Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EMPIRICAL INVESTIGATIONS TO PREDICT STOCK PRICE USING REGRESSION TO IMPROVE THE TRENDS OF INORMATION TECHNOLOGY

[1]G. RESHMA, [2]SUDARSHAN TUMKUNTA, [3]TGAYATHRI, [4]T.V. HYMA LAKSHMI, [5]N. NEELIMA, [6]GARIGIPATI RAMA KRISHNA, [7]BORRA BHAVITHA

[1]Dept. of Information Technology, PVP Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh,
[2]Technical Program Manager, Meta, Bellevue, WA, USA
[3]Dept. of CSE, Shri Vishnu Engineering College for Women, Bhimavaram, Andhra Pradesh, India
[4]Department of ECE, SRKR Engineering College, Bhimavaram, India
[5] Dept. of Information Technology, RVR&JC College of Engineering, Guntur, India
[6]Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
[7]Student, Indian Institute of Sciences, Bangalore

Email: borrabhavitha@gmail.com

ABSTRACT

accumulating riches by astute investment—which doesn't! In actuality, numerous trading, financial, and even technological firms have been actively researching stock market movements and stock price prediction. Machine learning techniques have been used to generate a number of algorithms for stock price prediction. Here, we'll concentrate on mastering a number of well-known regression methods, such as support vector regression, regression tree, regression forest, and linear regression, and using them to solve this billion-dollar problem.

The prediction's primary goal is to lessen the ambiguity surrounding investment decision-making. Following the random walk, the stock market suggests that the best forecast for tomorrow's value is today's value. Investors' beliefs are impacted by the stock market indices' extreme volatility. Because of the fundamental characteristics of the financial industry and, in part, the combination of known and unknown factors, stock prices are thought to be highly volatile and subject to abrupt fluctuations. There have been various attempts to predict stock price utilizing Machine Learning. Each research endeavour has a very different focus in three different ways. Target price changes might be short-term (tomorrow to a few days later), long-term (months later), or near-term (less than a minute). Less than ten specific companies, stocks in a specific industry, or almost all stocks can be included in the collection. A worldwide news and economic trend, specific firm attributes, or just time series data of the stock price can all be employed as predictors.

Keywords: *Stock, Regression, Trends, IT, Price.*

1. INTRODUCTION

In essence, highly wealthy quantitative traders purchase stocks and stock derivatives at low prices and then sell them at high prices. Although the tendency in stock market predictions is not new, several groups continue to discuss this problem. Before making an investment, investors can do one of two types of stock analysis. The first is called fundamental analysis, and it looks at the companies' inherent value as well as the performance of the economy, industry, political environment, etc. to determine whether or not to make an investment.

Technical analysis, on the other hand, describes how stocks have changed over time by examining the data produced by market activity, such as historical prices and volumes. The growing importance of machine learning across a range of industries in recent years has encouraged many traders to use machine learning techniques in their work, some of which have shown very encouraging outcomes.

The financial data predictor algorithm that will be developed in this research will use a dataset that contains all of the past stock prices as training data.

2. PROBLEM STATEMENT

To use machine learning regression techniques to forecast stock prices, particularly stock index prices.

Existing System: In the current system, we tend to suggest that internal communication patterns predict a company's performance, as measured by the movement of its stock value. We tend to believe that in order to avoid attainable adversities that an organization could face in the securities market and to protect stakeholders' interests as much as possible, it is essential to detect patterns in company communication networks earlier in order to predict serious stock worth movement and obtain early warning signals. Very little work has been done in this crucial area, despite the data's potential significance for business communication.

1.1Disadvantages:

• Setting more stock market movement levels, however, might reduce accuracy.
• For "two levels," "three levels," and "five levels," the average prediction accuracies utilizing Decision Tree as the classifier are 43.44%, 31.92%, and 12.06%, respectively.
•These findings suggest that using a standard classifier leads in unpredictable stock prices.

1.2Proposed System:

When predicting the stock market, accuracy is crucial. Even though there are numerous algorithms available for this purpose, the key to achieving the best results is still choosing the most accurate one. This research compares and analyzes the performance of several available algorithms, including Random Forest, SVM, and logistic regression, among others, in order to do this. This entails training the algorithms, running them, analyzing the outcomes, comparing their many performance metrics, and ultimately determining which is the most accurate.

The following are the advantages

• The buyer will gain the most from the accurate prognosis.

