2022 14th International Conference on Computational Intelligence and Communication Networks (CICN) | 978-1-6654-8771-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/CICN56167.2022.10008300

An Efficient Method to Predict the Tata-Motors Stock Price using Hybrid Machine Learning Methods

Abhishek Bajpai
Dept. of Computer Science and Engineering
Rajkiya Engineering College
Kannauj, India

Anjali Singh
Dept. of Computer Science and Engineering
Rajkiya Engineering College
Kannauj, India

Abhineet Verma
Dept. of Computer Science and Engineering
Rajkiya Engineering College
Kannauj, India

***Abstract*—Stock market analysis has always been an important aspect of every country's financial sector. As of now, various research has been done to predict the stock market prices but only considering the technical stock data. However, the problem lies in combining the technical data of stock prices and news sentiments from financial news data so that prediction can be done with much greater accuracy. In our paper, we have designed a stock price prediction system and proposed an approach in which technical stock Data is evaluated by technical means and news sentiment data is represented in the form of sentiment vectors using sentiment analysis. We have deployed Particle Swarm Optimization (PSO) to tune the hyper-parameters of the Support Vector Machine for regression (SVR), thus providing better results. We have done experiments on the Tata Motors stock price data and compared our approach with [1] who have deployed the SVM-PSO model with basic technical features taken into consideration. Our model SVR-PSO with financial news data gives a Mean Absolute Percentage Error of 0.29% as compared to the standard SVM-PSO which gives a Mean Absolute Percentage Error of 0.71%**

Keywords—Forecasting, Support vector Machine for Regression, Particle Swarm Optimization, Sentiment Analysis

I. INTRODUCTION

Stock market analysis continues to be an important aspect of every country's financial backbone and has attracted economists and researchers for a long [2]. Stock prices fluctuate based on facts about the stock and the market segment in which it is traded. However, as technology has advanced, the elements that influence this information have changed dramatically. In the beginning, information was sent over a telephone line or conveyed one-on-one; stock price data was even gathered one-on-one from every section of the country. Later on, government policies and media headlines began to have a role in influencing stock values. Along with market price data, news items began to have a significant effect on whether stock values would trend downwards or upwards. With the introduction of online technology, information is now reaching every part of the globe in nanoseconds. Furthermore, remarks made by eminent figures on their social media accounts have begun to influence stock values. For a better understanding of the statement, examine a recent example. Recently, a remark was made about cryptocurrency, and its prices skyrocketed overnight. Now, by looking at all of the characteristics together, we can observe how financial news data and social media trends are impacting prices in the crypto and stock markets. The models available have not taken into account the Financial News Data as a feature of their model and the ones which have taken this factor into account do not have a very accurate classifying model due to less precise values of their hyperparameters. Several machine learning techniques, including ANN, Gradient Boosted RegressionTrees, support vector machines for regression, and random forecast, have recently been refined using a combination of stats and learning models. These methods can show complicated non-linear patterns and some relationships that are impossible to uncover using linear techniques. These algorithms also outperform linear regressions in terms of efficacy and multicollinearity. But still, the problem lies in that these models do not include Financial News Data. Later on, seeing the trends in the stock market being affected by the financial news researchers focused their attention on the sentiments of the Financial news and verified that this data has an increasing impact on the stock market trends. But still, the problem lies in how to combine the technical data of stock prices and financial news data in the form of sentiment vectors to make the prediction more accurate. Let us understand how stock prices vary. Stock prices are highly volatile, and their value is determined by a variety of factors. Many intrinsic and extrinsic factors influence the stock price of a corporation. The development or fall of a sector as a whole is influenced by macroeconomic conditions. Company net profit, liabilities, demand stability, market competition, technically sophisticated assembly line, news, and social media sentiments, surplus cash for bad times, stakes in raw material suppliers and finished product distributors, and so on are some of the inherent elements. Extrinsic influences include crude oil prices, dollar exchange rates, political stability, and government policy decisions. Extrinsic influences are not under the control of the company. Now the parameters that we can evaluate that consider all these factors are Stock Opening Price, Stock Closing Price, Stock High Price, Stock Low Price, Stock Volume traded and Stock Average Price.These were the basic factors that are existing for a long period. With the invention of social media and internet the stock prices have started varying rigorously

