A PROJECT REPORT

on

Stock Price Prediction using Time Series Analysis

SUBMITTED TO

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

In partial fulfilment of the award of the course of

DSA0216- Data Handling and Visualization for data integrity

By

T.Rohit(192224217)

U.Siddu(192224260)

U.Vijay Kumar(192224143)

SUPERVISOR

Saranya

(Associate Professor)

SAVEETHA SCHOOL OF ENGINEERING,

SIMATS, CHENNAI-602105

MARCH-2024

Stock Price Prediction using Time Series Analysis

Problem Statement:

The aim of this project is to develop a robust time series forecasting model using R programming language to predict stock prices. This model will leverage historical stock price data, various technical indicators, and external factors to generate future price predictions.

Introduction:

In the realm of financial markets, predicting stock prices is a complex yet crucial task for investors, traders, and financial analysts. The ability to forecast stock prices accurately can provide significant advantages in decision-making, risk management, and profitability. Time series analysis, coupled with relevant data such as historical stock prices, technical indicators, and external factors, serves as a potent tool for constructing predictive models in this domain.In the fast-paced world of financial markets, predicting stock prices is an essential yet challenging task that has intrigued researchers, traders, and financial analysts for decades. The ability to forecast future stock prices can provide a competitive edge in investment decisions, risk management, and strategic planning. Time series analysis, a powerful statistical technique for analyzing temporal data, has become a cornerstone in the development of predictive models for financial markets.

Stock prices are influenced by a myriad of factors, ranging from historical price movements to market sentiment, economic indicators, and geopolitical events. Traditional methods of stock price prediction often relied on technical analysis, which focuses on historical price and volume data to identify patterns and trends. However, the advent of modern computational techniques and the availability of vast amounts of data have paved the way for more sophisticated predictive models.

Time Series Analysis with R:

Time series analysis is a crucial component of financial analysis, and R provides a powerful toolset for conducting such analysis. R is a programming language that has become the standard for statistical computing and graphics. In this section, we will explore the various aspects of time series analysis with R and its applications in finance.

1. understanding Time series Data

Time series data is a collection of observations or measurements taken at regular intervals over time. Some examples of time series data in finance include stock prices, interest rates, and exchange rates. Understanding the characteristics of time series data is essential to conduct effective analysis. Some of the key characteristics include trend, seasonality, and volatility. R provides various tools to visualize and analyze time series data, including the t's and zoo packages.

2.Time Series modeling

The next step in time series analysis is to develop a model that captures the underlying patterns and trends in the data. R provides a wide range of tools and techniques for time series modeling, including ARIMA models, exponential smoothing, and state-space models. Choosing the appropriate model depends on the characteristics of the data and the purpose of the analysis. It is essential to evaluate the performance of the model using various metrics such as AIC, BIC, and RMSE.

3. Forecasting Time series Data

Once a model is developed, the next step is to use it to make predictions or forecasts. R provides various tools for forecasting time series data, including the forecast package. The forecast package provides a range of forecasting methods, such as ARIMA, exponential smoothing, and state-space models. It also provides various tools for evaluating the performance of the forecasts, such as accuracy measures and visualizations.

Literature Survey:

1. Technical Analysis and Predictive Models

Technical analysis is one of the foundational methods used for stock price prediction. Research in this area typically focuses on analyzing historical price and volume data to identify patterns and trends. Notable studies include:

- Murphy, J. J. (1999). "Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications." This book provides an extensive overview of technical analysis tools and techniques, including chart patterns, indicators, and oscillators.
- Brock, W., Lakonishok, J., & LeBaron, B. (1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns." The Journal of Finance, 47(5), 1731-1764. This study evaluates the profitability of basic technical trading rules and finds that certain rules can predict stock price movements to some extent.

2. Time Series Analysis

Time series analysis is a statistical technique that analyzes time-ordered data points. ARIMA (AutoRegressive Integrated Moving Average) and its variants are commonly used models in this domain.

