iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40857

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Stock Market Prediction Using Machine Learning

Srihitha Podduturi[1], Geohnavi Marskatla[2], Dr. SN Chandrashekhar[3], Mr. P. Anvesh[4], Ms. E. Lavanya[5], Mr. P. Pradeep Kumar[6], Dr. S.P.V. Subba Rao[7]

[1, 2]*Student, Department Of Electronics & Communication Engineering Sreenidhi Institute Of Science & Technology Yamnampet, Ghatkesar, Hyderabad–501301, India.*

[3, 4, 5, 6]*Guide, Associate Professor, Department Of Electronics & Communication Engineering Sreenidhi Institute Of Science & Technology, Yamnampet, Ghatkesar, Hyderabad–501301, India.*

[7]*Professor& HOD, Department Of Electronics & Communication Engineering Sreenidhi Institute Of Science & Technology, Yamnampet, Ghatkesar, Hyderabad–501301, India.*

Abstract: ***Wise use of financial predictions plays a vital role in facilitating investment decisions for many investors. With the right information, investors can adjust their portfolio to maximize returns while minimizing risk. However, not all investments guarantee a good return, and this is mainly due to the fact that many investors have limited knowledge and skills to predict stock trends. However, the complexity and turmoil of the stock market, make any prediction efforts extremely difficult. This paper aims to provide a comprehensive review of the emerging research related to the use of Mechanical Learning and In-depth Learning models in the field of financial market forecast. To prepare for this task, more than sixty research papers have been thoroughly analyzed to extract the much needed information, application, and results of the various methods. It is found in this project that Intensive Reading is the most successful Automatic Reading in all the research papers collected, and is the most appropriate way to use the stock market prediction domain.***

I. INTRODUCTION

Predicting the stock index and stock price is difficult because of the uncertainty involved. There are two types of analysis performed by investors before investing in stocks. First is the basic analysis. In this regard, investors consider the value of the shares, the performance of the industry and the economy, the political situation etc. to decide whether to invest or not. On the other hand, technical analysis evaluates stocks by reading statistics generated by market activity, such as previous prices and volumes. Performance analysts do not attempt to measure internal security value but use stock charts to identify patterns and trends that may influence how the stock will behave in the future. Malkiel and Fama's active market hypothesis (1970) states that stock prices work best with information which means it is possible to predict stock prices based on trading data. This makes sense as in many uncertain aspects such as the political situation of the country, the image of the company in public will begin to appear in stock prices. So, if the information obtained from stock prices is pre- processed efficiently and appropriate algorithms are applied then trend of stock or stock price index may be predicted.

Over the years, a number of methods have been developed to predict stock trends. Initially ancient methods of retreat were used to predict stock trends. Since stock data can be categorized as incorrect time series data, indirect machine learning methods have also been used. Linear Regression (LR) and Random Forest Regressor (RFR) are two widely used machine learning algorithms to predict stock index movements and stock values. Each algorithm has its own way of learning patterns. Linear Regression (LR) mimics the activity of our brain so that we learn about building a network of neurons. Hassan, Nath, and Kirley (2007) proposed and implemented an integration model by combining the Hidden Markov Model (HMM), Linear Regression (LR) and Genetic Algorithms (GA) to predict financial market behavior. Through ANN, daily stock prices are converted into independent sets of values into HMM inputs. Wang and Leu (1996) developed a predictive system that is useful for predicting mid-term price trends in the Taiwan stock market. Their system was based on an ongoing neural network trained using features extracted from the ARIMA analysis.

II. METHODOLOGY

A. Problem Description

The results showed that networks trained using 4-year data were able to predict up to 6 weeks of market trends with acceptable accuracy. Soft computerized integrated strategies for automated stock market predictions and trend analysis were presented by Abraham, Nath, and Mahanti (2001). They used the Nas-daq-100 stock market index of Nasdaq stocks with a one-day pre-existing stock forecast and a neuro-fuzzy system to analyze the predictable stock price trend.

