Final Exam: Stock market forecasting using Time Series analysis.

Ahmed Shmels Muhe
M.Tech In Data Science
Indian Institute of Technology Madras, IIT Madras
Chennai, India
ge22m009@smail.iitm.ac.in

***Abstract*—In this study, we examine data from various industries, including banking and information technology, to glean insights from the data. Using the concepts of Time Series, we comprehensively obtain insights and analyse different stocks in a large amount of data. We then try to forecast our predictions that could help analyse an investment's risks and benefits. To comprehend the behaviour of the stock market, we employ specific measures and graphs from the financial world. We model stock market data sets using this information. Then we look at each stock's distinct behaviour, individually and across sectors. Due to the complex nature and the abstractness involved in Stock Market, it is often necessary to forego a detailed analysis before investments to make informed choices.**

***Index Terms*—stock market, data, Time Series**

I. INTRODUCTION

A stock market, also known as an equity market or a share market [1], is a gathering of buyers and sellers of stocks (also known as shares) representing ownership claims on businesses. Stores can be listed on a public stock exchange or privately traded, such as shares of private companies sold to investors through equity crowdfunding platforms.

The stock market allows numerous buyers and sellers of securities to meet, interact, and transact. Stock analysis is essential for investors and traders to make buying and selling decisions. By studying and evaluating past and current data, one could attempt to gain an edge in the markets by making informed decisions [1]. Stocks are classified according to the country in which the company is based.One97 Communications (Paytm) and Zomato, for example, are based in India and trade on the NSE and BSE; thus, they can be regarded as part of the Indian stock market; however, the stocks can also be traded on other exchanges, such as American depositary receipts (ADRs) on U.S. stock exchanges.

In the 16th and 17th centuries, the first stock markets arose in Europe, mainly in port cities or trading hubs such as Antwerp, Amsterdam, and London. Due to the tiny number of corporations that issued equity, these early stock markets were more analogous to bond exchanges [2].

This paper will study Time Series Analysis, a technique used to analyze data where the sequence is essential, usually data involving date and time. We will use this technique to forecast a stock of interest that was interpreted by weighing risks and returns. This analysis would aid in obtaining insights among a large dataset and hopefully uncover the trend behind a stock of interest.

II. TIME SERIES ANALYSIS AND FINANCIAL TOOLS

A time series is a sequence of data points occurring in successive order over time. Since Time series data have a natural temporal ordering, this can be contrasted with cross-sectional data, which captures a point in time and has no natural sequence of the observations [3]. In particular, a time series allows one to see what factors influence certain variables from period to period. Time series analysis can be helpful to know how a given asset, security, or economic variable changes over time. This helps in judging the future in the context of its past performance.

This section explains the fundamental mathematical and financial methods used to analyse stock market data. Because investing requires a significant amount of wealth, it is critical to grasp the fundamentals before entering the battlefield.

A. Stock market data jargon

- Open: The opening price of a stock is the price at which the stock first trades in the market. It is the initial price it is exchanged for a specific time frame. In the stock chart, it is symbolised by the letter "O".
- High: This is the stock's highest price during the day since trading began. In the stock chart, it is denoted by the letter "H."
- Low: This is the stock's lowest price during the day since trading began. In the stock chart, it is denoted by the letter "L."
- Close: The closing price of a stock is the final price at which it is traded during the trading session. It is the price at which a supply is closed for a specific period. In the stock chart, it is denoted by the letter "C."
- Volume: The quantity of an asset or security that changes hands over a specific period, usually a day, is referred to as volume. Stock trading volume, for example, refers to the number of shares of securities traded between its daily open and close.
- Adjusted Close: An adjusted closing price is a stock's closing price that has been changed to reflect its worth after accounting for any corporate activities. It is frequently

employed for analysing historical returns or performing a comprehensive examination of prior performance.

B. Simple Moving Average

With time series data, a moving average is widely used to smooth out short-term swings and emphasise longer-term trends or cycles. The application determines the short-term and long-term threshold, and the moving average parameters are adjusted accordingly. It is frequently employed in the technical analysis of financial data, such as stock prices, returns, or trade volumes.