978-1-6654-8771-9/22/$31.00 ©2022 IEEE
DOI: 10.1109/cicn.2022.71

depending on the financial news sentiments shared over any platform. Expectations drive financial markets [3], and investors purchase and sell stocks based on fear or desire triggered by a certain occurrence [4]. Even though the information is not always true, investors bank on the trustworthiness of media sites that broadcast rumours [5] and [6]. Investors aim for a trend in selected companies that is either extremely comparable to or greater than the related indexes. As a result, an investor needs tools that allow him or her to analyse the most relevant data for technical and fundamental research. As a result, he/she seeks to minimise risk and uncertainty while maximising rewards [7]. Investors' decisions to acquire or sell financial assets, particularly on the stock market, are influenced by information disseminated on social media [8]. Whether genuine or not, news posted on social media causes adjustments in global stock market indexes [9]. With the invention of online social media networks [10], this tendency is expected to continue. Investor responses, decision-making, learning, communication, and awareness are all influenced by emotions and sentiments [11]. Understanding emotions is crucial to personal development and progress, as well as the replication of human intelligence. The identification of polarity and emotions is a job strongly connected to the processing of emotions, which is vital for artificial intelligence growth [12]. The objective of this project is to find a more efficient and more accurate method to predict Tata Motors Stock Price for the next two months considering all the relevant features and automating the stock price prediction process. Stock market price prediction is a relatively difficult endeavour. Technological analysis is a well-liked kind of stock market research.In this paper, we are trying to integrate technical data and news sentiments as a feature vector to improve the accuracy of the existing SVM-PSO model.

The following is how the paper is organised: The previous work on this topic is mentioned in Section 2. The proposed technique is presented in section 3. The implementation specifics were detailed in section 4. The results are explained and the analysis is given in section 5. The conclusion of this study is offered in Section 6, and the scope of future work is considered.

