# A study on Prediction of Stock Market Prices Using SVM Machine Learning Technique

**Dr.P. VaraPrasad,**

*Professor of CSE, Sai Rajeswari institute of Technology, Proddatur, AP, India*

---

**Abstract -** Stock market prediction is a challenging task due to the inherent complexity and volatility of financial markets. Accurate predictions of stock prices can significantly benefit investors, traders, and financial analysts in making informed decisions. This study explores the application of Support Vector Machines (SVM), a supervised machine learning technique, for predicting stock market prices. SVM is known for its robustness in handling non-linear data and its ability to perform well in high-dimensional spaces, making it suitable for financial time series data. The study incorporates historical stock price data, including open, close, high, low, and volume, as features for model training. Various pre-processing techniques, such as data normalization and feature selection, are employed to enhance the model's performance. Additionally, the SVM model's performance is evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and prediction accuracy.

***Key Words*: Stock market prediction, Support Vector Machines (SVM), machine learning, financial forecasting, time series analysis.**

## 1.INTRODUCTION

The stock market is a cornerstone of the global economy, influencing investment strategies, financial planning, and economic policymaking. Accurately predicting stock market prices has long been a topic of interest for researchers, investors, and traders due to its potential for maximizing profits and minimizing risks. However, the prediction of stock prices is a complex task, as market movements are influenced by a myriad of factors, including historical price trends. This introduction sets the stage for an in-depth exploration of SVM as a machine learning technique for stock market prediction, outlining its strengths, challenges, and potential applications in the dynamic field of financial forecasting.The aim of this paper is to study the effectiveness of SVM in predicting stock prices using historical market data and identify which model metrics provide more accurate and reliable predictions.

Kyoung-jae Kim investigations it has been found that SVM provides a promising alternative to stock market prediction [3]. Traditional models have been used in stock market forecasting, including Support Vector Machines (SVM) [4].The use of SVM, NN, and Genetic Adversarial Network (GAN) machine learning techniques in stock market forecasting applications has been the subject of extensive research [7,8]. The data analyzer employed support vector regression (SVR) and neural network (ANN) machine learning algorithms to guess the stock market price index [9].

Kohara et al. [1] improved stock market prediction performance by incorporating prior knowledge. Extensive research [5,6] was conducted in the area of stock market forecasting applications in the using SVM, NN, and Genetic adversarial network (GAN) ML techniques. The authors examine ANN, SVM, and LSTM neural networks, stressing their unique features and useful applications, thereby emphasizing how machine learning is revolutionizing investment strategies [2].Machine learning methods, such as SVM, Random Forests, and k-Nearest Neighbours, have been explored for stock prediction.

## 2. METHODOLOGY

### 2.1 Dataset

The TESLA dataset used in this study consists of historical stock prices, from Kaggle, collected over a period of (2010-2022). The data set particulars is shown in Table 1 and Table 2.

**Table -1:**TESLA dataset

| S.No | Date | Open | High | Low | Close | Volume |
|---|---|---|---|---|---|---|
| 0 | 2010-06-29 | 3.800000 | 5.000000 | 3.508000 | 4.778000 | 93831500 |
| 1 | 2010-06-30 | 5158000 | 6.084000 | 4.660000 | 4.766000 | 85935500 |
| 2 | 2010-07-01 | 5.000000 | 5.184000 | 4.054000 | 4.392000 | 41094000 |
| ... | ... | ... | ... | ... | ... | ...1 |
| 2954 | 2022-03-23 | 979.940002 | 1040.699951 | 976.400024 | 999.109985 | 40225400 |
| 2955 | 2022-03-24 | 1009.729980 | 024.489990 | 988.799988 | 1013.919983 | 22901900 |

**Table -2:** Data columns (total 7 columns)

| # | Column | Non-Null Count | Dtype |
|---|---|---|---|
| 0 | Date | 2956 non-null | object |
| 1 | Open | 2956 non-null | float64 |
| 2 | High | 2956 non-null | float64 |
| 3 | Low | 2956 non-null | float64 |
| 4 | Close | 2956 non-null | float64 |
| 5 | Adj Close | 2956 non-null | float64 |
| 6 | Volume | 2956 non-null | int64 |

### 2.2 Data Preprocessing

The dataset was pre-processed by normalizing the stock prices to a scale between 0 and 1 to improve model training.

