PROJECT REPORT

DATA MINING

****

**SALMAN ZAIDI & DAWOOD TANVIR**

21 i 1738 - 21 i 1352

5.14.2024

SIR SAAD MUNIR / MA’AM AYESHA KAMRAN

# **DATA PREPROCESSING:**

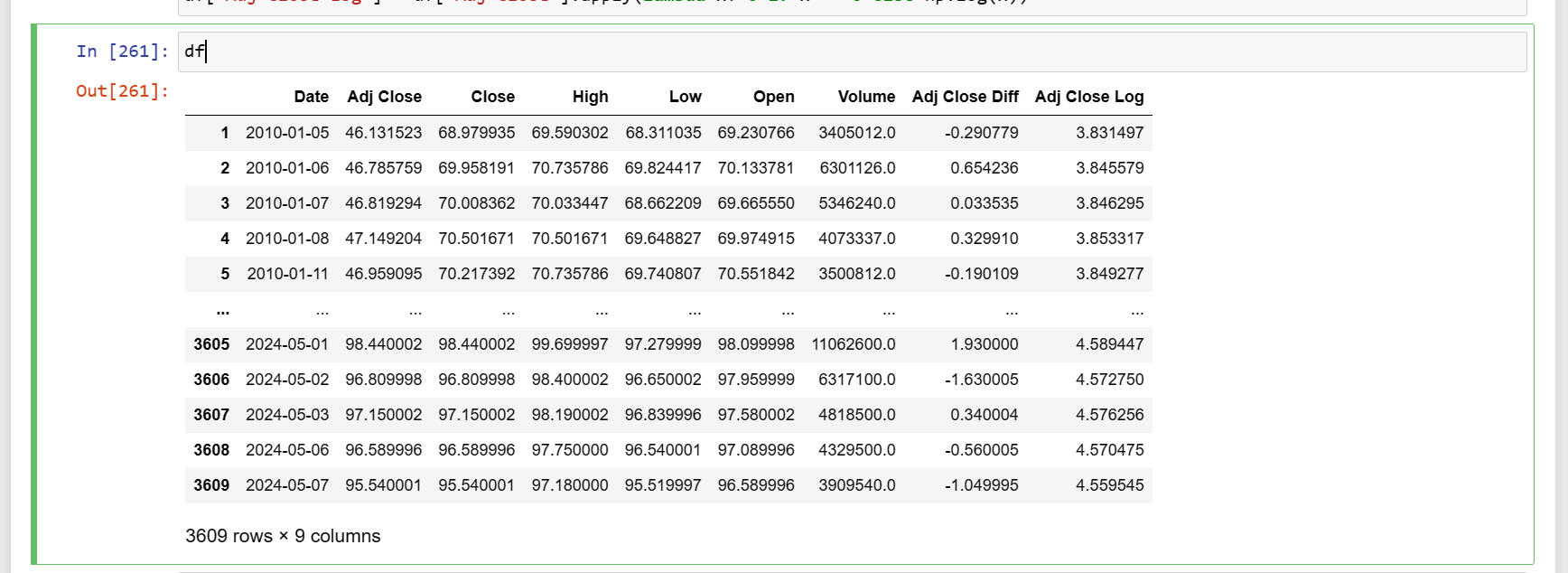
* DATA:

Data contains all critical information about SP500 Stocks, Such as, Close, Adj close, Date, Volume, High, Low etc



