2016 International Conference on Computer Communication and Informatics (ICCCI -2016), Jan. 07 – 09, 2016, Coimbatore, INDIA

# Gold Price Prediction using Support Vector Regression and ANFIS models

Akash Dutt Dubey

Department of Computing Sciences

SMCS, College of Engineering, Science and Technology,

Fiji National University, Fiji

akash.dubey@fnu.ac.fj

**Abstract**: In recent times, gold has been one of the prioritized commodities in terms of long term as well as short term investments since the investors consider gold as a hedgerow against the unforeseen events leading to chaos in the market. Consequentially, the price of gold in the market plays an important role. In this research work, time-series gold price prediction models have been developed using the support vector regression and anfis models for the prediction of daily gold prices. The support vector model was designed using epsilon support vector regression method while the adaptive neural fuzzy inference systems have been developed using grid partition and subtractive clustering methods. The gold prices obtained for the training and testing were obtained from Perth Mint of Australia. The evaluation criteria for the comparison of the models are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Nash–Sutcliffe model efficiency coefficient (E) and Mean absolute percentage error (MAPE). It was observed that the models obtained using support vector regression outperformed the ANFIS models. In the ANFIS models, it was observed that ANFIS-GP performed slightly better than the ANFIS-SC model.

***Keywords: Gold Price Prediction, Support Vector Regression, ANFIS, Australia Perth Mint, Grid Partition, Subtractive Clustering***

## 1. Introduction

The rate of gold price plays a pivotal role in the economic and monetary systems of a market. Often there has been a close correlation observed between the gold price and other assets [1]. While the relation between the gold price and oil is considered to be positive, the relation between the gold price and the equities is often negative. Moreover, gold has been one of the prime investments in recent times for both short term and long term. The accurate forecasting of gold prices, therefore, can be used efficiently to tackle the market according to the anticipated trends in the future. An accurate gold price forecasting model can be used by the clients to prevent or mitigate possible risks and accordingly reduce the risk of financial losses and bankruptcy.

The traditional forecasting methods for gold price prediction involve Autoregressive Integrated Moving Average (ARIMA) [2] and multi linear regression [3] [4] [5]. With the evolution of artificial intelligence, many soft computing methods have also been used for the forecasting of the gold price. Artificial Neural Networks has been one of the major techniques that have been used for the development of the prediction model [6] while others soft computing methods too have been put into practice for the prediction of the gold price.

The aim of this research paper is to develop and compare different models which are based on Support Vector Machines (SVM) and Adaptive Neural Fuzzy Inference System (ANFIS) for the purpose of daily gold price prediction for Perth Mint, official bullion of Australia. The data used for this research ranged from 1st January 2007 to 30th March 2015. This research paper has been divided into 6 sections. Section 2 describes the related works that have been done in the area of gold price prediction. Section 3 discusses the methodologies used in the work. Section 4 discusses the data set and the valuation criteria of the described model. Section 5 illustrates the results with discussions. Section 6 concludes the work.

## 2. Related Works

According to the works done by Ismail, Yahya and Shabri [4], the price of gold is less likely to be affected over a given period of time as compared to other commodities since the supply of gold is accumulated over the a long duration of time. They applied multiple regression method to predict the gold price and developed two models which varied in the terms of factors affecting the price of the gold, the best obtained accuracy in their was 96.92%.

978-1-4673-6680-9/16/$31.00 ©2016 IEEE

The gold price prediction was done using Radial Basis Function neural networks by Hussein et al [6]. In this work, time series prediction of gold prices was done on daily basis using RBF networks. Guixia Yuan [7] studied the factors affecting that gold price and then applied these factors as input to the neural network for prediction purposes. The Back propagation neural network was optimized using the genetic algorithms. This hybrid method provided better network convergence and simple neural network structure. The results obtained from this method yielded better results as compared to simple BP neural network.

Ping, Miswan and Ahmed [8] compared the two methods Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) for the gold price prediction of Kijang Emas, official gold bullion of Malaysia. The evaluation criteria of the models used in this study was Akaike's information criterion (AIC) and mean absolute percentage error (MAPE). The results concluded that the GARCH models had better predicting ability than the ARIMA models. However, a comparison of performances of GARCH models and artificial neural networks was done by Sarangi and Dublish [9] where they used 18 different specifications from the GARCH family of models and 20 different artificial neural network models. In this study, it was concluded that the artificial neural networks performed better than the GARCH models.

