PSYCHOLOGY AND EDUCATION (2018) 55(1): 752-768
ISSN: 0033-3077

An Empirical Research and Comprehensive Analysis of Stock Market Prediction Using Regression Analysis

[1]Sanam Sharma
Assistant Professor
Maharaja Agrasen Institute of Management Studies, New Delhi
Sanam.maims@gmail.com
[2]Manika Mittal
Assistant Professor
Maharaja Agrasen Institute of Management Studies, New Delhi
Manika16garg@gmail.com

ABSTRACT

Predicting the price of stocks on the stock market is a fascinating topic in the financial business because it involves a number of different variables and forces individuals to take a difficult approach to handling the problem. The financial markets are an essential component of every nation's economy and rank among its most significant components. Researchers have been putting in a lot of effort over the past several years to have a better understanding of how to anticipate the stock market. The very act of prediction is fraught with peril due to the myriad of factors that must be taken into consideration. This makes for an intriguing challenge in and of itself. During the course of this study, an exhaustive investigation of the strategies that are now at one's disposal for predicting the behavior of the stock market is carried out. There is also a discussion of the future route included, in addition to a comparative analysis of the different ways that are being used now, which includes a discussion of the many methodologies being used.

Keywords
Comprehensive Stock Market, Prediction, stock prices,risk & return.

Introduction

As a consequence of the progress that has been accomplished in the economy, the stock market has evolved into the sector of the financial industry that has the most significance. Investors are able to combine their resources and work together toward the common goal of expanding their companies with the help of the stock market, which acts as a meeting place for corporations and governments. Because it provides diversification across a number of various entities, a platform from which to profit from a professionally managed portfolio at a relatively low cost of investment, and high liquidity, the stock market is the sort of investment that is most suited for the typical individual. This is because the stock market offers diversification across a number of different entities. Real Gross Domestic Product, Inflation Rate, and Interest Rate are three economic indicators that have a substantial influence on the stock market [1]. RGDP is for Real Gross Domestic Product, INF stands for Inflation Rate, and INT stands for Interest Rate. Over the course of the last several years, forecasting the behavior of the stock market has been the focus of a substantial amount of interest. This issue is extremely intriguing because of its dynamic, non-linear, complicated, time-varying, and chaotic character, in addition to the fact that it is regularly impacted by economic and political variables [2].

Another reason why this subject is interesting is because it is frequently influenced by a variety of factors. The desire that many investors have to find a forecasting technique that would guarantee profit without the chance of loss and remove all hazards associated with the stock market is one of the driving forces behind the development of new predictive models by researchers [3]. Because of this purpose, academics have been hard at work developing new prediction models. It is vital to acquire a tool for forecasting, which can be done by examining the activity of the stock market using a range of ways and procedures. Doing so will allow you to obtain the forecasting tool you need. When investors use this technology, they will discover that it enables them to purchase and sell at prices that are more comfortable for them. Investors can find it helpful to get an early indicator of the estimated level of volatility in order to better prepare themselves to take risks. In recent years, a variety of research have been carried out with the intention of making projections on the price of the stock market. One approach to doing analysis and making projections regarding the price of the stock market is to make use of machine learning in addition to a number of other important algorithmic approaches. It is possible to make a prediction on the stock market for either the short term or the long term. On the other hand, it is far more challenging to make accurate short-term predictions than it is to make correct long-term predictions. A sizeable fraction of national governments all around the world are currently engaged in active trading on the stock market. These governments allocate a part of their resources into healthcare, employment, or retirement with the expectation that it would result in improved outcomes for their citizenry. Trading platforms that may be accessed online have brought about a fundamental shift in the way individuals transact in the equity markets. The financial markets have developed at a breakneck pace to become an interconnected marketplace on a worldwide scale. The research that has been done on the subject of predicting the stock market may be sorted into one of three primary groups based on the methodology that was used. These groups include the statistical approach, the artificial intelligence approach, and the hybrid approach. This article will give a critical analysis of the different approaches that are presently available for forecasting the stock market. The objective of this essay is to provide this review. In the next part, the many research methodologies that may be utilized for forecasting tools will be described and briefly reviewed. A comparison and evaluation of the several approaches that have been discussed may be found in the third part. This analysis contains a summary of the objectives of the investigations as well as the datasets on which the experiments were performed. A discussion on the potential range of future research into the topic that was investigated is offered as the final part of the study.

Related Study

In the recent days, a wide variety of alternative techniques have been created in an effort to estimate the future value of shares that are traded on the market. In order to anticipate the future price of a stock market, these people have used a range of methods, and at times they have utilized a mix of methods. These are the ways that they have gone about it. The subfield of