II. Literature Review

Ping-Feng Pai et al. (2005) suggested a hybrid technique for forecasting stock prices that takes advantage of the ARIMA and SVMs models' particular strengths. The suggested model's predicting accuracy was tested using data on stock prices. Computational experiments yielded highly promising prospects [13]. Siami-Namini and companion (2018) compared ARIMA and LSTM models and concluded that LSTM performed far better as compared to ARIMA. Adil Moghar et al.(2020) used RNN and, in particular, the LSTM model to forecast time stock market values. The data set in this approach are daily opening prices from yahoo finance for two companies on the New York Stock Exchange NYSE (GOOGL and NKE). For GOOGLE, the data series span from 8/19/2004 to 12/19/2019, whereas for NKE, the data span from 1/4/2010 to 12/19/2019. The primary goal of the article was to determine the precision with which an ML algorithm can forecast and how much the epochs can enhance the model [14]. Ghosh, Pushpendu et al. (2021) proposed a multiple feature set that included not only returns concerning close prices, but also returns concerning open prices and intraday returns, and concluded that their model provides a daily return of 0.64 per cent using LSTM networks and 0.54 per cent using random forests, before transaction costs which was a great score [15]. Xiaodong Li et al. (2020) develop a stock prediction system and provide a method for converting past prices into technical indicators. These technological indicators summarise pricing data, model news sentiments with various sentiment dictionaries, and represent textual news items with sentiment vectors. These even build a 2-layer LSTM neural network to learn the serial information within market snapshots series, as well as a fully connected neural network to learn the sequential information [16]. Mohammed Siddique et al. (2018) in their research proposed combining SVR with particle swarm optimization while making a forecasting model to reduce the noise (basically predicting the missing data in the past dataset). They did so as models proposed before using only SVM gave impaired performance due to the presence of noises. Here they collected a set of technical indicators often utilized in stock market investigations and ran them through the SVR and PSO algorithms. In this model, SVR is at the core of the prediction mechanism and PSO optimizes the free parameters of SVM. The suggested approach's performance is assessed using 18 years of daily transactional data from the Bombay Stock Exchange for Tata Steel stock prices (BSE). Results reveal that the suggested model improves on the prior prediction model's performance. This technique is compared to current models with real data sets and produces more accurate findings with a MAPE(Mean Absolute Percentage Error) of 0.7 per cent (approximately) [17]. Felipe Dias Paiva et al. (2019) present a novel choice-making representation for stock market day trading investments. The model was created in such a way by combining a machine learning-based classifier with the MV method and the SVM method for portfolio selection. The S. Paulo Stock Exchange Index assets were used in the model's experimental evaluation.The impact of brokerage fees on stocks on buying and selling was specifically analyzed in the Brazilian market. The performance of the classifier, the cardinality of portfolios, and the returns and hazards of models were all investigated in the study. But still, demand for trade value may be a limiting issue for its adoption [18]. Raghavendra Kumar et al. (2021) used a combination of differential evolution(DE) and the artificial bee colony method to suggest a hyperparameter selection technique for the ARIMA model. With the combination of evolutionary algorithms and stock market time series data, modified algorithms keep the exploration and exploitation tactics. In comparison to conventional ARIMA models, the improved ABC with DE Optimization promotes higher generalisation and efficient performance. From September 1, 2010, to August 31, 2020, they ran trials on a dataset of the NSE and BSE Oil Drilling & Exploration & Refineries sector. The obtained results show that the proposed approach based on a modified ABC-ARIMA hybrid model outperforms its competitors. In multiple-step time series forecasting, the proposed technique improves forecasting accuracy while still retaining data patterns [19]. P Jiao and colleagues (2020) investigate the impact of conventional and social media coverage on stock flickering and its yield. They discovered that conventional news media coverage predicts lower fluctuating and yield in the future, but social media coverage predicts more flickerability and yield in the future. They show that these patterns are compatible with an "echo chamber" concept, in which social networks repeat news, but some investors misinterpret repeated information as true latest information [20]. Khatua et al.(2014) looked at the Ebola outbreak data and Zika epidemic outbreak data and discovered that smaller domain-specific input linguistic data from the Twitter data extract significant semantic associations better than classic prior-trained Word2Vec or GloVe [21].

III. Proposed Work

The recommended model is built using particle swarm optimization (PSO) and support vector machine for regression (SVR). The SVR is at the heart of this model's prediction process, whereas PSO optimizes the free parameters of the SVM. The most significant elements to consider in SVR are the kernel type and regularisation parameter choices. We used a radial basis function (RBF) kernel in our proposed model because of the non-linearity of the dataset under consideration. As a result, our first objective is to integrate the technical and news indicators into a single feature set based on the date so that we may use them in our model. Based on the date, the merging is done. The next step is to divide the data set into two parts: the Training Set (October 2011 to September 2013) and the Validation Set (October 2011 to September 2013). The model development approach begins by initializing the PSO and SVR parameters on the training dataset after separating the dataset into training and testing datasets. PSO is used to find the best values for SVR's hyperparameters, and the procedure is repeated until the termination criteria are met. Finally, using the optimal parameters discovered throughout the search phase, SVR is constructed and applied to the testing dataset.

A. Support Vector Regression

Let us take a vector $\vec{w}$ perpendicular to the median of the street representing two parallel vectors. The street is called a hyperplane that separates the two classes of features. For a given unknown vector $\vec{u}$ we want to predict whether it falls on the right side of the plane or the left side of the perpendicular plane. So we will project the vector $\vec{u}$ on $\vec{w}$ and hence we will predict its class as $\vec{w}.\vec{u}$ c. Without the loss of generality, it can be stated that if

$\vec{w} \,.\, \vec{u}$ + b >= 0 (c = -b) ……………… (1)

then the sample is considered to be the positive sample. This is the Decision rule of our work. Further, we need to determine $\vec{w}$. For a positive sample $\overrightarrow{x_+}$ and a negative sample $\overrightarrow{x_-}$, we have:

$$\vec{w}.\overrightarrow{x_-} + b \geqslant 1.. \dots\dots\dots\dots\dots (2)$$

$$\vec{w}.\overrightarrow{x_+} + b \leqslant 1.. \dots\dots\dots\dots\dots (3)$$

To label the samples we define a variable y_i such that $y_i = +1$ designates for positive samples and $y_i = -1$ designates negative samples.

