TESLA STOCK PRICE PREDICTION USING REGRESSION MODEL

Aruna.K, Astalakshmi.G, Preethi.B

Student, Department of CSE, Panimalar Engineering College,Chennai-600123, India

Mail: arunakumar569@gmail.com, snekhaganesh87@gmail.com,preethiroshni2@gmail.com

Abstract: *Stock prices are initially set by an initial public offering (IPO), which occurs when a company first issues its shares on the market. Investment companies use a variety of metrics, including the total number of shares being offered, to calculate the price of a stock. The share price will then fluctuate as a result of the factors mentioned, with potential business earnings having a substantial impact. Using financial metrics including a company's earnings history, market movements, and the profit it is predicted to make, traders regularly evaluate a company's value. As a result, stock price forecasting has grown in significance as a field of study. Forecasting machine learning-based stock price prediction techniques is the aim. Univariate, bivariate, and multivariate analysis are used to analyse the dataset using SMLT's supervised machine learning technique (SMLT). to offer an approach based on machine learning for accurately forecasting stock price. The proposed machine learning algorithm technique can be compared to the best accuracy in terms of precision, recall, and F1 Score.*

Keywords: *Machine learning, Regression, tesla*

1.INTRODUCTION

The stock market ranks among the most challenging problems to observe and analyse. An investment's market value is influenced by a wide range of factors, including market volatility and several autonomous factors. Every stock market expert will be unable to accurately forecast when the market will rise or collapse due to these factors. Nonetheless, improvements in stock market forecasting have begun to employ algorithms, and it is accurate to analyse historical data. There is a good chance that many automakers employ machine learning algorithms to predict the stock market. In this article, I'll show you how to use a handy Python application that analyses and projects the stock price of a reputable outsourcing firm using a number of machine learning techniques.

2.RELATED WORK

Accurately estimating stock values is a tricky issue due to the complexity of balance sheets. Based on the Pearson coefficient of correlation (PCC) and the Wide Learning System, a multi-indicator features extraction strategy for stock price projection was developed (BLS). The ultimate goal of all this report is to aid traders in arriving at sound investment choices by accurately order to forecast price fluctuations using machine learning algorithms. In this report, we created a new framework out of PCC and BLS and utilized it to forecast stock prices mostly on Stock Exchange of Shenzhen or the Shanghai Exchange for Stocks in the near term.

3.PROPOSED WORK

In proposed system, the Datasets from different sources would be combined to form a generalized dataset. After the generalized dataset is prepared it is checked for cleanliness, and then trimmed dataset is analyzed. The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using machine learning algorithms are applied on the Training set and based on the test result accuracy, test set prediction is done. In addition, we are going to compare the forecasting results with four machine learning methods, like Adaptive Boosting (Adaboost), Decision Tree, Ridge regression and Lasso regression. Among all algorithms used for testing, the algorithm that provides the best performance is finally used. Then the system is deployed using flask. For predicting the tesla stock problem, ML prediction model is effective because it is strong in preprocessing of data, irrelevant variables, and a mix of continuous, categorical and discrete variables.

The proposed methodology has the following steps:

- Data Preprocessing
- Data visualization
- Algorithm Implementation
    - Adaptive Boosting (Adaboost)
    - Decision Tree Regression
    - Ridge Regression
    - Lasso Regression
- Deployment

DATA PREPROCESSING: Machine learning validation approaches are used to obtain the margin of error of such Machine Learning (ML) system, which would be close to the genuine error rate of something like the dataset. If the data is substantial enough to be considered accurately representing the population, validation approaches may not be required. Yet, in real-world circumstances, it is necessary to work with data samples that are not always accurately reflecting the population of the given dataset. To identify the standard error, duplicate value, and data type description, whether it is a float variable or an integer variable. For tuning model hyper parameters, a sample of data is employed to offer an empirical assessment of an observed performance on the training dataset.

When skill from the training data is used to configure the model, overall evaluation becomes more unbalanced. The training set can be utilized to test a particular model, however this is done frequently. The modelling hyper parameters are adjusted while using data by machine learning specialists. The process of gathering data, analysing it, and dealing with its structure, quality, and substance can take a lot of time. Understanding your data and its characteristics is helpful during the data evaluation stage since it will assist you decide which algorithm to employ to create your model. A variety of data cleaning jobs are underwent using Python's Pandas library, with a particular emphasis on the most important data cleaning work, missing values, and the ability to clean information more rapidly. It would like to devote almost no time cleaning data and far more length exploring and modelling.