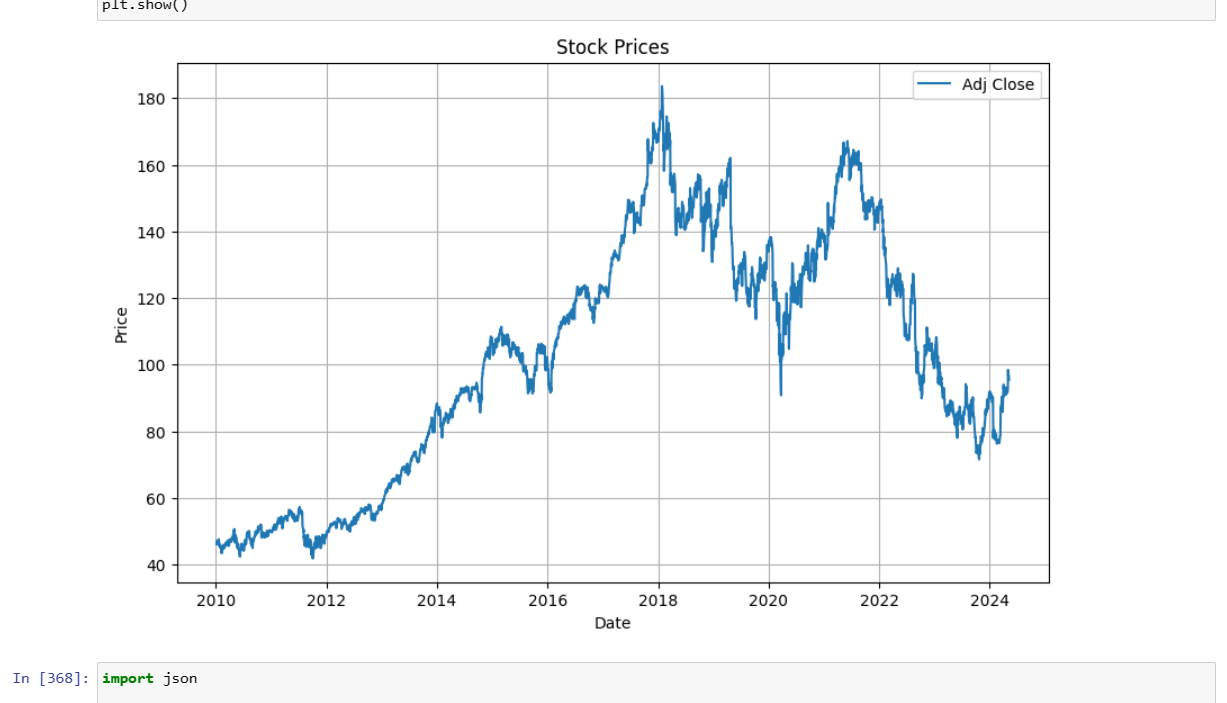
* DIMENSIONALITY REDUCTION : PREPROCESSING

The yearly stock data is converted into monthly stock data.



* DATA STATISTICS AND MODELING FOR BETTER UNDERSTANDING

The monthly stock data is visualized over trends.