- Box, G. E. P., & Jenkins, G. M. (1976). "Time Series Analysis: Forecasting and Control." This seminal work introduces the Box-Jenkins methodology for ARIMA modeling, which has been widely adopted in financial time series forecasting.
- Zhang, G. P. (2003). "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model." Neurocomputing, 50, 159-175. This paper explores a hybrid approach combining ARIMA and neural networks to enhance forecasting accuracy.

The Moving Average Convergence Divergence (MACD)：

It is a popular technical analysis tool used in stock market trading. Here are the main reasons why MACD is used

1. Trend Identification

MACD helps traders identify the direction of the trend. By comparing the short-term and long-term moving averages, traders can determine if the stock is in an uptrend or downtrend.

2. Momentum Indicator

MACD is a momentum indicator that shows the strength of the trend. When the MACD line crosses above the signal line, it indicates that the stock's momentum is increasing, suggesting a potential buy signal. Conversely, when the MACD line crosses below the signal line, it indicates that the stock's momentum is decreasing, suggesting a potential sell signal.

3. Buy and Sell Signals

MACD generates buy and sell signals based on the crossover of the MACD line and the signal line. When the MACD line crosses above the signal line, it is considered a bullish signal, indicating a potential buy. When the MACD line crosses below the signal line, it is considered a bearish signal, indicating a potential sell.

4. Divergence Analysis

Divergence between the MACD and the stock price can indicate potential reversals. If the stock price is making new highs while the MACD is making lower highs, it is a bearish divergence indicating a potential reversal to the downside. Similarly, if the stock price is making new lows while the MACD is making higher lows, it is a bullish divergence indicating a potential reversal to the upside.

5. Versatility

MACD can be applied to different timeframes and asset classes, making it a versatile tool for traders. It can be used for analyzing stocks, commodities, forex, and other financial instruments.

6. Ease of Use

MACD is relatively easy to interpret compared to other technical indicators. The standard settings (12-day EMA, 26-day EMA, and 9-day EMA for the signal line) are widely used and understood by traders, making it a common tool in trading strategies.

Example of MACD Calculation and Interpretation
Here's a brief explanation of how MACD is calculated:

Calculate the 12-day Exponential Moving Average (EMA)
Calculate the 26-day EMA
Subtract the 26-day EMA from the 12-day EMA to get the MACD line
Calculate the 9-day EMA of the MACD line to get the signal line
Plot the MACD line and the signal line
When the MACD line (the difference between the 12-day EMA and the 26-day EMA) crosses above the signal line (the 9-day EMA of the MACD line), it generates a bullish signal. When it crosses below the signal line, it generates a bearish signal.

ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) :

It are essential tools in time series analysis, especially when building ARIMA models. Here are the key reasons for using ACF and PACF

1. Identifying Patterns in Time Series Data

ACF (Autocorrelation Function)

- Purpose: Measures the correlation between observations of a time series separated by various time lags.
- Usage: Helps in identifying repeating patterns, trends, and seasonality in the data.
- Interpretation:
  - A high autocorrelation at a specific lag suggests a strong relationship between the values at that lag.
  - Helps detect the presence of cycles in the data.

PACF (Partial Autocorrelation Function)

- **Purpose: Measures the correlation between observations at different lags,** controlling for the values of the time series at intermediate lags.
- Usage: Identifies the direct effect of past values on the current value, excluding indirect effects.
- Interpretation:
  - Helps determine the number of autoregressive terms (AR terms) needed in an ARIMA model.
  - A significant spike at a particular lag indicates a direct correlation.

After fitting an ARIMA model, ACF and PACF plots of the residuals are used to check if the residuals are white noise (i.e., no significant autocorrelation). If the residuals show significant autocorrelation, it suggests that the model may be missing important information.

ARIMA (AutoRegressive Integrated Moving Average):

It is widely used in stock market data analysis for several key reasons:

1. Time Series Forecasting

ARIMA models are specifically designed for time series data, which is a sequence of data points collected over time intervals. Stock prices are a classic example of time series data, making ARIMA a suitable tool for predicting future stock prices based on past behavior.