	Date	Open	High	Low	Close	Adj Close	Volume
0	28-06-2010	3.800	6.084	3.508	3.840	3.840	246560000
1	05-07-2010	4.000	4.000	2.996	3.480	3.480	127753000
2	12-07-2010	3.590	4.300	3.380	4.128	4.128	77194500
3	19-07-2010	4.274	4.450	3.900	4.258	4.258	35878500
4	26-07-2010	4.300	4.300	3.910	3.988	3.988	15260000
5	02-08-2010	4.100	4.436	3.904	3.918	3.918	21998500
6	09-08-2010	3.980	3.996	3.478	3.664	3.664	21083000
7	16-08-2010	3.690	3.918	3.652	3.820	3.820	12050500
8	23-08-2010	3.818	4.078	3.712	3.940	3.940	15389500
9	30-08-2010	3.940	4.260	3.866	4.210	4.210	11752500
10	06-09-2010	4.122	4.210	3.952	4.034	4.034	6473000

	Date	Open	High	Low	Close	Adj_Close	Volume
0	28-06-2010	3.800000	6.084000	3.508000	3.840000	3.840000	246560000
1	05-07-2010	4.000000	4.000000	2.996000	3.480000	3.480000	127753000
2	12-07-2010	3.590000	4.300000	3.380000	4.128000	4.128000	77194500
3	19-07-2010	4.274000	4.450000	3.900000	4.258000	4.258000	35878500
4	26-07-2010	4.300000	4.300000	3.910000	3.988000	3.988000	15260000
...	...	...	...	...	...	...	...
603	17-01-2022	1026.609985	1070.790039	940.500000	943.900024	943.900024	105363500
604	24-01-2022	904.760010	987.690002	792.010010	846.349976	846.349976	208309200
605	31-01-2022	872.710022	943.700012	862.049988	923.320007	923.320007	132213500
606	07-02-2022	923.789978	947.770020	850.700012	860.000000	860.000000	103196000
607	14-02-2022	861.570007	898.880005	853.150024	875.760010	875.760010	22585500

608 rows × 7 columns

Figure 3.1 *Screenshot of Data Pre-processing*

DATA VISUALIZATION: Data visualization is a crucial ability in machine learning and applied statistics. In fact, the main focus of statistics is on numerical estimates and descriptions of data. An essential set of tools for obtaining a qualitative understanding is provided by data visualization. This can be useful for discovering trends, corrupt data, outliers, and much more while exploring and getting to know a dataset. Data visualizations can be utilised to convey and illustrate critical relationships in plots and charts that are more visceral and engaging to stakeholders than measurements of association or importance with a little subject knowledge. It will suggest a deeper look into some of the books suggested at the conclusion.

Data visualization and exploratory data analysis are entire fields in themselves**.** Data may not always make sense unless it is presented visually, such as through charts and graphs. Both applied statistics and applied machine learning value fast visualization of data samples and other objects. It will identify the many plot types you must be aware of while Python data visualization techniques and how to apply them to your own data.

- Learn to visualize categorical data using bar charts and time series data using line graphs.
- Using histograms and box graphs to condense data distributions.