artificial intelligence known as artificial neural networks provides forecasts regarding the price of stocks on the stock market by employing the artificial neural network back propagation algorithm in conjunction with the feed forward neural network. This is done in order to maximize accuracy. The author of "4" conducts research and makes projections on the stock market index using Markov characteristics, stock prices, and the state of the interval in light of the Markov model. This model provides investors with a proper reference model so that they may avoid participating in behavior that is irrational and illogical in order to maximize their potential returns. The main idea behind using a Markov chain to generate predictions is to construct a Markov forecasting model that predicts the state of an item in a certain period of time in the future by virtue of the probability vector of the starting state and the state transition probability matrix. This model anticipates the state of the item based on the probability vector of the beginning state. A Markov forecasting model is the name given to this particular model. The Markov process is a form of random process that does not display any after-effect features. This property makes the Markov process unique. The after-effect properties suggest that the state of at a time t larger than s will only rely on the state of at the moment s in some process if that state was known at the moment s in some process. This is because the after-effect properties entail that the state of at a time t greater than s will depend on the state of at the moment s in some process. If this is not the case, the state will not be dependent on the state that was before the instant s in the process. The following is the formula that should be used to determine a transition matrix: In a system that is thought to be in a balanced state, if it is determined that the probability of the system transitioning from state i to state j is Pij, then the collection of transition probability vectors that are discovered in each state of the system will create a transfer matrix. This matrix will be used to determine how likely it is that the system will transition from state i to state j. An novel approach for calculating the price of equities on the stock market is presented by the author of [5], which can be found here. An HMM is a kind of state machine that may be used to a system that follows the Markov process but has certain states that are hidden from view. If we refer to the hidden state at time t as x(t) and the observation at the same time as y(t), then the following propositions are always true with relation to the hidden Markov model: The only thing that determines the value of x(t) is x(t-1), and the only thing that determines the value of y(t) is x(t). Find the day in the stock market data for a particular day that is most similar to that day so that it may be used to forecast the close value for the day following it. This is the first thing that has to be done in order to make a prediction, which is the second thing that needs to be done in order to make a prediction. In order to achieve this, we must first determine the probability that the days before to the current one fell somewhere within the parameters that have been set. When there is just one day's worth of stock data available, calculating the HMM's probability for that specific day is a basic and easy operation. The last step of this experiment will be to make a prediction about which day will have a value that is the closest to the objective once all of the likelihood probabilities associated with the various days have been computed. They

achieved this by introducing a parameter known as probability tolerance, which defines the similarity neighbourhood within which it is acceptable to accept days that are similar to the day before. This allowed them to accept days that were similar to the day before. We use the probability tolerance to extract a list of days whose stock data is comparable to yesterday's data, and then we try to determine which of the many possible predictions is the most accurate by selecting the one that has the highest likelihood in comparison to the other options.

Data Collection

The returns, also known as the percentage change in the closing price of various businesses with Indian Rupees acting as the currency and categorised according to small, medium, and huge capitals as indicated in Table1, make up the bulk of the data. Table1 presents the data in tabular form. The market capitalization of a firm is an indication of how valuable its shares are to investors. The outstanding securities of companies with low market capitalization are referred to as "small cap" securities, which is a term that is used to characterize these securities. This term comes from the phrase "small cap" securities. A company is considered to have "small cap" status in India if its market value is less than 5000 billion rupees, which is the country's currency. One should frequently be prepared for a high amount of volatility when dealing with small size companies. If a firm's market worth is higher than 5000 crores but is less than 20,000 crores, then the company is regarded to be a mid cap. "large capitalization of the market" is an extensive phrase that has been simplified and shortened to "large cap." This is not the least important point. Companies are considered to have a "large cap" status if their market valuation is larger than 20,000 crores (or millions) of Indian rupees. [6] Regarded as the alternative that promotes the greatest equal distribution of benefits.

Table1. corresponding stocks, organized according to their respective market capitalization sectors

Small Caps	Mid-Caps	Large Caps
Ashoka	Adani	Adanisport
Bajaj	Ajantfarm	Asian
Bomdyeing	Amarajabat	Axis
Century	Apollo	Bajfinance
Fortis	Bergepaint	BPCL
Gujalkali	Castround	Cipla
IDFC	Cummins	Dreddy
Ircon	DHFL	Eicher
ITDC	Excide	Gail
Jetairways	Gimrin	HDFC

Calculation of the beta coefficient using the ordinary least squares (OLS) method of linear regression

In this section, the beta value was calculated by employing Linear Regression on the data received from the stocks LT and Nifty50. The results of this analysis were then used to estimate the beta value. The Ordinary Least Squares Regression is a kind of linear regression that is used rather

commonly in statistical methods for the goal of undertaking predictive analysis. Its name comes from the fact that it uses the least number of squares possible.

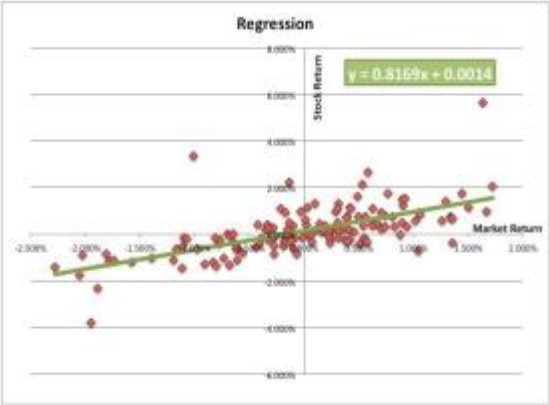

Figure 1. Regression References

The solution to the problem that involves linear regression is depicted in Figure 1. It gives us an equation in which our traits, which change depending on the variable that we are interested in, are regarded as independent variables. This allows us to more accurately predict the outcomes of our experiments. Before establishing the cost of equity, one of the stages that an investor has to do is to compute the beta of the company's shares. This is done in preparation for that step. The Regression (slope) Market Model is one tool that may be applied to the procedure of estimating the beta value of a publicly traded stock, as demonstrated by the first equation in this paragraph. Because of this method, the company's stock returns were altered in a way that made them appear lower when compared to the NIFTY50 [7].

$$Y = \theta_1 \times x_1 + \theta_2 \times x_2 + \ldots \theta_n \times x_n$$

Y is the dependent variable in this scenario, whereas x and are the independent variables and the slope of the line.