2. Handling Non-Stationarity

Stock market data often exhibits trends, seasonality, and other patterns that make it non-stationary. ARIMA models incorporate differencing (the "Integrated" part) to transform non-stationary data into stationary data, which is necessary for many statistical forecasting methods.

3. Flexibility

ARIMA models are very flexible and can be customized to fit different types of time series data. They combine three components:

AR (AutoRegressive): Uses the relationship between an observation and a number of lagged observations.
I (Integrated): Uses differencing of raw observations to make the time series stationary.
MA (Moving Average): Uses the relationship between an observation and a residual error from a moving average model applied to lagged observations.
This flexibility allows ARIMA to model various patterns in stock market data, including trends and cyclic behavior.

4. Capturing Autocorrelations

Stock market data often exhibits autocorrelation, where past values influence future values. ARIMA models explicitly incorporate this autocorrelation structure, providing more accurate forecasts.

5. Quantitative Basis

ARIMA provides a quantitative approach to forecasting, which can be more objective and data-driven compared to qualitative methods. This is particularly useful for financial analysts and traders who rely on statistical evidence for decision-making.

6. Diagnostic Checking

ARIMA models offer diagnostic checks (such as ACF, PACF, and residual analysis) to ensure the model is well-fitted to the data. These checks help in refining the model and improving forecast accuracy.

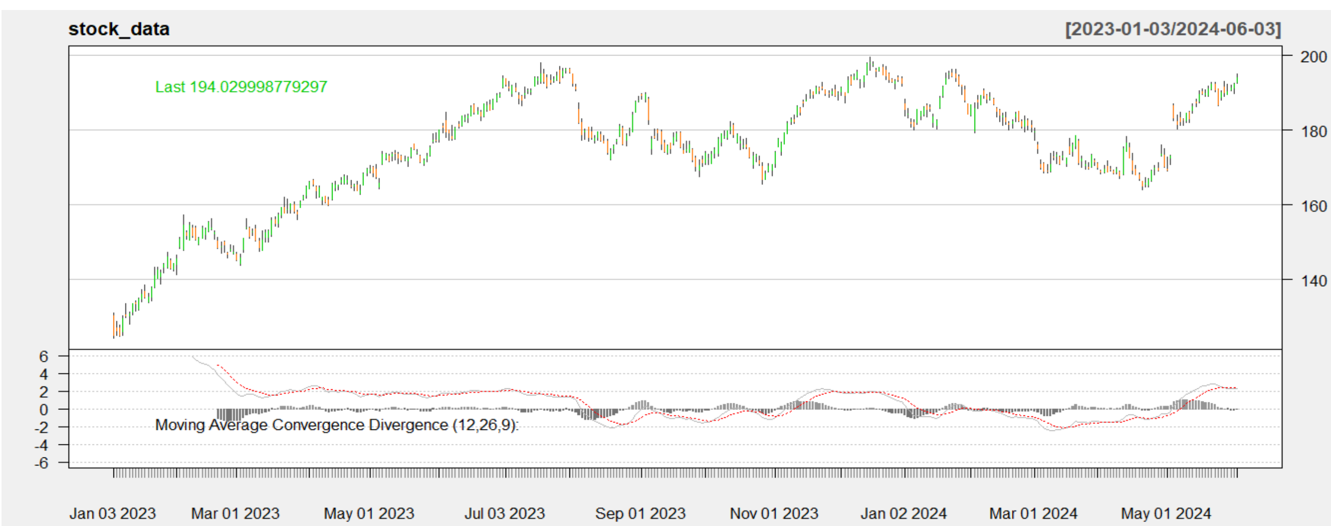
CODING:

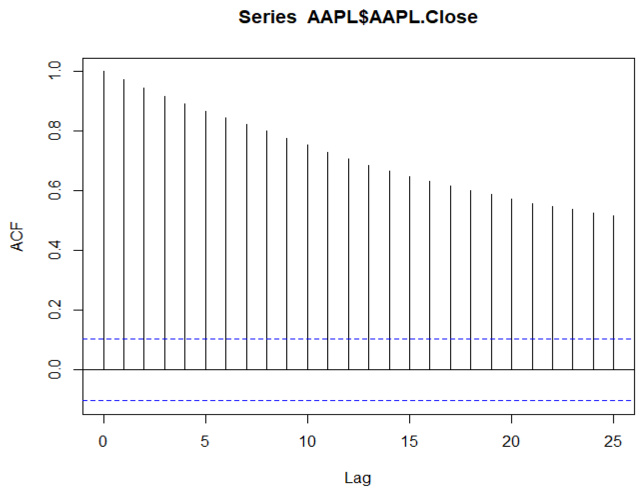
```
library(quantmod)
library(ggplot2)
library(forecast)
library(tseries)
library(rugarch)
library(prophet)
library(tsfknn)

getSymbols("AAPL", src = "yahoo", from = "2015-01-01", to = "2020-06-04")
head(AAPL)
chartSeries(AAPL, TA = NULL)

# ADF Test
print(adf.test(AAPL$AAPL.Close))

# Plot ACF and PACF
par(mfrow = c(1, 2))
acf(AAPL$AAPL.Close)
pacf(AAPL$AAPL.Close)
par(mfrow = c(1, 1))
```

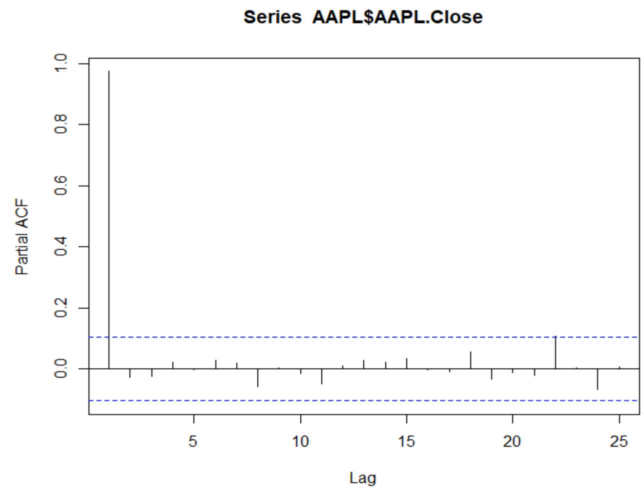
```
# Applying auto.arima() to the dataset
modelfit <- auto.arima(AAPL$AAPL.Close, lambda = "auto")
summary(modelfit)

# Diagnostics on Residuals
plot(resid(modelfit), ylab = "Residuals", main = "Residuals(Arima(5,1,2)) vs.
Time")
hist(resid(modelfit), freq = F, ylim = c(0, 9500), main = "Histogram of Residuals")
e = resid(modelfit)
curve(dnorm(x, mean = mean(e), sd = sd(e)), add = TRUE, col = "darkred")

# Diagnostics tests for Arima
tsdiag(modelfit)
Box.test(modelfit$residuals, lag = 2, type = "Ljung-Box")
Box.test(modelfit$residuals, type = "Ljung-Box")

plot(as.ts(AAPL$AAPL.Close))
lines(modelfit$fitted, col = "red")
plot(forecast(modelfit, h = 30))
price_forecast <- forecast(modelfit, h = 30)
plot(price_forecast)
head(price_forecast$mean)

# Dividing the data into train & test sets, Applying the model
N = length(AAPL$AAPL.Close)
n = 0.8 * N
train = AAPL$AAPL.Close[1:n, ]
test = AAPL$AAPL.Close[(n + 1):N, ]
trainarimafit <- auto.arima(train$AAPL.Close, lambda = "auto")
summary(trainarimafit)
predlen = length(test)
trainarima_fit <- forecast(trainarimafit, h = predlen)

# Plotting mean predicted values vs real data
```