©IJRASET: All Rights are Reserved

Predictions and forecast forecasts using the proposed hybrid system were promising. Chen, Leung, and Daouk (2003) investigated a neutral network (PNN) to predict index direction after training historical data. The results showed that PNN-based investment strategies received higher returns than other investment-tested investment strategies such as buy and hold and investment strategies guided by predictable estimates of random movement model and parametric GMM models.

B. Algorithm Used

1) Linear Regression

Linear Regression Model Representation

Linearity is a line model, e.g. a model that captures the linear relationship between input variables (x) and single output variations (y). Specifically, that y can be calculated from the line combination of input variables (x). If there is one single input (x), the method is called a simple line reverse. When there is a lot of variation in the input, letters from the numbers often refer to the method as a lot of line deflection.

Different techniques can be used to prepare or train the number of line deviations from the data, the most common being called Extraordinary squares. It is therefore common to refer to a model modified in this way such as the Ordinary Least Least Linear Regression or simply the Reduction of Small Squares.

a) Simple Linear Regression: With a simple line break if we have one input, we can use calculations to measure coefficient.

This requires you to calculate mathematical properties from data such as methods, standard deviations, correlations and compliance. All data must be available to cut and calculate statistics.

b) Ordinary Least Squares

If we have more than one input, we can use Common squares to estimate the value of coefficients.

The Ordinary Least Squares process seeks to reduce the total amount of square waste. This means that when looking at a data backbone we calculate the distance from each data point to the regression line, re- square, and then add all the square errors together.

c) Gradient Descent: If there is one or more inputs you can use the process to improve the coefficients values by repeatedly reducing the model error in your training data. This function is called the Gradient Descent and works first with random values in each coefficient. The total number of square errors is calculated for each pair of input and output values. The level of learning is used as a scale factor and the coefficient is revised when going towards reducing error. The process is repeated until a square sum error is reached or no further improvements are possible.

2) Regularization: There are extensions for line model training called orientation techniques. These seek to both reduce the number of square model error in training data (using smaller standard squares) but also to reduce the complexity of the model (such as the number or total size of all coefficients in the model).

Two popular examples of regularization procedures for linear regression are:

a) Lasso Regression

b) Ridge Regression

3) Preparing Data for Linear Regression

Line retrieval is read at great length, and there are many books on how your data should be structured to use the best model.

a) Line Guess: Drop down line assumes that the relationship between your input and output is linear. It does not support anything else. This may seem like a lot, but it's good to remember when you have a lot of qualities.

b) Remove Noise: Line reversal assumes that your input and output output are no audio.

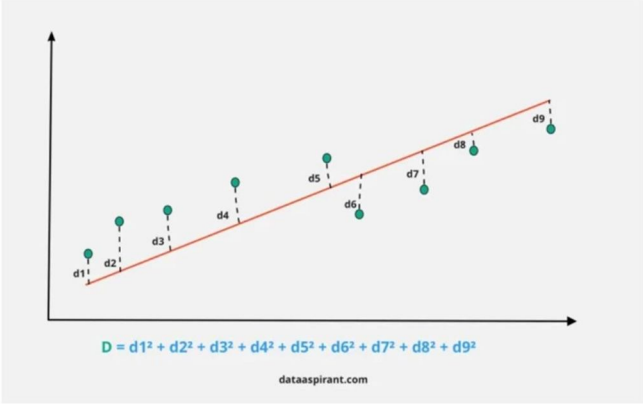
c) Remove Collinearity: Line retrieval will greatly measure your data if you have a variety of closely related input.

d) Gaussian Distribution: The reversal of the line will make reliable predictions if your input and output variations have a Gaussian distribution.

