Faculty of Natural and Applied Sciences Journal of Computing and Applications
Print ISSN: 3026-8133
www.fnasjournals.com
Volume 2; Issue 1; September 2024; Page No. 38-50.

Application of Machine Learning Techniques for Stock Price Prediction

[*1]Ogunsanwo, G.O, [2]Kuti. A.A, [3]Aiyelokun O.O., & [4]Alaba, O.B.
[1,2,4]Tai Solarin University of Education, Nigeria
[3]University of Ibadan, Ibadan, Nigeria

***Corresponding author email**: ogunsanwogo@gmail.com

Abstract
This study highlights the robustness of selected Machine Learning models in Stock Closing Price prediction, which effectively mitigates several defects and hazards resulting from its indeterminable nature. From different literature consulted, so many tools and techniques have been used widely to tackle the prediction of the stock market In this paper, our focus is to develop a machine learning predictive model for Stock Closing Price using three different machine learning (ML) algorithms. The data inputted into the predictive model was obtained from the Yahoo Finance online repository. The dataset used contains five hundred and eighty-eight instances. Neural Networks (ANN), SVM, and KStar Algorithm were employed for the stock market price predictive models. Stimulated on Waikato Software for Knowledge Analysis (WEKA). Comparative analysis for Stock Closing Price and predictions were performed using MAE, RMSE & RAE. The results of the predictive models revealed that the correlation coefficient of the model developed for SVM, ANN, and KStar Algorithm have values of 0.0104, 0.7184, and 0.4819 respectively. So also, the Root Relative Squared Error of 107.6185%, 96.8064%, and 88.71% for SVM, ANN, and KStar respectively. This shows that ANN has a strong positive relationship frame followed by SVM and KStar. The KStar shows a faster learning time of the three algorithms followed by SVM and ANN. The study concluded that Artificial Neural Network (ANN) performs better than SVM and KStar in predicting Stock Market Prices.

Keywords*:* Stock, Prediction, SVM, ANN, KStar

Introduction

The concept of stock can be explained as a unit of ownership an individual or corporate body has in a company. Stock can also be referred to as equity. There are many reasons why companies decide to sell part of their shares to other people. Most companies sometimes sell stock to acquire more funds to expand their business, sometimes to settle debt, or for any other reason. Stock may be classified as Common Stock (CS)or preferred stock. CS as the name implies is generally referred to as an ownership of a few shares of that company while Preferred stock is the type of share owned by a person or company that gave them preference to receive part of the profit made by the issued company. It has higher priority when compared to common stock. Trading in stock usually takes place in the stock market. A place where all the transactions that involve shares, stocks, government bonds, debentures, and other approved securities take place is called the Stock Market (Okobiah, 2000). This market avails the opportunities for large and small investors to buy and sell stocks or shares of companies and government agencies with the assistance of stock brokers. Since individual and cooperate organisations do engage in the stock business this made stock investment popular and lucrative. Stock trading plays a key role in the global economy because of its numerous benefits. However, despite its huge influx of investors into the stock market, it comes with some flaws and risks due to its unpredictable behaviour Therefore, any variation in this market affects personal and corporate financial lives, and as a whole determines the economic health of a country. Stock market prediction have attracted different tools and techniques. In the past, there were two major methods used in stock prediction. They are fundamental analysis (FA) and technical analysis (TA). The FA are more concerned with the company's performance rather than the actual stock. They used the past performance of the company to establish their decision and forecasts. On the other hand, TA is carried out by Technical Analysts, and the stock price is determined using time series based on the past patterns of the stock.

MLTs are often engaged in building real-time application model which yields better results than other traditional methods (Mehar et al., 2019). The use of ML techniques in stock data prediction is more concerned with Technical

Cite this article as:
Ogunsanwo, G.O, Kuti. A.A, Aiyelokun O.O., & Alaba, O.B. (2024). Application of machine learning techniques for stock price prediction. *FNAS Journal of Computing and Applications*, *2*(1), 38-50.

Analysis in order to establish that the algorithm used in the prediction has the capability to learn the underlying patterns in the stock time series. Some of the new work carried out in this field uncovered that ML Techniques can be used greatly in stock market price prediction, for instance, SVM Random Forest (RF), Artificial Neural Network (ANN) (Adebayo et al., 2012; Asagba et al., 2013), Convolutional Neural Network (CNN), Chen (2020) and DNN like LSTM, Chen (2020) also have shown promising results. Kundu, et al., (2020) used ARIMA,. Sathya et al., (2020) used Reinforcement Learning & Features Extraction, and Behravan and Razavi (2020) tested Swarm Intelligence. In this paper, three machine techniques namely: ANN, SVM, and KStar were tested for the stock closing price prediction of an organization. The input for the predictive models was a financial dataset that comprised close, open, low and high. The performance of the models developed would be tested using standard metrics, such as RMSE and MAPE. The rest of the work is: section 2 which clearly explains the methodology used, section 3 also discusses the outcome of the study carried out section 4 concludes the paper.

Framework for The Stock Price Prediction

The Methodology adopted in this study includes: data collection, data pre-processing, development of machine learning models, and performance evaluation of the models as revealed in figure 1