We have covered a number of prediction algorithms in this work.
•In this research, we trained various machine learning algorithms using stock data of five businesses from the Huge Stock Market dataset, which includes data from 2011 to 2017. Thus, we evaluated the precision of various machine learning techniques.

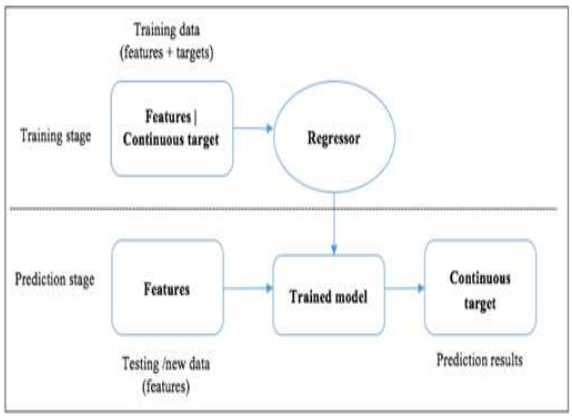

Fig 1. Training and Prediction

3. Design Functionalities

The Design Functionalities can be viewed as

- Importing the dataset.
- Getting the different regression techniques.
- Choosing one regression technique.
- Predicting the stock price.
- Checking for the efficiency.
- Performing error matrices on the Regression techniques.
- Proving the regression technique, we took works efficiently

Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

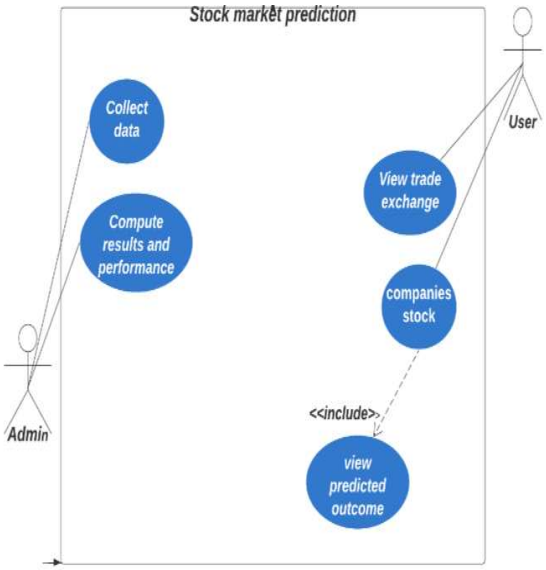

Fig 3.1 Design of Stock market prediction

4.IMLEMEMNTATION STEPS:

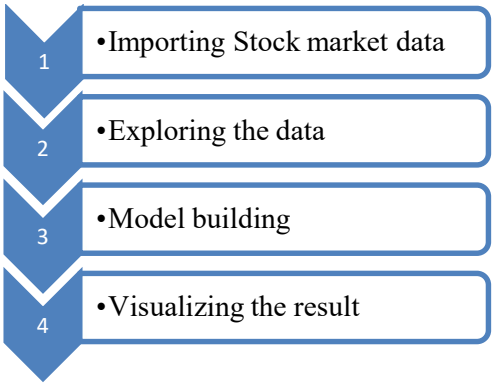

Fig 4.1. Steps of Implementation

4.1: Importing the stock price dataset:

- We used Stock Price history data from the Quadl and apply a regression analysis method.

```
start_date = datetime.date(2009, 3,8)
end_date = datetime.date.today()
# Load data from Quandl
data = quandl.get('FSE/SAP_X',
start_date=start_date, end_date=end_date)
# Save data to CSV file
data.to_csv('data/sap_stock.csv')
```

4.2: Exploring the data:

Sort the datasets into training and testing sets, allocating 20% to the testing set and 80% to the training set. The data set is now analysed to look for the following graph:

4.3: Model Building:

The data at hand and presumptions about different training procedures must be carefully considered when developing machine learning models that can generalize well on future data. The final assessment of a machine learning model's quality necessitates the proper choice and interpretation of evaluation criteria.

Algorithms used in machine learning are capable of automating the process of creating analytical models. Machine learning models enable computers to uncover hidden insights from Big Data without explicit programming by using algorithms that iteratively learn from data. As a result, numerous machine learning-based applications have emerged.