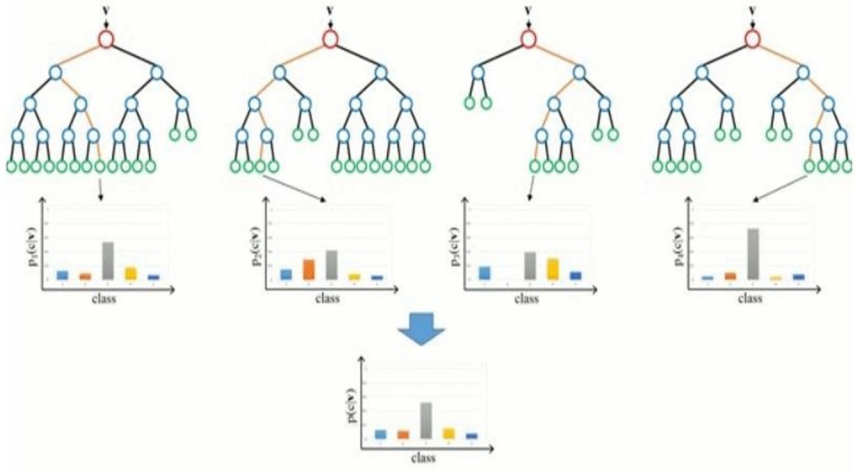
e) Relocation Input: Line reversal will usually make reliable predictions when you re-adjust variable inputs using stop or normalization.

4) *Lasso Regression:* The decrease in Lasso is similar to the decline of the line, but uses a process "decrease" in which the cut-off coefficients decrease towards zero. The linear regression gives you the drop-down coefficients as seen in the database. Lower lasso allows you to reduce or make these coefficients the same to avoid overcrowding and make them work better on different databases.

The Statistics of Lasso Regression:



5) *Ridge Regression:* Ridge regressor is actually a standard version of Linear Regressor. that is, in the initial cost function of the line regressor we can add a common term that forces the learning algorithm to fit the data and help keep the weights as low as possible. The standard term has an 'alpha' parameter that controls the normal operation of the model i.e., it helps to reduce the variability in measurements.
6) *Random Forest Regressor:* As machine learning professionals, we encounter a variety of machine learning algorithms that we can use to build our model. In this article I will try to give you an intuition of ideas on how a random forest works from the beginning.
7) *Ensemble Learning:* Suppose you want to watch a web series on Netflix. You will simply log in to your account and watch the first webisode appear or you will browse a few web pages, compare ratings and make a decision. Yes. That is very likely you will go with the second option again instead of making a direct conclusion you will consider other options as well.



As machine learning practitioners, we come across a wide array of machine learning algorithms that we may apply to build our model. In this article I will try to give you an intuition of the concepts of how a Random Forest model works from scratch.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

III. MODELING AND ANALYSIS

A. *Data Set*

1) *Series:* Type of security
2) *Prev Close:* Previous day's close price
3) *Open:* Open price of day
4) *High:* Highest price in day
5) *Low:* Lowest price in day
6) *Last:* Last traded price in day
7) *Close:* Close price of day
8) *VWAP:* Volume Weighted Average Price

B. *Contet*

Stock market data is widely analyzed for educational business and personal interests.
Content
Update Frequency
Data on price history and trading volumes of 50 stocks on the NIFTY 50 index from the NSE (National Stock Exchange) India. All data sets are day-to-day where prices and trading prices are divided into all .cvs files per stock and a metadata file with some great information about the stocks themselves. Data runs from January 1, 2000 to April 30, 2021.
Since new stock market data is generated and made available every day, in order to have thelatest and most useful information, the dataset will be updated once a month.

C. *Acknowledgements*

NSE India: https://www.nseindia.com/
Thanks to NSE for providing all the data publicly.

D. *Inspiration*

Various machine learning techniques can be applied and explored to stock market data,especially for trading algorithms and learning time series models.

