Stock Price Prediction Using Genetic Neuro-Fuzzy Model

AKINTOLA K.G[1], OLATUNDE O.V.[2]
[1, 2] *Department of Software Engineering, Federal University of Technology, Akure, Ondo State, Nigeria*

Abstract- The need to accurately predict the price of stock before trading is important as it will minimize loss and maximize profit. A lot of approaches have been used to do this but the results obtained have not been satisfactory. This paper therefore presents a hybrid of genetic algorithm and the adaptive neuro-fuzzy systems for stock price prediction. A neuro-fuzzy model was optimized using Genetic Algorithm which is an evolutionary approach. The model was tested using stock dataset of First bank Nigeria PLC. The model was trained using 2001 data items consisting of 3 attributes obtained during feature selection. The model's parameter obtained from the training are then saved. For testing, 228 data items are used to test the model. Out of this, 172 are classified correctly while 56 were misclassified. The model was compared with a neural networks model, a decision tree model and a neuro-fuzzy model. The model outperforms these models by having the lowest mean square error.

Indexed Terms- Stock, neuro-fuzzy, genetic, evolutionary, model

I. INTRODUCTION

Stock is a general term used to describe the ownership certificates of any company. A share on the other hand, refers to the stock certificate of a particular company. Holding a particular company's share qualifies one a part owner (shareholder) of the company (Coleman, 2016). Stock market is a place where company stocks, bonds and other securities are traded at an agreed price. Stock exchange is the body that runs a stock market. The stock market in Nigeria is run by the Nigeria Stock Exchange (NSE) (Nwaiwu, 2004; kamich, 2003). The stock market is essentially a non-linear, nonparametric system that is extremely hard to model with any reasonable accuracy (Wang, 2002). Stock price prediction is the act of trying to determine the future value of a company stock or other financial instruments traded on a financial exchange.

To efficiently trade in stock, there is the need to forecast the future prices of stock in order to know whether to sell or buy more of the stock. Investors have used several techniques to predict stock prices and to find the right stocks and right timing to buy or sell. The methods can be categorized as fundamental analysis and technical analysis. In fundamental analysis, trading rules are developed based on the information associated with macroeconomics, industry, and company. On the other hand, technical analysis refers to the various methods that aim to predict future price movements using past stock parameters such as prices and volume information. It is based on the assumption that history repeats itself and that future market directions can be determined by examining historical price data (Ritchie, 1996; Murphy, 1999). Most researchers believe that fundamental analysis is a good method only on a long-term basis.

Several machine learning algorithms have been developed to solve this problem. Akintola et. al. (2009) presents a linear model for stock market prediction. The stock data however is not linear. Therefore, in Akintola et. al. (2011) a model that can learn the non-linearity of stock data was developed. This model however still underperforms the prediction task. Thus in 2018 Akintola developed a neuro-fuzzy model which is a hybrid of neural netwoks and fuzzy logic models with the aim of improving the prediction accuracy. The technique used to train the ANFIS model which is the hybrid of least square and back-propagation algorithm can get stuck at local maxima. In this paper therefore, the parameter optimization of the neuro-fuzzy model using Genetic Algorithm (GA) with the hope of improving the prediction accuracy of ANFIS model is presented.

II. RELATED WORKS

Abbasi and Abouec (2008) presents Stock Price Forecast by Using Neuro-Fuzzy Inference System, in this research, a model to forecast the current trend of stock price of the "IRAN KHODRO corporation" at Tehran Stock Exchange using using an Adaptive Neuro - Fuzzy Inference system is proposed. Two models were developed; one for the Long term and one for the short-term period. For the Long-term Period, a Neuro-Fuzzy with two Triangular membership functions and four independent Variables including trade volume, Dividend Per Share (DPS), Price to Earnings Ratio (P/E), and also closing Price and Stock Price fluctuation as a dependent variable are selected as an optimal model. For the short-term Period, a neureo –fuzzy model with two triangular membership functions for the first quarter of a year, two trapezoidal membership functions for the Second quarter of a year, two Gaussian combination membership functions for the third quarter of a year and two trapezoidal membership functions for the fourth quarter of a year were selected as an optimal model for the stock price forecasting. In addition, three independent variables including trade volume, price to earnings ratio, closing Stock Price and a dependent variable of stock price fluctuation were selected as an optimal model. The model performs well at forecasting prices but no efforts were made to select the best features for the model.

