Comparitive analysis of Stock Price Prediction: A study of Linear, Lasso, Ridge Techniques

Rachana N S

*Artificial Intelligence and Data science Global Academy Of Technology* Bengaluru,India [rachanans08@gmail.com](mailto:rachanans08@gmail.com)

Pankaja H Y

*Artificial Intelligence and Data science Global Academy Of Technology* Bengaluru,India [pankajadas9114@gmail.com](mailto:pankajadas9114@gmail.com)

Rohini B R

*Artificial Intelligence and Data Science Global Academy of Technology* Bengaluru,India

[rohinibr@gat.ac.in](mailto:rohinibr@gat.ac.in)

Trupthi Rao

Artificial Intelligence and Data Science *Global Academy of Technology* Bengaluru,India

[trupthirao@gat.ac.in](mailto:trupthirao@gat.ac.in)

Ashwini Kodipalli

Artificial Intelligence and Data Science Global Academy Of Technology Bengaluru,India [dr.ashwini.k@gmail.com](mailto:dr.ashwini.k@gmail.com)

***Abstract*- The prediction of stock prices has garnered significant attention from investors, financial analysts, and researchers alike. Accurate forecasting of stock movements not only mitigates investment risks but also supports strategic decision-making and enhances potential returns. In this study, we explore the application of three regression techniques—Linear Regression (LR), Ridge Regression, and Lasso Regression—for forecasting stock prices based on historical data. Key features incorporated into the models include opening price, closing price, highest and lowest prices, trading volume, and adjusted closing prices. These models are trained and validated on past market data to uncover patterns conducive to prediction. Performance evaluation is conducted using error metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), and Root Mean Square Error (RMSE) to assess the accuracy and robustness of each model.**

***Keywords- Stock price prediction, loss functions, linear regression, ridge regression, lasso regression.***

1. Introduction

Indian stock market is rapidly growing and widely spread offering both good chances and obstacles for predicting stock prices. India has mainly two primary stock markets: Bombay Stock Exchange(BSE) being the oldest stock exchange and National Stock Exchange(NSE) being the largest financial market. Stock markets are import aspect for economy. ‘Stock’ are not characterized as constant, as it is dynamic and volatile. It has equal probability of pros and cons for people investing. It is like an investment for investors . In the past, many traditional techniques were used for prediction of the stock price ,but these approaches or technique did not provide appropriate accuracy. As of now, there are so many machine learning algorithms, deep learning model and statistical methods that are used for predicting stock prices. Some machine learning(ML) algorithms and techniques like, LSTM(long short-term memory) that is the best suitable for prediction of stocks as it analyses time series patterns. SVR(Support Vector Regression)mainly predicts using training values and target variables. Similarly, SVM (Support Virtual Machine)is used for predicting with higher precision.

The concept of stock market comes into the picture when a public limited company wants to raise the fund for it’s further expansion. First of all initially it enters the market by IPO(initial public offering) where people apply for the considerable amount of shares through NSE or BSE resulting it ends up in secondary market. The major difference between

primary market and secondary market is the transaction of shares inbetween two, In primary market the transaction takes place between issuer company and the investor whereas in secondary market the transaction takes place inbetween two investors, such that companies like Zerodha, Angel one and other online apps act like broker where the transaction of stocks happens under depositories. Investors are broadly classified into 4, RII(Retail individual investors who invests less than 2 lakh, HNI(high networth individual) who invests more than 2 lakh, DII(Domestic institutional investors like ICC, LIC and FII(Foreign portfolio investors.