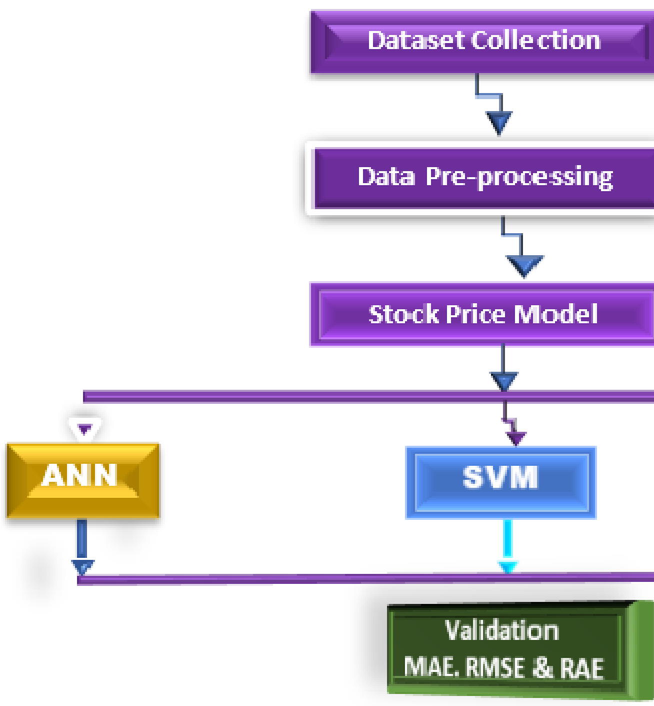


Figure.1: ***framework of the model rework***

Methodology

The past data for the two firms were retrieved from Yahoo Finance. The dataset spans 5 years dated 18/09/2016 to 18/09/2021 of Nike and JP Morgan Chase and Co. The data includes information about the stock. We extracted the closing price of the stock as shown in Table 1.

Table 1: Details of the Dataset used

	Dataset	**Training Dataset**	**Testing Datasets**
Time Interval	18/09/2016 – 18/09/2021	18/09/2016 – 17/09/2019	18/09/2019–18/09/2021

Cite this article as:
Ogunsanwo, G.O, Kuti. A.A, Aiyelokun O.O., & Alaba, O.B. (2024). Application of machine learning techniques for stock price prediction. *FNAS Journal of Computing and Applications*, *2*(1), 38-50.

Data Pre-processing

The data collected for the prediction contain attributes with continuous numeric values. In order to have a better prediction. The standard used for transforming the numeric values to discrete values for the attributes depends on the closing price of the stock. For example, in case values of the properties open, min, max, and last were higher than the value of the preceding for the same trading day, then the numeric values of the attributes will receive the value positive. But if the values of the properties for the same trading day were lower than the attribute previous one, then numeric values of the attributes will be negative. Also, if they are equal, then the numeric values will have equal value.

Model Development

ANN Model

ANN is a smart data mining technique that comes up with results by observing a basic trend from given data. ANN generally can simulate complex patterns from unstructured data. It uses the basic structure of a network having neurons with various layers. The input, the hidden, and the output are the three layers. The input layers for the ANN model developed are made up of High, Low, Open, Close, and Volume variables. The product of the input load and weight is added and the total summation is sent to the second layer where the processing is done. The calculated weight is displayed in the output. This layer consists of one neuron which is the stock price predicted value. Figure 2 displays the ANN Model used in this study for stock prediction.

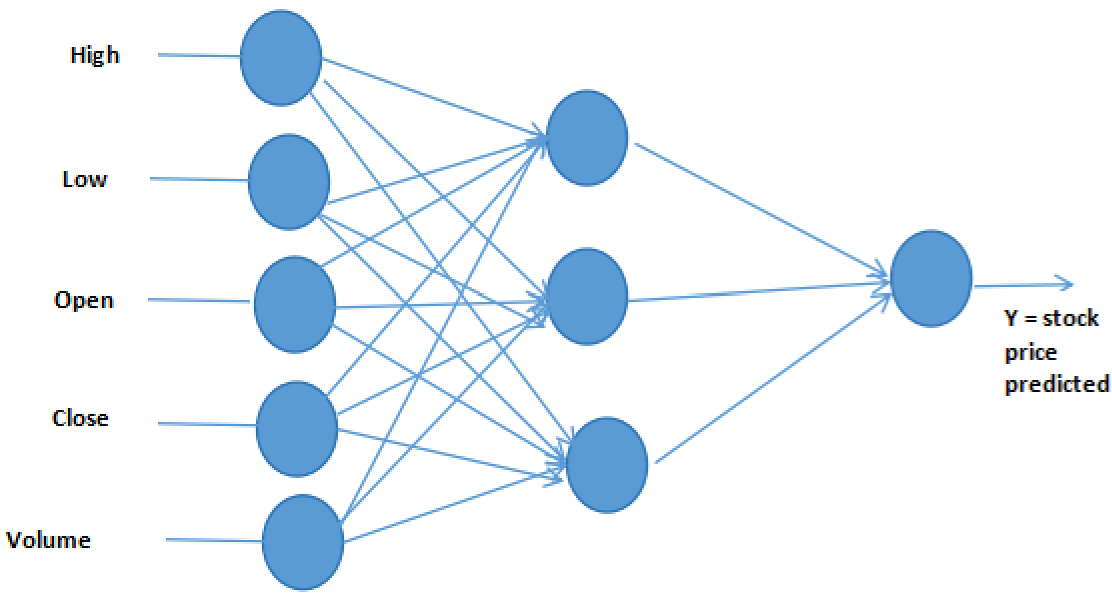


Figure 2: Neural Network Model

The Neuron accepts the input quantity and adds a bias with the total assigned weight. The value of the outputs can be discovered by introducing a transfer function. The number of neurons in the input and the output must correspond. Mathematically, the expected result of neuron SP can be described as follows:

Usp = $\sum_{i=1}^{n} W_{pi} X_i$ (1a)

Ysp = ϕ(uSP + bSP) (1b)

where $x_1,\ldots,x_n$ represents the input parameters; $W_{P1},\ldots,W_{Pn}$ denotes the connection weights of neuron SP; while the Usp is the input combiner; bs_P is the bias; ϕ is the activation function, and ys_P is the output of the neuron.

Activation Function

The input received the predicting parameters, the relationship between weights and biases that the neurons perform on this input is shown in equation (equation2a)

x = (weight * input) + bias (2a)

The sigmoid activation function used is represented by

$$S(x) = \frac{1}{1+e^{-x}}$$ (2b)

Sigmoid functions usually show a return value (y-axis) in the range of 0 to 1.