A measure of the extent to which an asset's earnings are variable in proportion to a price benchmark (often an index of stock prices), the beta of an asset is referred to in this context. A stock that has a beta value that is larger than 1.0 and has a level of volatility over time that is greater than that of the market as a whole is considered a stock with a high beta. When the price of a stock drops below its prior level, the beta of that stock drops below the level of 1.0. It is essential to bear in mind that high-beta shares are projected to be more volatile, but they also have the potential for higher profits, whereas low-beta firms entail less risk but also have lower potential returns [8]. This dichotomy should be kept in mind when comparing the two types of companies. The following is a list of potential challenges that might be encountered while performing calculations using Beta:

- Calculated time duration On the basis of previous observations, most people estimate that the beta value will be between two and nine years in the future. The selection of the measurement period has a role in the determination of the Beta value. It is a reflection of the present dynamics of the company if the amount of time needed for the estimation is relatively short.
- Calculated time duration On the basis of previous observations, most people estimate that the beta value will be between

two and nine years in the future. The selection of the measurement period has a role in the determination of the Beta value. It is a reflection of the present dynamics of the company if the amount of time needed for the estimation is relatively short.

- Market index selection: In order for the analyst to provide an accurate estimate of the return, they are responsible for picking a market index after giving it a great deal of thought. For the purpose of this computation, the Nifty50 index was utilized.
- Small-cap stocks often have a greater risk of financial loss, but they also typically offer a greater potential for financial gain. Additionally, it's probable that industry professionals will have the desire to raise the beta of small size businesses. The stock with the ticker symbol LT was selected for this endeavor; the company that produced it had a substantial market capitalization.

The second equation in the previous statement provides the equation for the regression line, which may be found in the following sentence.

$$\gamma_i = \alpha + \beta\gamma_m$$

where,

γ are the stock's returns

α represents the intercept

γm are the market returns

β is represented by the slope of the regression line.

The K-NN classifier was utilised to make a prediction of Call based on Bollinger columns.

Within the scope of this part, we had talked about the definition of categorization. The classifier model was developed using three Bollinger columns, the stock price (close price), and TITAN business performance data with 'Calls.'

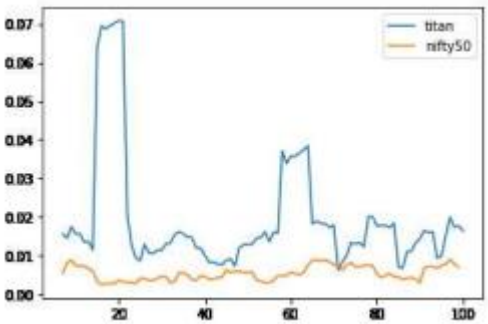

Figure 2. Volatility between Titan & Nifty50

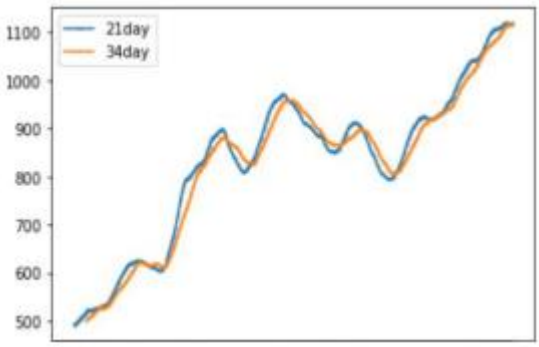

Figure 3. Moving average of Titan

In the first place, the volatility of both the Titan index and the Nifty index was measured and compared, as can be seen in Figure 2. The formula that is used to compute the call price is shown in Figure 3. This formula is derived from an average price that is determined by employing a moving average that is comprised of 21 and 34 days.