1. LITERATURE SURVEY

The research about the stock prediction is conducted by[1] used Gaussian process(GP) and SMOreg to predict stock prices and achieved efficient results by applying GP and SMOreg algorithms to predict stock prices of five companies in the telecommunication sector. The best result is produced by SMOreg than gaussian process with RMSE value of 0.000005, MAPE 1.88% and MBE 0.00025. A similar study by [7],where researchers have used linear regression,SMO regression and random forest algorithms . After the experimental analysis of each model, it depicted that SMOreg has smallest error value compared to other models. It is concluded that SMOreg ha spoor behaviour in studying sequential data and causing overfitting. The researched conducted by [2] used SVM algorithm. It has mainly centralized on enhancing and improving model according to the overall trend of stock prices. The accuracy obtained is higher using this model and finds applications in the stock market. In [3],researchers have adapted techniques like LSTM(long short -term memory) and SVR(Support Vector Regression).It is stated that,LSTM is compared with Svr using various stock index data such as S&P 500,NSE,NYSE,BSE,NASDAQ.The evaluation metrics can be used for evaluation of prediction accuracy. It is deduced that LSTM has better accuracy compared to SVR.Furthermore, the research conducted by[4] has done prediction of stock prices by comparing Gradient Boosting Algorithms and Naïve Bayes Algorithm. The Gradient Boosting Algorithms novel loss function, is based on the prior stock prices that in turn helps in reducing total prediction error. The accuracy obtained are very accurate where Gradient Boosting algorithm is 4.6% more accurate than Naïve Bayes

Algorithm. A study by[5] used regression and classification algorithms. It has been stated that, the regression model is used to predict the closing price of the stock ,where classification is used to predict whether the closing price of the stock will increase or decrease. Maximum mean of accuracy is obtained by using Logistic Regression model.The research performed by[6] used the classification techniques, Linear Regression and Long short-term memory .It is stated that,the performance of the models is measured in terms of Bombay Stock Exchange(BSE).After obtaining experimental results,it is inferred that linear regression model performed better than Long short- term memory.[8]perception of investor’s profit is seen through company’s stock prices which changes in interval of time periodically using d event-driven LSTM model and recent, upcoming news and media drives increment and decrement in stock prices. [9] uses Deep Reinforcement Learning (DRL) algorithms to predict the profit by analysing the company’s assets and describes about solving Partially Observed Markov Decision Process (POMDP) problems. [10] Broad learning system used enables efficient and fast learning whereas other deep learning models and Pearson correlation coefficient finds the relationship between two variables which increases the prediction accuracy of the above algorithm since stocks are highly volatile. [11] Here linear Regression has been employed to figure out the relationship between predictors and stock prices which is linear and has been compared with support vector regression which is non-linear and exists between other factors influencing stock prices and respective stock prices which results in optimal solution. ELM(Extreme Learning Machine)’s speed of training superfast which makes model to learn in one test used by [13] ensures continuous learning from available online that is very much required in fields which changes rapidly as new data arrives. Sometimes traditional algorithms cannot be practical and very slow so to overcome that metaheuristic search algorithms are designed to search through large solution spaces which has been employed by [15] to find optimal solution as it deals with time consuming problem and high cost problems in financial data analysis. In [14] some of used algorithms are Deep Reinforcement Learning (DRL), Markov Decision Process (MDP), Twin Delayed Deep Deterministic Policy Gradient (TD3), Q-learning as they kind of set up autonomous agent to make real time decision and even sentiment analysis has been deployed to know the present news about stocks through text data. In [12] Morphological Similarity Clustering, Hierarchical Temporal Memory (HTM) has been used to discover the trend of the stock price not just prices from the distance metric and then stock price is predicted using 2nd algorithm after that K-means clustering is used to classify the similar data into different partitions.

Above discussed and many more are used to deal with data analysis and machine learning minimizing manual effort with unique insights and increases accuracy of prediction and enables trading automation as dealing with technical indicators and real time charts, candlesticks are complex.

1. PROBLEM STATEMENT

Accurate and precise stock price prediction remains problematic due to inconsistency and influencing factors such as economic, social factors. Traditional approaches often exhibited limited prediction strength, mainly in capturing short-term fluctuations. This study mainly aims for applying various machine learning models which are efficient and robust in improving stock price prediction accuracy.

1. METHODOLOGY

We have implemented the selected dataset using various ML(machine learning) algorithms. We applied all these algorithm’s in python using sklearn.linear\_model.