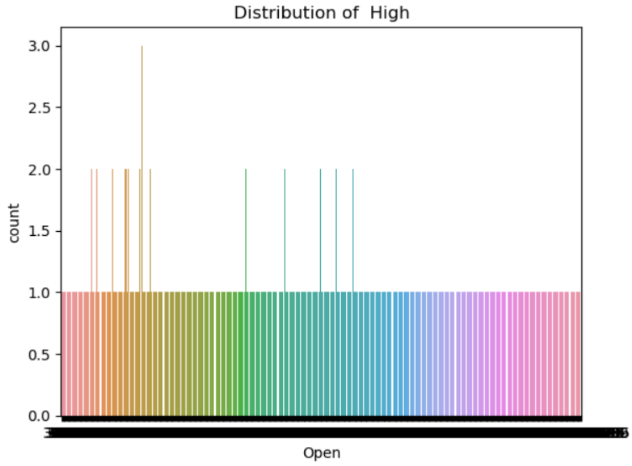

Figure 3.2 *Screenshot of Data Visualization*

ALGORITHM IMPLEMENTATION: It is crucial to regularly compare the results of various machine learning algorithms, and it will be found that using scikit-learn and Python, it is possible to develop a test harness for this purpose. You can apply this test harness as a model for your own machine learning issues and include additional and various algorithms to contrast. There will be variations in the performance attributes of each model. You may estimate each model's potential accuracy on unobserved data by using resampling techniques like cross validation. It must be able to select one or two of the best models from the group of models you have developed using these estimates**.** In order to view a fresh dataset from various angles, it is a good idea to visualize the data using a variety of ways. The choice of models follows the same logic. In order to select the one or two that will be used for finalization, you need consider a lot of various angles when evaluating the machine learning algorithms' predicted accuracy. One way to achieve this is to demonstrate the average accuracy, variance, and other characteristics of the distribution of model accuracies using various visualization techniques.

You will learn precisely how to achieve that in Python using scikit-learn in the following section. Making sure that each algorithm is evaluated uniformly on the same data is essential for conducting a fair comparison of machine learning algorithms, and this may be done by requiring that each algorithm be tested using a uniform test harness.

Train set + Test set = Total Dataset

1.ADABOOST ALGORITHM: AdaBoost is adaptive in that it adjusts succeeding weak learners in favour of those situations when earlier regressors made mistakes. It may be less prone to the overfitting issue in particular situations than other learning methods. It can be demonstrated that the final model converges to a strong learner even if the performance of each individual learner is just marginally better than random guessing. Although decision stumps and other weak base learners are commonly combined using AdaBoost, it has been demonstrated that strong base learners, such as deep decision trees, may also be combined effectively. This results in a model that is even more accurate.

The Adaboost algorithm is trained used the “AdaBoostRegressor” module form sklearn.ensemble library. The dataset is split into train and test ie. “X_train, X_test, y_train, y_test”. The test dataset sizeis 30% of the total dataset.

2.DECISION TREE REGRESSION : The decision tree is a supervised problem solving method that can be used to solve classification and regression problems. Each leaf node in the tree-structured regressor represents the outcome, and the innermost nodes act as placeholders for the dataset's constituents, decision-making criterion, and decision-making procedures. A decision tree has two nodes: a decision node as well as the leaf node. A leaf node is the outcome of all that decision and has no extra branches, whereas a choice is arrived at to use a decision node, that has numerous branches.
The "DecisionTreeRegressor" modules from the sklearn.tree library is used to train the decision tree algorithm. The test dataset makes up 30% of the overall dataset in size. The dataset is divided into training and testing segments, denoted by the notation "X train, X test, y train, y test".

3.RIDGE REGRESSION: Regularization methods like ridge regression are employed to make the model less complex. It also goes by the name L2 regularization. This method modifies the cost function by including a penalty term. Ridge Regression penalty refers to the amount of bias that is added to the model. If there are numerous significant parameters that are roughly the same value, Ridge performs well. Ridge regression introduces some bias into the regression estimates in an effort to lower the standard error.
The ridge regression algorithm is trained used the “Ridge” module form sklearn. linear_model library.The dataset is split into train and test ie. “X_train, X_test, y_train, y_test”. The test dataset size is 30%of the total dataset.

4.LASSO REGRESSION: A regularization method is called least absolute shrinkage and selection operator (LASSO) regression. For a more accurate forecast, it is preferred over regression techniques. Shrinkage is used in this model. When data values shrink towards the mean, this is referred to as shrinkage. Models with little details are encouraged by the lasso process (i.e. models with fewer parameters). If you wish to automate some steps in the model selection process, such as variable selection and parameter removal, or if your models exhibit a high degree of multicollinearity, this particular sort of regression is a good choice. L1 regularization is used in Lasso Regression. Because it does feature selection automatically, it is employed when there are more features.
The lasso regression algorithm is trained used the “Lasso” module form sklearn. linear_model library. The test dataset size is 30% of the total dataset. The dataset is split into train and test ie. “X_train, X_test, y_train, y_test”.