It is also used in economics to study GDP, employment, and other macroeconomic time series [4]. A moving average is a sort of convolution in mathematics, and as such, it may be considered an example of a low-pass filter used in signal processing. When applied to none time series data, a moving average selects higher frequency components without regard for time. However, some order is usually inferred.

It may be thought of as simply smoothing the data when viewed simplistically. A simple moving average (SMA) is the unweighted mean of the past k data points in financial applications. In science and engineering, however, the standard is generally calculated from an equal number of data points on either side of a centre value.

This guarantees that mean variations are matched with data variations rather than being moved in time. The mean over the past k entries of a data set with n entries is an example of a basic evenly weighted running mean. Those data points are $p, p, ..., p$. This might be a stock's closing 12n pricing. SMAK is the mean of the latest k data points (days in this case) and is computed as:

$$SMA_k = \frac{p_n - k + 1 + p_n - k + 2 + ... + pn}{k} \quad (1)$$

When calculating the next mean $SMAk$, next with the same sampling width k the range from $nk+2$ to $n+1$ is considered. A new value $p_n + 1$ comes into the sum and the oldest value $p_n^k + 1$ drops out. This simplifies the calculations by reusing the previous mean SMA_k, $prev$.

$$SMA_{k,next} = \frac{1}{k} \sum_{i=n-k+2}^{n+1} p_i$$
$$= SMA_(k, prev) + \frac{1}{k}(p_n + 1 - p_n - k + 1) \quad (2)$$

This means that the moving average filter can be computed quite cheaply on real-time data with a FIFO / circular buffer and only 3 arithmetic steps.

III. THE PROBLEM

The data set provided explains the stock prices of ICICI, HDFC, SBI, Infosys, Cognizant, and HCL, as well as the USD-INR exchange rate. In this part, we give the findings from a detailed examination of stock market data.

A. Analysis of individual stock prices

We plot the starting, high, low, and closing values of all stocks using visualisation tools. The following conclusions may be drawn from Figure-3:

- Due to the COVID-19 epidemic, all stocks dropped in March-April 2020.
- Cognizant stock witnessed two large drops, but it regained momentum when the lockdown was lifted.
- HDFC stocks plummeted the most, but showed bullish behaviour and surged back up in the previous fiscal year.
- Following the epidemic, all equities provided consistent returns with the exception of Cognizant, whose price stayed nearly constant.

On the other hand, Figure 1 depicts the USD-INR exchange rate. It can be seen that the dollar rose when COVID-19 hit the world. After hitting a peak of around seventy-eight rupees, the dollar has been on a downward trend and is currently around the seventy-five rupees mark.

Fig. 1. USD-INR exchange rate

B. Sector-wise stock price analysis

From Figure 2, it can be observed that the banking sector stocks have seen a similar trend. As expected, a lot of panic selling happened during the COVID-19 outbreak in India, which caused a price dip in stocks. Since then, all the stores in the banking sector, especially HDFC, have shown steady growth.

Figure 4 depicts the volume of equities traded in the banking industry. The figure clearly shows several situations where the volume transacted is exceptionally high. One of the reasons for this is the quarterly reports issued by banks. These reports publish information on how the business is functioning and its future goals, which is essential information for investors since it helps them decide whether or not to invest in the stock. We can also see that ICICI's traded stock volume increased significantly in November 2019, owing to the announcement of a cut in corporate tax rates.

From Figure 5 it can be clearly observed that the IT sector stocks have seen a similar trend. Since the data given is after the upliftment of the lockdown, one can clearly observe steady