SVM Model

SVMs are learning algorithms that make use of linear learning machines in kernel-induced feature mode. SVM works with the best margin standard to optimize separating hyperplanes between binary classes. Hyperplane is best applied to linearly separable data which is used in binary decision classes (Cortes &Vapnik; 1995) as shown in (equation 3)

$$\boldsymbol{y = W_0 + W_1X_1 + W_2X_2}$$ (3)

Where: y: outcome, W_1: the attribute values, X_1: Weights determined by the learning algorithm.

Equation 4 represents the maximum margin hyperplane.

$$y = b + \sum_{i=1}^{n} \propto_i y_i x(i) .. x$$ (4)

y : class value of training
$x(i)$, the vector x represents a test
vector $x(i)$ are the support vectors and represent the dot product.
b are parameters used for the hyperplane.

SVM uses a Kernel function to produce high-magnitude feature space from the given inputs as shown in (equation 5).

$$y = b + \sum_{i=1}^{n} \propto_i y_i k(x(i) .. x)$$ (5)

3.2.3 K-STAR MODEL

K-star makes use of an entropic which depends on the probability of changing one instance to another form by randomly selecting from all possible outcomes. The changes are achieved in a manner: By defining a finite set that will map one instance to another one. The second one is achieved using a program in a finite sequence to change instance (a) to (b)

Assume we have a set of infinite points and stated transformations T,
P, the set of all prefix codes from T*.
Therefore, Members of T* and of P uniquely define a transformation as shown in equation (6a)

$$\bar{t}(a) = t_n(t_{n-1}(\cdots t_1(a)\cdots))$$ (6a)

Where t = t_1...tn P represents the probability function on T*. Satisfies the following conditions:

$$o \leq \frac{p(\bar{t}\,u)}{p(\bar{t})} \leq 1$$ (6b)

$$\sum_{u} p(\bar{t}u) = p(\bar{t})$$
$$p(\wedge) = 1 \tag{6c}$$

it satisfies the following:

$$\sum_{rep} p(\bar{t}) = 1 \tag{7}$$

The probability paths from instance a, to instance b is calculated as the probability function P*:

$$p^*\left(\frac{b}{a}\right) = \sum_{rep\, t(a)\to b} P(t) \tag{8}$$

P* satisfies the these conditions

$$\sum_{b} p^*\left(\frac{b}{a}\right) = 1 \tag{9a}$$

$$0 \le P^*\left(\frac{b}{a}\right) \le 1 \tag{9b}$$

The K* is calculated as:

$$K^*\left(\frac{b}{a}\right) = -\log_2 P^*\left(\frac{b}{a}\right) \tag{10}$$

Models Evaluation

The performance of the stock price prediction models developed was determined using the standard metrics: MAE, RMSE, RAE and Time Taken for each of the models.

Correlation Coefficients measure the relationship between two variables. The output returns a value between -1 and 1

- When the value returns 1 it shows a strong positive relationship
- When the value returns -1 it shows a strong negative relationship
- When the value returns a result of zero it shows no relationship at all.

MAE

The MAE does neglect direction when measuring. it measures the average magnitude of the errors in a given prediction. It is good at measuring continuous variables. The MAE is a linear score which means that all the individual differences are weighted equally in the average.

$$\textbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{(O_i - F_i)}{O_i}\ x\ 100 \tag{11}$$

'*Oi*': the original closing price,
'*Fi*': the predicted closing price and
'n': the total window size.

RMSE

RMSE is useful in measuring the accuracy, but best at comparing the prediction errors and model configurations for a specific variable and not between variables, RMSE is scale-dependent.

Cite this article as:
Ogunsanwo, G.O, Kuti. A.A, Aiyelokun O.O., & Alaba, O.B. (2024). Application of machine learning techniques for stock price prediction. *FNAS Journal of Computing and Applications*, *2*(1), 38-50.

$$\textbf{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(O_i - F_i)^2}{n}} \qquad (12)$$

Where '*Oi*' is the first closing price, '*Fi*' is the predicted closing price, and 'n' refers to the entire window size.

RAE

RAE measures the performance of predictive models, It's majorly used in the area like operations management, data mining, and machine learning. RAE is expressed as a ratio of a mean error which is also called a residual to errors produced by an unimportant or naïve model.

Evaluation

The dataset used for the stock prediction was collected in CSV (comma delimited) format and was transformed and saved in Arff format as shown in Figure 3