* FINALIZED DATASET

The finalized monthly SP500 dataset…



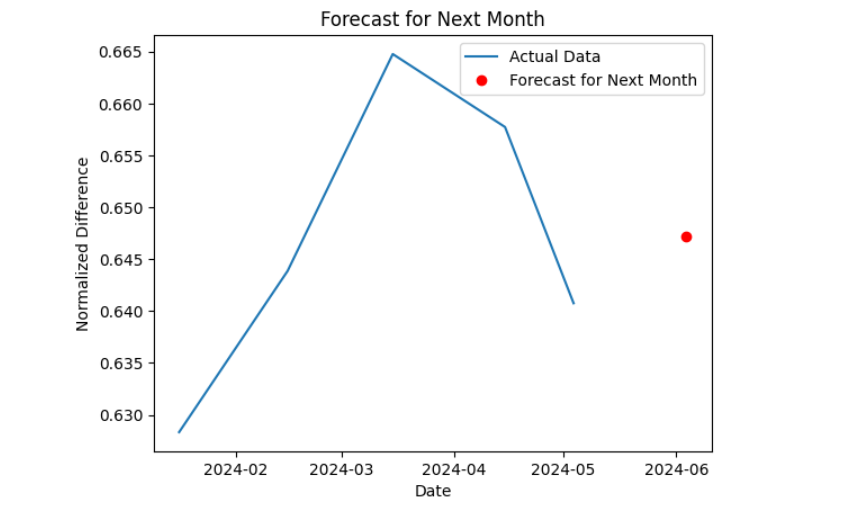
# **SYSTEM ARCHITECTURE:**

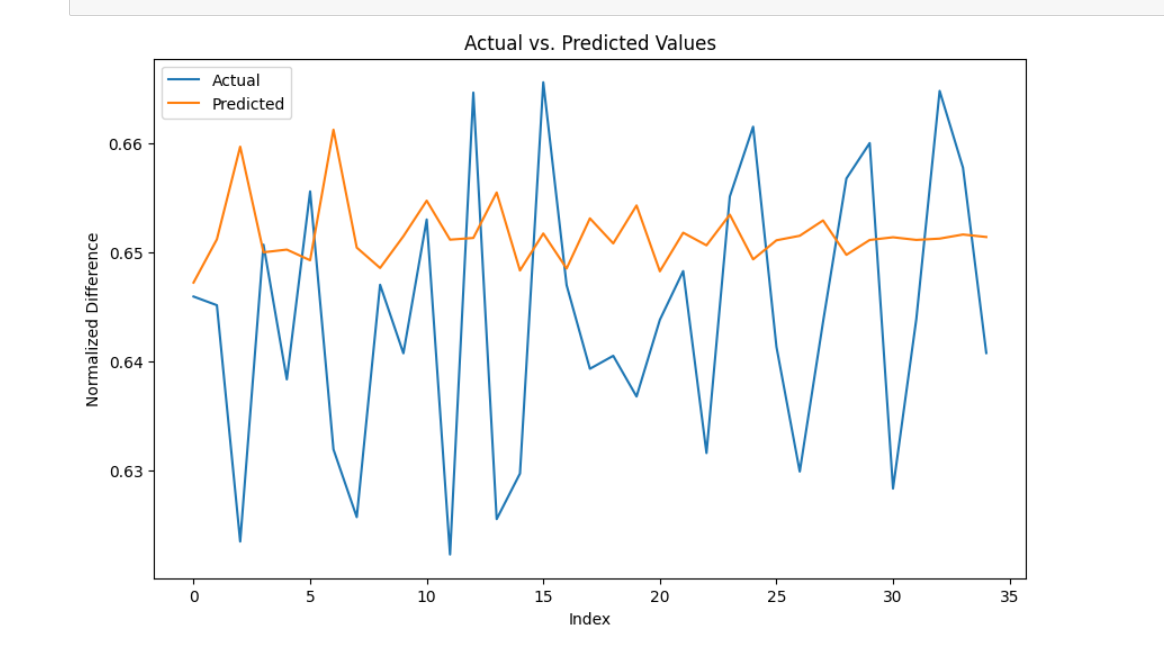
* **ARIMA**:

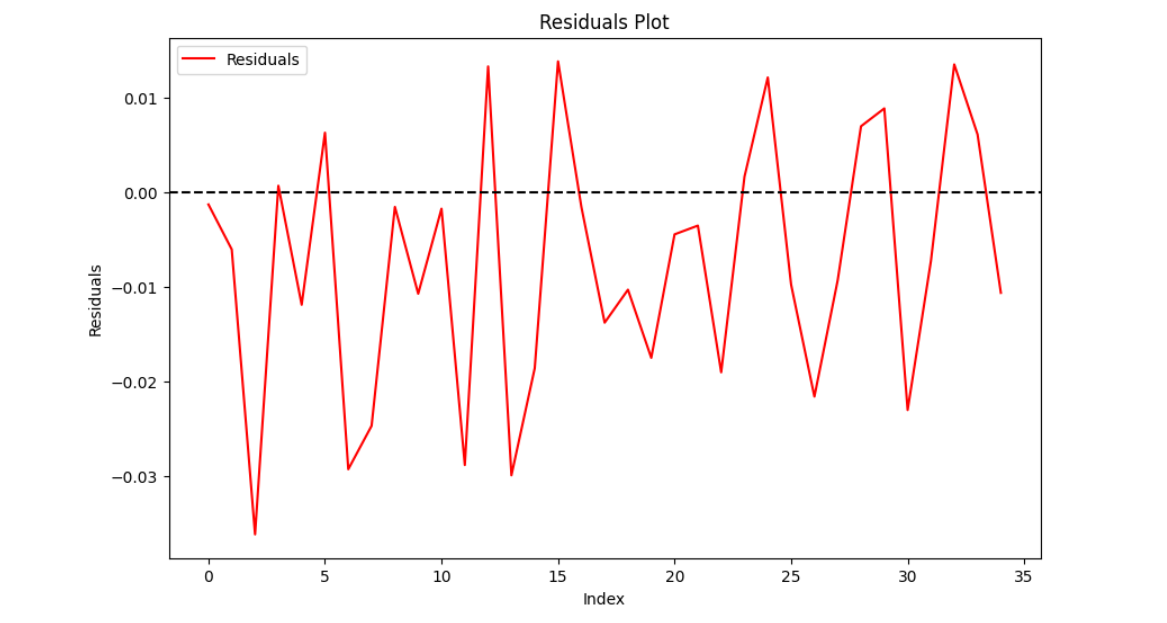
Aim to forecast stock price movements using the ARIMA model.

