Stock Market Prediction Using LASSO Regression

[1]Divya S, [2]SowmyaR, [3]Shnega sri T, [4]Shreenidhi T A

[1]Professor, [2,3,4] Students
Department of Information Technology,
K.L.N College of Engineering,
Sivagangai, Tamil Nadu, India

***Abstract*—Accurate prediction of stock market returns is a very challenging task due to the volatile and non-linear nature of the financial stock markets. With the introduction of artificial intelligence and increased computational capabilities, programmed methods of prediction have proved to be more efficient in predicting stock prices. The financial data: Open, High, Low, and Close prices of stock are used for creating new variables used as inputs to the model. The system is developed by a machine learning algorithm such as lasso regression. The models are evaluated using standard strategic indicators: RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error). The low values of these two indicators show that the models are efficient in predicting stock closing prices.**

***Index Terms*—Artificial Intelligence, Lasso regression, Root Mean Squared Error, Mean Absolute Error.**

I. INTRODUCTION

The stock market is characterized as dynamic, unpredictable, and non-linear in nature. Predicting stock prices is a challenging task as it depends on various factors, including political conditions, the global economy, a company's financial reports and performance, etc. Data mining is the computational process of discovering patterns in large datasets. It involves methods at the intersection of machine learning, statistics, and database systems. Prediction is important in the sense that it provides concrete data for investment decisions. To deal with the variety of data, an efficient model that can identify the hidden patterns and complex relations in the large data set is needed. Given its role in the creation of successful strategies for stock exchange deals, stock price prediction is a key consideration.

The country's prosperity heavily depends on the stock market. This is because the stock market contributes to the growth of commerce and industry, which in turn affects the economy of the nation. There are two choices available to a company whenever it needs money to grow or launch a new venture. Either a loan from a financial institution or the issuances of shares on the stock market are options. A business may give partial ownership shares of its stock. A company must list on the stock market before it can issue shares for stock investments, and after that, it can raise the money required for its operations. A generic platform for sellers and buyers of stocks posted on the stock market is another crucial role that the stock market performs. Investors in general—both retail and institutional—make up the purchasers and sellers. These individuals are traders who invest in stocks to provide capital for companies. If the stock's future price can be predicted, it may be possible to avoid suffering sizable losses while also boosting earnings.

Investors have recently shown a strong interest in stock price prediction as well as in the incorporation of variable historical data into computer algorithms to generate estimates of expected price fluctuations. The noisy environment makes it very difficult to forecast the stock price. Technical indicators based on stock data that can be gathered every day are frequently relied upon by traders. Even though they use these indicators to get some price information, it is still challenging to forecast daily to weekly trends with any degree of accuracy. Stock dealing is risky for someone who is not experienced. But if one is fixated on identifying market patterns, one can quickly make a tidy pile of intraday deals. There was a generalized mindset in recent times when depending on the beliefs of people Trading was regarded as a game of buying and selling stocks. By using a technique known as technical analysis to forecast future prices using past price data, investors have now developed a few new tools. Technical indicators are the foundation of broad technical analysis. A technical stock price indicator is a function that gives a value for a given stock price over a specified period of time in history. In addition to technical analysis, there is a technique called basic analysis that focuses on the business, which is the basis for the stock itself. It evaluates a company's historical success as well as the veracity of its financial statements. A stock's rationality can be assessed using a variety of success ratios. Stock market forecasting has advanced into the realm of technology since the advent of the digital computer.

To forecast the stock market, we suggest a system built on a generalized linear regression model. In this article, we show a model that we developed based on the LASSO method for stock price prediction that is more accurate than the Ridge method or an artificial neural network model.