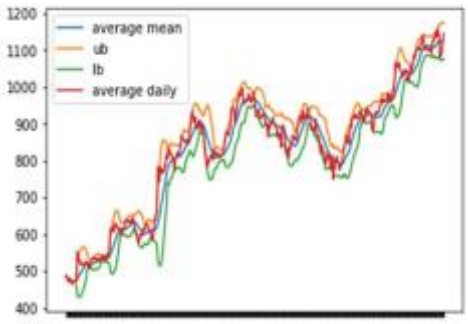

Figure 4. Titan's closing price's Bollinger bands

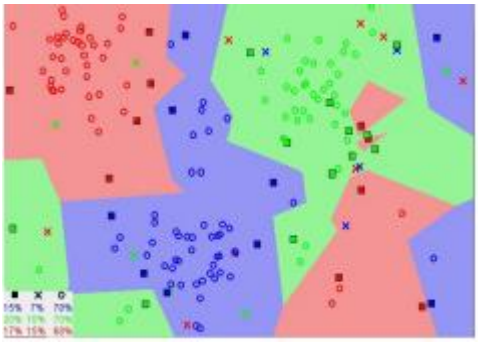

Figure 5. K-NN Classification

See Figure 4 for an explanation of how the Moving Average (MA) generates the price range's average by dividing the number of time periods that fall within that range by the range's regular closing prices. One form of technical indicator is known as the moving average (MA), and it may be utilized to help identify whether or not a price trend will continue or turn around. The stocks that will make up the TITAN and the NIFTY50 have been chosen. Following the creation of the moving average graph, the choice on whether or not to call may then be made based on their behavior. For instance, a call option trader should buy when the shorter moving average (21) crosses over the longer moving average (34), and a call option trader should sell when the shorter moving average (21) crosses under the longer moving average (34). The following data points were utilized during the construction of the Bollinger bands:

- The weighted moving average of the price at which a security closed during the last 14 days
- Upper band, which represented the rolling mean plus two standard deviations from the average of the values; and lower band, which represented the rolling mean only.
- The rolling mean was represented by the lower band, which was positioned two standard deviations below the average value.
- The stock price calculated on a daily basis and averaged out.

Because Bollinger bands are assessed using Moving Average (MA), as shown in equations (3.1) and (3.2), they analyze older consumer data in the same way as more recent data, which means that stale data will dilute down fresh information. This is because Bollinger bands are evaluated using equations (3.1) and (3.2). Between these two bands is where about 90 percent of all of the commercial activity takes place [9]. Every time the bands are reorganized higher or lower, there is an important event that takes place. The breakout does not offer any trading insights at this time. Many people are under the mistaken impression that an alert to buy or sell arrives when the price strikes or divides one of the bands, but this is not the case. This is a mistake that many people make. Breakouts do not convey any clues as to the possible shift in price, either in terms of the direction or the amount of the change [10]. Calculating the Bollinger band may be done with the help of the following formula:

$$U = MA\,(TP) + m \times \sigma\,[TP, n]$$
$$L = MA\,(TP, n) - m \times \sigma\,[TP, n]$$

where:

U = Upper Bollinger Band L = Lower Bollinger Band Both MA and TP stand for "moving average," whereas TP refers to the "typical price." (Typical price) = (High + Low + Close) 3, and n stands for the number of days in the smoothing period, which is typically regarded to be 20. m = The number of standard deviations, which is commonly considered to be 2. [TP, n] = the standard deviation over the course of the TP measured over the past n periods.

As can be seen in Figure 5, the K-NN approach is predicated on the idea of categorizing data points according to how closely they are related to one another. Take notice that the majority of the time, connections that are similar to those you just saw are created between data points that are related to one another in the image that you just seen. Because the KNN algorithm is dependent on this presumption to supply it with meaningful information, the presumption is critical to the operation of the KNN algorithm.

Proposed trading model: LSTM (Long-short term memory)

Because of the interaction between the knowledge and the LSTMs, a system that is known as cell states is generated as a result of this interaction. After that point, it will be much simpler for LSTMs to recall or forget information [11]. As can be seen in Figure 6, there are three unique requirements that the data at every cell level need to meet. These dependencies are on the data itself. Dependencies of this sort are amenable to generalization in answer to any enquiry because:

- The state that the cell was in before the relevant time step, which is the information that was stored in memory after the step in question.
- The feedback at the present time stage, which refers to the new information that is being fed in at that particular instant; this is the same as the output from the cell that came before it.
- The prior secret state. This indicates that this is the same as the output from the cell that came before it.

An illustration of this is included for your perusal, should it be necessary. Consider the exercise of forecasting the prices of individual items for a certain stock at some point in the future. The price of a stock for today will be determined by a number of factors, including the pattern that the market has seen in the previous days, which could be an uptrend or a decline, as well as the stock price from the day before, because traders frequently compare the price of a stock from the previous day before purchasing it [12]. The pattern that the market has seen in the previous days could be an uptrend or a decline.

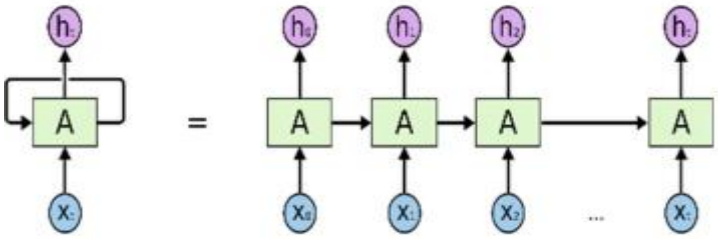

Figure 6. Architecture of LSTM

There are a number of different elements that might have an effect on the price of the stock at the moment. It might be a brand-new corporate strategy that is greeted with

major opposition, a decline in the revenue that is brought in by the product, or even a sudden shift in the top management team of the organization that caused the problem. LSTMs offer a lot of benefits, but traditional neural feed-forward networks and RNNs have a few drawbacks [13]. This is in comparison to the benefits that LSTMs have. This is due of their ability to recover patterns in a discerning fashion over long periods of time, which accounts for why they can do this.