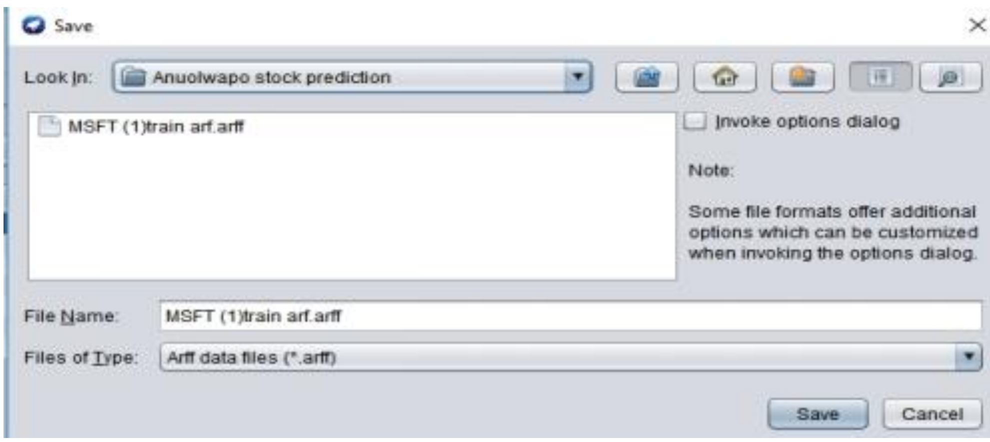
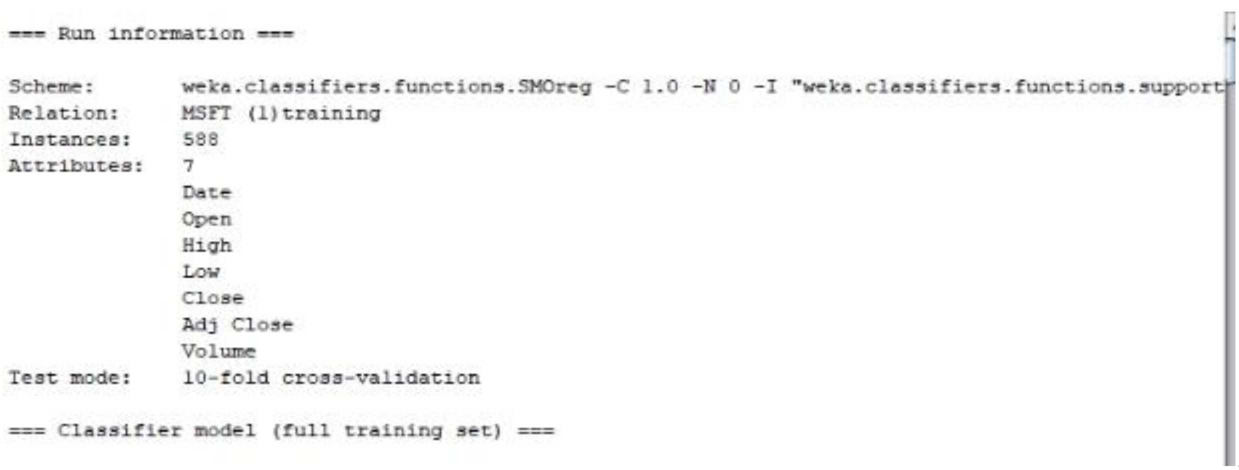


Figure 3 the sample of the pre-processed dataset used for the stock prediction

Data Pre-Processing

The pre-processed dataset was divided into two groups at the ratio of 75:25. The training set represents 75% see Figure 4 and the testing represents 25%. In the stock prediction, a comparison is made between SVM, ANN, and KStar Algorithms.



```
=== Run information ===

Scheme:       weka.classifiers.functions.SMOreg -C 1.0 -N 0 -I "weka.classifiers.functions.support
Relation:     MSFT (1)training
Instances:    588
Attributes:   7
              Date
              Open
              High
              Low
              Close
              Adj Close
              Volume
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===
```

Figure 4. Training data for the stock prediction

Predictive Model

Stock Prediction model using ANN as shown in Figure 5 and Figure 6. While figure 7, 8, and 9 show the predicted

