**FORECASTING FUTURE PRICES OF NIFTY FIFTY STOCKS USING TIME SERIES MODELS**

* ***A guide to stock market analysis***

**M.Sc. Big Data Analytics**

**Semester -1**

**BP1P2 – Project Report**

**Submitted By - Team Stockers**

Alexei Beuno H - 21BDA01

Rashmi S - 21BDA02

Sneha T - 21BDA30

Shivanshi Maheshwari - 21BDA33

Murrel Miranda - 21BDA36

Shrimanta Satpati - 21BDA45

**ABSTRACT**

Stock market also called as the equity market is the aggregation of sellers and buyers. It is concerned with the domain where the shares of various public listed companies are traded. For predicting the growth of economy, stock market acts as an index. Due to the nonlinear nature, the prediction of the stock market becomes a difficult task. But the application of various time series models has been becoming a powerful source for the prediction. These techniques employ historical data of the stocks for the training of time series models and help in predicting their future behavior.

**PROBLEM STATEMENT**

1. To forecast closing prices of 30 companies under nifty fifty stocks for about 90 days.
2. To understand trends and patterns underlying in the historical data of each of the 30 companies.

Consider a situation wherein SRK wants to sponsor the IPL team Kolkata Knight Riders. He is in short of funding and he would like to invest in a company which has potential growth in order to gain enough returns. Which company should he invest in?

**SOURCE OF DATA**

We have obtained the datasets from yahoo.finance.com. We use real time data that shows the updated price or volume information on a tick-by-tick or minute-by-minute basis for stocks that trade on the exchange.

The companies listed under Nifty-fifty stocks that we are using for analysis are as given below:

1. TCS (Tata Consultancy Services Limited)
2. BAJAJFINSV (Bajaj Finserv Limited)
3. WIPRO (Wipro Limited)
4. TITAN (Titan Company Limited)
5. ONGC (Oil and Natural Gas Corporation Limited)
6. INFY (Infosys Limited)
7. CIPLA (Cipla Limited)
8. HCLTECH (HCL Technologies Limited)
9. RELIANCE (Reliance Industries Limited)
10. TECHM (Tech Mahindra Limited)
11. M&M (Mahindra and Mahindra Limited)
12. BAJAJ-AUTO (Bajaj Auto Limited)
13. HEROMOTOCO (Hero MotoCorp Limited)
14. EICHERMOT (Eicher Motors Limited)
15. MARUTI (Maruti Suzuki India Limited)
16. LT (Larsen & Toubro)
17. ICICIBANK (Industrial Credit and Investment Corporation of India)
18. AXISBANK (Axis Bank Limited)
19. KOTAKBANK (Kotak Mahindra Bank Limited)
20. INDUSINDBK (INDUSIND Bank Limited)
21. BAJFINANCE (Bajaj Finance Limited)
22. HINDUNILVR (Hindustan Unilever Limited )
23. NESTLEIND (Nestle India Limited)
24. PIDILITIND (Pidilite Industries Limited)
25. BERGEPAINT (Berger Paints India Limited)
26. DABUR (Dabur India Limited)
27. COLPAL (Colgate-Palmolive India Limited)
28. GODREJCP (Godrej Customer Products Limited)
29. HAVELLS (Havells India Limited)
30. AAPl (Apple)

**METHODOLOGY**

The simplest approach to understand the stock market time series data that we have is to do so by data visualization.

Data Visualization and Technical Analysis:

1. A simple time series plot:

Since we are analyzing close prices of the companies, this particular graphical representation is the plot of closing prices versus time. We can visualize dips and peaks of the closing prices of all the 30 companies over the past 10 years from this plot.

1. MACD (Moving Average Convergence Divergence):

Moving average convergence divergence is a trend following momentum indicator that shows the relationship between two moving averages of a security’s price.​ The MACD is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period exponential moving average (EMA).​ MACD is often displayed with a histogram which graphs the distance between the MACD and its signal line.

1. Decomposition of Time series Plot:

Decomposition of a time series might lead to either a multiplicative time series or an additive time series. The various components of the time series are trend, cyclic, random and seasonal components. The additive model assumes that all the four components of the time series operate independently of each other so that none of these components has any effect on the remaining three components. Whereas, in a multiplicative time series all the components operate proportionately to the general level of the series.

1. Candle Stick Chart:

Candlesticks are useful while trading as they show four price points (open, close, high, and low) throughout the period of time the trader specifies. ​Let us consider a weekly candlestick chart. Then, open will be Monday’s first trade in and close will be Friday’s last trade in before the market closes for the week. High and low will be high and low of the week respectively.