To summarize equations (2) and (3):

$$y_i(\vec{x_i}.\vec{w_i} + b) \geqslant 1$$

$$y_i(\vec{x_i}.\vec{w_i} + b) - 1 \geqslant 0$$

and

$$Width = (x_+ - x_-) * \frac{\vec{w}}{||\vec{w}||}$$

With the substitution of values of x_- and x_+

$$Width = (1 - b + 1 + b)) * \frac{\vec{w}}{||\vec{w}||}$$

$$Width = \frac{2}{||\vec{w}||} \dots\dots\dots\dots\dots\dots (4)$$

The goal here is to maximize the width for which we can maximize, $Width = \frac{1}{||\vec{w}||}$ minimize $Width = ||\vec{w}||$ or $Width = \frac{1}{2||\vec{w}||}^2$

With the use of Lagrange Multiplier we have:

$$L = \frac{1}{2}||\vec{w}||^2 - \sum \alpha[y_i(\vec{w}.\vec{x_i} + b) - 1] \dots\dots\dots\dots (5)$$

Now differentiating w.r.t. $\vec{w}$, we get

$$L = \frac{1}{2}||\vec{w}||^2 - \sum \alpha[y_i(\vec{w}.\vec{x_i} + b) - 1]$$

$$\frac{dL}{d\vec{w}} = 0$$

$$\vec{w} - \sum \alpha_i y_i x_i = 0$$

$$\vec{w} = \sum \alpha_i y_i x_i$$

$$\sum \alpha_i y_i \$ = 0.. \dots\dots\dots\dots\dots (6)$$

Substituting in L equations obtained, we get

$$L = \frac{1}{2}(\sum \alpha_i y_i \vec{x_i})(\sum \alpha_i y_i \vec{x_j}) - (\sum \alpha_i y_i \vec{x_i})(\sum \alpha_j y_j \vec{x_j}) - \sum \alpha_i y_i b + \sum \alpha_i$$

$$L = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \vec{x_i}.\vec{x_j}$$

The value of the Kernel Function K(x_i, x_j) is equal to $\vec{x_i}.\vec{x_j}$ in the feature space $\phi(x_i)$ and $\phi(x_j)$ such that

$$K(x_i, x_j) = \phi(x_i).\phi(x_j) \dots\dots\dots\dots\dots\dots (7)$$

B. Particle Swarm optimization

PSO is an evolutionary and stochastic optimization approach inspired by nature for solving computationally difficult optimization problems. PSO is a stochastic optimization approach based on the movement and intelligence of swarms that are resilient (it solves problems in a variety of settings). It was created by James Kennedy and Russ Eberhart in 1995. It's been used to solve a wide range of search and optimization issues. PSO Algorithm implementation:

- The main idea behind PSO is to use random weighted acceleration to accelerate each particle towards the best location it has discovered so far (p-best) and the global best position (g best) acquired so far by any particle.
- This is done by simply adding the $\vec{v}$-vector to the $\vec{x}$-vector to get another $\vec{x}$ $(\vec{x_i} = \vec{x_i} + \vec{v_i})$.
  $$v_{i+1} = W * v_i + c_1 * rand(0,1)(p_{best} - x_i) + c_2 * rand(0,1) * (g_{best} - x_i)$$
  where W is the initial weight,
  g_{best} is the global best position, p_{best} is the self-best position,
  c_1 and c_2 are the acceleration coefficient.
- The particle then evaluates to its new location after computing the new X_i.
- $p_{best} = X_i$ and p-fitness = x-fitness if x-fitness is better than p-fitness.