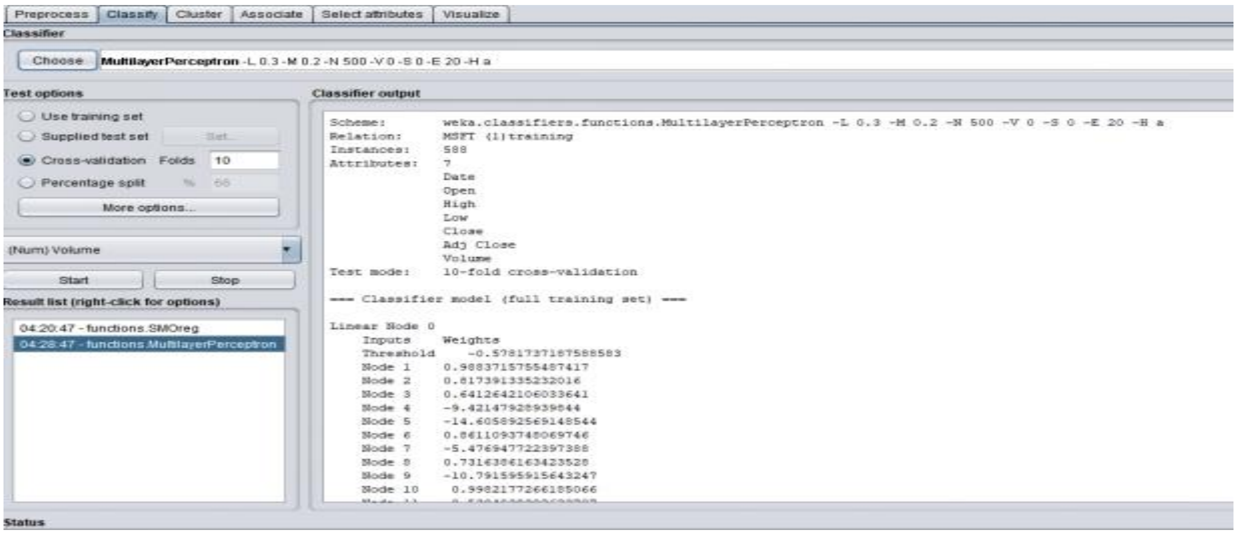


Figure 5 ANN for the stock prediction

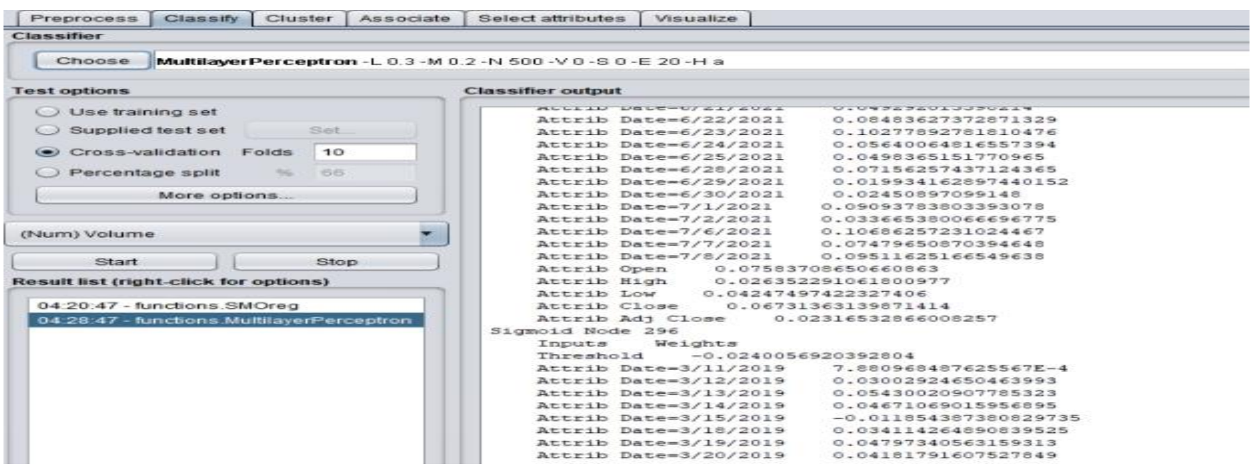


Figure 6 e ANN for the stock prediction

Cite this article as:
Ogunsanwo, G.O, Kuti. A.A, Aiyelokun O.O., & Alaba, O.B. (2024). Application of machine learning techniques for stock price prediction. *FNAS Journal of Computing and Applications*, *2*(1), 38-50.

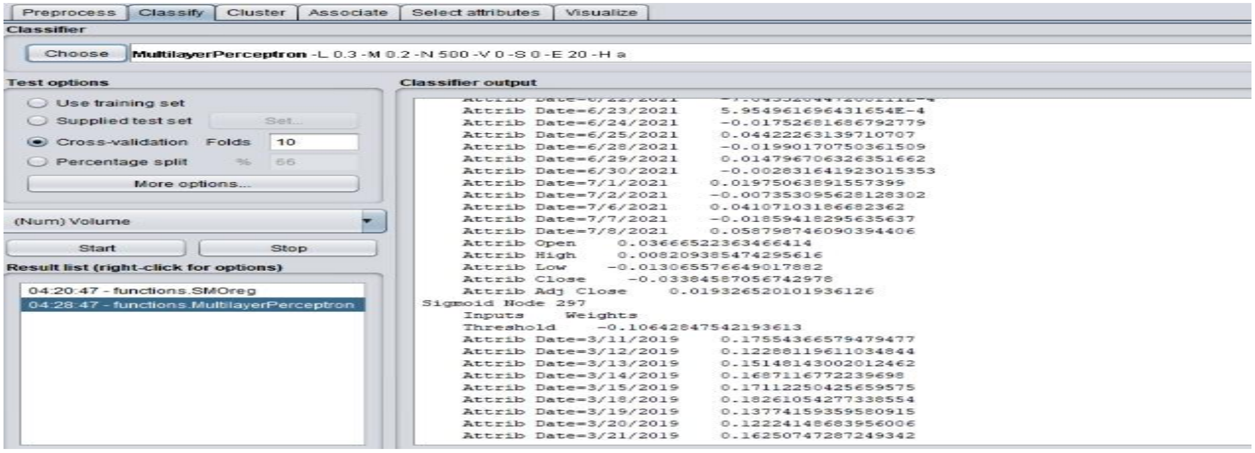


Figure 7 shows the ANN model for the stock prediction.