Investment Portfolio Optimization

A concept that describes how risk-averse investors could create portfolios to maximize expected returns given a specified degree of market risk is known as the Modern Portfolio Theory, which is also frequently referred to as MPT for short. According to the conventional portfolio method, the risk and return characteristics of an investment should not be understood in isolation; rather, they should be evaluated based on how the investment affects the risk and return of the entire portfolio [14]. This is because the risk and return characteristics of an investment are not meant to be understood in isolation from the other characteristics of the portfolio. According to this strategy, the risk and return parameters of an investment should not be evaluated in a vacuum at any point throughout the analysis process. The Multi-Asset Portfolio Theory (MPT) illustrates that an investor may design a multi-asset portfolio in order to maximize the returns for any given degree of risk. This can be accomplished by diversifying the investor's holdings across many asset classes. When an investor requires a particular level of return on their investment, they may create a portfolio that has the least amount of risk possible. This is similar to the previous example. The success of an individual investment is less essential than how it affects the portfolio as a whole, according to statistical parameters like as variance and correlation [15]. The assessment of the portfolio's estimated return, which is produced from this number, is based on the weighted value of the return on each individual security. This value serves as the foundation for the evaluation. If the portfolio is made up of four assets that are all given the same amount of weight and each asset has a different anticipated return, such as 4, 10, or 14%, then the return on the portfolio will be approximately as follows:

(4% x 25%) + (6% x 25%) + (10% x 25%) + (14% x 25%) = 8.5%

A portfolio may be conceptualized while thinking about five different equities. In an ideal scenario, the participating businesses would represent a wide range of business sectors and geographical areas, but they would all be given the same weight of twenty percent of the total vote.

Clustering for diversification analysis

The K-means Clustering method was used to group 30 different stocks into clusters, and the following is a list of objectives that have been met in order to do so:

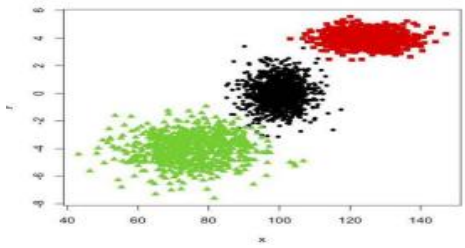

Figure 7. K-means Clustering

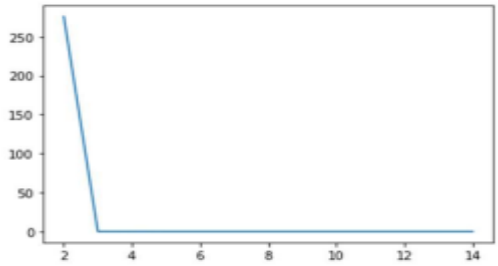

Figure 8. Elbow Curve.

- Constructed a data frame that includes the closing prices of thirty different stocks, with ten equities from each capitalization included in the frame.
- Calculated the average annual percentage return as well as the volatility of each of the 30 stocks over the course of a fictitious period of one year.
- K-Means Clustering was applied to 30 distinct stocks after calculating the average yearly volatility and return of each of those shares. The results of this analysis may be found below.
- We were successful in identifying the optimum number of clusters by use the elbow curve methodology.
- Constructed a distinct data frame in order to indicate which equities are members of the same group. This was done in order to show which equities belong to the same group.

Clustering K-means is a type of unsupervised learning that is utilized in circumstances in which there is availability of unlabelled data [16]. Unlabelled data refers to data that does not have any predetermined categories or groupings. Figure 7 may be used to display this kind of information. The goal of this approach is to recognize patterns of grouping that exist within the data, and the value of the variable K represents the total number of patterns that exist inside the dataset. The fact that the approach assigned each data point to one of the K groups depending on the features of those categories contributed to the fact that the method was effective. The results of the clustering procedure for the K-means algorithm are the K-cluster centroids. Because each data point is assigned to a particular cluster, these centroids may be used to label new data as well as identify training data, and the findings can be used to determine which K-clusters exist. The elbow technique, which can be seen in Figure 8, is one of the algorithms that is utilized in the process of cluster analysis. This strategy is used for estimating the number of clusters that are present in a data set. The approach entails charting the variance described as a function of the number of clusters and picking the number of clusters that corresponds to the curve elbow as the one to be utilized in the study. Additionally, the method requires charting the variance described as a function of the

number of clusters. As can be seen in Figure 9, an ideal number of clusters has been established to be five, and this conclusion was reached on the basis of the graph that was just presented. The core concept that underpins it is to identify the value of k at which the amount of error begins to diminish in a very short amount of time. When plotting the elbow curve, the degree of inaccuracy has a tendency to decrease as 'k' becomes increasingly big. This is because when the number of clusters increases, there is a corresponding reduction in the amount of distortion, which in turn causes the clusters to become more compact.