**DATA GATHERING**

**DATA PREPROCESSING**

**CHOOSING MODEL**

**TRAINING MODEL**

**HYPERPARAMETER**

**TUNING**

**EVALUATION**

Fig 1: Fundamental steps of machine learning

1. *Data Gathering*

Data is gathered from Kaggle dataset. The dataset comprises of 1091 samples with 7 features, where 6 are input features and 1 output feature. Six input attributes are used to depict the output attribute. The features used in this dataset are: Input features:

* + Date
  + High- Highest trading value at specific window.
  + Low- Lowest trading value at specific window.
  + Open- Starting transaction value at specific window.
  + Volume- Number of transactions at specific window.
  + Adj Close- Closing value after adjustment for different events.

Output feature:

* + Close- Last transaction value at specific window

1. *Data pre-processing*

Data preprocessing is a vital step in machine learning before selecting a suitable model. The dataset used in this study exhibited no missing values. Outliers were detected within the ‘Volume’ attribute.

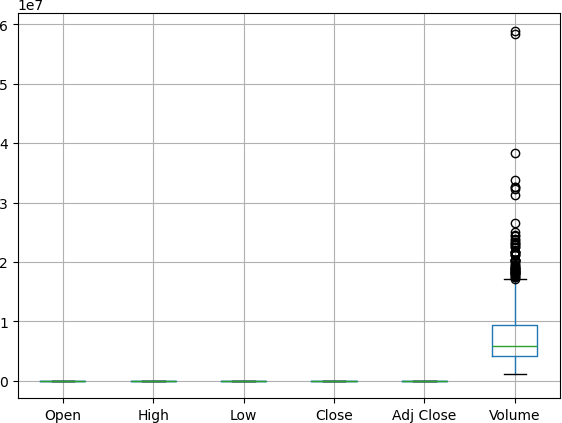


Fig 2 : Outliers Representation

It is detected using Interquartile range (IQR) method, visually represented through boxplot to highlight outliers present in the dataset.

1. *Choosing the model*

Three ML algorithms: 1) Linear Regression 2) Lasso Regression 3) Ridge Regression are used as distinct ML models in this research.

1. *Linear Regression*

This model is used to find the best linear relationship between a dependent or output variable and one or more independent variables. Hence, the value of the dependent variable can be predictable using independent variables.

The equation is given by,

y=β0 + β1x1 + β2x2 + …. + βnxn + ϵ

in underfitting so while choosing λ going for moderate one would be better. Along with predicting Lasso interprets the data with high volume and takes out the meaningful insights.

1. *Ridge Regression*

Ridge regression mainly focuses on multicollinearity i.e. when correlation between independent variables are quite high. It does shrinkage as that of Lasso Regression but features of some correlations aren’t set to zero thus sustains all the features. The only difference is adding penalty which is proportional to sum of squared co- efficient as represented in the below equation:

Minimize (∑(yi-(β0+∑βjxij))2+λ∑β 2)

j

This type of shrinkage is also referred as L2 norm where, yi dependent variable

Xij  independent variable

Β0 intercept form

βj  coefficient of independent variable

λ regularization parameter (penalty controller) sum of squared co -efficient

∑ β 2sum of squared co-efficient

j

Unlike, linear regression it reduces unstable coefficients but not exactly to zero as that of Lasso regression.

1. *Elastic Net Regression*

Elastic Net is a hybrid regularization technique that combines the penalties of both Lasso (L1) and Ridge (L2) regression. It is particularly useful when there are multiple correlated features, as it allows for both feature selection and coefficient shrinkage.

The objective function is given by:

Minimize [∑𝑛 (𝑦i –(𝛽0 + ∑𝑝 𝛽jxij))2 + λ1∑𝑝 |βj| +

Here,

ydependent variable

x1, x2, …, xnindependent variable

Here,

𝑖=1

𝑗=1

λ2∑𝑝 𝛽 2]

𝑗=1 j

𝑗=1

β1,β2,…,βn relationship between dependent and independent variables

Therefore, linear regression predicts values from the relationship between two variables which increases the accuracy of value y.