IV. RESULTS AND DISCUSSION

Most paper only considers a model in a single market place and despises the app in international markets and some papers did not compare both machine reading and deep learning methods in the same market with the same database. After weighing the advantages and disadvantages of both methods, it can be concluded that Deep Learning has more advantages compared to Mechanical Learning models. Moreover, to date, the development of 80% or 90% accuracy detection methods is still a myth, this is not only due to static, non-linear data set in a complex and complex environment but also relevant. due to the influence of external factors such as national finance, monetary policies and many unforeseen and unexpected external factors.

	Open	High	Low	Last	Close	VWAP	Volume	Turnover
0	74.35	74.75	71.40	74.75	74.75	73.20	286260.0	2.095440e+12
1	73.05	78.50	NaN	73.25	73.05	73.38	296264.0	2.173890e+12
2	70.00	73.50	67.50	70.00	69.50	70.85	227624.0	1.612790e+12
3	71.00	74.00	69.55	69.75	70.05	72.04	275149.0	1.982120e+12
4	69.00	72.50	66.00	67.00	67.40	68.72	138809.0	9.538880e+11

©IJRASET: All Rights are Reserved

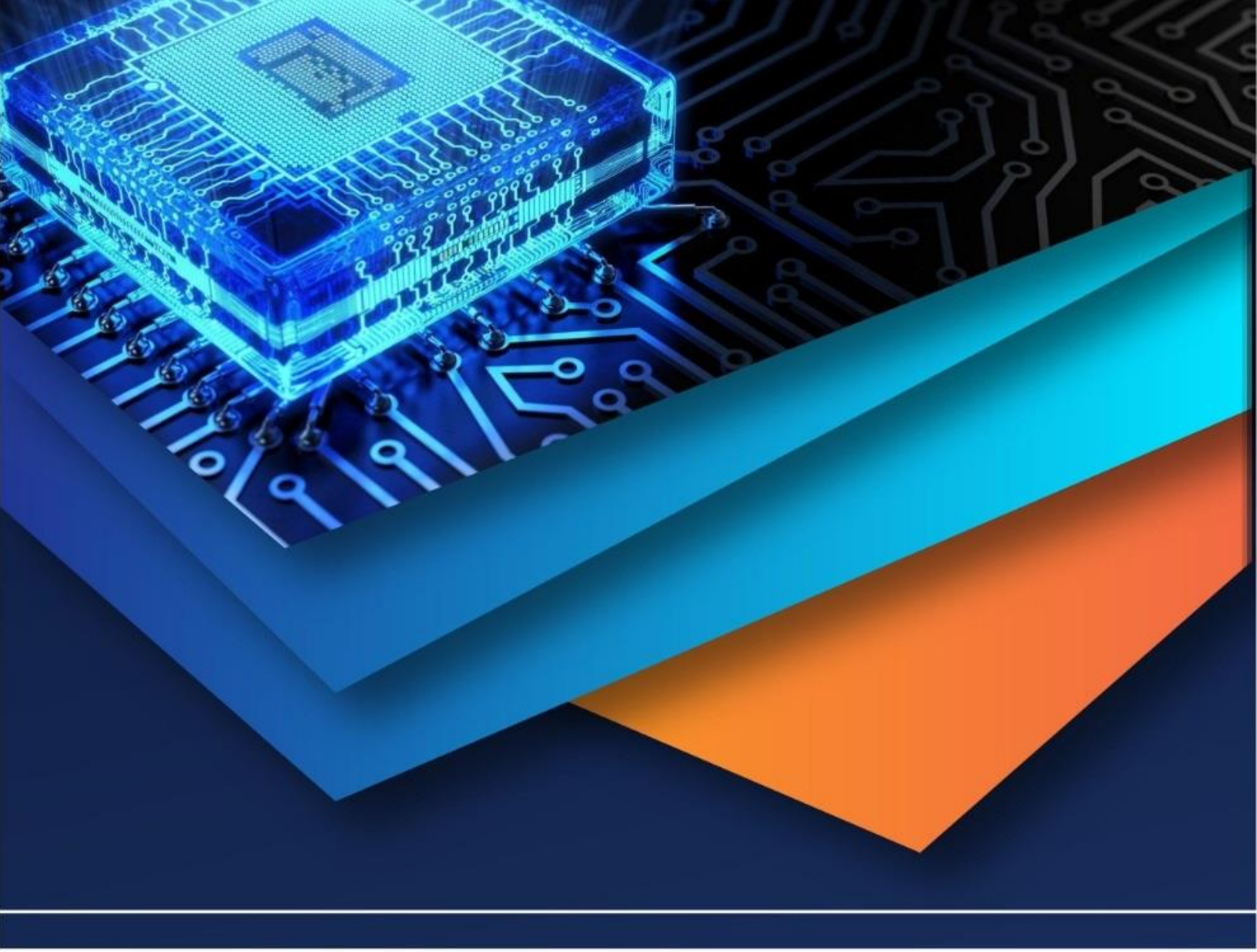
	Algorithm	Score	R2_Score
0	Linear Regression	0.912581	0.912581
1	Ridge Regression	0.912581	0.912581
2	Random Forest Regressor	0.992057	0.912581
3	Lasso Regression	0.912075	0.912075