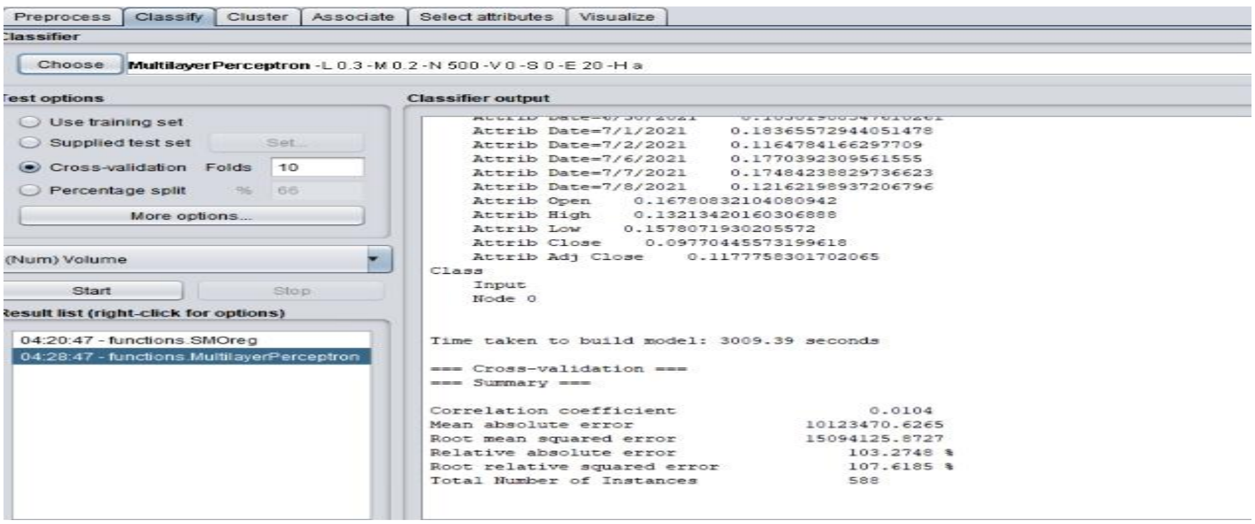


Figure 8 ANN model for the stock prediction

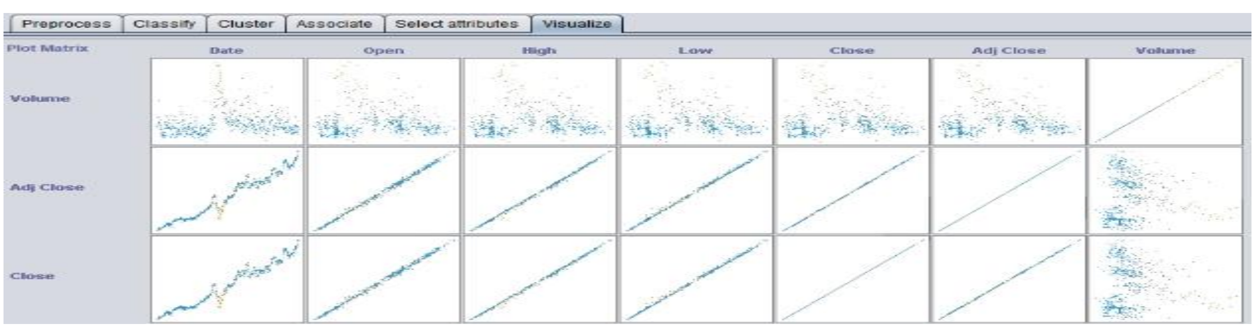


Figure 9 ANN Visualisation for the Stock Prediction

Stock Prediction model using SVM figure 10 and Figure 11, while Figure 12 shows the predicted.

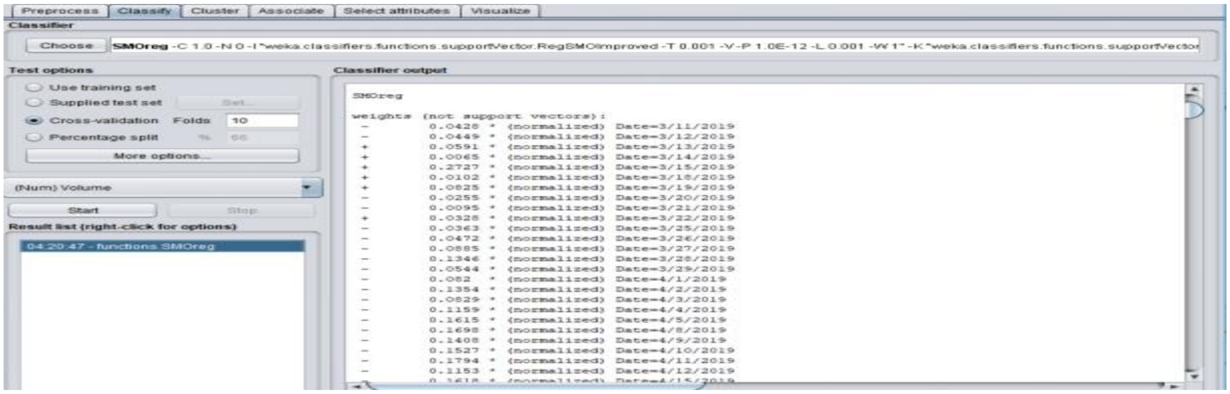


Figure 10 SVM for the stock prediction

```
Number of kernel evaluations: 173166 (92.122% cached)

Time taken to build model: 1.32 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.7184
Mean absolute error                 9504703.7538
Root mean squared error            13577664.2217
Relative absolute error                 96.9624 %
Root relative squared error             96.8064 %
Total Number of Instances              588
```

Figure 11 SVM for the stock prediction

Cite this article as:
Ogunsanwo, G.O, Kuti. A.A, Aiyelokun O.O., & Alaba, O.B. (2024). Application of machine learning techniques for stock price prediction. *FNAS Journal of Computing and Applications*, *2*(1), 38-50.

Figure 12 The SVM for the stock prediction

Stock Prediction model using KStar Algorithm as shown in Figure 13 and figure 14. while the figure 15 shows the predicted