Related works

S. Feng, C. Xu, Y. Zuo, G. Chen, F. Lin, and J. XiaHou, ''Relation-aware dynamic attributed graph attention network for stocks recommendation,'' Pattern Recognit., vol. 121, Jan. 2022, Art. no. 108119

The inherent properties of the graph structure of the financial market and the correlation attributes that actually exist in the system inspire us to introduce the concept of the graph to solve the problem of prediction and recommendation in the financial sector. In this paper, we are adhering to the idea of recommending high return ratio stocks and put forward an attributed graph attention network model based on the correlation information, with encoded timing characteristics derived from time series module and global information originating from the stacked graph neural network(GNN) based models, which we called Relation-aware Dynamic Attributed Graph Attention Network (RA-AGAT). On this basis, we have verified the practicality and applicability of the

application of graph models in finance. Our innovative structure first captures the local correlation topology information and then introduces a stacked graph neural network structure to recommend Top-N return ratio of stock items. Experiments on the real China A-share market demonstrate that the RA-AGAT architecture is capable of surpassing the previously applicable methods in the prediction and recommendation of stock return ratio.

W. Lu, J. Li, J. Wang, and L. Qin, ‘‘A CNN-BiLSTM-AM method for stock price prediction,’’ Neural Comput. Appl., vol. 33, no. 10, pp. 4741–4753, May 2021.
In recent years, with the rapid development of the economy, more and more people begin to invest in the stock market. Accurately predicting the change in stock price can reduce the investment risk of stock investors and effectively improve the investment return. Due to the volatility characteristics of the stock market, stock price prediction is often a nonlinear time series prediction. The stock price is affected by many factors. It is difficult to predict through a simple model. Therefore, this paper proposes a CNN-BiLSTM-AM method to predict the stock closing price of the next day. This method is composed of convolutional neural networks (CNN), bi-directional long short-term Memory (BiLSTM), and attention mechanism (AM).

E. Hoseinzade and S. Haratizadeh, ‘‘CNNpred: CNN-based stock market prediction using a diverse set of variables,’’ Expert Syst. Appl., vol. 129, pp. 273–285, Sep. 2019.
This dataset contains several daily features of the S&P 500, NASDAQ composite, Dow Jones Industrial Average, RUSSELL 2000, and NYSE composite from 2010 to 2017. It covers features from various categories of technical indicators future contracts price of commodities, important indices of markets around the world, the price of major companies in the U.S. market, and treasury bill rates. Sources and descriptions of features have been mentioned in the paper “CNNpred: CNN-based stock market prediction using a diverse set of variables.”

1. J. M.-T. Wu, Z. Li, G. Srivastava, M.-H. Tasi, and J. C.-W. Lin, ‘‘A graph-based convolutional neural network stock price prediction with leading indicators,’’ Softw., Pract. Exp., vol. 51, no. 3, pp. 628–644, Mar. 2021.
The stock market is a capitalistic haven where the issued shares are transferred, traded, and circulated. It bases stock prices on the issue market, however, the structure and trading activities of the stock market are much more complicated than the issue market itself. Therefore, making an accurate prediction becomes an intricate as well as highly difficult task. On the other hand, because of the potential benefits of stock prediction, it attracts generation after generation of scholars as well as investors to continuously develop various prediction methods from different perspectives, a myriad of theories, a multitude of investment strategies, and different practical experiences. In this article, aiming at the task of time series (financial) feature extraction and prediction of price movements, a new convolutional novel neural network that can be called a framework to improve the prediction accuracy of stock trading is proposed. The method that is proposed is called SSACNN, a short form of stock sequence array convolutional neural network. SSACNN collects data including historical data of prices and its leading indicators (options/futures) for a stock to take an array as the input graph of the convolutional neural network framework. In our experimental results, five Taiwanese and American stocks were used as a benchmark to compare with the previous algorithms and proposed algorithm, the motion prediction performance of SSACNN has been improved significantly and proved that it has the potential to be applied in the real financial market.

II. METHODOLOGY

The LASSO Regression type of linear regression algorithm used in the stock price dataset was taken as input from the dataset repository. Then, the data pre-processing step takes place for processing the data previously. In this step, we have to handle the missing values to avoid wrong predictions. After that, the feature selection for selecting the best features from our dataset by using the Pearson coefficient. Then, the data has to be split into two types as data into test and train. In this step, the test is used to predict the model, and the train is used to evaluate the model. To implement the machine learning regression algorithms such as linear regression is used.