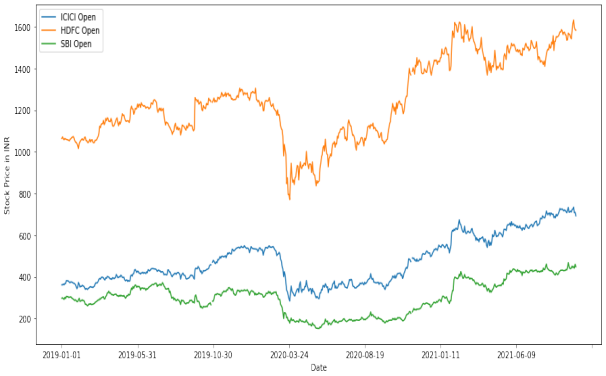

Fig. 2. Opening price of banks- ICICI, HDFC and SBI

Fig. 3. Open-High-Low-Close chart of company stocks for ICICI

Fig. 4. Open-High-Low-Close chart of company stocks for SBI

Fig. 5. Open-High-Low-Close chart of company stocks for HDFC

Fig. 6. Open-High-Low-Close chart of company stocks for Cognizant

Fig. 7. Open-High-Low-Close chart of company stocks for HCL

growth in the stock prices. But something seems unusual in the stock price of Cognizant. From Figure 3 (d) we can conclude that the stock hasn't shown much growth when compared to its competitors. Figure 11 depicts the volume of equities traded in the IT sector. The figure clearly shows several situations where the volume transacted is exceptionally high. One of the reasons for this is the quarterly reports issued by banks. These reports publish information on how the business is functioning and its future goals, which is essential information for investors since it helps them decide whether or not to invest in the stock.

We can also see that the traded stock volume of Infosys soared in October 2020, owing to the announcement of a successful second quarter. Even while it may appear strange that the stock price dropped by a tiny amount after announcing strong results, this is common when many investors sell the company, anticipating that the stock has peaked. Throughout the year, we can observe that the volume of traded Cognizant stock was low because maybe most of the investors were waiting for the store to shoot up so that they could sell it at a higher price than what they bought at.