The output for the above code snippet appears as:

	Date	Close	Prediction
14	2009-02-27	25.52	22.290642
24	2009-03-13	25.73	22.603368
40	2009-04-06	27.65	23.103731
93	2009-06-18	28.84	24.761183
258	2010-02-04	33.80	29.921175
301	2010-04-08	35.94	31.265900
385	2010-08-04	35.40	33.892805
417	2010-09-17	36.96	34.893530
530	2011-02-25	43.33	38.427343
631	2011-07-20	41.21	41.585883
1010	2013-01-14	61.06	53.438228
1176	2013-09-06	53.81	58.629493
1283	2014-02-11	57.21	61.975669
1406	2014-08-06	57.67	65.822208
1484	2014-11-25	56.90	68.261477
1550	2015-03-04	63.07	70.325474
1631	2015-07-01	63.63	72.858561
1774	2016-01-18	71.22	77.330554

Fig 4.1 Generate array with predicted values

4.4 : Visualizing the output:

We take a 25 random integers and estimate the closing cost.

It is done as follows

```
# Generate 25 random numbers
randints = np.random.randint(2550, size=25)
# Select row numbers == random numbers
```

```
df_sample = df[df.index.isin(randints)]
```

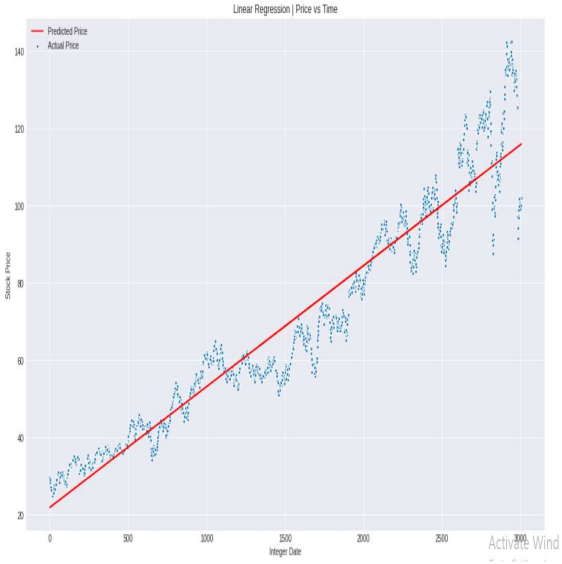

Fig 4.2 Estimate the closing cost.

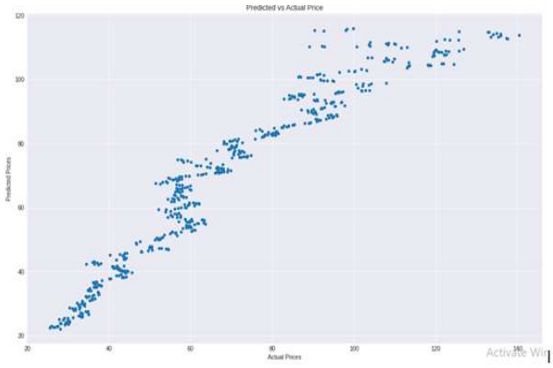

Fig 4.3 Predicted vs Actual Price

5: TESTING THE OUTPUT:

When referring to machine learning models, the term "testing" is often used to refer to evaluating the model's accuracy and precision. It should be highlighted that the definitions of "testing" in traditional software development and machine learning model creation are different. From the standpoint of quality assurance, machine learning models would also need to be tested like traditional software development. Therefore, methods like black box and white box testing would also be applicable to machine learning models in order to conduct quality control checks on them. It appears that test engineers and QA specialists may find employment in the artificial intelligence sector.

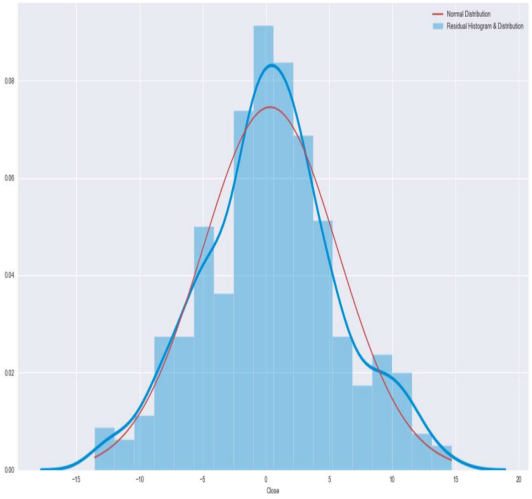

Fig 5.1Normal distribution to the data

	Date	Close	Prediction
0	2009-02-09	29.64	21.852824
1	2009-02-10	28.88	21.884097
2	2009-02-11	29.00	21.915369
3	2009-02-12	28.45	21.946642
4	2009-02-13	28.58	21.977915
...	...	...	...
3003	2020-11-25	99.04	115.764674
3004	2020-11-26	99.74	115.795947
3005	2020-11-27	100.10	115.827219
3006	2020-11-30	101.70	115.858492
3007	2020-12-01	101.90	115.889765