Kangarani Farahani and Mehralian [10] have studied and compared the artificial neural networks and adaptive neural fuzzy inference system (ANFIS) models for the purpose of gold price prediction and found out that the predicted results of ANFIS were better than ANN with comparatively low prediction error. It was also found that the wavelet de-noising of the data resulted in lower prediction results due to the chaotic nature of the gold price. Li Bai [11] applied wavelet neural network with the artificial bee colony algorithm to predict the gold price. In this study, the conventional roulette wheel strategy was discarded and it was found out that this method yielded higher efficiency as compared to simple WNN performance.

## 3. Methodology

**Support Vector Machine:**

Support Vector Machines have been studied extensively in the area of machine learning and data mining. Support Vector Machines were developed by Cortes and Vapnik [12] in 1992 for the purpose of the classification of supervised learning framework. The SVMs have been extended for the purpose of regression and rank learning [13] [14]. While initially, the support vector machines were basically used as a binary classifiers used to classify data objects into positive and negative categories, its extended versions can also be used for multiclass classification.

In a support vector machine, the data sample X is mapped into a high dimensional feature space with the help of the kernel functions followed by the projection of an optimal hyper plane in the new space which then separates the different classes. Support vector machines, therefore is more reliable than the other conventional learning methods since it avoids the risk of local minima and producing a global optimum result. The Support Vector Regression (SVR) combines regression with support vector machine and the dependencies between the input dataset and the target datasets is calculated.

In epsilon support vector regression, a training set is given as $\{(x_1,y_1)\ldots(x_t,y_t)\} \subset R^n$, where $x_i$ is the set of input features while $y_x$ is the set of target values. The purpose of the support vector regression is to find a function f(x) such that the variation of this function is at most ε from all the target values and is as flat as possible. In other words, the function can be defined as

$$f(x) = w * \Phi(x) + b \tag{1}$$

Where $w \subset R^n$, $b \subset R$ and $\Phi$denotes the nonlinear transformation from $R^n$ to high-dimensional space. The main target in this function is to find the appropriate values of w and b which will allow minimum regression risk in the evaluation of the value of x.

$$R_{reg}(f) = C\sum_{i=0}^{l}\tau(f(x_i)-y_i) + \frac{1}{2}\left|\left|w\right|\right|^2 \tag{2}$$

Where $\tau$ is the cost function and C is the constant which determines the adjustment between the deviations from target larger than ε and the flatness of the function. The vector w can be described in the terms of data points as

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\Phi(x_i) \tag{3}$$

On substituting the value of w in the equation 1,

$$\begin{aligned} f(x) &= \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\big(\Phi(x_i).\Phi(x)\big) + b \\ &= \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(x_i, x) + b \end{aligned} \tag{4}$$

The dual functions $\Phi(x_i).\Phi(x)$ in this equation form the basis of the non-linear formulation of the SVR. Known as the kernel trick, these two functions can be effectively replaced by the kernel functions which map the input data into a higher dimension space whose linear regression will be equal to the non-linear regression in the original space. The transformation of the function from low dimension to high dimension does not need any knowledge of $\Phi$, thereby facilitating the transformation. The kernel function used in this study is the RBF (Radial Basis Function). Some of the common kernels are as below:

**Table 1.** Kernels used in the Support Vector Machine

| Kernel | Function |
|---|---|
| Linear | x*y |
| Polynomial | $[(x*x_i)+1]^d$ |
| Radial Basis Function (RBF) | $\mathrm{Exp}\{-\gamma\lvert x-x_i\rvert^2\}$ |

**Adaptive Neural Fuzzy Inference System:**

Adaptive neuro-fuzzy inference system is a hybrid technique which uses fuzzy inference system within the adaptive neural network framework, used for the recognition and control of complex non-linear systems. Although artificial neural networks are considered to be a powerful technique for the modeling of different real world problems, it has its own limitations. ANFIS provide a better option on the problems which deal with the problems where input data is either ambiguous or is subjected to high uncertainty. ANFIS also overcomes the limitations of a fuzzy inference system like the requirement of an expert for fuzzy rule generation and designing non-adaptive fuzzy set.

The ANFIS model is a sugeno fuzzy model which can be described by a forward network structure. The sugeno fuzzy model is framed into a neural network model to facilitate the adaptation thereby creating a fuzzy rule set depending on the input and the target data. The Sugeno model is basically based on the if-then-else framework which can be described as below:

$$\text{if } x_1 \text{ is A and } x_2 \text{ is B then } y= f(x_1,x_2) \quad (5)$$

Where A and B defined the fuzzy sets and the function f is function which is associated with the facts of $x_1$ and $x_2$ which correspond to A and B respectively.