Fig. 8. Open-High-Low-Close chart of company stocks for Infosys

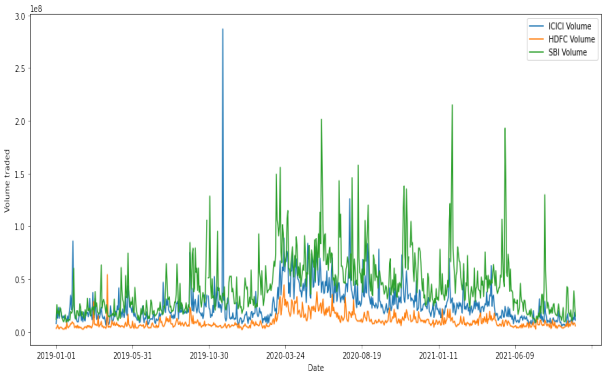

Fig. 9. Volume of stocks traded- ICICI, HDFC and SBI

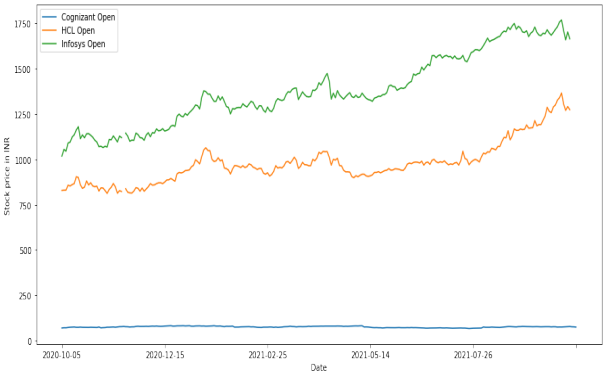

Fig. 10. Opening price of banks- Cognizant, HCL and Infosys

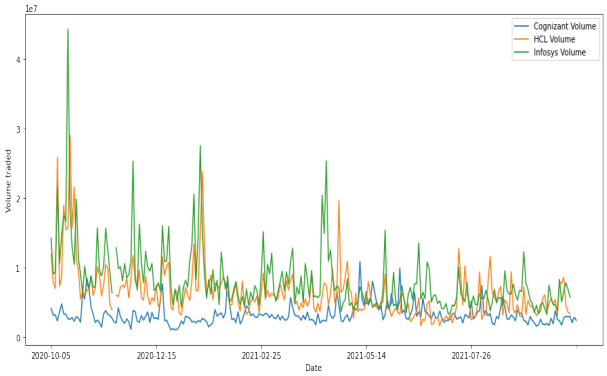

Fig. 11. Volume of stocks traded- Cognizant, HCL and Infosys

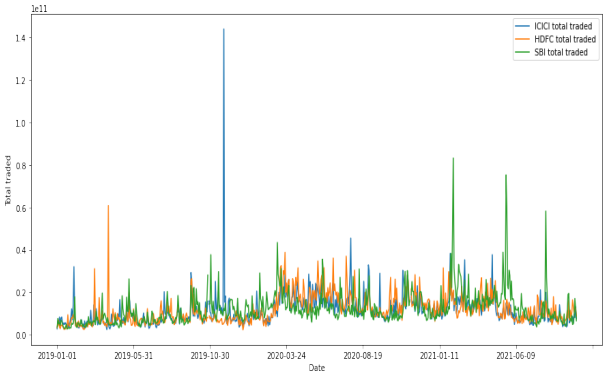

Fig. 12. Total money traded trend- Cognizant, HCL and Infosys

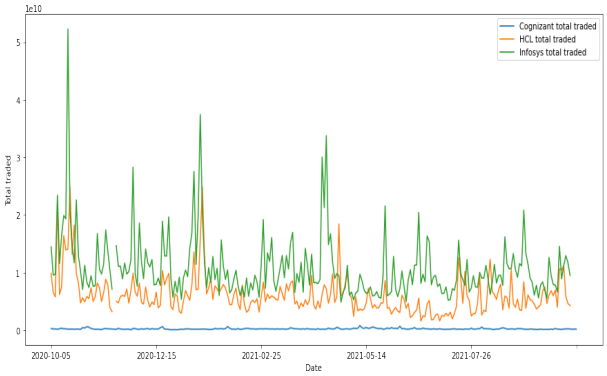

Fig. 13. Total money traded trend- Cognizant, HCL and Infosys

C. Total money traded trend

The essence of total traded value may be represented by multiplying the closing price by the total volume traded. Some spikes can be seen in the banking plots. Economic data, geopolitical events, and market mood are just a few reasons that may cause the stock market to move drastically in one direction or the other. For example, the ICICI bank stock had the most significant increase towards the end of 2019 due to the government's announcement of a cut in corporate tax rates.

D. Sector-wise Scatter Matrix

In our database, we have three banks and three IT entry prizes. We assume that the increase and decrease in stock prices of firms in the same industry will be connected. To demonstrate this, we may create scatter plots for each sector. From Figure 16, we can see a high correlation among the banking sector bonds. It can be understood as the government policy change, and the economy's ups and downs affect all the banks. The same goes for the information technology sector.