Akintola et al. (2009) presents computer decision support system for stock prices forecasting. The model is a regression model using the least square technique. Stocks of Intercontinental bank were collected from January to May 2008. The weekly prices of the stock were calculated the data was used to build a regression model. The parameters of the model were obtained using the least square method. The limitation of the model is that it could only predict linear forecasts.

Atsalakis and Valavanis (2009) developed a neuro-fuzzy system which composed of an Adaptive Neuro Fuzzy Inference System (ANFIS) controller used to control the stock market process model. Real case studies using data from emerging and well-developed stock markets – the Athens and the New York Stock Exchange (NYSE) – to train and evaluate the proposed system were carried out. Results obtained demonstrated much improved and better predictions compared to other approaches, of short-term stock market trends, and in particular the next day's trend of chosen stocks.

Akintola *et al.* (2011) presents a neural network model for forecasting the time series of stock prices of intercontinental bank of Nigeria. The aim of the work is to develop a neural network solution that can be used to predict the values of shares so that buyers especially short time operators can know which share to acquire or sell at the appropriate time. A feed-forward neural Network model was developed to predict stock prices for short time duration. Stock Prices of Intercontinental Bank Nigeria were used to test the model. The stock data were collected for the period of a year and three months and grouped into average weekly prices. Normalization of the stock prices were done which gives the range [0 . 1] in order to standardize the input data. A neural network with four inputs and two hidden layers with four neurodes each and a single (neurode) in the output layer, that is, [4:4:4:1] was trained to learn the data. The result obtained shows that the model is promising to predict stock prices. The limitation of the model is that Neural Network can get stuck at local maxima.

Hegazy et. al. (2014) investigates the power of Quantum Genetic Algorithm in a neuro-fuzzy system composed of an Adaptive Neuro Fuzzy Inference System (ANFIS) controller used in prediction of stock market. A double chains quantum genetic algorithm was developed to optimize the parameters of an ANFIS. The optimization algorithm finds the optimal value for optimization variable in ANFIS using a double chains quantum genetic algorithm. The proposed model is tested with actual financial data and it demonstrates an improved and better predictions.

Olanrewaju et al. (2017) present a Regression Tree algorithm for predicting stock prices from historical stock database of a financial institution, First Bank Nigeria Plc. The model was first trained using some of the data points called the training data. The model obtained from this training is called a Regression Tree. From the Regression Tree rules were generated that guides the prediction of the stock prices when supplied with hypothetical data. The model is then tested using the remaining data points called the testing data. The performance of the model on the testing data has 97% prediction accuracy which shows that Regression Trees is suitable in predicting stock data. The limitation of this model is getting the efficient rules.

Akintola (2018) presents the development of an adaptive Neuro-Fuzzy model for predicting the values of stock prices in order to improve on the previous studies. First Bank stock data collected for the period of eleven years was used in this research. The model was first trained using some of the data points called training data. The model was then tested using the remaining data points called the testing data. The accuracy of the model on the testing data was 98.09% which shows that the model is suitable in predicting stock data. The limitation of this work is that the training method used can get stuck at local maximum due to the algorithm used for the parameter optimization.

III. METHODOLOGY

In this work, ANFIS model was developed for the stock prediction. The parameters of the model were optimized using Genetic Algorithm.

3.1 The ANFIS Model

Figure 1 denotes the ANFIS architecture. In this figure, a circle indicates a fixed node, whereas a square indicates an adaptive node.

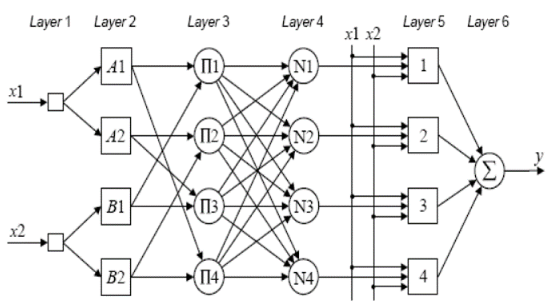


Figure 1. ANFIS model

Two typical fuzzy rules from the ANFIS model based on first order Takagi-Sugeno model are given as follow:

Rule 1: IF x_1 is A_1 AND x_2 is B_1 THEN $f_{1=}p_1x_1 + q_1x_2 + r_1$.
Rule 2: IF x_1 is A_2 AND x_2 is B_2 THEN $f_{2=}p_2x_1 + q_2x_2 + r_2$.

where x and y are the inputs,
A_i and B_i are the fuzzy sets.
f_i are the outputs within the fuzzy region specified by the fuzzy rule
$p_i, q_i and r_i$ are the consequent parameters of the model.