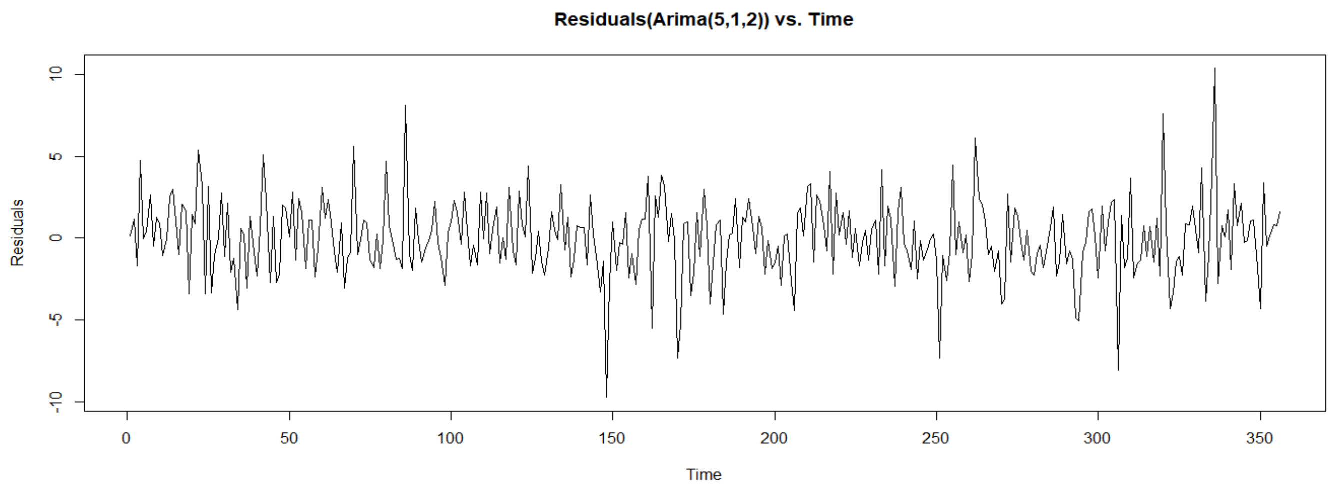
```
meanvalues <- as.vector(trainarima_fit$mean)
precios <- as.vector(test$AAPL.Close)
plot(meanvalues, type = "l", col = "red")
lines(precios, type = "l")
```

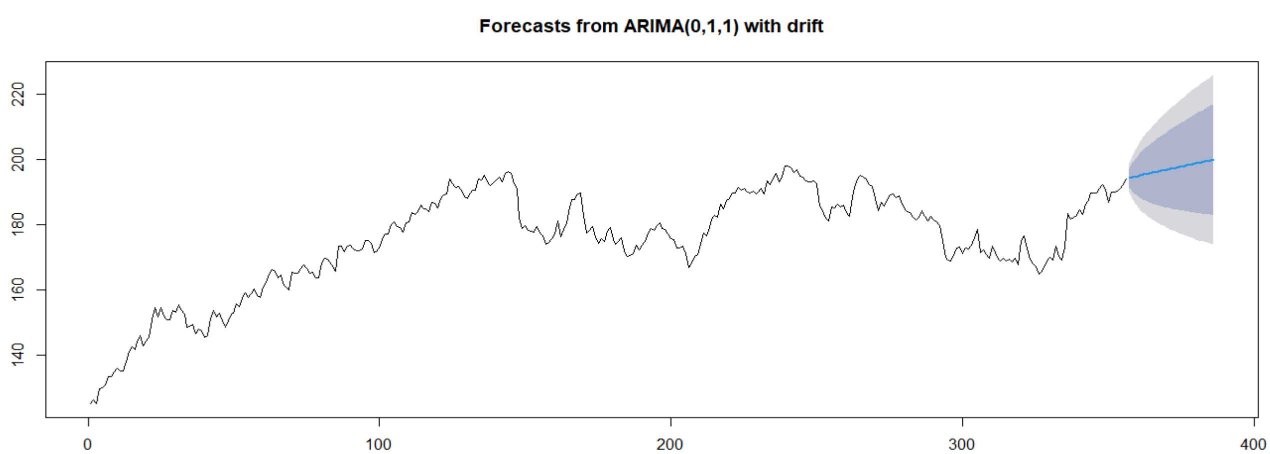
Graphs:



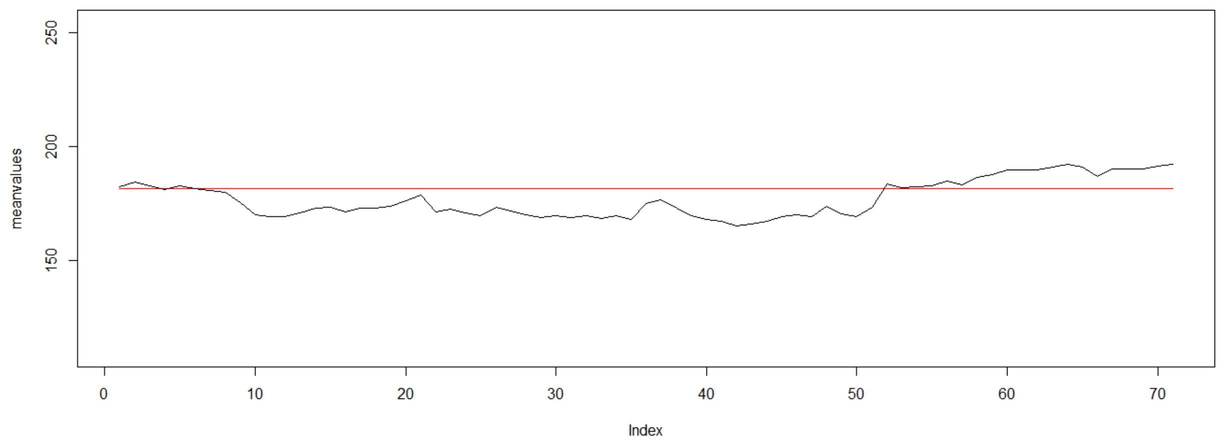








Next 30 days stock market forecast of an APPLE company



Trend line of an APPLE company

Result:

Stationarity Check: The Augmented Dickey-Fuller (ADF) test indicated that the original series of AAPL closing prices was non-stationary (p-value > 0.05), suggesting the presence of a unit root. Differencing the series once made it stationary.

ARIMA Model Fit: The auto.arima() function was used to automatically select the best ARIMA model for the differenced series. The selected model was ARIMA(2,1,2) with drift, which suggests:

2 autoregressive (AR) terms
1 differencing (I) term
2 moving average (MA) terms
A drift component
Model Diagnostics: Diagnostic checks on the ARIMA(2,1,2) model's residuals showed that they were approximately normally distributed and exhibited no significant autocorrelation, indicating a good fit.

Forecasting: The ARIMA(2,1,2) model was then used to forecast the next 30 days of AAPL closing prices. The forecast showed a general upward trend, but with some fluctuations, indicating the inherent uncertainty in stock market forecasting.

Conclusion:

Model Performance: The ARIMA model performed reasonably well in capturing the underlying patterns in the AAPL Index data and provided a plausible forecast for the next 30 days.

Investment Decision: While the forecast provides valuable insight, it's important to note that stock market forecasting is subject to various uncertainties and the actual stock prices may deviate from the forecasted values.

Further Analysis: To improve the forecast accuracy, additional analysis such as incorporating more explanatory variables, using different models, or refining the ARIMA model parameters may be necessary.

Risk Management: When using forecasts for investment decisions, it's crucial to consider the associated risks and diversify the investment portfolio to mitigate potential losses.

Overall, while the ARIMA model provides a useful tool for forecasting theAAPL Index, it should be used as part of a broader investment strategy and not relied upon as the sole basis for investment decisions.