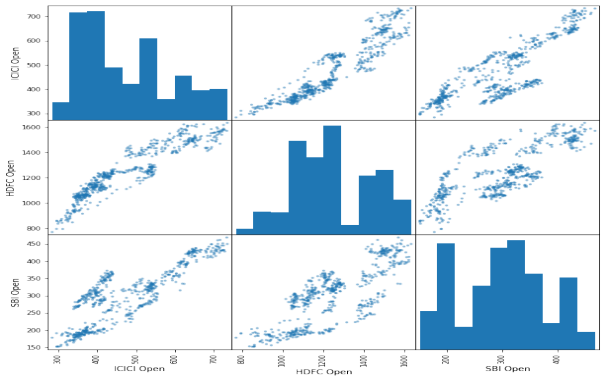

Fig. 14. Scatter Matrix- ICICI, HDFC and SBI

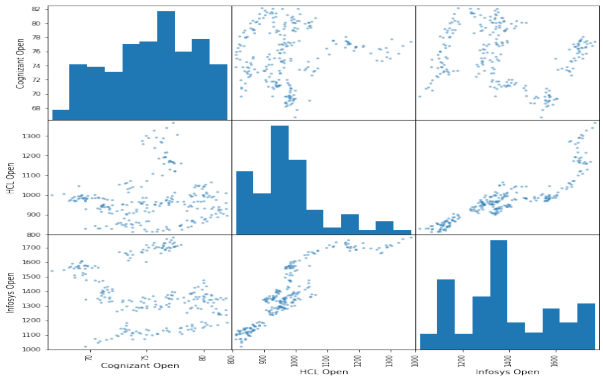

Fig. 15. Scatter Matrix- Cognizant, HCL and Infosys

E. Sector-wise daily returns

The returns on a stock may be computed as the stock's price change over time, which can be expressed as a percentage change or price change. We can calculate the returns of the supplied stocks by computing the percentage change in the stock's closing value. The returns are plotted in the line plot below.

The line plot makes it difficult to assess and quantify the returns. To better understand, we may determine the volatility of the returns. Volatility reflects the extent to which price moves. A highly volatile stock has a fee that varies wildly—hits new highs and lows or moves unpredictably. Low volatility is associated with a store that maintains a reasonably constant price. We can calculate volatility by plotting the distribution of returns. SBI is less volatile than HDFC and ICICI in the banking sector and provides consistent returns. HCL is less volatile than Infosys equities in the IT industry.

Box charts are another approach to seeing the results. Box plots show variance in statistical population samples without making any assumptions about the underlying statistical distribution. According to the box plots, ICICI bank has the greatest mean return, followed by SBI and HDFC, respectively. According to the IT sector box plot, the mean return of HCL and Infosys is nearly identical, although HCL returns are more spread out.

F. Cumulative returns

A cumulative return on an investment is the total amount gained or lost over time, regardless of the length of time involved. It is the overall change in the price of an investment over a specific period. We can plot the cumulative returns for both the banking and IT sector.

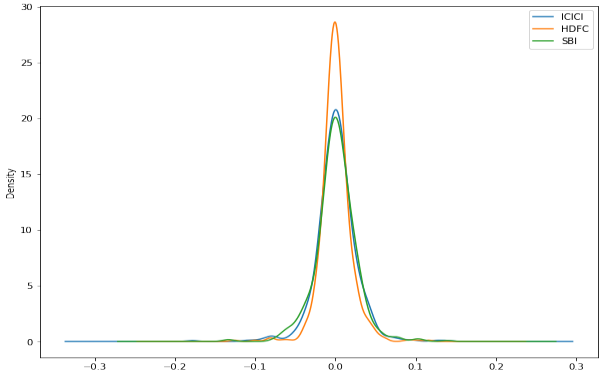

Fig. 16. Returns plot of Banking sector stocks

IV. CONCLUSIONS

Only roughly 2,000 pulsars have been detected so far. For their celestial research, scientists and astronomers benefit from establishing whether a star is a pulsar.

The provided data was used to describe the link between various qualities and determine which ones were the most advantageous for this issue statement.

The association between predictors and regressors may be better understood using the EDA and Support Vector classifier approach. It aids in the explanation of the underlying pattern and the prediction of accurate outcomes.

In the future, GridSearchCV and RandomSearchCV can be utilised to find the best imputation approach and make a hyper-parameter selection.

REFERENCES

[1] Roberts, H. V. (1959). Stock-market" patterns" and financial analysis: methodological suggestions. The Journal of Finance, 14(1), 1-10.
[2] Granger, C. W., Morgenstern, O. (1963). Spectral analysis of New York stock market prices 1. Kyklos, 16(1), 1-27.
[3] Aly, H. Y., Mehdian, S. M., Perry, M. J. (2004). An analysis of day-of-the-week effects in the Egyptian stock market. Available at SSRN 551101.
[4] Sapankevych, N. I., Sankar, R. (2009). Time series prediction using support vector machines: a survey. IEEE computational intelligence magazine, 4(2), 24-38.
[5] Weigend, A. S. (2018). Time series prediction: forecasting the future and understanding the past. Routledge.
[6] Connor, J. T., Martin, R. D., Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. IEEE transactions on neural networks, 5(2), 240-254.
[7] Farmer, J. D., Sidorowich, J. J. (1987). Predicting chaotic time series. Physical review letters, 59(8), 845.