1. *Lasso Regression*

Least absolute shrinkage and selection operator regression adds penalty to the regression model which avoids overfitting of the data and it is also known for its regularization and feature selection when it comes to dealing with multiple features in huge dataset

The equation is given by,

Minimize (∑(yi-(β0+∑βjxij))2+λ∑|βj|)

Here, the penalty proportional is lasso which avoids large coefficients by adding penalty (L1 norm) and λ is shrinkage which brings the co - efficient of less important features to zero in order reduce the noise which enables model to capture underlying data. If λ has very less value it indicates that there’s little chance of penalty for the co - efficient where nearing to 1 may indicate huge penalty which might end up

yi actual output for sample i

xij  value of feature j in sample i

𝛽0 intercept

βj coefficient of feature j λ1penalty for L1(Lasso) λ2penalty for L2(Ridge)

1. *Polynomial Regression*

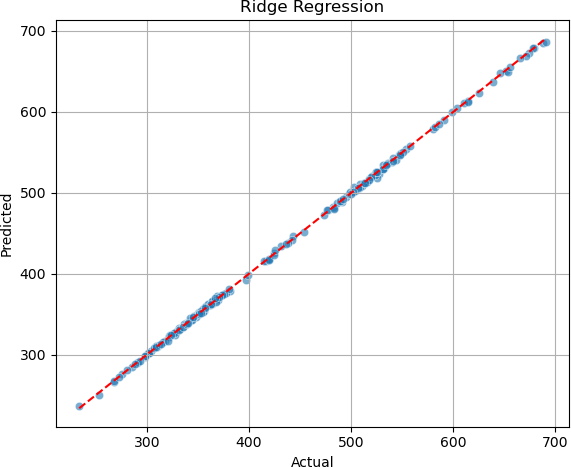
Polynomial Regression is a type of regression analysis where the relationship between the independent variable x and the dependent variable y is modeled as nth-degree.

The general formula of a polynomial regression of degree n is ,

y= β0 + β1x + β2x2 + β3x3 + …. +βnxn + ϵ

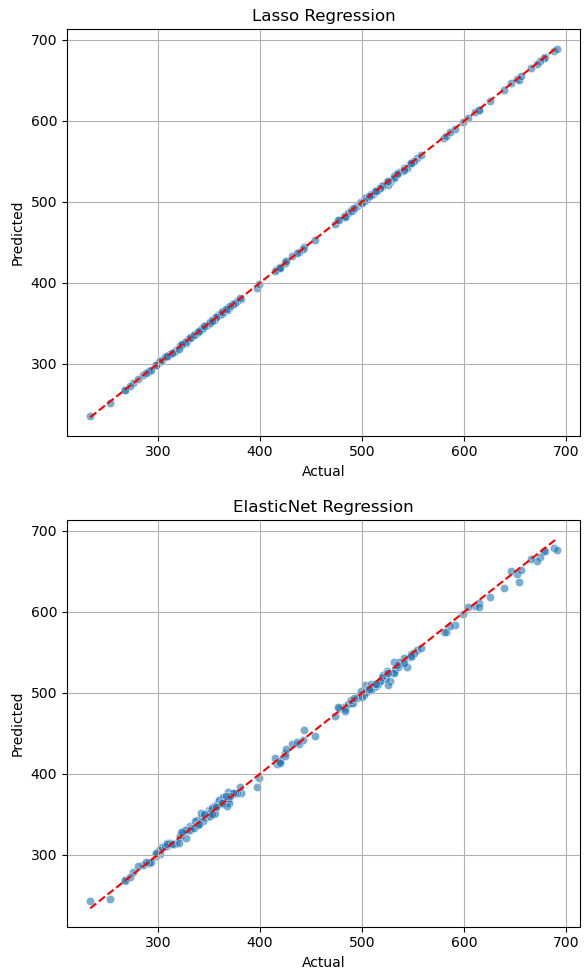
where,

y dependent variable xindependent variable β0,β2,…,βnregression coefficients ϵ error term

1. ***Training the selected models*

Training the model is a crucial step in machine learning. Based on our study, the models are trained using training dataset which comprises 70% of the total dataset, and the remaining 30% dataset is used as validation and test datasets. The main criteria for training the model relies on reducing the loss functions like Mean square error (MSE), Mean Absolute Error (MAE) etc. between the predicted and actual stock prices.