3.2 ANFIS Model Parameters Optimization

The ANFIS parameters both premise and consequent parameter are optimized using Genetic Algorithm. Using the Genetic algorithm, the two categories of parameters of the ANFIS model can be optimized. The process of optimizing these parameters is shown in Figure 2

a. Initialization. The first step in GA implementation is the determination of a genetic encoding scheme, that is, to denote each possible point in the problem's search space as a characteristic string of defined length. The genes represent the values of premise and consequent parameters concatenated to form a chromosome. Several of these chromosomes are randomly generated to form the initial population of the solution space.

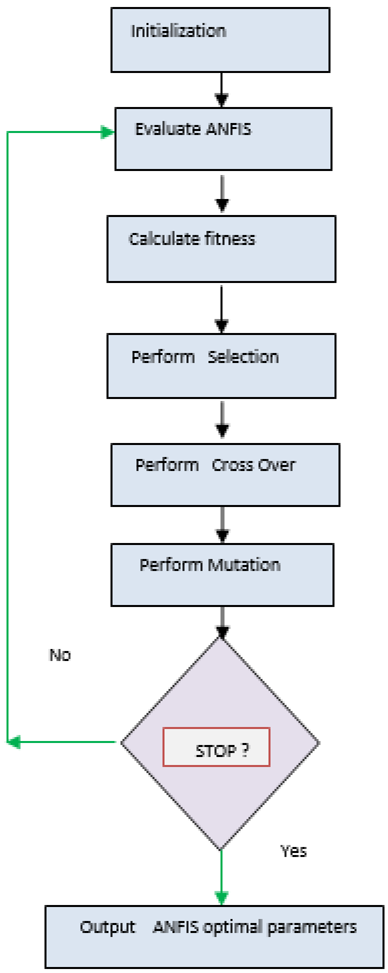


Figure 2 Flowchart of GA for ANFIS parameters optimization

Figure 3 shows a typical chromosome. The length of chromosome is obtained by concatenating the bits representing each gene as shown in Figure 3.

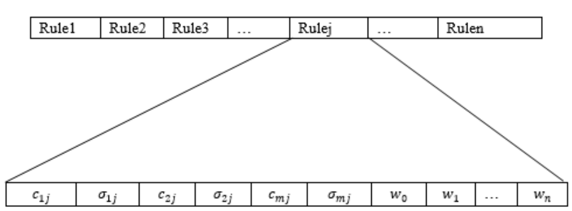


Figure 3: Sample Chromosome Encoding

In this research work, c_{1j}, σ_{1j} represents the centre and standard deviation of the Gaussian membership function. A chromosome represents one ANFIS model. The length of chromosome is obtained by concatenating the bits representing each gene as shown in Figure 3.

Typical premise parameters generated from the model are given as follows:

0.5076 2.4753 0.6723 2.7363 0.4141 1.8869 from Rule1

0.4665 2.4644 0.5930 11.8448 11.1719 59.7104 from Rule2

3.7522 23.4939 3.7474 17.2202 0.4003 2.5076 from Rule3

0.4784 2.5974 0.3325 2.2425 0.3787 2.5030 from Rule4

0.6850 1.6214 0.4720 2.1502 0.7191 2.6074 from Rule5

0.6280 1.7092 0.1890 2.6744 0.2114 2.8088 from Rule6

0.2889 2.5835 0.2407 1.1949 0.2384 1.2793 from Rule7

0.1346 1.1927 0.1577 1.0915 0.1670 1.1759 from Rule8

0.0297 1.0026 0.0417 1.1445 0.0856 1.0979 from Rule9

Typical consequent parameters generated from the model are given as follows:

-0.1000 0.0060 -0.0593 1.7686 from Rule1
-0.1000 0.0060 -0.0593 1.7686 from Rule2
-0.1000 0.0060 -0.0593 1.7686 from Rule3
-0.1000 0.0060 -0.0593 1.7686 from Rule4
-0.1000 0.0060 -0.0593 1.7686 from Rule5
-0.1000 0.0060 -0.0593 1.7686 from Rule6