V. CONCLUSION

This paper reads a very interesting topic in the financial market, which is the stock price forecast. Because of the many external factors ranging from unpredictable to unpredictable, it is almost impossible to make predicting future price movements. Fortunately, in recent years, the field of forecasting analysis, in particular, the field of Machine Learning and Intensive Learning has exploded with many new models and techniques that can be used to predict stock market trends. It has been found that in the field of machine learning, the leading runner is SVM due to its ability to convert low-dimensional space data into high-resolution and linear fragmentation. As research shows that the stock data set is sound and linear, SVM is the best solution although the design and algorithm can be complex and time consuming. In recent years, in the field of Deep Learning, various models developed based on the structure of Neural Networks have been presenting at the financial market platform. Models like Linear Regression, Random Forest Regressor, Lasso and Ridge regressor are well known among all financial market analysts. These neural network models have been proven to have better performance and predictable accuracy. Outflow displays on different charts show predictions similar to targeted stock movements with a very low error rate.

REFERENCES

[1] Abraham, A., Nath, B., & Mahanti, P. K. (2001). Hybrid intelligent systems for stock market analysis. In Computational science-ICCS 2001 (pp. 337–345). Springer.
[2] Ahmed, S. (2008). Aggregate economic variables and stock markets in India. International Research Journal of Finance and Economics, 141–164.
[3] Araújo, R. d. A., & Ferreira, T. A. (2013). A morphological-rank-linear evolutionary method forstock market prediction. Information Sciences, 237, 3–17.
[4] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees.CRC press.
[5] Chen, A.-S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: Forecasting and trading the taiwan stock index. Computers & OperationsResearch, 30,901–923.
[6] Garg, A., Sriram, S., & Tai, K. (2013). Empirical analysis of model selection criteria for genetic programming in modeling of time series system. In 2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr) 739 (pp. 90–94). IEEE
[7] Han, J., Kamber, M., & Pei, J. (2006). Data mining: Concepts and techniques. Morgankaufmann.
[8] Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of hmm, ann and ga for stock market forecasting. Expert Systems with Applications, 33, 171–180.
[9] Hsu, S.-H., Hsieh, J., Chih, T.-C., & Hsu, K.-C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. Expert Systems with Applications, 36, 7947– 7951.
[10] Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement 748 direction with support vector machine. Computers & Operations Research, 32, 749 2513–2522.
[11] Kara, Y., Acar Boyacioglu, M., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange. Expert systems with Applications, 38, 5311–5319.
[12] Khemchandani, R., Chandra, S., et al. (2009). Knowledge based proximal support vectorKim, (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55, 307–319.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)