C. Sentiment Analysis

We have studied various sentiment dictionaries, such as SenticNet 5, Vader and Loughran-McDonald Financial Dictionary 2018, some of them were built manually and some were semi-automated and in our modules, we have implemented SenticNet 5.SenticNet 5 is capable of doing very fine sentiment analysis, which includes sentiment polarity values and four very fine sentiment dimensions. Positive, neutral, or negative sentiment polarity values show whether words are positive, neutral, or negative. When the emotion polarity value is more than 0, the word's sentiment inclination is positive; when the emotion polarity value is zero, it is considered to be neutral; and when the emotion polarity value is less than 0, it is taken to be negative. If a word's polarity is greater than zero, we add the value to the positive aspect; if the word's polarity is less than zero, we add the value to the negative aspect; if the word's polarity is equal to zero, we add $\frac{1}{w_t}$ to the neutral aspect, where the length of the word vector of news NI_i is $|w_t|$.

D. Normalization

Min-max normalisation was performed at the pre-processing step to minimise numerical issues during the modelling process and to avoid the priority of features with larger numerical ranges over features with smaller numerical ranges. The following equation has been used to normalise the data by linearly scaling it to [0, 1]:

$$Norm_j = \frac{D_j - D_{min}}{D_{max} - Dmin}$$ for j = {1,2,3,........N} where D_j is the actual value of the j[th] feature, N is the total number of data sets available,
D_max, D_min are the maximum and minimum values respectively, and the Norm is the corresponding normalized value.

IV. Implementation Details

Particle swarm optimization (PSO) and support vector regression (SVR) is used to create the suggested model. SVR is at the heart of this model's prediction process, while PSO optimises SVM's free parameters. The right selection of kernel type and regularisation parameter is the most important factor to consider in SVR. Due to the nonlinearity of the dataset under consideration, we adopted a radial basis function (RBF) kernel in our suggested model. So our first task is to merge the technical and the news indicator into a combined feature set on the basis of data so that we can apply these features to our model. The combining is done based on the date. Now our next step is to split the data set into two parts i.e, Training Set (0ct 2011 to Sept 2013) and Validation Set (0ct2011 to Sept2014). After splitting the dataset into training and testing datasets, the model development procedure begins by initialising the PSO and SVR parameters on the training dataset. PSO is used to search for the optimal values of SVR's hyperparameters, and the process continues until the termination requirements are fulfilled. Finally, SVR is created and applied to the testing dataset using the optimum parameters found throughout the search procedure.

A. Technical Data

We have acquired the StockPrice Data of Tata Motors From the National Stock Exchange from 1 December 2012 to 1 December 2015 i.e., for 3 years.

B. Technical Indicators

The data set acquired comprises certain features which are described in detail in the given table (Table 1). We have taken features for each day for the building of the time-series forecasting model. We have 6 features taken from the two dimension transaction data provided by NSE which can be described in the form of the set

$$T_{NSE} = \{PrevClose_j, OpenPrice_j, HighPrice_j, LowPrice_j, Volume_j, Average_j\}$$

d = {1, 2, 3,......j}, where d is the trading day.

TABLE I. DESCRIPTION OF TECHNICAL INDICATO

S.No	Feature	Description
1	ClosePrice	The price at which the stock last trades upon closing of an exchange on a trading day.
2	OpenPrice	The price at which the stock first trades upon opening of an exchange on a trading day.
3	HighPrice	The highest price of a share on a trading day.
4	LowPrice	The lowest price of a share on a trading day.
5	Volume	Total quantity of shares traded on a trading day.

6	Average	The average price of stock is calculated by dividing the total amount invested with shares purchased.

To increase the precision of our model we have included four derived features from our side $T_{Derived} = \{M5_j, M10_j, M15_j, DIFF_j\}$, $d = \{1, 2, 3, \ldots\ldots j\}$, where d is the trading day. The Description of the indicators mentioned is given in table 2. The indicators, designated by $T_{Derived}$ are configurable by various factors such as time durations, are supposed to reflect stock patterns or swings from many perspectives, providing rich market signals to the model.