3008 rows × 3 columns

Fig 5.2Normal distribution to the data

5.2: Error Evaluation:

Metrics:

- **Mean Absolute Error (MAE)** is the mean of the absolute value of the errors:
- **Mean Squared Error (MSE)** is the mean of the squared errors:
- **Root Mean Squared Error (RMSE)** is the square root of the mean of the squared errors:
-

Journal of Theoretical and Applied Information Technology
31st December 2024. Vol.102. No. 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Mean Absolute Error (MAE):

It is a measurement of the discrepancies between two observations that show the same thing. Examples of Y against X include comparisons of predicted versus observed, subsequent time versus starting time, and one technique of measuring versus another technique of measurement.

Mean Squared Error (MSE)

MSE of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value. MSE is a risk function, corresponding to the expected value of the squared error loss. The fact that MSE is almost always strictly positive (and not zero) is because of randomness or because the estimator does not account for information that could produce a more accurate estimate. The MSE is a measure of the quality of an estimator—it is always non-negative, and values closer to zero are better.

Root Mean Squared Error (RMSE)

The difference between values (sample or population values) predicted by a model or estimator and the observed values is commonly measured using RMSE. The square root of the average of squared errors is known as RMSE.

```
[135] # Calculate and print values of MAE, MSE, RMSE
      print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
      print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
      print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))

      Mean Absolute Error: 5.838195968709102
      Mean Squared Error: 56.498889606672944
      Root Mean Squared Error: 7.516574330150201

[136] print('R2: ', metrics.r2_score(y_test, y_pred))

      R2:  0.9234959960240222
```

Fig 5.2 Metrics

6. RESULTS:

Stock Market graph of a company:

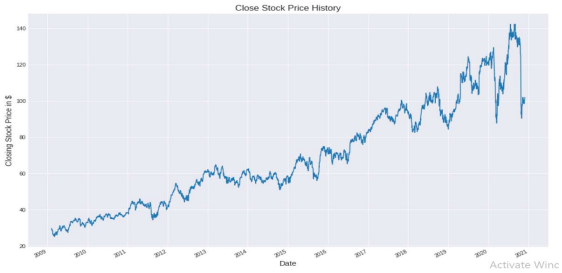

Figure6.1: Stock Market

Predicted outcomes for 25 values:

	Date	Close	Prediction
14	2009-02-27	25.52	22.290642
24	2009-03-13	25.73	22.603368
40	2009-04-06	27.65	23.103731
93	2009-06-18	28.84	24.761183
258	2010-02-04	33.80	29.921175
301	2010-04-08	35.94	31.265900
385	2010-08-04	35.40	33.892805
417	2010-09-17	36.96	34.893530
530	2011-02-25	43.33	38.427343
631	2011-07-20	41.21	41.585883
1010	2013-01-14	61.06	53.438228
1176	2013-09-06	53.81	58.629493
1283	2014-02-11	57.21	61.975669
1406	2014-08-06	57.67	65.822208
1484	2014-11-25	56.90	68.261477
1550	2015-03-04	63.07	70.325474
1631	2015-07-01	63.63	72.858561
1774	2016-01-18	71.22	77.330554

Fig 6.2.Predicted outcomes for 25 values

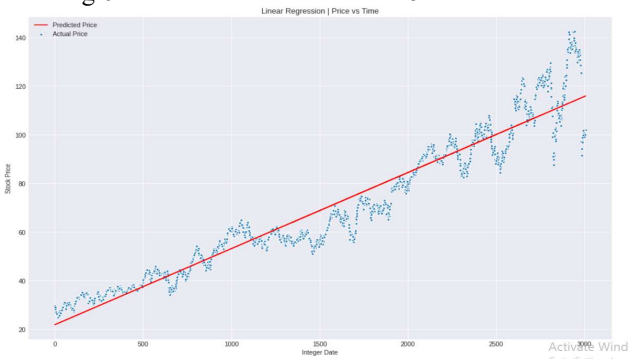

Fig 6.3. Linear Regression

7: CONCLUSION AND FUTURE SCOPE:

Thus based on the available training sets given to the system our system will be able to predict the stock prices of the given new set of data, which would be helpful to the investor in knowing the advantages and disadvantages before investing. So, this serves the purpose of our requirement and our system will also give the previous graph of the

particular dataset also because of which the investor is aware of the company's previous record and could easily predict the future stock prices of the company which helps him in getting profits as he knows everything prior to his investment on a particular company.