-0.1000 0.0060 -0.0593 1.7686 from Rule7
-0.1000 0.0060 -0.0593 1.7686 from Rule8
-0.1000 0.0060 -0.0593 1.7686 from Rule9

b. Evaluate ANFIS: Each chromosome is evaluated by calculating their fitness function using equation 6. The root mean square error (RMSE) was used as the fitness function. The RMSE is computed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_i (d_i - o_i)^2} \quad (2)$$

where d_i is the actual value and o_i is the predicted value.

It is obvious that individuals with lower RMSE values are more likely to be reproduced during the next generation.

c. Selection. The fitness of the new offspring is calculated and sorted in the descending order. So, chromosomes of highest fitness values are selected for the next generation. In this research work, the roulette wheel method is adopted for selection. The probability of selection is given by:

$$p_i = \frac{1}{\sum_{i=0}^{N} f_i} = \frac{f_i}{f_{sum}} \quad (7)$$

in which; f_i is the fitness value of individual i, f_{sum} is the total fitness value of population; P_i is the selective probability of individual. It is obvious that individuals with high flexibility values are more likely to be reproduced during the next generation.

d. Crossover. In this research work, one-point crossing is adopted. The specific operation is to randomly set one crossing point among individual strings. When crossing is executed, partial configuration of the anterior point and posterior point are exchanged, and this gave birth to two new offspring
e. Mutation. As for two-value code strings, mutation operation is to reverse the gene values within a random number generated between zero and one.
f. Termination Criterion: The termination condition is the maximum number of generations.

IV. EXPERIMENT AND RESULT

The implementation of the GA-ANFIS model was done using MATLAB (R2013), a high-performance language for technical computing. The system was implemented in the following stages:
a. Data collection
b. Data preparation
c. Data partitioning
d. Model Development
e. Implementation and Results

4.1 Data collection

The dataset used for the prediction was downloaded online from Capital Assets Ltd, a stock brokering firm with the Nigerian Stock Exchange (NSE) with URL: www.capitalassetsng.com. The dataset was stored in an excel file format. The dataset attributes include Date, opening price, Closing price, previous close price, High, Low, Average, Deals, Volume, and Value. The dataset is for almost 11 years and contains data from the period of 12th February 2005 to 4th May 2016. The dataset includes the attributes: Date, opening price, Closing price, previous close price, High, Low, Average, Deals, Volume, and Value. Table 3 shows the attributes of the dataset with their descriptions as shown in table 3.

Table 3: Attributes with their descriptions.

Attribute	Description
Close	The last price traded during the day.
Open	The first price traded during the day or in the morning.
Previous	The Previous day close price of the stock
High	The highest traded price during the day.
Low	The lowest price traded during the day.
Average	The average price of the stock during the day.
Deals	Deals made.
Volume	No of volume of stock
Value	Value of stock

Date	The date for which the above attributes are collected.

4.2 Data preparation

Initially, all the records with missing values were removed from the dataset in order to improve the accuracy of the prediction. Then the data was further partitioned into two parts: the training data and the testing data. The training data is the data used to train the models. The total number of the training data is 2001 The testing data **i**s the data used for the purpose of testing the models. It is used to find the accuracy rate of the model. The total number of the testing data is 228.

4.3 Variable selection

The dataset contained variables that majorly influence the close price and some that has little influence. A Regression model was fitted into the data using SPSS to determine the degree of effect of each variable on close price. The comparison of the R^2 Value(s) and Standard Error value(s): Comparison of the R^2 Value(s) and Standard Error value(s) of the independent variables and each of the independent variables in predicting the outcome variable (close price). Implication in table 2 is that the independent variables were made to predict the outcome variable (close price}. Similarly, each of the independent variable elucidated above were also used as predictors. Result of the R^2 Value(s) and Standard Error value(s) obtained clearly demonstrate that Previous Close Price and Opening Price are a better predictor of the outcome variable than the other variables. However, if the variables are combined to predict the outcome variable, it slightly edges both variables (Previous Close Price and Opening Price), as it boasts a better Standard Error value.

