modeling**

* Moving Average (MA) Model:
  + The MA model uses the average of past observations within a specific window to forecast future values. A lower AIC/BIC indicates a better model fit. This model considers only past observations and does not account for trends or seasonality.
* AutoRegressive (AR) Model:
  + The AR model uses a linear combination of past observations to forecast future values. It captures the relationship between an observation and several lagged observations. Similar to MA, lower AIC/BIC values suggest a better model fit.
* AutoRegressive Moving Average (ARMA) Model:
  + The ARMA model combines both autoregressive and moving average components. It captures both the linear relationship between past observations and the average behavior over time. Higher AIC/BIC values compared to AR and MA models indicate a slightly poorer fit
* AutoRegressive Integrated Moving Average (ARIMA) Model:
  + ARIMA is an extension of ARMA with an additional differencing step to make the time series stationary. It integrates differencing into the model. The lower AIC/BIC values compared to other models suggest a better fit.
* Seasonal AutoRegressive Integrated Moving Average (SARIMA) Model:
  + SARIMA is an extension of ARIMA that includes seasonality in the model. It's suitable for time series data with seasonal patterns. The considerably higher AIC/BIC values compared to other models suggest a poorer fit, possibly indicating that the seasonal component might not be adequately models
* Facebook Prophet Model:
  + Prophet is a forecasting tool developed by Facebook that is designed for analyzing time series data that display patterns on different time scales.
  + AIC and BIC were not provided.

Interpretation and Conclusion:

* AIC and BIC: AIC and BIC are information criteria used for model selection. Lower values indicate better models, balancing goodness of fit with model complexity.
* Best Model Selection: Based on the AIC and BIC values, the SARIMA model seems to be the best fit for your data, as it has the lowest AIC and BIC values among the models you've tried.
* Reasoning: SARIMA models are particularly useful for capturing both non-seasonal and seasonal patterns in time series data, which might be present in your Amazon stock price data.

**Evaluation**

1. Moving Averages (MA):

· AIC (Akaike Information Criterion): 3246.632

· BIC (Bayesian Information Criterion): 3254.410

2. Autoregressive Model (AR):

· AIC: 3248.458

· BIC: 3260.124

3. Autoregressive Moving Averages (ARMA):

· AIC: 3250.417

· BIC: 3265.972

4. Autoregressive Integrated Moving Averages (ARIMA):

· AIC: 3242.061

· BIC: 3253.719

5. Seasonal Autoregressive Integrated Moving Average Model (SARIMA):

· AIC: 4454.614

· BIC: 4462.324

Interpretation and Conclusion:

* AIC and BIC: AIC and BIC are information criteria used for model selection. Lower values indicate better models, balancing goodness of fit with model complexity.
* Best Model Selection: Based on the AIC and BIC values, the SARIMA model seems to be the best fit for your data, as it has the lowest AIC and BIC values among the models you've tried.
* When the SARIMA model is compared to the prophet model The Prophet model captures the underlying patterns and dynamics of the stock market data more effectively than the SARIMA model. This could be due to Prophet's ability to handle complex seasonal patterns and trend changes in the data more intuitively. Therefore prophet model is performing better in terms of predictive accuracy for the given dataset and forecasting horizon thus the best model

**Deployment**

Deployment was done using the Streamlit. This service enabled us to create a webpage that a user can interact with when they intend to predict.

The page contains a drop-down where the user can select the company of interest and a slider to choose the number of years the user can predict

The selected company is scrutinized and the output is given to the user Contents of the output are:

1. The forecasted components.
2. The graph shows the trend of the forecast price in the year, the trend weekly, and the seasonal trend for the whole year.

NB: The deployment sometime causes changes