DEPLOYMENT: The model with high accuracy is converted to .pkl file in order to use it for webpage. Thus the best model obtained from the four algorithm is Decision Tree . This model is deployed as a webpage using Flask framework. Using this framework, we will connect the saved model of decision tree model and HTML,CSS code to create a webpage.

4.RESULTS & DISCUSSIONS

Finding the R2 score

```
In [22]: R2 = (r2_score(y_test, predictD)*100)
         print("The Accuracy of adaboost regressor:", R2)
         print("")

The Accuracy of adaboost regressor: 99.06166312809103
```

***Figure 4.1** Screenshot of Accuracy using Adaboost Regression*

Find the R2_score

```
In [23]: R2_SCORE = (r2_score(y_test, predictR)*100)
         print("The Accuracy of Ridge regressor :", R2_SCORE)
         print("")

The Accuracy of Ridge regressor : 99.82611262743984
```

Figure 4.2 *Screenshot of Accuracy using Ridge Regression*

Find the r2score

```
In [21]: R2_SCORE = (r2_score(y_test, predictLR)*100)
         print("The Accuracy of Lasso:", R2_SCORE)
         print('')

The Accuracy of Lasso: 99.74023051964012
```

Figure 4.3 *Screenshot of Accuracy using Lasso Regression*

Find Accuracy Score

```
In [20]: R2=(r2_score(y_test,predictD)*100)
         print('Accuracy result of Decision Tree regressor is :',R2)
         print("")

Accuracy result of Decision Tree regressor is : 99.83258573196214
```