However, distinct technical indicators do not necessarily start from the same timestamp due to varying time length setups. As a result, T_j (day j) is maintained. Thus now our final feature is derived by combining T_{NSE} and $T_{Derived}$ set becomes T with 10 features.

$$T_j = \{PrevClose_j, OpenPrice_j, HighPrice_j, LowPrice_j, Volume_j, Average_j, MA5_j, MA10_j, MA20_j, DIFF_j\}$$

$d = \{1, 2, 3, \ldots\ldots j\}$, where d is the trading day.

TABLE II. DESCRIPTION OF DERIVED INDICATORS

S.No	Feature	Description
1	M5	Moving-average of 5-day closing price
2	M10	Moving-average of 10-day closing price
3	M15l	Moving-average of 15-day closing price
4	EMA6	Exponential moving-average of 5-day closing price
5	EMA10	Exponential moving-average of 10-day closing price
6	DIFF	Difference between EMA12 and EMA6

C. News Data

Corresponding news data of the same period is also taken into account. We have collected news data from Google News. Each news text is labelled with its date of publication which makes it easier to analyse news text by date.

D. Sentiment Analysis

Using sentiment dictionaries, news stories are translated into depictions. While emotion dictionaries can allow searches for specific phrases or words, news phrases on online media are typically extensive documents. Tokenization is thus required as the initial stage in the pre-processing of news items. Every news item NI is tokenized and transformed into a word vector NI = [n_1, n_x,...,n_m], with n_x being the article's x^{th} unique word. Second, it looks to see whether there are any phrases in the tokenized word vector. Using original phrases rather than tokenized words might enhance sentiment retrieval recall from dictionaries. As a result, during tokenization, word combinations are verified against keys in emotion dictionaries. Finally, the module adds negation tags to terms like "not," "no," and "neither." Finally, the module lemmatizes word variants and eliminates stop words. A particular dictionary D then maps each word vector NI into an emotion space,

$$f_D : NI- > E, \ldots\ldots\ldots (8)$$

The word vector space is NI, while emotion vector space is E. Assume that every value in the emotion vector deviates from linear superposition; each word n_x can be depicted by an emotion vector e_x.

$$e_x = f_D(n_x), \ldots\ldots\ldots\ldots (9)$$

and each NI_a can be depicted by an emotion vector E_a by adding up all e_x of each n_x because the dimension of the e_x is fixed,

$$E_k = \sum_{n_x \in NI_a} e_x \ldots\ldots\ldots\ldots (10)$$

Finally, because there is usually more than one piece of news per day, the daily news emotion vector E_y (day y) is given by proportioning $E_a'e$ from the identical day. Both the $T_y'e$ and the $E_y'e$ are calculated on a daily basis, allowing vectors to be matched by their timestamps from the identical trading day.

V. RESULT ANALYSIS

A. Evaluation Criterion

We employed three conventional statistical criteria to assess the suggested regression model's performance. The mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) are the three types of errors, and their specifics are listed in Table. Because MAE, RMSE, and MAPE all represent different types of discrepancies between actual and anticipated values, it's vital to remember that the smaller the error value, the better the performance.

1) Mean Absolute Error (MAE)

Sum of absolute differences between the actual value and the forecast divided by the number of data points available.

$$MAE = \frac{1}{N}\sum_{j=1}^{N}|y_i - d_i|$$

2) Root Mean Squared Error (RMSE)

The square root of the sum of the squared errors is divided by the number of data points available.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - d_i)^2}$$

3) Mean Absolute Percentage Error (MAPE)

The average percentage of absolute errors is divided by actual observation.

$$MAPE = \frac{1}{N}(\sum_{i=1}^{N}|\frac{y_i - d_i}{d_i}|)100$$

where, N is the total number of data points under consideration,

d_i the desired output value,

d_i and the predicted value obtained from the model.

B. Comparison Of Results

The performance of our suggested hybrid model, Modified PSO-SVR, is compared to that of the standard PSO-SVR model in this study.