	0	0
0	(ashoka,)	4
1	(bajaj,)	3
2	(bombdyeing,)	0
3	(century,)	3
4	(fortis,)	4
5	(gujalkali,)	3
6	(idfc,)	3
7	(ircon,)	3
8	(itdc,)	4
9	(jetairways,)	4
10	(adani,)	0
11	(ajantfarm,)	3
12	(amarajabat,)	3
13	(apollo,)	3

Figure 9. Prediction of cluster groups

Evaluation and Performance Measures

The many measures that were utilised in the assessment of our performance are listed below.

Root Means Square Error (RMSE)

The Root Means Square Error value, which is frequently referred to by its abbreviated form, RMSE for short, is a measurement that is utilized frequently for determining the differences in values that may be anticipated by an estimator or a mode. It provides an explanation of the sample standard deviation of the disparities between the predicted and observed values, which can be located in equation (4) lower down on this page. You can find equation (4) here.

$$\text{RMSE}=\sqrt{\sum_{i=1}^{N} (predicted_i - actual_i)^2 \div N} \quad (4)$$

where,
$predicted_i$- are the predicted values.
$actual_i$- are the actual values.

Slope of the regression line- Beta (β)

$$\beta = \frac{Covariance}{Variance}$$

The phrase "covariance" refers to a measurement that is taken of a stock's relative return, whereas the term "variance" refers to a measurement that is taken of how the price changes in reference to its "mean" [17]. Covariance and variance are both measurements that may be taken.

The K-NN classification model's accuracy score may be predicted using the following formula

The accuracy for K-NN model is given by equation (6)

$$Score = \frac{TP+}{TP+TN+FP+F}$$

where,
TP→ $True\ positive$
TN→ $True\ negative$
FP→ $False\ positive$
FN→ $False\ negative$

Sharpe Ratio:

The Sharpe Ratio, which is expressed by the equation (7), is utilized when computing the prior output of a portfolio, and the computation takes use of real returns. This allows for more accurate results. When the Sharpe ratio is larger, it means that the efficiency of the adjusted risk is higher [18].

The following equation may be used to describe the slope of the regression line: (5)

(5)

$$Sharpe\ Ratio = \frac{Rp-Rf}{\sigma p}$$

where
Rp = Return of Portfolio
Rf = Risk free rate
σp = Standard deviation of portfolio excess's return.

Result and Analysis

The following is a list of the results that came about as a direct result of the application of a number of different models.

The results of the model while calculating beta with the ordinary least squares (OLS) method of linear regression

Figure 10 displays the values that were used to calibrate the model; these were the percentage changes in the closing prices of LT and NIFTY50. In this scatter graph, the real values are shown as blue dots, while the set of anticipated values are shown as a red line. The set of anticipated values was calculated using model [19], which is shown as the source of the red line.

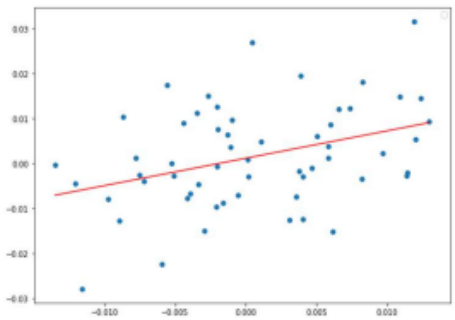

Figure 10. OLS linear regression output

This article provides an analysis of the model's effectiveness based on the RMSE created as 0.013, which indicates that the algorithm was quite accurate, and the Beta generated as 0.609, which indicates that the LT stock would have been 60.9% less volatile compared to the nifty50 during the last three months. This study's objective is to offer a more in-depth look at the impact that the model has had on the world by providing a more complete examination of that impact.

Call against Bollinger Bands assessment based on a simulation of the model output and results

As can be seen in Figure 11, Bollinger bands were created for the selected stocks (TITAN and NIFTY50) with a period of 14 days and 2 standard deviations away from the average [20]. At this juncture, the question is whether or:

- If the price of the stock is now lower than the lower Bollinger band on the chart, you should "buy" the stock.
- If the stock price is between the lower and middle bands, either buy the stock with the "Hold" option or buy it with the liquidation option.
- "Hold short/liquidate buy" when the stock price is between the middle and higher bands of the banking system. This strategy is used when the stock price is between these two bands.
- The stock price is considered "Short" if it is currently trading at a level that is below the lower Bollinger band but above the middle Bollinger band.

```
['Hold Buy/Liquidate Short' 'Hold Short/Liquidate Buy'
 'Hold Buy/Liquidate Short' 'Hold Buy/Liquidate Short'
 'Hold Buy/Liquidate Short' 'Hold Short/Liquidate Buy'
 'Hold Buy/Liquidate Short' 'Short' 'Hold Short/Liquidate Buy
 'Hold Short/Liquidate Buy' 'Hold Buy/Liquidate Short'
 'Hold Buy/Liquidate Short' 'Hold Buy/Liquidate Short'
 'Hold Buy/Liquidate Short' 'Hold Buy/Liquidate Short'
 'Hold Short/Liquidate Buy' 'Hold Buy/Liquidate Short'
 'Hold Buy/Liquidate Short' 'Hold Buy/Liquidate Short'
 'Hold Short/Liquidate Buy' 'Hold Buy/Liquidate Short' 'Short
 'Hold Short/Liquidate Buy' 'Hold Buy/Liquidate Short'
 'Hold Short/Liquidate Buy' 'Hold Buy/Liquidate Short'
 'Hold Buy/Liquidate Short' 'Hold Short/Liquidate Buy'
 'Hold Short/Liquidate Buy' 'Hold Short/Liquidate Buy' 'Buy'
 'Hold Short/Liquidate Buy' 'Hold Buy/Liquidate Short'
 'Hold Short/Liquidate Buy' 'Hold Short/Liquidate Buy'
 'Hold Buy/Liquidate Short' 'Hold Buy/Liquidate Short'
 'Hold Buy/Liquidate Short' 'Hold Buy/Liquidate Short'
```