And the future scope of this system is on predicting the Stock prices using LSTM, which when used produce the better results as compared to the existing ones.

REFERENCES:

[1] B. Liu, Sentiment Analysis(Introduction and Survey) and Opinion Mining. 2012.

[2] X. Lei, X. Qian, and G. Zhao, "Rating Prediction Based on Social Sentiment from Textual Reviews," IEEE Trans. Multimed., vol. 18, no. 9, pp. 1910–1921, Sep. 2016, doi: 10.1109/TMM.2016.2575738.

[3] Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," in Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications, ICIEA 2016, Oct. 2016, pp. 1383–1388, doi: 10.1109/ICIEA.2016.7603801.

[4] R. Majethia, V. Mishra, A. Singhal, K. Lakshmi Manasa, K. Sahiti, and V. Nandwani, "PeopleSave: Recommending effective drugs through web crowdsourcing," in 2016 8th International Conference on Communication Systems and Networks, COMSNETS 2016, Mar. 2016, doi: 10.1109/COMSNETS.2016.7440000.

[5] R. C. Chen, Y. H. Huang, C. T. Bau, and S. M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection," Expert Syst. Appl., vol. 39, no. 4, pp. 3995–4006, Mar. 2012, doi: 10.1016/j.eswa.2011.09.061.

[6] J.-C. Na and W. Y. M. Kyaing, "Sentiment Analysis of User-Generated Content on Drug Review Websites," J. Inf. Sci. Theory Pract., vol. 3, no. 1, pp. 6–23, Mar. 2015, doi: 10.1633/jistap.2015.3.1.1.

[7] M. E. Basiri, M. Abdar, M. A. Cifci, S. Nemati, and U. R. Acharya, "A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques," Knowledge-Based Syst., vol. 198, p. 105949, Jun. 2020, doi: 10.1016/j.knosys.2020.105949.

[8] S. Vijayaraghavan and D. Basu, "Sentiment Analysis in Drug Reviews using Supervised Machine Learning Algorithms," arXiv, Mar. 2020, Accessed: Nov. 20, 2020. [Online]. Available: http://arxiv.org/abs/2003.11643.

[9] V. Doma et al., "Automated Drug Suggestion Using Machine Learning," in Advances in Intelligent Systems and Computing, Mar. 2020, vol. 1130 AISC, pp. 571–589, doi: 10.1007/978-3-030-39442-4_42.

[10] A. A. Hamed, R. Roose, M. Branicki, and A. Rubin, "TRecs: Time-aware twitter-based drug recommender system," in Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012, 2012, pp. 1027– 1031, doi: 10.1109/ASONAM.2012.178.

[11] C. Chen, L. Zhang, X. Fan, Y. Wang, C. Xu, and R. Liu, "A epilepsy drug recommendation system by implicit feedback and crossing recommendation," in Proceedings - 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCo, 2018, doi: 10.1109/SmartWorld.2018.00197.

[12] D. Chen, D. Jin, T. T. Goh, N. Li, and L. Wei, "ContextAwareness Based Personalized Recommendation of AntiHypertension Drugs," J. Med. Syst., vol. 40, no. 9, pp. 1– 10, Sep. 2016, doi: 10.1007/s10916-016-0560-z.

[13] A. Gottlieb, G. Y. Stein, E. Ruppin, R. B. Altman, and R. Sharan, "A method for inferring medical diagnoses from patient similarities," BMC Med., vol. 11, no. 1, p. 194, Sep. 2013, doi: 10.1186/1741-7015-11-194.

[14] K. Shimada, K. Fujikawa, K. Yahara, and T. Nakamura, "Antioxidative Properties of Xanthan on the Autoxidation of Soybean Oil in Cyclodextrin Emulsion," 1992. Accessed: Jul. 29, 2020. [Online]. Available: https://pubs.acs.org/sharingguidelines.

[15] Q. Zhang, G. Zhang, J. Lu, and D. Wu, "A framework of hybrid recommender system for personalized clinical prescription," in Proceedings - The 2015 10th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2015, Jan. 2016, pp. 189–195, doi: 10.1109/ISKE.2015.98.