### 2.3 Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm that finds the hyperplane that best separates data points into distinct categories. In this study, SVM is used for regression (SVR), where it attempts to predict stock prices based on historical data. A radial basis function (RBF) kernel is used for the SVR to capture nonlinear relationships between stock prices and time.

### 2.4 Evaluation Metrics

The most common types of evaluation metrics for Machine Learning models are MSE, RMSE, MAE, and MAPE. Let's explain what each acronym means.

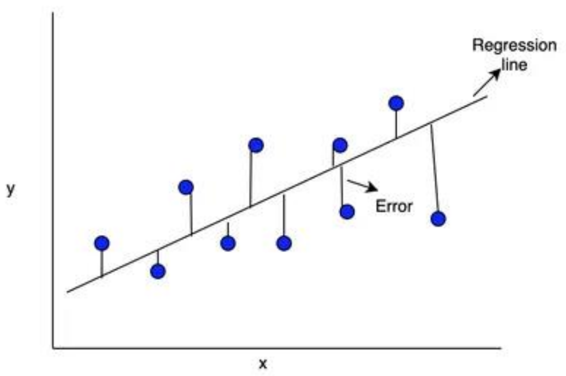


Fig 1. Linear Regression(source -Google)

Mean Squared Error (MSE): Measures the average of the squares of the errors between predicted and actual in Fig.1

$$\text{MSE}=1/n \sum_{i=1}^{n}(yi - Y\hat{}i)^2$$

n= number of observations

Yi = actual value

Y^i =Predicted or forecasted value

This metric is essential for identifying outliers in models. Due to its easy comprehensibility, RMSE is widely used to compare the performances of different models. From a geometric perspective, RMSE and MAE represent mean forms of the L2 and L1 norms, which correspond to the Euclidean distance and the Manhattan distance, respectively.

The Euclidean distance is commonly used in algorithms such as k-nearest neighbors (classification) and k-means (clustering) to find the closest points to a sample, determining their proximity based on differences along the variables.

Root Mean Squared Error (RMSE): Square root of MSE to bring the error back to the original scale of the stock prices.

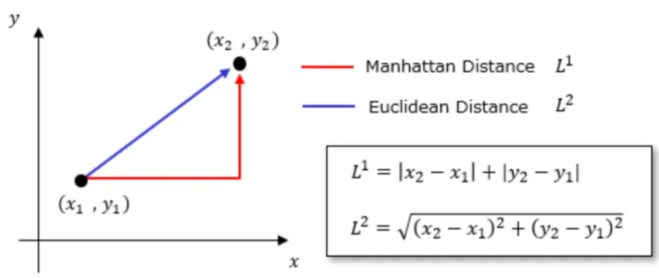


**Figure 2** Euclidean distance

practical application for the common use of RMSE, without a doubt, lies in the performance of Machine Learning models.

In simple terms, the Manhattan distance is the sum of the absolute differences between the horizontal and vertical coordinates of the points.

### Mean Average Error (MAE)

Similarly, MAE is the calculation of the average absolute differences between the expected and actual values. MAE is not as sensitive to outliers as MSE and RMSE.

A very important metric for measuring regressions and frequency distributions. Similarly to other metrics, the lower the value of MAE, the more accurate the model is.

**Mean Absolute Percentage Error (MAPE)**

MAPE calculates the average of the absolute percentage differences between the model's predictions and the actual values.

MAPE = $1/n \sum_{i=1}^{n} |(yi - Y\hat{}i)/yi|$

MAPE penalizes negative errors more heavily (when the predicted value exceeds the actual). Consequently, MAPE tends to favour models that underestimate rather than overestimate.

However, in this study we focused on MSE, RMSE and MAPE metrics in evaluation of performance of model.

## 3. DISCUSSION

The findings of this study highlight the potential of Support Vector Machines (SVM) as an effective machine learning technique for stock market price prediction. However, the results also underline several considerations, challenges, and opportunities for improvement, which are discussed in this section.