The dataset used in this analysis is the 'normalize\_diff' column, which represents the normalized difference in stock prices. The dataset is split into training and test sets, with the training set comprising 80% of the data.:

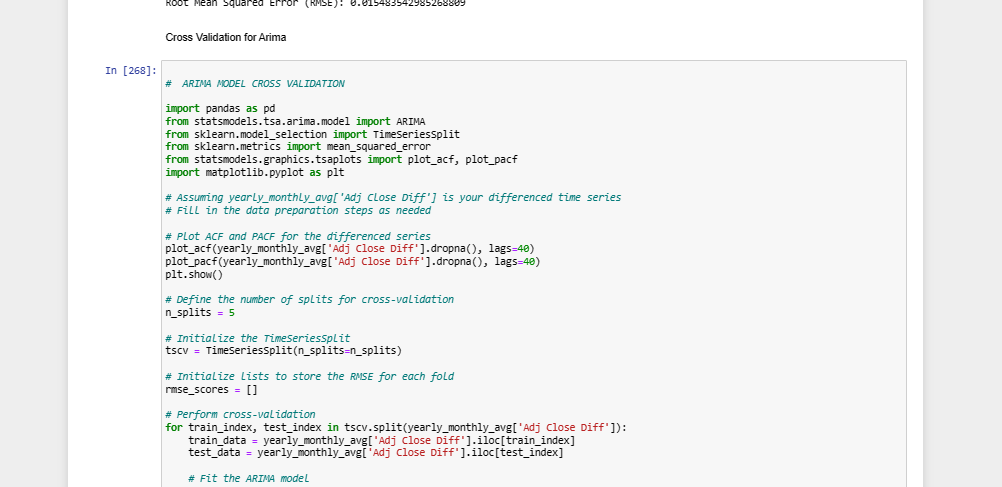
We begin by fitting an ARIMA model to the training data. The ARIMA model is specified with an order of (22, 1, 22), indicating the number of lag observations, differences, and moving average terms to use in the model, respectively.





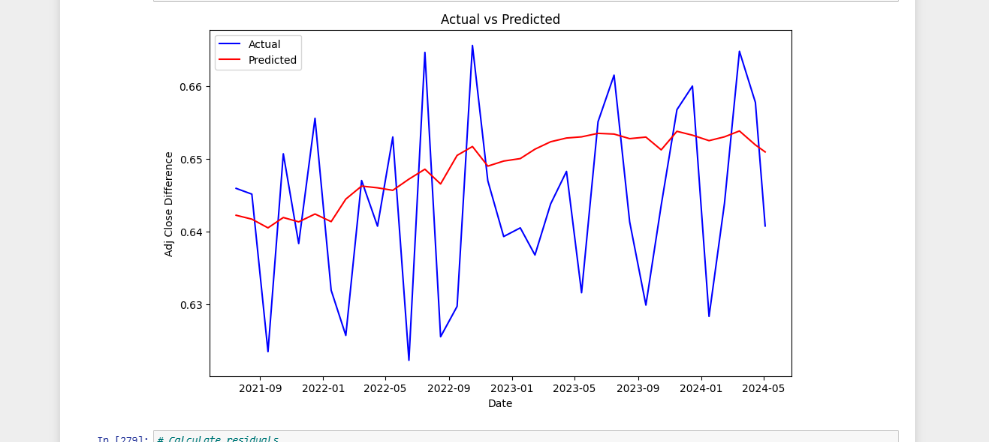


* CROSS VALIDATION OF ARIMA MODEL



* **ANN:**

Use of ANN to predict stock price movements based on historical data. The dataset used in this is the 'normalize\_diff' column representing the normalized difference in stock prices. The dataset is split into training and test sets, with 80% of the data used for training and 20% for testing.