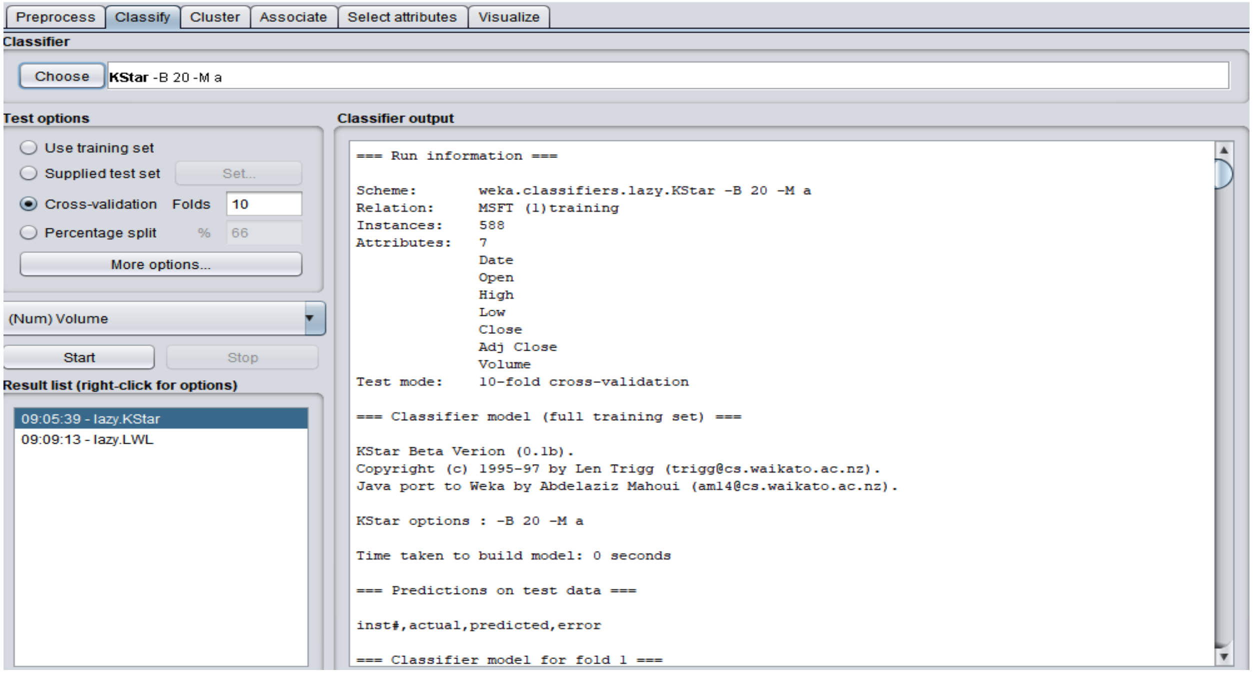


Figure 13 The KStar for the stock prediction

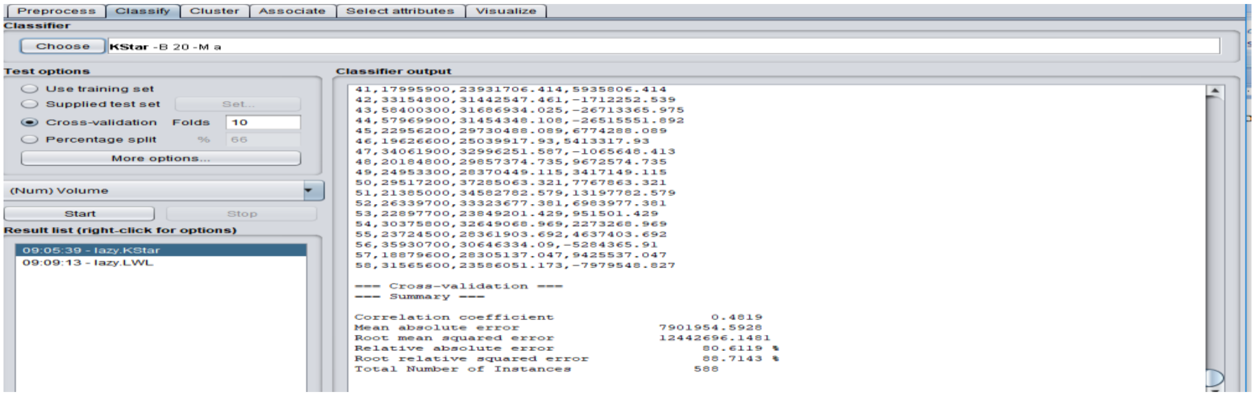


Figure 14 The KStar

Cite this article as:
Ogunsanwo, G.O, Kuti. A.A, Aiyelokun O.O., & Alaba, O.B. (2024). Application of machine learning techniques for stock price prediction. *FNAS Journal of Computing and Applications*, *2*(1), 38-50.

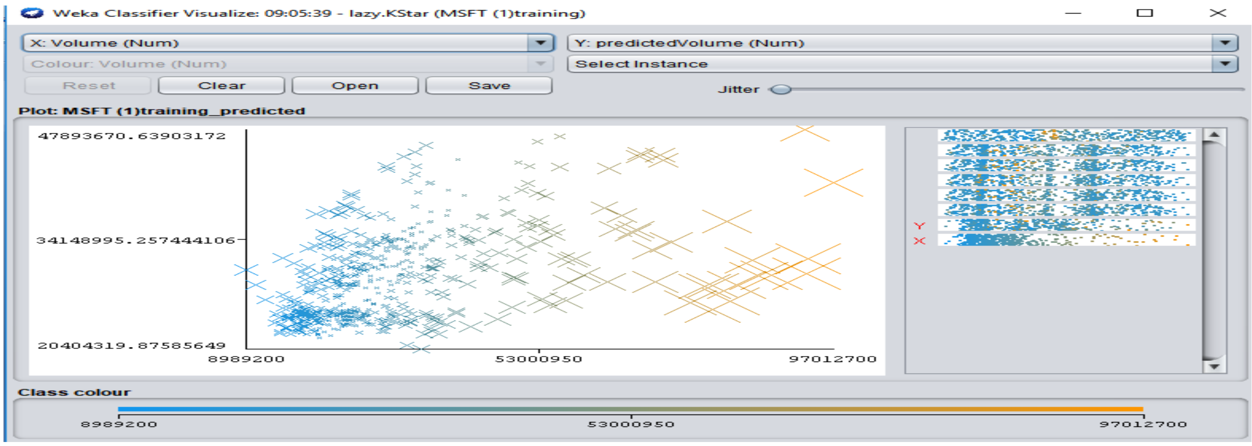


Figure 15 The KStar

Evaluation of the predictive model

The effectiveness of the models is evaluated by doing a comparison between three machine learning techniques which are: SVM, ANN, and KStar models. Predicted closing prices are subjected to RMSE, MAPE, and MBE for finding the final minimized errors in the predicted price. As shown in figures 16, 17, and 18 respectively.

```
Number of kernel evaluations: 173166 (92.122% cached)

Time taken to build model: 1.32 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.7184
Mean absolute error                9504703.7538
Root mean squared error           13577664.2217
Relative absolute error                 96.9624 %
Root relative squared error             96.8064 %
Total Number of Instances              588
```

Figure 16 Evaluation of the SVM model

```
04:28:47 - functions.MultilayerPerceptron

    Node 0

Time taken to build model: 3009.39 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.0104
Mean absolute error               10123470.6265
Root mean squared error           15094125.8727
Relative absolute error                103.2748 %
```

Figure 17 Evaluation of the ANN

Cite this article as:
Ogunsanwo, G.O, Kuti. A.A, Aiyelokun O.O., & Alaba, O.B. (2024). Application of machine learning techniques for stock price prediction. *FNAS Journal of Computing and Applications*, *2*(1), 38-50.

```
=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.4819
Mean absolute error                7901954.5928
Root mean squared error           12442696.1481
Relative absolute error                 80.6119 %
Root relative squared error             88.7143 %
Total Number of Instances              588
```