The figure 1 illustrates the architecture of the ANFIS model which has been divided into five layers [15].. These five layers can be described as follows:

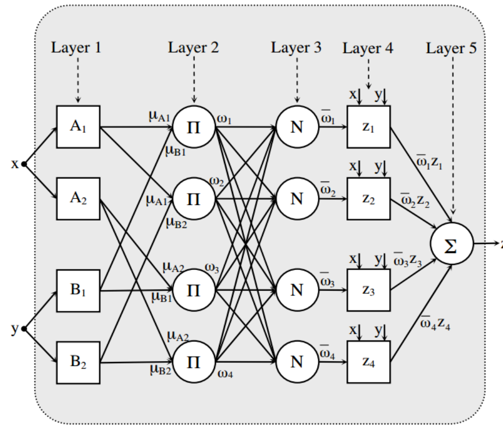


Figure. 1. Basic architecture of ANFIS

Layer 1:

This layer consists of adaptive nodes which have a node function associated with it. The input to this layer is considered to be x for node i, along with the linguistic labels $A_i$ with them. The output of this layer will be a membership rank of fuzzy set A(=$A_1$, $A_1$, $B_2$ and $B_2$) and can be represented as $O_{1,i}$ . The Gaussian 2 parameter used for this layer will be

$$\mu_a(x) = exp\left(-\frac{1}{2}\left(\frac{x-c_i}{a_i}\right)^2\right) \quad (6)$$

given $c_i$ and $a_i$ as the premise or antecedent parameters.

Layer 2:

This layer consists of fixed nodes which are used to multiply the incoming signals and produce the outputs. The outputs produced by this layer carry a firing strength of the rule for each node.

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \quad (7)$$

Layer 3:

The number of nodes in this layer is also fixed and it is used to normalize the firing strengths of the nodes using the following equation:

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \tag{8}$$

Where $\overline{w_i}$ is the normalized firing strength.

Layer 4:

This layer consists of adaptive nodes with a node function which can be given as follows:

$$O_{4,i} = \overline{w_i} f = \overline{w_i}(p_i x + q_i y + r_i) \tag{9}$$

The parameters $p_i$, $q_i$ and $r_i$ are the parameters of the nodes.

Layer 5:

The overall output of this network is calculated by a single node in this layer which is the summation of the incoming signals. The calculation can be done as below:

$$overall\ output = O_{5,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i \overline{w_i} f_i}{\sum_i \overline{w_i}} \tag{10}$$

A detailed study of ANFIS has been done in [16-17]. For the development of the ANFIS sugeno model, different identification methods such as grid partitioning and subtractive clustering can be adopted.

## 4. Data Sets and Evaluation Criteria

The data set used for this study has been collected from the official bullion of Australia, Perth Mint. The data extends from 1st January 2007 to 30th April 2015 and has total 2206 data samples for the simulation of the models. In order to evaluate the efficiency and prediction errors of the models, four evaluation criteria were used root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and Nash–Sutcliffe model efficiency coefficient (E) were used as the statistical evaluation criteria.

The Root Mean Square Error measures the standard deviation between actual value and the predicted value. The mean absolute error (MAE) is the measure of the average of the errors in the prediction of the model. The accuracy of a prediction model can be expressed in terms of Mean Absolute Percentage Error (MAPE) which is a common measure of forecast error. Nash–Sutcliffe model efficient coefficient is also a criterion to assess the predictive power of the model. The value of E can range from - ∞ to 1, the closer its value is to 1, the more accurate model is considered to be.

The formula for the RMSE, MAE, MAPE and E can be

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{12}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{13}$$

$$E = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_0})^2} \tag{14}$$

$$R = \frac{n(\sum y_i \hat{y}_0) - \sum y_i \sum \hat{y}_0}{\sqrt{[n\sum y_i^2 - (\sum y_i)^2][n\sum \hat{y}_0^2 - (\sum \hat{y}_0)^2]}} \tag{15}$$

## 5. Results and Discussions

The data sets were used to train and test three different models based on support vector regression, adaptive neural fuzzy inference system grid partitioning method and adaptive neural fuzzy inference system subtractive clustering method. The models were developed and compared to each other on the basis of RMSE, MAE, MAPE and E.