Figure 4.4 *Screenshot of Accuracy using Decision Tree Regression*

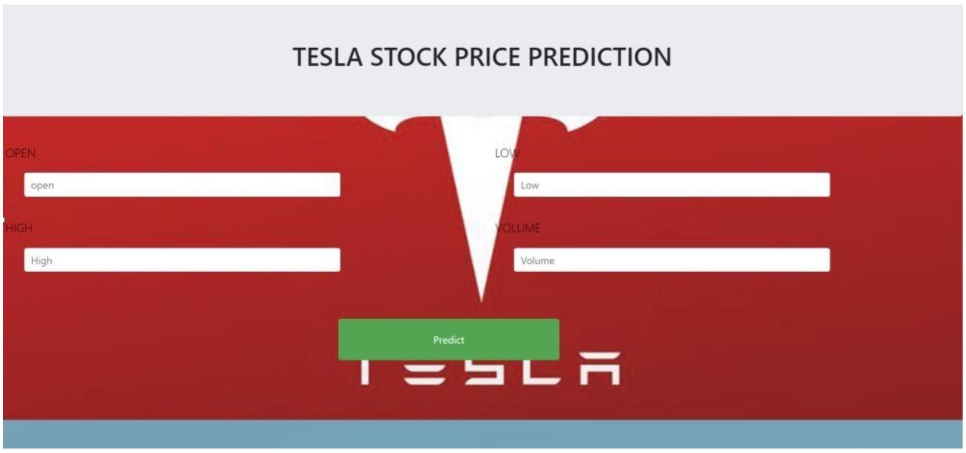

Figure 4.5 *Screenshot of web page for Tesla stock price prediction*

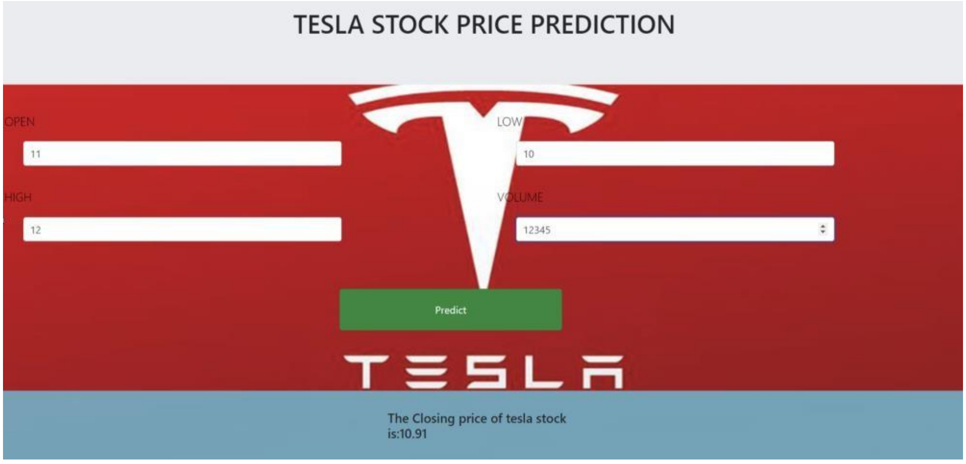

Figure 4.6 *Screenshot of the predicted stock price is displayed*

The output screen includes a heading that describes the purpose of the application, which is to predict the tesla stock price. There are four input boxes where users can enter the values of Open, Low, High, Volume to predict the stock price. A "Predict" button is provided to initiate the prediction process. After the analysis is completed, the output displays the closing price of tesla stock. The example shows how the output screen will work. Depending on the specific implementation of the application and the machine learning model behind it, the output could predict closing price for different set of inputs.

5.EVALUATION METRICS

	Mean Absolute Error	**Mean Absolute Percentage Error**	**Explained Variance Score**	**Mean Squared Error**	**Median Absolute Error**	**R2 score**
Adaptive boosting	19.57	1.66	99.07	611.39	19.17	99.06
Ridge Regression	4.00	0.04	99.82	87.95	0.95	99.82
Lasso Regression	5.15	0.07	99.75	199.66	1.19	99.74
Decision Tree Regression	4.06	0.04	99.83	92.38	1.03	99.83

6.COMPARITIVE ANALYSIS

Thus, we have built four models based on four algorithms and each has yielded different accuracies. The first model which was built using Adaboost regression yielded an accuracy of 99.06%., the Ridge regression yielded an accuracy of 99.82%, the Lasso regression yielded an accuracy of 99.74% and Decision Tree regression model yieldedan accuracy of 99.83%. From this we can say that Decision tree algorithm yields the highest accuracy. It is thenfollowed by Ridge regression. The least accuracy was yielded by Adaboost regression algorithm. These models were built using Machine Learning algorithms.

S.No	ALGORITHM	ACCURACY
1.	ADABOOST ALGORITHM	99.06
2.	RIDGE REGRESSION	99.82
3.	LASSO REGRESSION	99.74
4.	DECISION TREE REGRESSION	99.83

