Stock Prediction for Seritage Growth Properties (SRG) Using LSTM Networks

Introduction

This report presents a project aimed at developing a machine learning model to predict the stock prices of Seritage Growth Properties (SRG). The model uses historical stock price data, financial indicators, and potentially sentiment data from news and social media, aiming to provide investors with a tool to make informed decisions about buying, holding, or selling SRG stocks.

Problem Statement and Context

The primary challenge in predicting stock prices lies in the myriad of factors influencing stock price movements. For a REIT like SRG, these factors range from the company's financial performance and the general health of the real estate market to broader economic indicators. The goal of this project is to create a model that can accurately predict future SRG stock prices and outperform naive benchmark models.

Data Collection and Preprocessing

The project initially uses data collected from Yahoo Finance, but as future models go, it will likely utilize data sources including from APIs such as Alpha Vantage or Yahoo Finance, financial reports, databases like Bloomberg Terminal or FactSet, and sentiment data sources like GDELT for news media and tweepy for social media. The data was preprocessed by checking for missing values, converting the 'Date' column to a datetime type, and creating new features such as lagged prices, moving averages, and returns.

Exploratory Data Analysis

An extensive exploratory data analysis was performed to understand the distribution of the data, visualize the time series of the stock prices, and conduct a correlation analysis to understand the relationships between different features.

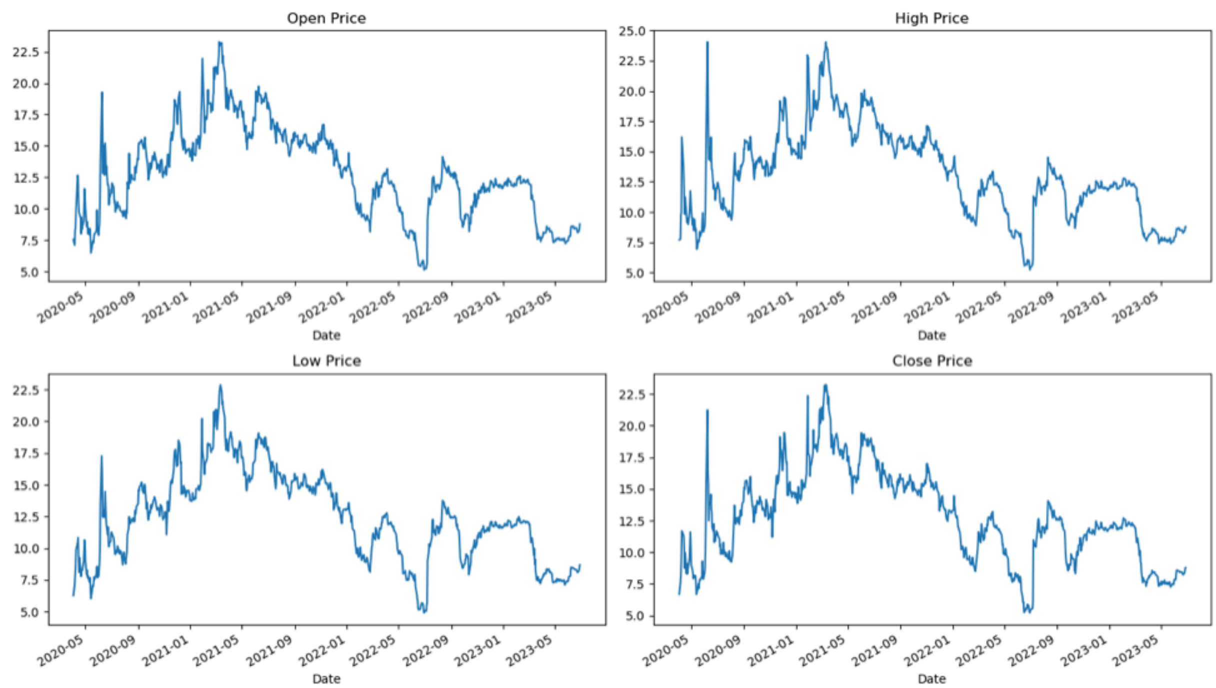
The Exploratory Data Analysis (EDA) phase is a crucial step in any data science project. It involves preparing and exploring the dataset to uncover underlying patterns, trends, and relationships that are critical for subsequent predictive modeling. For this project, EDA was applied to the stock prices of Seritage Growth Properties (SRG), a publicly traded real estate investment trust (REIT), with the goal of conducting an in-depth analysis of the historical stock prices and preparing the data for future time series forecasting.

The EDA phase involved the following steps:

- **Data Inspection**: The dataset was inspected for missing values and appropriate data types. There were no missing values, eliminating the need for imputation or row/column dropping. The Date column, initially an object type, was converted to a datetime type to facilitate time series analysis.
- **Descriptive Statistics**: Descriptive statistics were computed to understand the distribution of the data. For instance, the mean opening price of the SRG stock was

found to be approximately $12.77, with a standard deviation of $3.62, indicating some variability in the opening prices.