Table 1: Comparison of the R^2 Value(s) and Standard Error value(s) of the independent variables

	Multiple Regression equation	Previous Close Price	Average Price	Deals Made	Volume Sold	Value	Opening Price
R^2 Value	0.997	0.996	0.002	0.208	0.011	0.007	0.997
Standard Error	0.659	0.75314	11.77	10.48	11.74	11.25	0.666

Furthermore, a correlation analysis of the variables of the dataset was carried out. Table 2 presents the result of the analysis.

Table 2 Correlation Coefficients Table

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.065	.036		1.800	.072
	PreviousClosePrice	.186	.031	.186	5.992	.000
	AveragePrice	-1.743E-5	.000	.000	-.319	.750
	DealsMade	.000	.000	-.005	-3.858	.000
	VolumeSOLD	1.045E-10	.000	.000	.138	.890
	Value	1.869E-12	.000	.000	.049	.961
	OpeningPrice	.814	.031	.815	26.250	.000

a. Dependent Variable: ClosePrice

In table 2, the Column under standardized coefficients is considered. It shows the significance value of each of the predictor variable x-raying their predictive power. More importantly, it shows the predictor that is statistically important contributor to the outcome variable. Predictors with values more than 0.05 are simply not good predictors.

The Column Beta under standardized coefficients reflects the predictor that makes the strongest contribution of explaining the outcome variable. The negative sign is simply ignored and the greater the value, the stronger the predictor. The following predictors were finally selected.

a. Previous close price
b. Opening price
c. Deals

4.4 Data partitioning

After variable reduction from the dataset, the dataset was divided into two parts: Training data and testing data. The whole dataset was completely partitioned into two sets making the training data 1650 and the testing data 1065.

4.5 Model Development

The genetic adaptive Neuro-Fuzzy model was implemented in MATLAB environment. The system is tested using First bank stock. The parameters of the model are shown in Table 3.

Table 3 The Models Parameters

The ANFIS model Parameters	
Network type	ANFIS (Sugeno type)
Number of layers	6
Iteration	20
Input membership function	Guassian
Output function	Constant
Number of membership functions	3-3-3
Number of fuzzy rules	9
Optimization algorithm Hybrid	GA
"and" method	prod
"or" method	probor
Clarification method	wtaver
Error Function	RMSE

The GA Parameters	
Number of population	50
Iteration	20
Selection method	Roulette wheel
Crossing probability	0.8
Mutation probability	0.5
Fitness Function	RMSE

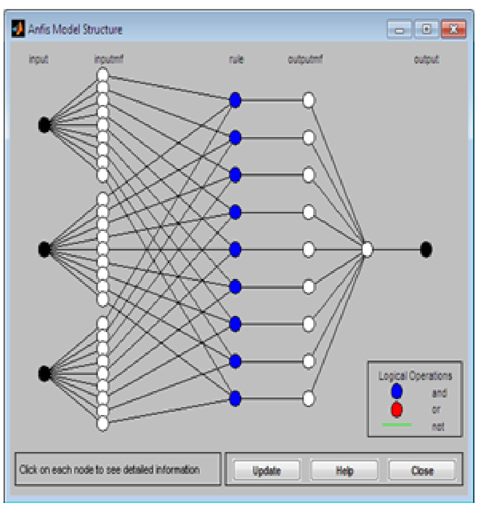
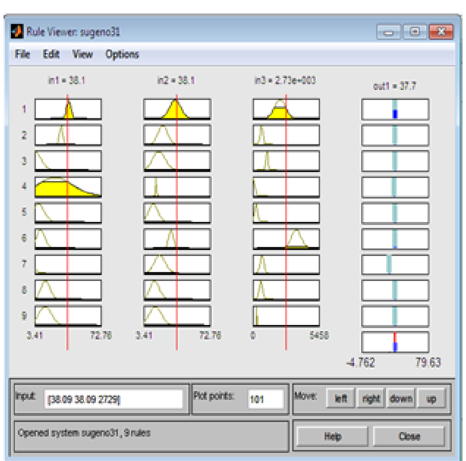
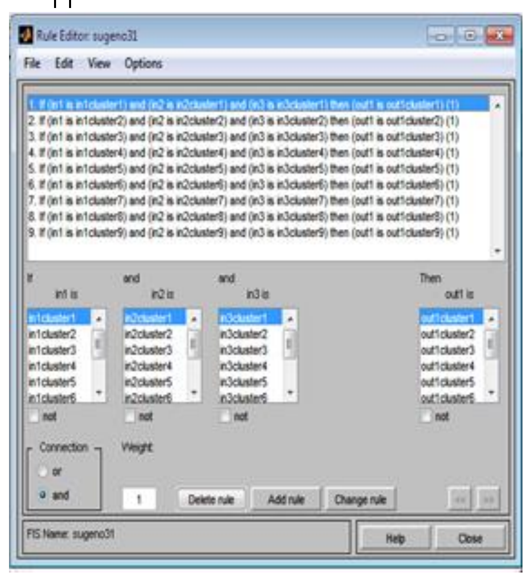