1. *Hyperparamter Tuning*

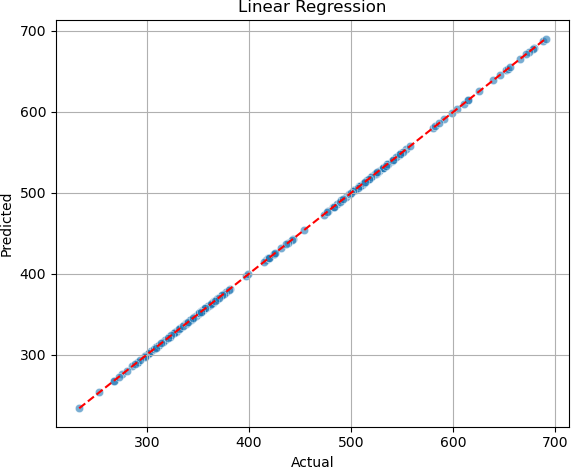
For optimizing the set of hyperparameters of the ML models, hyperparameter tuning is approached. Hyperparameter tuning is essential for improving the model’s performance. For linear, lasso, ridge regression models, the goal is selecting the appropriate ’alpha’ value that balances model’s prediction accuracy of closing price of the stock.

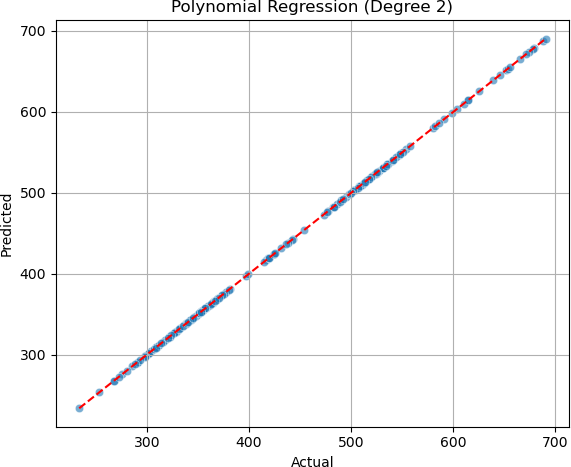
1. *Evaluation*

Evaluating the model’s performance on the unseen data is very much significant for model to be best. For evaluation, model is subjected to the unseen data, which is test dataset. . Evaluation metrics such as MSE, MAE, RMSE etc. are the performance measures for obtaining prediction accuracy. The output of the model is to predict the closing price of stock with contributing features such high, low, volume, open, adj close etc

1. Results

**Polynomial Regression (Degree 2)** performed the best. Its predicted values almost perfectly align with the actual values along the reference line, indicating minimal error and excellent generalization. **Linear Regression** also performed very well, slightly behind Polynomial, due to the largely linear nature of the dataset. **Ridge** and **Lasso** showed slightly more deviation, suggesting that regularization did not offer significant improvement here. **ElasticNet** had the most scattered points, making it the least accurate in this comparison. Therefore, **Polynomial Regression** is the most suitable model for this dataset, especially in capturing subtle non-linear patterns while maintaining high accuracy.





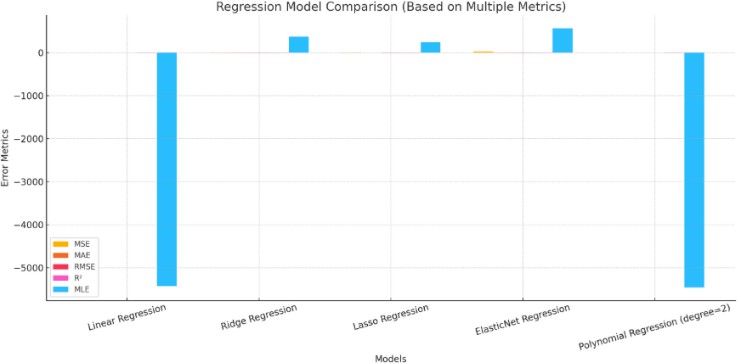
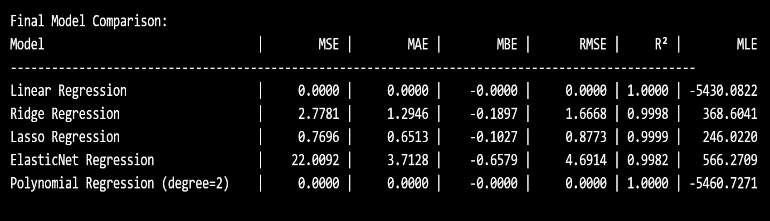