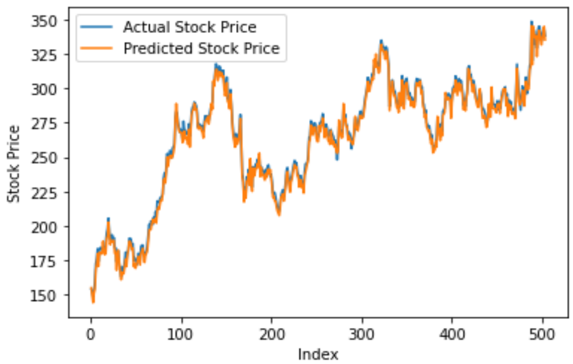

Figure 1 Actual Data Verses Predicted Data of PSO-SVR on Training Set

The PSO-SVR model is built around the Support Vector Machine for Regression (SVR) and uses Particle Swarm Optimization (PSO) to optimise the hyperparameters of SVR and takes technical as well as news indicators as features which improve its accuracy. The datasets in this study are divided into training and testing datasets, which are then used to the models for the training and testing stages of the models, respectively, in order to forecast the following day's opening price. Three-quarters of the data from Tata Motors is utilised to generate the training dataset as in (Figure 1), while the remaining one-fourth is used to build the testing dataset as in (Figure 2). The graph combing the testing and the training dataset is in (Figure 3).

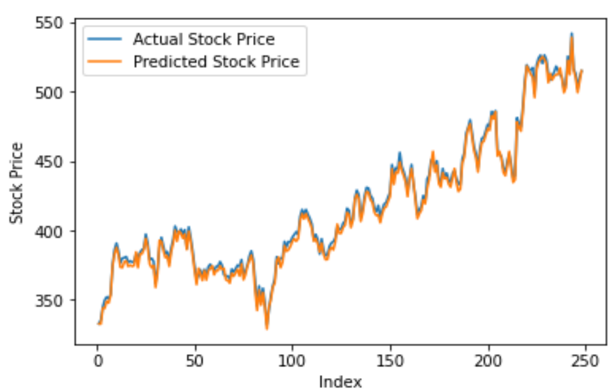

Figure 2 Actual Data Verses Predicted Data of PSO-SVR on Testing Set

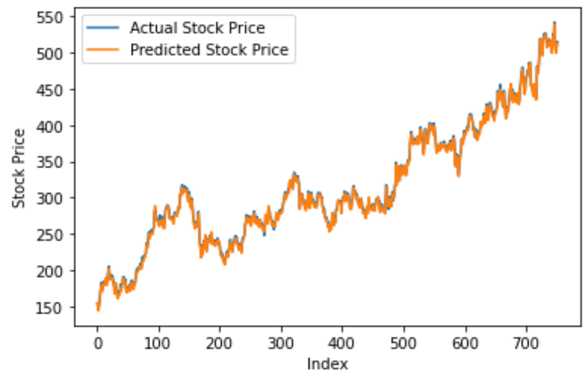

Figure 3 Actual Data Verses Predicted Data of PSO-SVR on Complete Data Set

The MAE, RMSE, and MAPE errors in the training phase are, whereas the errors in the testing phase are 2.126783923, 6.002637493, and 0.287352432%. The error measurements discovered for both models, basic PSO-SVR and PSO-SVR, are shown in Table. According to this study, PSO-SVR beat Standard SVR in all three assessment categories.

TABLE III. COMPARISON RESULTS OF DIFFERENT MODELS

	Models	Standard PSO-SVR	Modified PSO-SVR
Training	MAE	2.760213993	2.109825752
	RMSE	5.741340821	5.132452735
	MAPE	0.68994578%	0.3345268231
Testing	MAE	2.929112587	2.126783923
	RMSE	6.494903279	6.002637493
	MAPE	0.708516926%	0.287352432%

VI. Conclusion

In comparison to the conventional PSO-SVR model, our suggested methodology for solving the problem of forecasting Tata Motors stock prices has proven to be more accurate. On all three comparison metrics, MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), and RMSE, our model outperforms standard PSO-SVR (Root Mean Square Error). The standard PSO-SVR returns 0.708516926% Mean Absolute Percentage Error, however, our model returns 0.287352432% Mean Absolute Percentage Error, which is significantly better than the prior one. To summarise, we believe that our model has performed better on theoretical grounds, and we hope that our model will show to be extremely useful in the field of finance in the future.