Figure 4 a The ANFIS Model Structure



b. The ANFIS Model Rule operation



c. The ANFIS Model Rules

4.6 Price Trend Prediction using the GA-ANFIS

A neuro-fuzzy model is trained using 2001 data items consisting of 3 attributes obtained during feature selection. The model's parameter obtained from the training are then saved. For testing, 228 data items are used to test the model. Out of this, 172 are classified correctly while 56 were misclassified. The Root Mean Square Error in forecast is shown in table 3.

Table 4 RMSE values of the classifiers

S/N	MODEL	RMSE
1	DTREE	0.448702
2	ANN	0.337101
3	ANFIS	0.30862
4	GAANFIS	0.302895

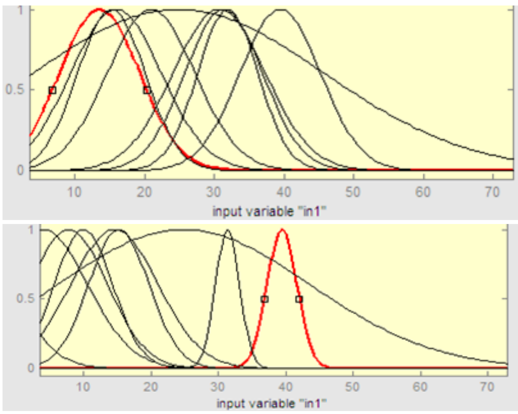


Figure 5 a. Previous close price before and after optimization

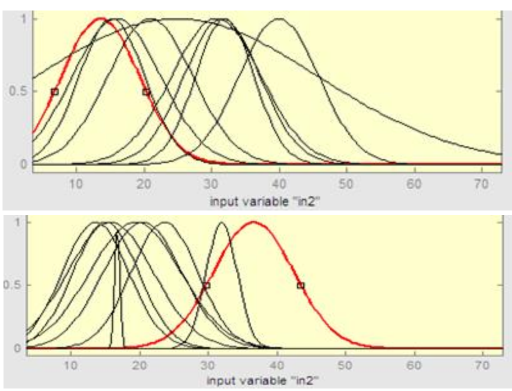


Figure 5 b. Opening price before and after optimization

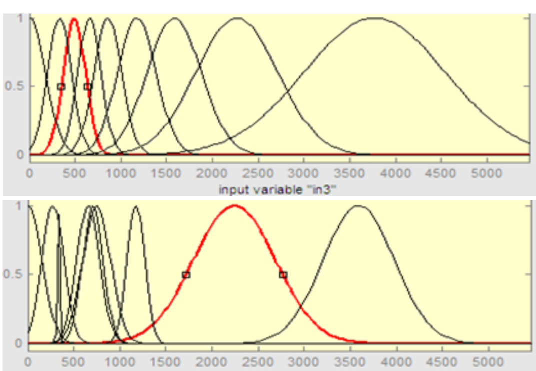


Figure 5c. Deals before and after optimization

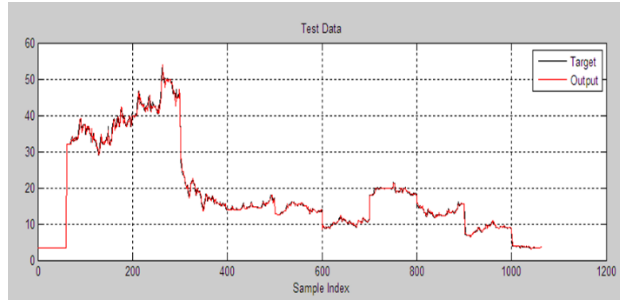


Figure 6: plot of predicted versus actual stock price.