Figure 11. Prediction of Daily Calls

This particular KNN model has an accuracy of 84.3%.

Model output and Performance for LSTM

The result of the stock price is shown graphically in Figure 12, which you may get here. The orange line makes a forecast about the stock's close price by using the outcomes of the tests as the basis for its analysis. The test data are represented by the orange line, and the projected close price for the stock is shown by the green line in the graph. It was established that the efficiency of the model with a 1/1 epoch was 58.44, which is good considering that it is much lower than 180. It was observed that the efficiency of the Model while using an epoch of 1/1 is 58.44, which is good given that it is substantially lower than 180.

Figure 12. LSTM prediction output

Results and Analysis

The following graphic illustrates how annual returns and volatility were calculated by applying the mean and covariance values of the entire portfolio, respectively, to the data. These results are displayed in the following table. The following companies' shares have been chosen for investment: Raymond, Ashoka, LT, Voltas, and Godrej. Both of these locations are traversed by the effective frontier curve, sometimes known as the number 21, which may be found here. The term "productive frontier" can either refer to the category of standardized portfolios that has the highest expected return for a level of risk that has already been determined or it can refer to the category of standardized portfolios that has the lowest risk for a given expected return. Both of these interpretations of the term are valid. Portfolios that have a value that is lower than the active boundaries are not optimal due to the fact that they do not provide an adequate return on the risks that they focus [22]. The efficient frontier is a graphical depiction of the set of investments that produces the best potential return proportionate to the amount of risk taken. This collection of investments is called the "efficient frontier." A protection standard deviation is another term that may be used interchangeably with the word fear. The link does not take the direct route that one might expect. To phrase it another way, increasing the amount of risk that a portfolio is exposed to does not result in returns that are comparable. It would appear that optimal boundary portfolios have a greater level of variety compared to sub-optimal portfolios, which are often less robust [23]. This is the case since optimal boundary portfolios are designed to maximize returns.

	ret	stdev	sharpe
0	0.024086	0.229386	0.105003
1	-0.095910	0.226200	-0.424003
2	-0.061470	0.220521	-0.278748
3	-0.037959	0.240082	-0.158109
4	-0.002026	0.218533	-0.009270
...	...	...	...
24995	-0.079336	0.232080	-0.341850
24996	0.018155	0.218613	0.083046
24997	-0.020662	0.216312	-0.095518
24998	-0.069234	0.221966	-0.311914
24999	-0.070254	0.211173	-0.332685

Figure 13. Returns on investment on an annual basis, volatility, and the Sharpe ratio for the entire portfolio.

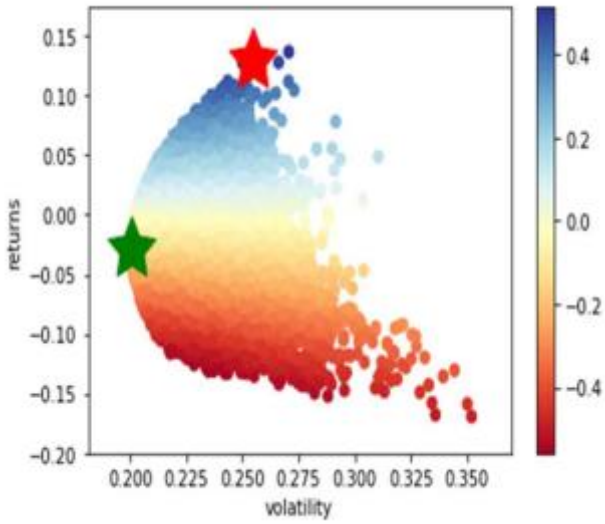

Figure 14. Scatter plot of portfolio

A bright red star appears in Figure 14 to denote that the Sharpe ratio has been elevated to its highest conceivable value. The condition shown by the green star as having the lowest level of volatility is one that is least likely to change. It is calculated as the ratio of the average return achieved over the risk-free rate to the entire amount of volatility in the market.

