Netflix Stock Price Forecasting using Time Series Models

1.Introduction:

This project focuses on forecasting Netflix (NFLX) stock prices by employing a synergistic approach that combines Seasonal Autoregressive Integrated Moving Average (SARIMA) with Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. As a major player in the streaming entertainment industry, accurately predicting Netflix stock prices is crucial for making informed investment decisions. The SARIMA component addresses seasonality and trends, while the GARCH model accounts for volatility, creating a comprehensive framework for nuanced forecasting.

Throughout this project, we will navigate the construction and fine-tuning of the SARIMA-GARCH model, leveraging historical data for validation and assessing real-time predictive accuracy. By offering insights into both systematic patterns and conditional volatility, the SARIMA-GARCH model aims to provide investors, financial analysts, and decision-makers with a reliable tool to anticipate Netflix stock movements and navigate the intricacies of financial markets.

2.Objective:

The major goal of this research is to create a reliable forecasting model for Netflix (NFLX) stock prices using a SARIMA-GARCH technique. We aim to create a comprehensive framework that provides accurate and timely predictions by integrating Seasonal Autoregressive Integrated Moving Average (SARIMA) to capture underlying trends and seasonality and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) to address volatility dynamics. The initiative aims to provide investors, financial analysts, and decision-makers with a strong tool for forecasting Netflix stock movements through painstaking model creation, fine-tuning, and validation against historical data. Finally, we want to help you better comprehend market patterns, make more educated investment decisions, and traverse the volatile terrain of financial markets with confidence.

3.Methodology:

1. Data Collection and Pre-processing:

We have collected the data regarding to Netflix stocks from the Kaggle website, that contains information of the following columns from the time period May, 2002 to June, 2022 –

Open:

The "Open" column represents the opening price of Netflix (NFLX) stock for a particular trading day. This is the price at which the first transaction occurred when the market opened on that day.

Close:

The "Close" column indicates the closing price of Netflix stock for the same trading day. This value represents the final traded price at the end of the market session.

High:

The "High" column records the highest traded price of Netflix stock during the trading day. It reflects the peak value reached by the stock within the given timeframe.

Low:

Conversely, the "Low" column represents the lowest traded price of Netflix stock on a specific trading day. It indicates the minimum value reached by the stock during the trading session.

Adjusted Close:

The "Adjusted Close" column takes into account corporate actions such as dividends, stock splits, or other adjustments that can impact the stock price. This adjusted value provides a more accurate representation of the stock's true market value over time.

Volume:

The "Volume" column denotes the total number of shares of Netflix stock traded during a given trading day. Volume is a key indicator of market activity and liquidity, providing insights into the level of interest and participation in the stock. High volume often accompanies significant price movements and can signal the strength or weakness of a trend.

Upon inspecting the dataset, it was observed that there were missing observations. To address this, the backward filling technique was applied to fill in the gaps within the dataset. Following this, all columns, with the exception of the adjusted closing price, were subsequently removed. The adjusted closing price was singled out as the focus for predicting stock prices in the subsequent analysis.

2.Exploratory Data Analysis (EDA):

Upon exploring the dataset, several key findings were identified:

1. The time series exhibited non-stationarity, as confirmed by the Dickey-Fuller test.
2. To address the non-stationarity, a first difference of the series was plotted, revealing stationarity. The first difference order indicated the presence of a discernible trend in the dataset.
3. Furthermore, the first-order difference highlighted the existence of volatility within the dataset. This volatility, discerned through changes in the first differences, provides insights into the fluctuating nature of the observed values over time.

3. SARIMA Model Development:

Following an examination of the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots of the dataset, we initially embarked on modelling with an ARIMA(7,1,7) configuration. However, after several rounds of fine-tuning the model parameters through trial and error, it became apparent that a more suitable representation for the data was achieved with a SARIMA$(4, 1, 3)x(2, 1, 1)_{12}$ model.

Subsequent to implementing the SARIMA model, a detailed scrutiny of the residuals ensued. Both the residuals and their squared counterparts were meticulously plotted. Interestingly, the squared residuals exhibited a notable degree of stationarity, and the plot unveiled distinctive spikes, signifying the presence of substantial volatility within the dataset. This insightful observation prompted a strategic move to address this volatility by introducing a Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) model, strategically designed to effectively model and encapsulate the observed patterns of volatility in the dataset.

4. GARCH Model Development:

After undergoing multiple iterations of trial and error, a strategic decision was made to employ a GARCH(1,3) model to capture the volatility patterns within the data. This GARCH model was then seamlessly integrated with the previously established SARIMA model, aiming to enhance the overall predictive capacity.

Upon fitting the residuals obtained from the GARCH model, a notable observation surfaced. Specifically, the residuals demonstrated a diminished level of heteroskedasticity, indicating that the GARCH model effectively contributed to stabilizing the volatility patterns within the dataset.

Furthermore, a detailed examination of the distribution of these residuals revealed a distinctive Laplace distribution pattern. This pattern provides valuable insights into the nature of the remaining errors, showcasing a specific distributional form that aligns with the underlying characteristics of the data. Overall, this dual-model approach, combining SARIMA and GARCH, not only addressed volatility concerns but also added a layer of sophistication by acknowledging the unique distributional patterns in the residuals.

6. Model Validation:

The computed Mean Absolute Percentage Error (MAPE) value for the model on the test dataset stands at 0.249. This metric, indicative of the average percentage difference between the predicted and actual values, serves as an evaluation measure for the model's accuracy. The MAPE of 0.249 reflects the model's performance in terms of its ability to forecast with a relatively low average percentage error on the test dataset. A lower MAPE value suggests a more accurate predictive model, and in this context, the obtained result indicates a reasonably favourable level of accuracy in predicting the observed values.

4.Conlusion:

To summarize, our study successfully forecasted Netflix (NFLX) stock values using a combination SARIMA-GARCH modelling technique. Beginning with a dataset investigation, non-stationarity was addressed using SARIMA, while volatility was treated using GARCH. On the test dataset, the integrated model displayed strong predictive ability, with a Mean Absolute Percentage Error (MAPE) of 0.249, indicating impressive accuracy. The approaches used in the experiment demonstrated the effectiveness of capturing both temporal patterns and volatility for accurate stock price predictions. This technique can help stakeholders navigate financial markets by providing insights about Netflix stock movements and contributing to the growing landscape of time series forecasting and financial modelling.