Fig 3: Linear regression, Lasso regression, Ridge regression models Fig 4: Loss functions



1. Conclusion

Machine learning techniques was used in this paper for improving the prediction of stock price. In this paper, the training and testing data size taken was 70:30. Model performance is evaluated using loss functions such as Mean square error (MSE), Root mean square error (RMSE), Mean absolute error (MAE), Mean bias error (MBE), Mean logarithmic error (MBE) etc. Out of all models, the polynomial regression model outperformed well than the others.

References

[1]J. H. Moedjahedy, R. Rotikan, W. F. Roshandi and J. Y. Mambu, "Stock Price Forecasting on Telecommunication Sector Companies in Indonesia Stock Exchange Using Machine Learning Algorithms," 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), Manado, Indonesia, 2020, pp. 1-4, doi: 10.1109/ICORIS50180.2020.9320758. keywords: {Companies;Prediction algorithms;Stock markets;Forecasting;Predictive models;Share prices;Gaussian processes;Forecasting;Stock Price;Gaussian Process;SMOreg;Telecommunication Sector},

1. Z. Liu, Z. Dang and J. Yu, "Stock Price Prediction Model Based on RBF- SVM Algorithm," 2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC), Chongqing, China, 2020, pp. 124-127, doi: 10.1109/ICCEIC51584.2020.00032. keywords: {Support vector machines;Predictive models;Prediction algorithms;Market research;Reliability;Stock markets;Investment;support vector machine;stock price;prediction model},
2. G. Bathla, "Stock Price prediction using LSTM and SVR," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 2020, pp. 211-214, doi: 10.1109/PDGC50313.2020.9315800. keywords: {Stock markets;Deep learning;Long short term memory;Support vector machines;Indexes;Measurement;Logic gates;Stock Price prediction;Deep Learning;LSTM;RNN;SVR},

[4]P. V. Reddy and S. Magesh Kumar, "A Novel Approach to Improve Accuracy in Stock Price Prediction using Gradient Boosting Machines Algorithm compared with Naive Bayes Algorithm," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 695-699, doi: 10.1109/ICAC3N56670.2022.10074387. keywords: {Supply and demand;Pricing;Boosting;Prediction algorithms;Naive Bayes methods;Reliability;Stock Price;Prediction;Naive Bayes;Gradient Boosting Algorithm},

1. S. Ravikumar and P. Saraf, "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154061. keywords:

{Companies;Indexes;Support vector machines;Machine learning;Stock markets;Regression tree analysis;Stock prices;stock market;machine learning;regression;classification;confusion matrix},

1. C. Ebenesh and K. Anitha, "A Novel Approach to Minimize the Mean Square Error in Predicting Stock Price Index using Linear Regression in Comparison with LSTM Model," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 1365-1370, doi: 10.1109/ICSCDS53736.2022.9760764. keywords: {Analytical models;Linear regression;Mean square error methods;Predictive models;Market research;Prediction algorithms;Robustness;Linear Regression;Long Short-Term Memory;Novel Stock Price Prediction;Machine Learning;Classification},
2. M. Lutfi, S. P. Agustin and I. Nurma Yulita, "LQ45 Stock Price Prediction Using Linear Regression Algorithm, Smo Regression, And Random Forest," 2021 International Conference on Artificial Intelligence and Big Data Analytics, Bandung, Indonesia, 2021, pp. 1-5, doi: 10.1109/ICAIBDA53487.2021.9689749. keywords:

{Measurement;Fluctuations;Linear regression;Predictive models;Big Data;Prediction algorithms;Artificial intelligence;Stock Price;Predictions;Algoritma linear regression;SMO regression;dan random forest},

1. Q. Li, J. Tan, J. Wang and H. Chen, "A Multimodal Event-Driven LSTM Model for Stock Prediction Using Online News," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 10, pp. 3323-3337, 1 Oct.