III. IMPLEMENTATION

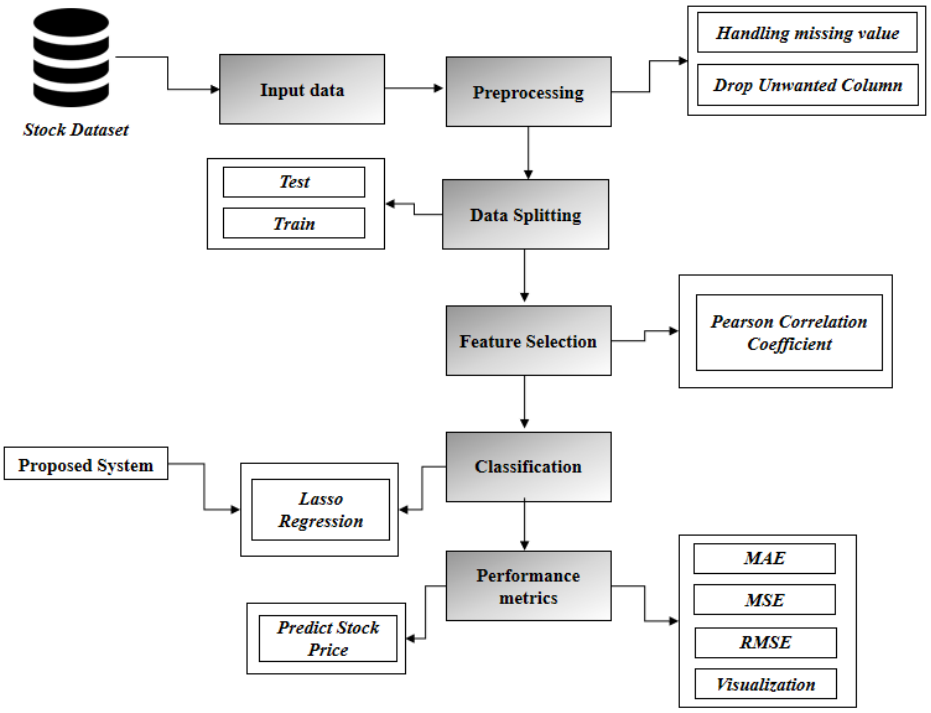

Figure 1: Flow Chart.

IV. I Module

- Data Selection
- Pre-processing
- Feature Selection
- Data splitting
- LASSO Regression
- Performance Analysis

IV.II Module Description

- **Data Selection**

Data selection is the process of selecting the data for predicting the stock. .The dataset was collected from a dataset repository like UCI. The dataset is in a format like '.csv'. In this system, the time series dataset is used for predicting the stock.The dataset contains information about the high, low, open, and close prices. With the help of the panda's package, we can read or load our input dataset.

- **Pre-processing**

Data pre-processing is the process of removing unwanted data from the dataset. Data pre-processing allows for the removal of unwanted data with the use of data cleansing, this allows the user to have a dataset to contain more valuable information after the pre-processing stage for data manipulation later in the data mining process. Missing data removal: In this process, the null values such as missing values and Nan values are replaced by 0.

- **Feature Selection**

In this step, we can select the features from pre-processed data by using Pearson's correlation. One of the measures used for feature selection is dependency measures. Many dependency-based methods have been proposed. The main measure is the Correlation-based method. Pearson's Correlation method is used for finding the association between the continuous features and the class feature. Features with high correlation are more linearly dependent and hence have almost the same effect on the dependent variable.

- **Data splitting**

Data splitting is the act of partitioning available data into two portions, usually for cross-validator purposes. One Portion of the data is used to develop a predictive model and the other to evaluate the model's performance. Separating data into training and testing sets is an important part of evaluating data mining models. Typically, when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing.