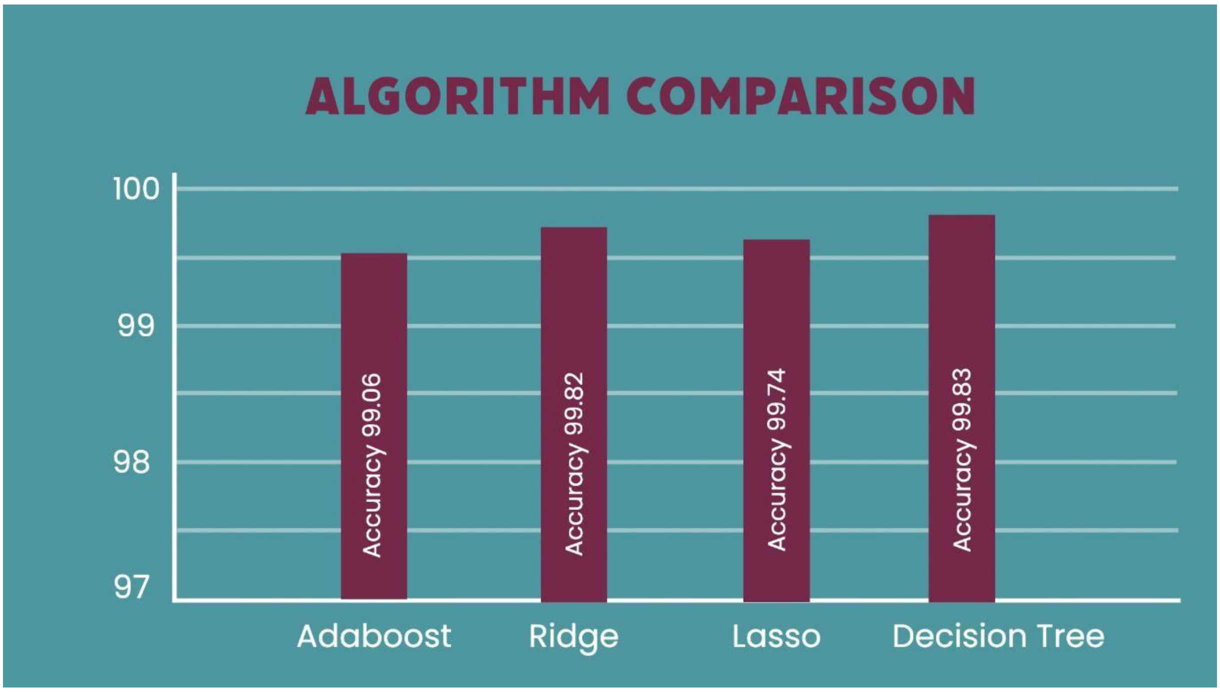

Figure 6.1 *Comparison of four algorithms in bar graph format*

From the above details, it is obvious that Decision tree algorithm performed well compared to other three algorithms based on the accuracy yielded.

7.CONCLUSION AND FUTURE WORK

CONCLUSION:

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The performance of these models largely depends on the quality and size of the training data, feature engineering, and the choice of the regression algorithm. The best accuracy on public test set of higher accuracy score algorithm will be find out. The founded one is used in the application which can help to find the tesla stock price.

FUTURE WORK:

- Deploying the project in the cloud.
- To optimize the work to implement in the IOT system.

8.REFERENCES

[1] 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, 2021.

[2] G. Chen, Y. Lou, and L. Wang, “A comparative study on controllability robustness of complex networks,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 66, no. 5, pp. 828–832, May 2019.

[3] J. Dong, W. Dai, Y. Liu, L. Yu, and J. Wang, “Forecasting chinese stock market prices using baidu search index with a learning-based data collection method,” Int. J. Inf. Technol. Decis. Making, vol. 18, no. 5, pp. 1605–1629, 2019.

[4] L. Bai, Y. Zhao, and X. Huang, “A CNN accelerator on FPGA using depthwise separable convolution,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 65, no. 10, pp. 1415–1419, Oct. 2018.

[5] B. N. G. Koneru and V. Vasudevan, “Sparse artificial neural networks using a novel smoothed LASSO penalization,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 66, no. 5, pp. 848–852, May 2019.

[6] A. Thakkar, D. Patel, and P. Shah, “Pearson correlation coefficientbased performance enhancement of vanilla neural network for stock trend prediction,” Neural Comput. Appl., vol. 33, pp.

[7] D. Liu, C. K. Tse, and X. Zhang, “Robustness assessment and enhancement of power grids from a complex network’s perspective using decision trees,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 66, no. 5, pp. 833–837, May 2019.

[8] J. Yang, R. Rao, P. Hong, and P. Ding, “Ensemble model for stock price movement trend prediction on different investing periods,” in Proc. 12th Int. Conf. Comput. Intell. Security (CIS), 2016, pp. 358–361.

[9] C. L. P. Chen and Z. Liu, “Broad Learning System: An effective and efficient incremental learning system without the need for deep architecture,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 1, pp. 10–24, Jan. 2018.

[10] L.-P. Ni, Z.-W. Ni, and Y.-Z. Gao, “Stock trend prediction based on fractal feature selection and support vector machine,” Expert Syst. Appl., vol. 38, no. 5, pp. 5569–5576, 2011

[11] Q. Li et al., “Integrating reinforcement learning and optimal power dispatch to enhance power grid resilience,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 69, no. 3, pp. 1402–1406, Mar. 2021.

[12] Q. Zhuge, L. Xu, and G. Zhang, “LSTM neural network with emotional analysis for prediction of stock price,” Eng. Lett., vol. 25, no. 2, pp. 167–175, 2017.

[13] J. Yang, R. Rao, P. Hong, and P. Ding, “Ensemble model for stock price movement trend prediction on different investing periods,” in Proc. 12th Int. Conf. Comput. Intell. Security (CIS), 2016, pp. 358–361