2021, doi: 10.1109/TKDE.2020.2968894.

keywords: {Companies;Media;Tensors;Stock markets;Predictive models;Solid modeling;Computational modeling;Stock prediction;tensor;multimodality;deep learning;LSTM},

1. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K.

P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," 2017 International Conference on Advances in

Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, pp. 1643-1647, doi: 10.1109/ICACCI.2017.8126078. keywords:

{Companies;Time series analysis;Logic gates;Predictive models;Forecasting;Data models;Machine learning;Time series;Stock market;RNN;LSTM;CNN},

1. G. Li, A. Zhang, Q. Zhang, D. Wu and C. Zhan, "Pearson Correlation Coefficient-Based Performance Enhancement of Broad Learning System for Stock Price Prediction," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 69, no. 5, pp. 2413-2417, May 2022, doi: 10.1109/TCSII.2022.3160266.

keywords: {Correlation;Stock markets;Predictive models;Machine learning;Learning systems;Forecasting;Training;Broad learning

system;machine learning;Pearson correlation coefficient;time series forecasting;complex system},

1. B. Panwar, G. Dhuriya, P. Johri, S. Singh Yadav and N. Gaur, "Stock Market Prediction Using Linear Regression and SVM," 2021 International

Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 629-631, doi: 10.1109/ICACITE51222.2021.9404733. keywords: {Support vector machines;Training;Machine learning algorithms;Linear regression;Supervised learning;Prediction algorithms;Data models;SVM(State Vector Machine);SVR(State Vector Regression);Linear Regression(LR)},

1. X. Wang, K. Yang and T. Liu, "Stock Price Prediction Based on Morphological Similarity Clustering and Hierarchical Temporal Memory," in IEEE Access, vol. 9, pp. 67241-67248, 2021, doi: 10.1109/ACCESS.2021.3077004.

keywords: {Predictive models;Data models;Time series analysis;Prediction algorithms;Clustering algorithms;Task analysis;Robustness;Machine learning;kmeans;morphological similarity distance;hierarchical temporal memory;stock prediction},

[13]C. Xiao, D. Sutanto, K. M. Muttaqi, M. Zhang, K. Meng and Z. Y. Dong, "Online Sequential Extreme Learning Machine Algorithm for Better Predispatch Electricity Price Forecasting Grids," in IEEE Transactions on Industry Applications, vol. 57, no. 2, pp. 1860-1871, March-April 2021, doi: 10.1109/TIA.2021.3051105.

keywords: {Training;Forecasting;Electricity supply industry;Uncertainty;Prediction algorithms;Extreme learning machines;Artificial neural networks;Electricity market;electricity price;extreme learning machine;online training;regression analysis},

1. T. Kabbani and E. Duman, "Deep Reinforcement Learning Approach for Trading Automation in the Stock Market," in IEEE Access, vol. 10, pp. 93564-93574, 2022, doi: 10.1109/ACCESS.2022.3203697.

keywords: {Stock markets;Mathematical models;Supervised learning;Markov processes;Approximation algorithms;Q- learning;Portfolios;Deep learning;Reinforcement learning;Sentiment analysis;Autonomous systems;Autonomous agent;deep reinforcement learning;MDP;sentiment analysis;stock market;technical indicators;twin delayed deep deterministic policy gradient},

1. M. Alsulmi, "Reducing Manual Effort to Label Stock Market Data by Applying a Metaheuristic Search: A Case Study From the Saudi Stock Market," in IEEE Access, vol. 9, pp. 110493-110504, 2021, doi: 10.1109/ACCESS.2021.3101952.

keywords: {Labeling;Stock markets;Task analysis;Predictive models;Search problems;Manuals;Simulated annealing;Automatic labeling;machine learning;metaheuristic search algorithms;stock data labeling;stock market trading},