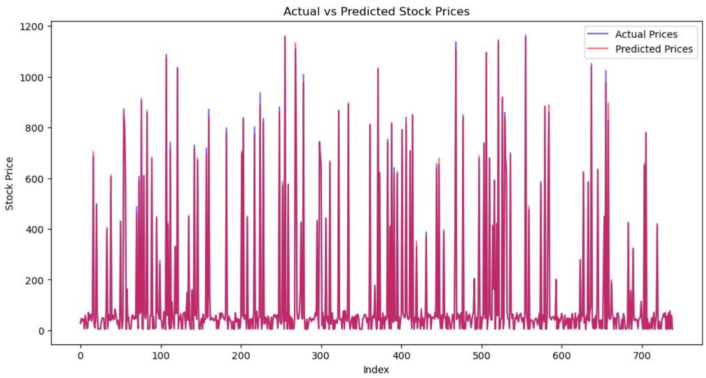


**Figure 3** The variations between Actual and Predicted Stock Prices for different index values

**Table -3:** SVR Model

| Metric | Mean Squared Error | Mean Absolute Error | Root Mean Squared Error | R-Squared ($R^2$) |
|---|---|---|---|---|
| Value | 36.46 | 1.99 | 6.03 | 0.99 |

The SVM model demonstrated notable accuracy in capturing stock market trends and predicting future prices based on historical data. Key performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared ($R^2$) indicated that SVM can effectively handle the non-linear and dynamic nature of financial data. However, the model's predictive capability was limited when applied to highly volatile stocks or during periods of sudden market fluctuations.

The Fig. 3 depicts the variations in actual price and Predicted stock price for different index values. The error values are moderate, hence the SVM model is well suitable for prediction of stock price.

The Mean Squared Error measures the average squared differences between predicted and actual stock prices. This value indicates that the squared deviations are relatively low, suggesting that the model performs well in minimizing errors.

## 4. CONCLUSION

A mean square error value of 36.46 suggests a strong performance, especially when the stock price range is high. The model's performance shows moderate accuracy but also some room for improvement, as both MAE and RMSE indicate prediction errors that could be significant depending on the stock's typical price range.

The Mean Squared Error measures the average squared differences between predicted and actual stock prices. This value indicates that the squared deviations are relatively low, suggesting that the model performs well in minimizing errors.

The Root Mean Squared Error translates the MSE back to the original units of the stock prices, making it easier to interpret. An RMSE of 6.04 means that the typical error magnitude is about 6 units. The slightly higher value compared to MAE reflects the influence of occasional larger errors, but overall, the RMSE is still very low.

The Root Mean Squared Error translates the MSE back to the original units of the stock prices, making it easier to interpret. An RMSE of 6.04 means that the typical error magnitude is about 6 units. The slightly higher value compared to MAE reflects the influence of occasional larger errors, but overall, the RMSE is still very low.

## REFERENCES

[1] K. Kohara, T. Ishikawa, Y. Fukuhara, Y. Nakamura, Stock price prediction using prior knowledge and neural networks, Int. J. Intell. Syst. Accounting Finance Manage. 6 (1) (1997) 11–22.

[2] Chhajer, P.; Shah, M.; Kshirsagar, A. The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. Decis. Anal. J. 2021, 2, 100015.

[ 3] Kyoung-jae Kim, 2003. Financial time series forecasting using support vector machines, Neurocomputing 55 (1-2) 307–319.

[4] R. Tsaih, Y. Hsu, C.C. Lai, Forecasting S& P 500 stock index futures with a hybrid AI system, Decision Support Syst. 23 (2) (1998) 161–174.

[5] Ahmad Waqas, Analyzing different ML techniques for stock market prediction, Int. J. Comput. Sci. Inform. Sec. 12 (2014) 12–17.

[6] H. Yang, L. Chan, I. King. Support vector machine regression for volatile stock market prediction. In International Conference on Intelligent Data Engineering and Automated Learning 2002 Aug 12 (pp. 391-396). Springer, Berlin, Heidelberg.

[7] Ahmad Waqas, Analyzing different ML techniques for stock market prediction, Int. J. Comput. Sci. Inform. Sec. 12 (2014) 12–17.

[8] H. Yang, L. Chan, I. King. Support vector machine regression for volatile stock market prediction. In International Conference on Intelligent Data Engineering and Automated Learning 2002 Aug 12 (pp. 391-396). Springer, Berlin, Heidelberg.

[9] H.R. Patel, S.M. Parikh, D.N. Darji. Prediction model for the stock market using news based different Classification, Regression, and Statistical Techniques: (PMSMN). In2016 International Conference on ICT in Business Industry & Government (ICTBIG) 2016 Nov 18 (pp. 1-5). IEEE.

## BIOGRAPHY

Dr. P. Vara Prasad has worked as a teacher for more than 20 years. He is now working as a professor of CSE at Sai Rajeswari Institute of Technology in Proddatur, AP. He received his PhD in CSE from Glocal University after completing his B.Tech and M.Tech degrees in CSE at JNT University. His areas of interest in research are deep learning and machine learning.