Figure 18 Evaluation of the KStar stock prediction

Results

The comparison between the three models for stock prediction as shown in Table 2

Table 2: Result of three models developed for the stock price prediction

S/NO	Particulars	ANN	SVM	KStar Algorithm
1	Correlation Coefficient	0.0104	0.7184	0.4819
2	Mean Absolute Error	10123471	9504704	7901955
3	Root Mean Squared Error	15094126	13577664	12442696
4	Relative Absolute Error(%)	103.27	96.96	80.61
5	Root Relative Squared Error(%)	107.62	96.81	88.71
6	Total Number of Instances	588	588	588
7	Time Taken	3009.39 seconds	1.32 seconds	0.44 seconds

Discussion

Table 2 shows the results of the model developed using SVM, ANN and KStar Algorithm based on training by inputting 75% of the cleaned dataset, and validation using the remaining 25% dataset. The table describes the results of the three models using 588 instances. The result revealed that the ANN model developed has a 0.0104 correlation coefficient which is close to 1 which indicates that there is a relation between the Prediction Model. Also with 10123470.6265, 1509412.8727, 103.27%, and 107.6185% values for MAE, RMSE, and RAE respectively. The result further revealed that SVM has a 0.7184 correlation coefficient which indicates that there is a strong relationship between the Prediction Model, 9504703.7538, 13577664.2217, 96.9624%, 96.8064% values for MAE, RMSE, RAE respectively. The result of the KStar Algorithm as shown in Table 4.1 shows that the algorithm has a 0.4819 correlation coefficient, 7901955,12442696,80.61%,88.71% values for MAE, RMSE, and RAE respectively.

Comparison of Developed Models: We can deduce from the comparative analysis of the models developed using Table 2 which shows that ANN has a strong positive relationship frame followed by SVM and KStar as shown in Figure 19. The KStar shows a faster learning time of the algorithm followed by SVM and ANN.

Note: MAE and RMSE values were converted to percentage

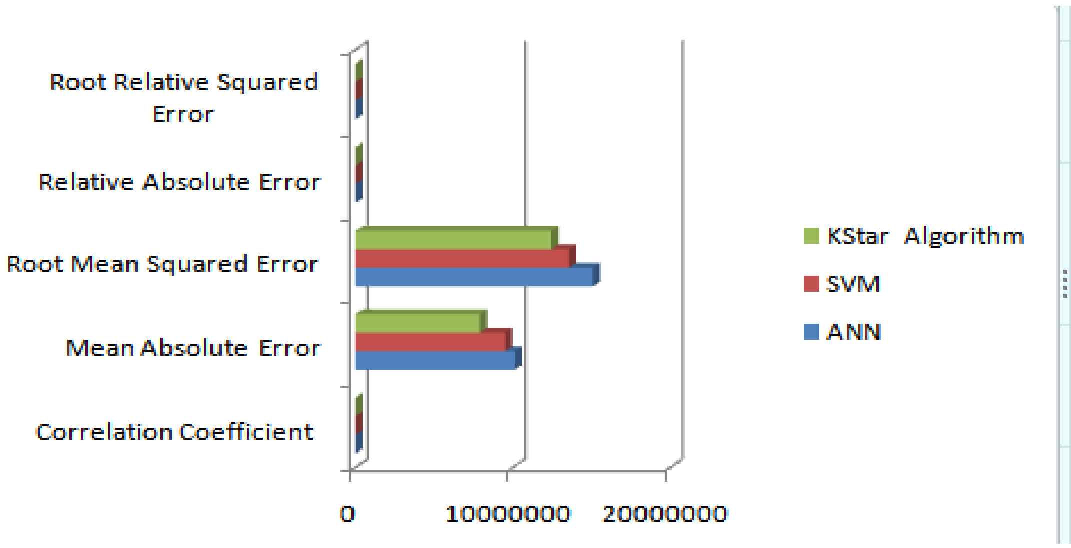


Figure 19 the comparison between the developed model

Conclusion

The study revealed that ANN, SVM, and K-Star classification algorithms are effective methodologies that can be used in the field of stock prediction. The study concluded that ANN performs better than SVM and KStar in predicting Stock Market Prices. The study recommends that the model can be used by stock brokers and the outcome of this study can be used by researchers working in stock price prediction. The study suggested that Deep Learning algorithms can be used to develop the predictive model for stock price prediction to have better performance comparison.

References

Adebayo, P., Osakwe, C., & Kayode, A. A. (2012). Prediction of stock market in Nigeria using artificial neural network. *International Journal of Intelligent Systems Technologies and Applications, October 2012*. https://doi.org/10.5815/ijisa.2012.11.08

Asagba, P. O., Neenwi, S., & Kabari, L. G. (2013). Predicting the Nigerian stock market using artificial neural network. *European Journal of Computer Science and Information*, *1*(1), 30–39.

Behravan, S. M., & Razavi, S. M. (2020). Stock price prediction using machine learning and swarm intelligence. Journal of Electrical and Computer Engineering Innovations (JECEI), 8(1), 31–40.

Chen, L. P. (2020). Using machine learning algorithms on prediction of stock price. *Journal of Modeling and Optimization*, *12*(2), 84–99.

Cortes, C., & Vapnik, V. N. (1995). Support-vector networks. Machine Learning, 20, 273–297. https://doi.org/10.1007/BF00994018

Ghani, U., Awais, M., & Muzammul, M. (2019). Stock market prediction using machine learning (ML) algorithms. *Advances in Distributed Computing and Artificial Intelligence Journal*, *8*(4), 97–116.

Kumar, V. B., Sai, K. B., Singh, K., & Srikar, B. (2020). Stock price prediction using machine learning. *International Journal of Creative Research Thoughts (IJCRT)*, *8*(2), 1–4.

Kundu, A., & Goyal, A. K. (2020). ARIMA and Indian stock market forecasting. *Journal of Xi'an University of Architecture & Technology*, (ISSN: 1006-7930).

Sathya, R., Kulkarni, P., Khalil, M. N., & Nigam, S. C. (2020). Stock price prediction using reinforcement learning and feature extraction. International Journal of Recent Technology and Engineering (IJRTE), 8(6). (ISSN: 2277-3878).