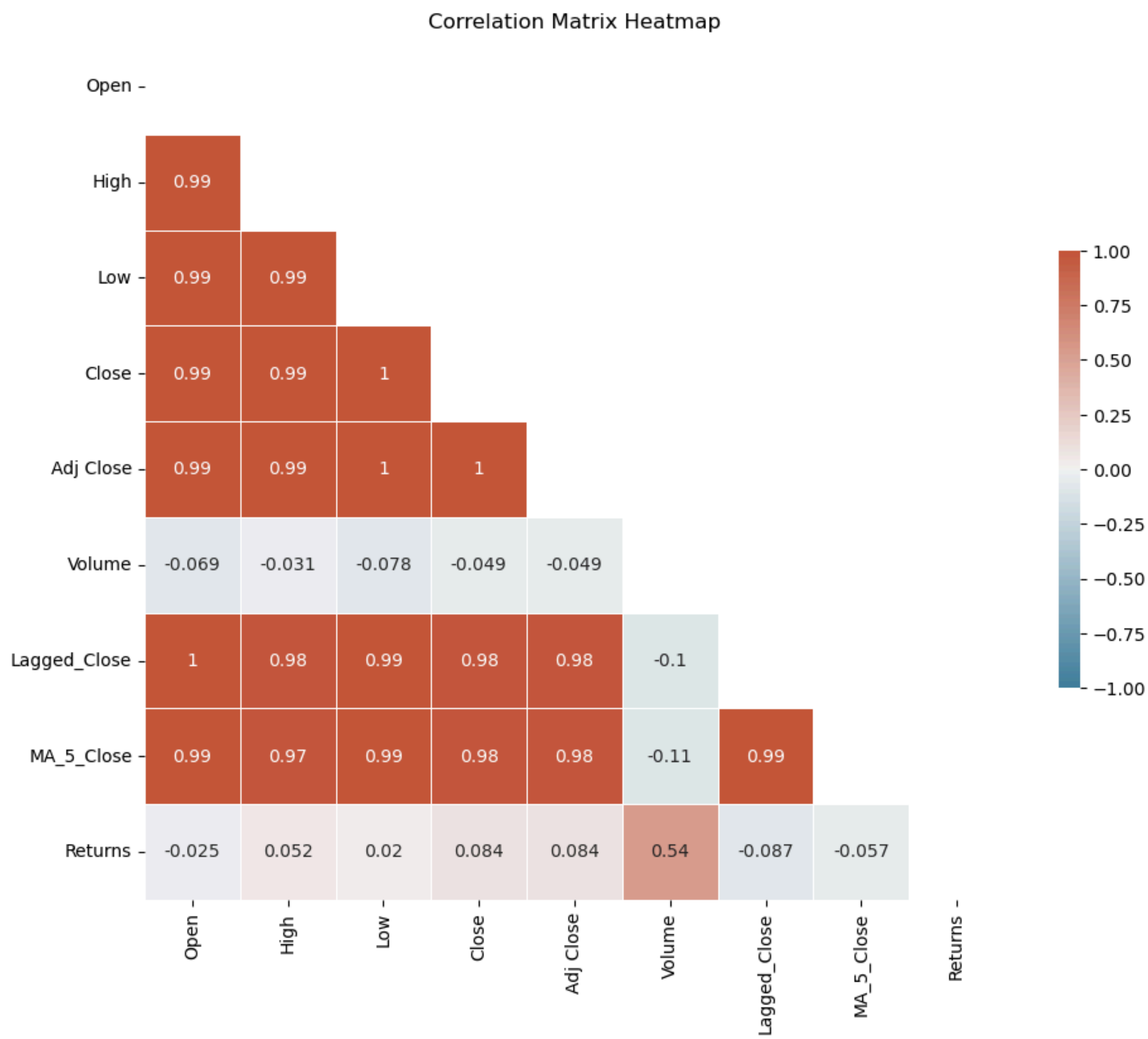
- **Visualization**: Time series plots for the opening, high, low, and closing prices of SRG stocks were created to visualize the trends over time. In addition, a histogram was created to visualize the distribution of the trading volumes, revealing a right-skewed distribution due to occasional periods of high activity.



- **Feature Engineering**: New features such as lagged closing prices, moving averages over the last 5 days, and daily returns (percentage change in closing price from the previous day) were created to capture temporal dependencies and trends in the data.

- **Stationarity Testing**: The Augmented Dickey-Fuller test was used to check for stationarity in the closing prices and returns. The test revealed that while the closing prices were not stationary, the returns were, providing crucial information for the choice of the time series model.