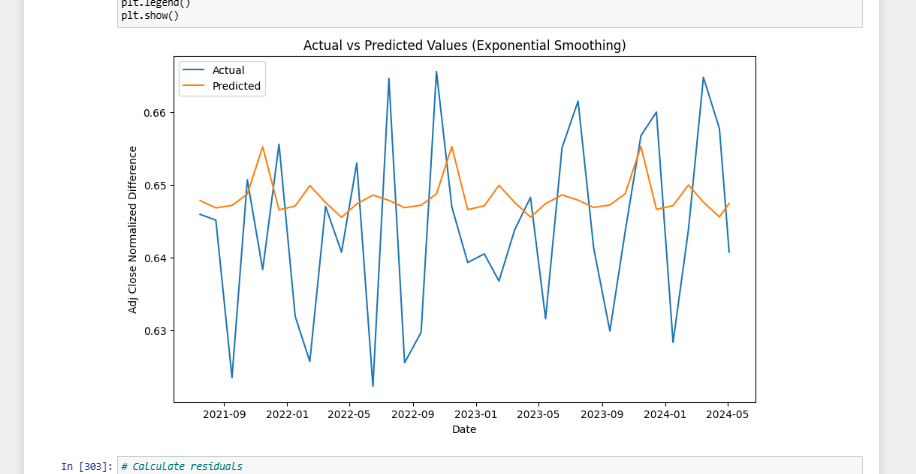


* **ETS:**

Exponential Smoothing is a popular technique for time series forecasting that assigns exponentially decreasing weights to past observations.

The dataset used is the 'normalize\_diff' column representing the normalized difference in stock prices. The dataset is split into training and test sets, with 80% of the data used for training and 20% for testing.

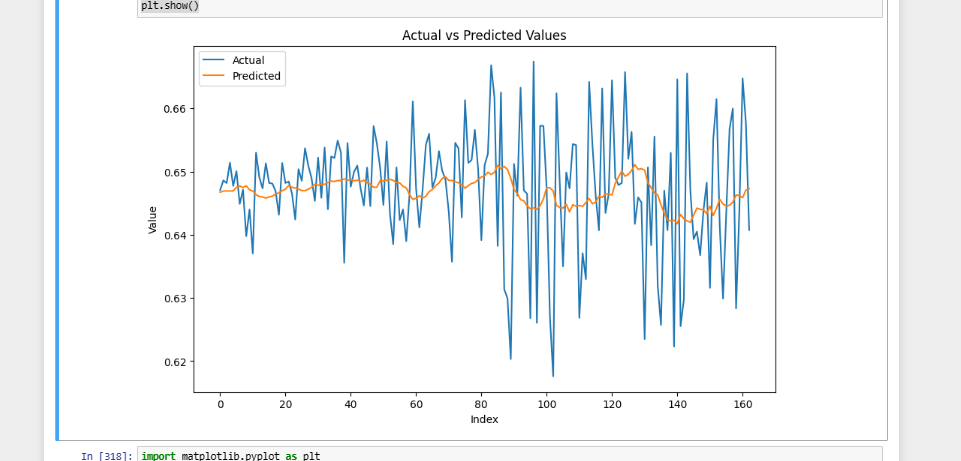
The Exponential Smoothing model is implemented using the Holt-Winters method with additive trend, additive seasonal components, and a seasonal period of 12 months. The model is trained on the training data and then used to make predictions for the next two months