References

[1] M. Siddique, S. Mohanty, D. Panda, A hybrid forecasting model for prediction of stock value of tata steel using support vector regression and particle479swarm optimization, International Journal of Pure and Applied Mathematicsics 119 (14) (2018) 1719–1727.481

[2] G. V. Attigeri, M. P. MM, R. M. Pai, A. Nayak, Stock market prediction: A big data approach, in: TENCON 2015-2015 IEEE Region 10 Conference,483IEEE, 2015, pp. 1–5.484

[3] R. J. Shiller, Measuring bubble expectations and investor confidence, The Journal of Psychology and Financial Markets 1 (1) (2000) 49–60.486

[4] A. Kurov, Investor sentiment, trading behavior and informational efficiency in index futures markets, Financial Review 43 (1) (2008) 107–127.488

[5] T. Oberlechner, S. Hocking, Information sources, news, and rumours in financial markets: Insights into the foreign exchange market, Journal of490economic psychology 25 (3) (2004) 407–424.491

[6] F. M. d. Cruz, M. Y. F. S. d. F. Gomes, The influence of rumors in the stock market: a case study with petrobras, Transinformation 25 (2013) 187–193.493

[7] G. Li, Information quality, learning, and stock market returns, Journal of494Financial and Quantitative Analysis 40 (3) (2005) 595–620.495

[8] C. Kearney, S. Liu, Textual sentiment in finance: A survey of methods and models, International Review of Financial Analysis 33 (2014) 171–185.497

[9] P. Jiao, A. Veiga, A. Walther, Social media, news media and the stock498market, Journal of Economic Behavior & Organization 176 (2020) 63–90.499

[10] B. Doerr, M. Fouz, T. Friedrich, Why rumors spread so quickly in social networks, Communications of the ACM 55 (6) (2012) 70–75.501

[11] S. Poria, E. Cambria, R. Bajpai, A. Hussain, A review of affective computing: From unimodal analysis to multimodal fusion, Information Fusion 37503(2017) 98–125.504

[12] E. Cambria, D. Das, S. Bandyopadhyay, A. Feraco, et al., A practical guide505to sentiment analysis, Springer, 2017.506

[13] P.-F. Pai, C.-S. Lin, A hybrid arima and support vector machines model in stock price forecasting, Omega 33 (6) (2005) 497–505.508

[14] S. Siami-Namini, A. S. Namin, Forecasting economics and financial time series: Arima vs. lstm, arXiv preprint arXiv:1803.06386.510

[15] A. Moghar, M. Hamiche, Stock market prediction using lstm recurrent neural network, Procedia Computer Science 170 (2020) 1168–1173.512

[16] P. Ghosh, A. Neufeld, J. K. Sahoo, Forecasting directional movements of513stock prices for intraday trading using lstm and random forests, Finance514Research Letters (2021) 102280.515

[17] X. Li, P. Wu, W. Wang, Incorporating stock prices and news sentiments516for stock market prediction: A case of hong kong, Information Processing517& Management 57 (5) (2020) 102212.518

[18] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka, W. M. Duarte, Decision-519making for financial trading: A fusion approach of machine learning and520portfolio selection, Expert Systems with Applications 115 (2019) 635–655.521

[19] R. Kumar, P. Kumar, Y. Kumar, Multi-step time series analysis and foreasting strategy using Arima and evolutionary algorithms, International Journal of Information Technology (2021) 1–15.52421

[20] P. Jiao, A. Veiga, A. Walther, Social media, news media and the stock525market, Journal of Economic Behavior & Organization 176 (2020) 63–90.526

[21] G. V. Attigeri, M. P. MM, R. M. Pai, A. Nayak, Stock market prediction:527A big data approach, in TENCON 2015-2015 IEEE Region 10 Conference,528IEEE, 2015, pp. 1–5.529