- **Correlation Analysis**: A correlation matrix was generated and visualized as a heatmap to understand the relationships between different features. The analysis revealed high positive correlation amongst stock prices and between them and the engineered features, while lower correlation was observed with trading volumes and returns.



The insights gained from the EDA provided a better understanding of the SRG stock prices and helped prepare the data for the subsequent modeling phase.

Model Building and Evaluation

The choice of the model was influenced by the nature of our data—time series, which inherently consist of sequential observations. Traditional machine learning algorithms may not be the best choice for such data as they treat each data point independently. Hence, a recurrent neural network (RNN) variant, Long Short-Term Memory (LSTM) networks, were used for the task.

Why LSTM?

LSTMs are a special kind of RNN, capable of learning long-term dependencies, making them a suitable choice for time-series predictions such as stock price forecasting. They were developed to deal with the vanishing gradient problem in traditional RNNs, where the model loses information over time. LSTM networks contain a system of 'gates' that control the flow of information to and from memory cells in the network. These gates determine what information gets stored, what gets discarded, and what gets read, thus allowing the model to capture and retain long-term dependencies in the data.

Data Preparation

The 'Close' price column was selected and reshaped for input into the neural network. This data was then normalized using a MinMaxScaler, enhancing the learning capability of the neural network. The normalized data was then split into an 80% training set and a 20% testing set. A function was written to convert an array of values into a dataset matrix optimized for training an LSTM model.

Model Architecture and Training

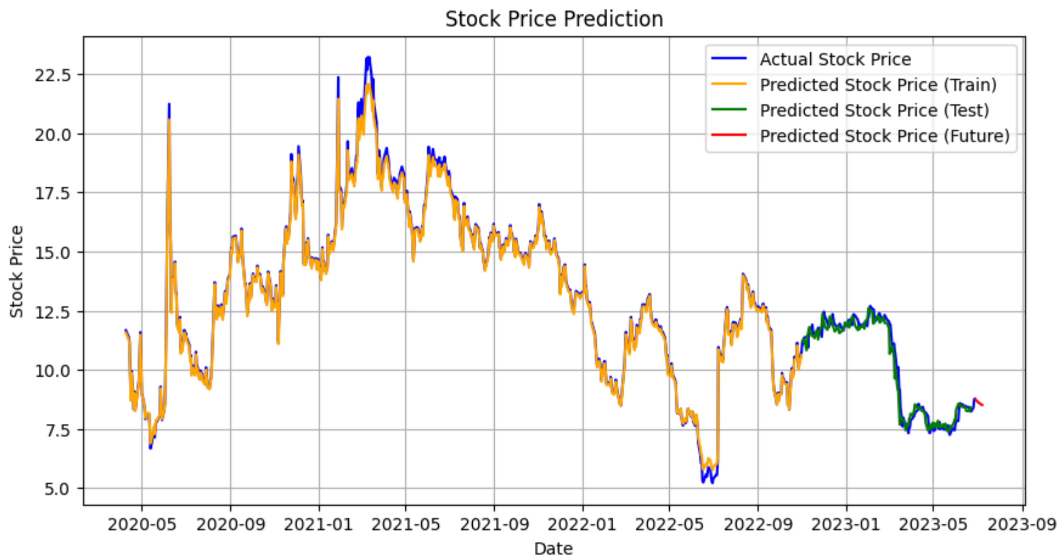
The LSTM model was constructed using the Sequential API from Keras. The model architecture included an LSTM layer due to its ability to remember long-term dependencies. The model used Mean Squared Error as the loss function and Adam as the optimizer. It was then trained on the training data for 100 epochs, where one epoch is when the entire dataset is passed forward and backward through the neural network once.

Model Evaluation

The trained model was used to make predictions on both the training and testing data. These predictions were scaled back to their original range using the inverse_transform function of the scaler. The model's performance was evaluated using the Mean Squared Error (MSE) metric.

Future Predictions

Finally, the trained model was used to make future price predictions using the last 'look_back' days of data. These predictions were made iteratively, with each prediction appended to the input data for the next prediction.



The LSTM model demonstrated high precision in its predictions, suggesting a low level of deviation between the model's predicted values and the actual stock closing prices. However, it's important to consider the inherent unpredictability of the stock market and the performance of the model under different market conditions or with other stocks.

Conclusion and Recommendations

While the LSTM model has shown high accuracy in predicting SRG's stock prices, it's important to note that its performance may vary under different market conditions or with other stocks due to the inherently unpredictable nature of the stock market. To make the model more useful, providing confidence intervals along with point predictions could be beneficial, giving a measure of the potential variability in the predicted prices.

Further Research

Possible areas for further research include exploring other features that could improve the model's accuracy, such as more detailed sentiment analysis or the incorporation of other economic indicators. Testing the model on different stocks and under different market conditions could also provide valuable insights.