- **LASSO Regression**

In our process, we have to implement the two machine learning algorithm such as lasso regression. Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).

- **Performance and Analysis**

Performance is measured using via the following methods,
The Performance analysis will get generated based on the overall classification and prediction.
The performance of this proposed approach is evaluated using some measures like MAE: In statistics, the mean absolute error (MAE) is a way to measure the accuracy of a given model

- MAE = (1/n) * Σ|yi – xi|
- Σ: A Greek symbol that means "sum"
- yi: The observed value for the ith observation
- xi: The predicted value for the ith observation
- n: The total number of observations
- MSE: The mean squared error (MSE) is a common way to measure the prediction accuracy of a model. It is calculated as:
- MSE = (1/n) * Σ (actual – prediction) 2

Where:

- Σ – a fancy symbol that means "sum"
- n – sample size
- actual – the actual data value
- forecast – the predicted data value

V. RESULT AND CONCLUSION

In our project, lasso regression is used to predict the stock market prices. The stock price dataset was processed using the LASSO Regression type of linear regression algorithm, with input from the dataset repository. The data is then processed in the subsequent stage of data pre-processing. To prevent inaccurate predictions, we must deal with the missing values in this phase. The feature selection process then uses the Pearson coefficient to choose the finest features from our dataset. The data must then be divided into two categories, test data and train data. In this stage, the model is predicted by the test, and its performance is assessed by the train. Regression methods for machine learning, like linear regression, are implemented.

REFERENCES:

1. S. Feng, C. Xu, Y. Zuo, G. Chen, F. Lin, and J. XiaHou, ''Relation-aware dynamic attributed graph attention network for stocks recommendation,'' Pattern Recognit., vol. 121, Jan. 2022, Art. no. 108119.
2. W. Lu, J. Li, J. Wang, and L. Qin, ''A CNN-BiLSTM-AM method for stock price prediction,'' Neural Comput. Appl., vol. 33, no. 10, pp. 4741–4753, May 2021.
3. E. Hoseinzade and S. Haratizadeh, ''CNNpred: CNN-based stock market prediction using a diverse set of variables,'' Expert Syst. Appl., vol. 129, pp. 273–285, Sep. 2019.
4. J. M.-T. Wu, Z. Li, G. Srivastava, M.-H. Tasi, and J. C.-W. Lin, ''A graph-based convolutional neural network stock price prediction with leading indicators,'' Softw., Pract. Exp., vol. 51, no. 3, pp. 628–644, Mar. 2021.
5. J. Zhao, D. Zeng, S. Liang, H. Kang, and Q. Liu, ''Prediction model for stock price trend based on recurrent neural network,'' J. Ambient Intell. Hum. Comput., vol. 12, no. 1, pp. 745–753, Jan. 2021
6. R. K. Dash, T. N. Nguyen, K. Cengiz, and A. Sharma, ''Fine-tuned support vector regression model for stock predictions,'' Neural Comput. Appl., pp. 1–15, Mar. 2021.
7. H. Wang, S. Li, T. Wang, and J. Zheng, ''Hierarchical adaptive temporal relational modeling for stock trend prediction,'' in Proc. IJCAI, 2021, pp. 3691–3698
8. W. Chen, M. Jiang, W.-G. Zhang, and Z. Chen, ''A novel graph convolutional feature based convolutional neural network for stock trend prediction,'' Inf. Sci., vol. 556, pp. 67–94, May 2021.
9. C. Lohrmann and P. Luukka, ''Classification of intraday S&P500 returns with a random forest,'' Int. J. Forecasting, vol. 35, no. 1, pp. 390–407, Jan. 2019.
10. J. Liu, H. Lin, X. Liu, B. Xu, Y. Ren, Y. Diao, and L. Yang, ''Transformerbased capsule network for stock movement prediction,'' in Proc. 1st Workshop Financial Technol. Natural Lang. Process., 2019, pp. 66–73.