* **LSTM:**

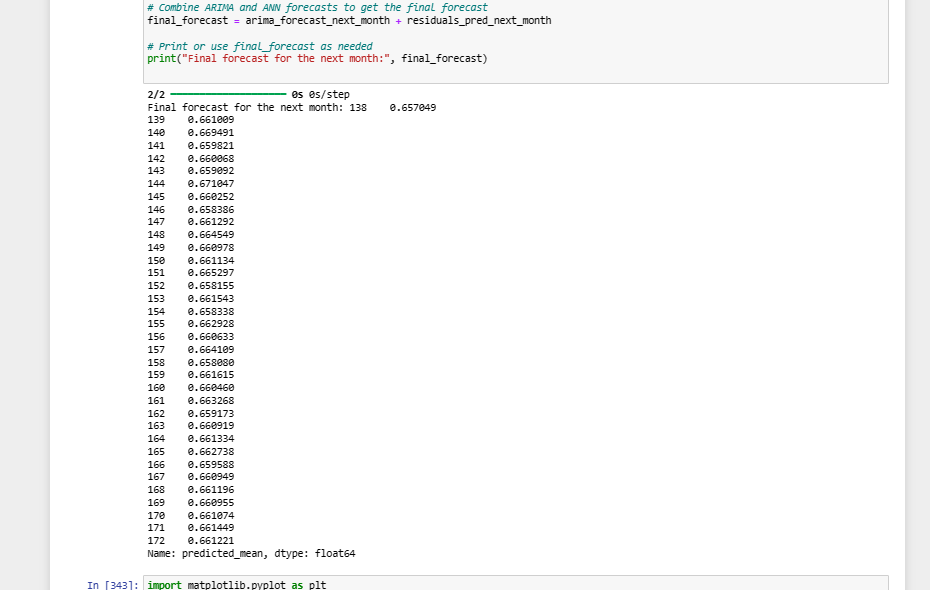
Explores the LSTM neural networks for forecasting stock price movements based on historical data.

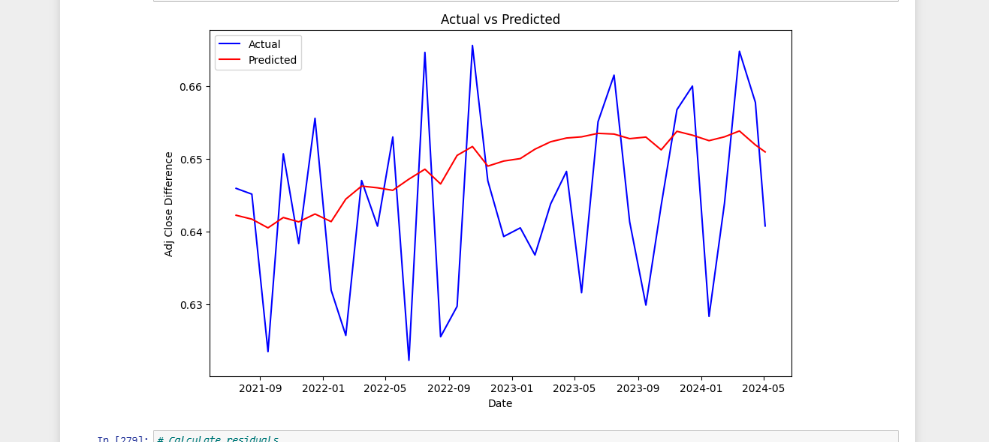
The 'normalize\_diff' column represents the normalized difference in stock prices. The data is preprocessed and scaled using a MinMaxScaler to ensure compatibility with the LSTM MODEL.



* **HYBRID-ANN**

Explores the hybrid-ann ml model for this task…





# 

# **USER INTERFACE DESIGN:**

# **FRONTEND:**

#### **Features**

1. Model Selection: Users can select from a list of machine learning models including ARIMA, ANN, SARIMA, ETS, Prophet, LSTM, and HYBRID ARIMA.
2. Predictions Display: Upon selecting a model, the app retrieves predictions made by that model from the backend and displays them on the web page.

#### **Backend Implementation**

1. Flask Routes: The app defines Flask routes to handle different HTTP requests. For example, the / route serves the main HTML page, /get\_predictions route fetches predictions from the backend, and so on.

#### **Frontend Implementation**

1. HTML Templates: HTML templates are used to structure the frontend of the web application. These templates are rendered dynamically using Flask's templating engine.
2. Styling: CSS stylesheets are applied to enhance the visual appeal of the web pages and ensure a consistent user interface.



THE END!