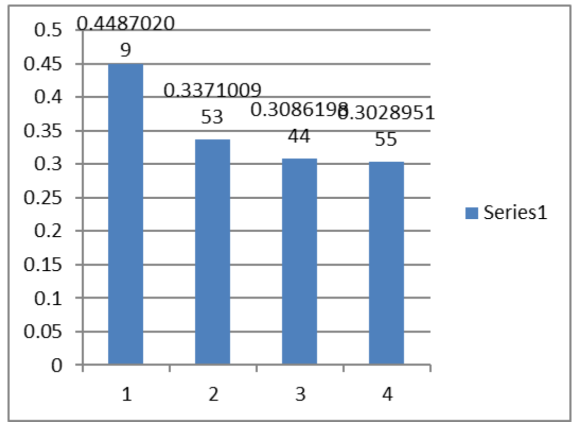


Figure 7 RMSE of the classiers

CONCLUSION

In this paper, a report of Genetic ANFIS approach to forecast stock prices is presented. For this purpose, an analysis was carried out using the First Bank of Nigeria Plc. (FBN) stock dataset. As a result, the price of the stock was predicted with RMSE of 0.303. The result shows that Nigerian stock market prices based on historical trading data can be modeled using GA-ANFIS. A comparative analysis with some other models such as ANFIS, ANN Decision tree shows that the new model has the lowest error value. In future, ensemble models will be studies to show if a better performance could be achieved.

REFERENCES

[1] Abbasi E. Abouec A. (2008) Stock Price Forecast by Using Neuro-Fuzzy Inference System, World Academy of Science, Engineering and Technology 46 2008

[2] Adewole, K.S., Olaniyi, A.S., Jimoh, R. G., (2011), Stock Trend Prediction Using Regression

Analysis –A Data Mining Approach, ARPN Journal of Systems and Software, vol.1, pp 154-155 International Journal of Trend in Research and Development, Volume 5(3), ISSN: 2394-9333 www.ijtrd.com IJTRD | May – Jun 2018 Available Online@www.ijtrd.com 346

[3] Akintola K.G., Boyinbode O.K., Ajose M. (2009), Computer Decision Support System for Stock Prices Forecasting: Towards forestalling Economic Downturn. The proceedings of the Nigerian Computer society, MACGLOBE 2009, Abuja, Nigeria.

[4] Akintola K.G. Alese B.K. Thompson A.F. (2011). Time Series Forecasting with Neural Network: A case study of stock prices of Intercontinental Bank Nigeria. IJRRAS 9 (3), December 2011.

[5] Akintola K.G. (2018) "An Adaptive Neuro-Fuzzy Inference System (ANFIS) for Forecasting Stock Price: A case study of Prices of First Bank Nigeria". International Journal of Trend Research and Development Vol. 5 Number 3 pp 340-346.

[6] Atsalakis, G. S., Valavanis, K.P. (2009) "Surveying stock market forecasting techniques – Part II: Soft computing methods", Expert Systems with Applications, vol 36, pp. 5932-5941.

[7] Murphy, J.J. (1999) Technical Analysis of the Financial Markets: a Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance.

[8] Nwaiwu, I.C. (2004), "The Beginner's Guide to Investing in the Nigerian Stock Market", Stockmarketnigeria.com.

[9] Olanrewaju O.E Akintola K.G Akinduyite C.O. (2016). Stock Prediction using Regression Tree Model, Proceedings of 2nd International conference on Intelligent Computing and Emerging Technologies (ICET, 13th -15th November, 2016, Computer Science Department, Babcock university. Ilishan Remo, Ogun State, Nigeria.

[10] Osman Hegazy, O. Soliman O. S. Toony A. A. (2014) Hybrid of neuro-fuzzy inference system and quantum genetic algorithm for prediction in stock market. Issues in Business Management and Economics Vol.2 (6), pp. 094-102, June 2014 Available online at http://www.journalissues.org/IBME/ ISSN 2350-157X

[11] Ritchie, J.C. (1996) Fundamental Analysis: a Back-to-the-Basics Investment Guide to Selecting Quality Stocks. Irwin Professional Publishing.

[12] Wang, Y.F. (2002) "Predicting stock price using fuzzy grey prediction system", Expert Systems with Applications, 22, pp. 33-39.