Conclusion

The research was carried out to explore and test each of the components, with the end goal of determining how accurately stock market forecasts may be made. There are a lot of companies that are having difficulties at the present time as a direct result of their failure to correctly foresee or anticipate potential problems that may arise in the future in order to make decisions that are suitable. Throughout the course of our work on this project, a number of different approaches have been utilized. Some of these approaches include linear regression, K-means Clustering, K nearest neighbor, LSTM, and others. The use of algorithms in the forecasting of stock prices has been demonstrated to be essential, which has led to its integration into effective market tactics due to the fact that it has led to improved accuracy. Bollinger bands are a tool that traders may find beneficial since they enable them to monitor the volatility of their position and provide them advice on the appropriate times to enter and exit a position. When seeking to make financial investments in the stock market, these techniques provide brokers and investors a major advantage since, prior to being applied to the stock market, they are validated using representative samples of historical data. This gives brokers and investors a considerable competitive edge. In addition, the analytics may be used to conduct an analysis of the content of public opinion and, as a consequence of this, can be used to construct patterns and links between members of the general public and those working in business. As a consequence of this, when we assess the findings of this study through the lens of the outcomes, we are able to see that technology still has a great distance to travel before it is completely capable of deciphering the riddle of the stock markets.

Reference

[1] Chen, C., Liu, Y. and Hsu, T., 2017. AN ANALYSIS ON INVESTMENT PERFORMANCE OF MACHINE LEARNING: AN EMPIRICAL EXAMINATION ON TAIWAN STOCK MARKET. International Journal of Economics and Financial Issues, 9(4), pp.1-10.

[2] Puri, N., 2015. A Survey on Machine Learning Approach for Stock Market Prediction. HELIX, 8(5), pp.3705-3709.

[3] Kanade, P., 2010. Machine Learning Model for Stock Market Prediction. International Journal for Research in Applied Science and Engineering Technology,

[4] Deju Zhang, Xiaomin Zhang, "Study on Forecasting the Stock Market Trend Based on Stochastic Analysis Method" in International Journal of Business and Management, Vol 4, No. 6, 2009.

[5] M. R. Hassan and B. Nath, "Stock market Forecasting using Hidden Markov Models: a new Approach", Proceedings in 5th IEEE conference on Intelligent Systems Design and Applications, ISDA 2005.

[6] Finance, Markets and Valuation, 2012. Forecasting stock market trend: a comparison of machine learning algorithms. 2012(1).

[7] Henrique, B., Sobreiro, V. and Kimura, H., 2017. Literature review: Machine learning techniques applied to financial market prediction. Expert Systems with Applications, 124, pp.226-251.

[8] J, P., 2016. Stock Market Analytics: Statistical and Machine Learning Techniques. International Journal of Psychosocial Rehabilitation, 24(1), pp.1828-1833.

[9] KOMPELLA, S. and CHILUKURI, K., 2015. STOCK MARKET PREDICTION USING MACHINE LEARNING METHODS. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND TECHNOLOGY, 10(3).

[10] Lee, T., Cho, J., Kwon, D. and Sohn, S., 2017. Global stock market investment strategies based on financial network indicators using machine learning techniques. Expert Systems with Applications, 117, pp.228-242.

[11] International Journal of Recent Technology and Engineering, 2016. Prediction of Option Price using Ensemble of Machine Learning Algorithms for Indian Stock Market. 8(2), pp.3231-3241.

[12] International Journal of Innovative Technology and Exploring Engineering, 2016. Optimizing Stock Market Prediction using Long Short Term Memory. 9(1), pp.2819-2823.

[13] Finance, Markets and Valuation, 2014. Forecasting stock market trend: a comparison of machine learning algorithms. 2014(1).

[14] International Journal of Advance Engineering and Research Development, 2017. STOCK MARKET PREDICTION USING MACHINE LEARNING. 04(5).

[15] Vignesh CK, 2017. APPLYING MACHINE LEARNING MODELS IN STOCK MARKET PREDICTION. EPRA International Journal of Research & Development (IJRD), pp.395-398.

[16] Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., Min, H. and Deng, X., 2014. Empirical analysis: stock market prediction via extreme learning

machine. Neural Computing and Applications, 27(1), pp.67-78.

[17] Vijay, N., Singh, S. and Malhotra, G., 2018. Sentiment Analysis: Gauging the Effect of News on Stock Prices in Indian Stock Market. International Journal of Trade, Economics and Finance, 9(4), pp.148-152.

[18] Kulkarni, M., Jadha, A. and Dhingra, D., 2014. Time Series Data Analysis for Stock Market Prediction. SSRN Electronic Journal,

[19] Nti, I., Adekoya, A. and Weyori, B., 2013. Efficient Stock-Market Prediction Using Ensemble Support Vector Machine. Open Computer Science, 10(1), pp.153-163.

[20] Vijh, M., Chandola, D., Tikkiwal, V. and Kumar, A., 2015. Stock Closing Price Prediction using Machine Learning Techniques. Procedia Computer Science, 167, pp.599-606.

[21] HUANG, C., YANG, D. and CHUANG, Y., 2008. Application of wrapper approach and composite classifier to the stock trend prediction. Expert Systems with Applications, 34(4), pp.2870-2878.

[22] Lee, T., Cho, J., Kwon, D. and Sohn, S., 2009. Global stock market investment strategies based on financial network indicators using machine learning techniques. Expert Systems with Applications, 117, pp.228-242.

[23] Montenegro, C. and Molina, M., 2008. Improving the Criteria of the Investment on Stock Market Using Data Mining Techniques: The Case of S&P 500 Index. International Journal of Machine Learning and Computing, 10(2), pp.309-315.