1. Golden Crosses:

A golden cross is a technical chart pattern indicating the potential for a major breakout. ​

The golden cross appears on a chart when a stock's short-term MA crosses above it's long-term MA.​

1. Ichimoku Cloud:

In Japanese, Ichimoku translates to one look. The traders just have to take a single look at the chart to determine how the price varies.​ There are five indicators in total i.e Tenkan, Kijun, Senkou A, Senkou B and Chiku Span.​

**The Conversion Line (Tenkan-Sen)** – can be perceived as the short-term line and it represents the average of the high and the low for the 9-period (9-period high + 9-period low / 2).

**The Base Line (Kijun-Sen)** – is the long-term line and it is calculated as the average of the high and low for the 26-period (26-period high + 26-period low / 2).

**The Lagging Span (Chickou Span)** – is a lagging line which represents the closing price for the previous 26 periods. This line enables you to easily compare the current price movements with the movements from 26 periods ago.

**The Leading Span A (Senkou Span A)** – is used as a leading indicator defined for future 26 periods. The values for this indicator are obtained as the middle point of Tenkan-sen and Kijun-sen based on the past last 26 periods (Conversion Line + Base Line) / 2).

**The Leading Span B (Senkou Span B)** – also used as a leading indicator since it is calculated for 26 periods ahead and it is based on the average of the 52 periods high and 52 periods low (52-period high – 52-period low / 2).

1. Relative Strength Index:

The Relative Strength Index is a momentum indicator that evaluates overbought or oversold conditions by measuring the magnitude of recent price changes for various assets. When the RSI surpasses the horizontal 30 reference level, it is a bullish sign and when it slides below the horizontal 70 reference level, it is a bearish sign.​

1. Bollinger Bands:

A Bollinger Band is a technical analysis tool defined by a set of trendlines plotted two standard deviations (positively and negatively) away from a simple moving average (SMA) of a security's price, but which can be adjusted to user preferences.​ When stock prices continually touch the upper Bollinger Band, the prices are thought to be overbought; conversely, when they continually touch the lower band, prices are thought to be oversold, triggering a buy signal.

Methods used for forecasting close prices of a company:

1. ARIMA Model (Auto Regressive Integrated Moving Average Model):

ARIMA model is a time series forecasting tool which makes use of auto regression and moving average methods along with a differencing parameter to handle seasonality complex to make forecasts. ARIMA models are generally denoted by ARIMA(p,d,q) where p, d, and q are the parameters with positive values.

AutoArima is a library in R that we use to find the order of p, d and f. Instead we can find the order of p and q using the ACF and PACF plots manually.

One of the underlying assumptions of the ARIMA model is the fact that the time series we have considered is stationary. Stationarity can be ensured using the Augmented Dickey Fuller(ADF) test. If the p-value is > 0.05, then the time series is not stationary and the differencing parameter d comes into play to make it stationary. ​

1. Prophet Model:

Prophet model is a time series forecasting tool developed by Facebook.​ The Prophet model uses a decomposable time series model with three model components being trend, seasonality and holidays.​ The three components are combined in the following equation​.

                                 y(t) = g(t) + s(t) + h(t) + e​

Here g(t) is the trend function which models non-periodic changes, s(t) represents periodic changes(daily, weekly, yearly,) and h(t) represents the holiday effects.​

Test to check accuracy of our forecasts:

1. Box Ljung Test:

This test is a way to test for absence of serial autocorrelation between the residuals up to a specified lag 'k'. Essentially Box Ljung test is a test to check lack of fit.​ We can conclude the results of the test by looking at the p-value. If p-value is greater than 0.05 then the time series is not auto correlated and the model shows good fit.

**DEVELOPMENT OF USER INTERFACE**

**SUMMARY**

**FUTURE SCOPE:**

Few potential improvements of our project can be:

* It takes time for the graphs to load – This can be resolved using various optimization methods.​
* Clubbing multiple graphs together – Example: Candle-Stick Chart​
* The front end part of the app can be more attractive – Using JavaScript components.​
* Add an additional news section for news grabbers.
* Add a tab separately to read about the company profiles.

**REFERENCES**

1. Datasets are obtained from,

<https://finance.yahoo.com/>

1. <https://fundsbase.com/importance-of-stock-market-in-the-economy/#:~:text=1%20It%20helps%20in%20the%20growth%20of%20company,leads%20to%20a%20better%20economy.%20More%20items...%20>
2. <https://www.investopedia.com/terms/i/ichimoku-cloud.asp>
3. <https://www.investopedia.com/terms/g/goldencross.asp#:~:text=A%20golden%20cross%20is%20a%20technical%20chart%20pattern,a%20death%20cross%20indicating%20a%20bearish%20price%20movement>
4. <https://www.investopedia.com/trading/candlestick-charting-what-is-it/>
5. <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>
6. <https://www.investopedia.com/terms/a/annual-return.asp>
7. <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc442.htm>
8. <https://www.statisticshowto.com/ljung-box-test/>