The support vector regression method used was epsilon SVR while the ANFIS models were trained and tested using two Gauss Membership Functions. The number of epochs was set to 100 to minimize the risk of overtraining of the network. For the SVR model, the kernel type selected was Radial Basis Function. For the optimal SVR, the values of the epsilon, cost parameter and the value of the gamma for the SVR were set to .00005, 200 and .0025 respectively.

Table 2: Performance of the Prediction Models

| | ANFIS-GP | ANFIS-SC | SVR |
|---|---|---|---|
| **RMSE** | 15.93184 | 16.24604 | **14.85914** |
| **MAE** | 10.82450 | 10.88973 | **8.021052** |
| **MAPE** | 0.008353 | 0.008374 | **0.0063055** |
| **E** | 0.996431 | 0.996289 | **0.9970015** |

The results obtained while developing the SVR, ANFIS-GP and ANFIS-SC models have been given in the table 2. According to the results, it was concluded that the Nash Sutcliffe efficiency model was closest to 1 in the case of support vector regression model, thereby suggesting higher efficiency ability of this model. The root mean square error, mean absolute error and the mean absolute percentage error were lowest in the case of support vector regression models which further support the deduction that the SVR model was more efficient in predicting the gold price. For the ANFIS models, the results of ANFIS-GP showed a slight improvement over the performance of the ANFIS-SC models.

The illustration of the actual price and the price predicted by the different models has been given in the figure 1. This chart shows the values for the month of March 2015 and April 2015. Similarly, the table 2 has been given to indicate the gold price for the month of April 2015 and the values predicted by the three different models for the same period. It can be seen in the chart and the table that the values predicted by the SVR model is better than the ANFIS-SC and ANFIS-GP models, thereby suggesting that SVR have better predicting capabilities in this case.

## 6. Conclusions

In this research paper, three different models have been developed for the gold price prediction in the Pert Mint, official bullion of Australia. The models used Support Vector Regression, ANFIS-GP and ANFIS-SC methods. The data samples used for this study ranges from 1st January 2007 to 30th March 2015. From the results obtained, it was concluded that the support vector regression method had better prediction ability for this purpose while the ANFIS- GP had a slight improvement over ANFIS-SC method.

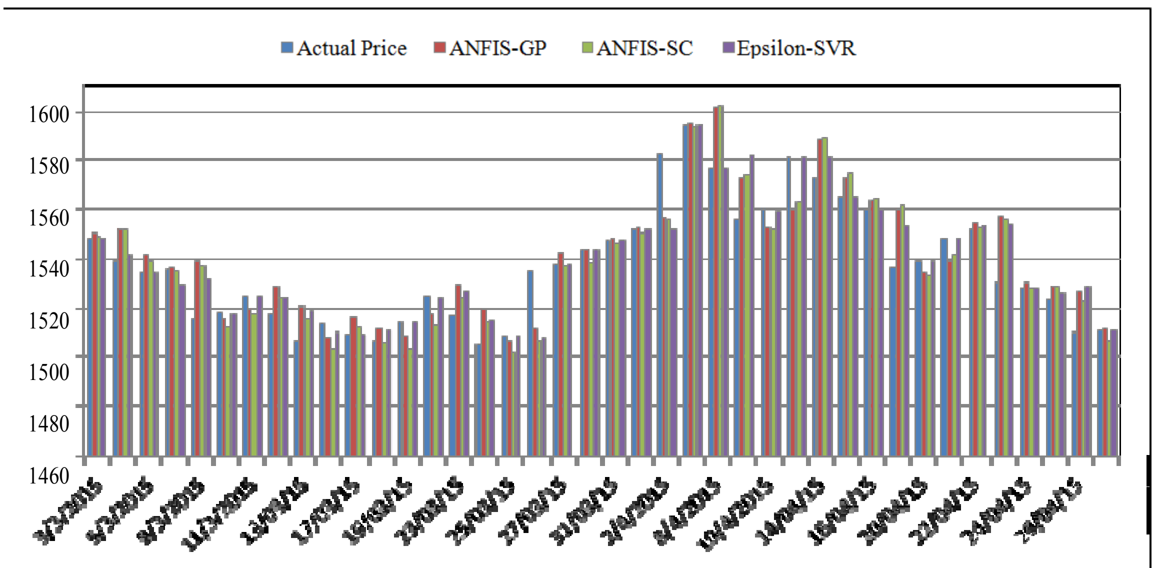


Figure 2. Prediction of the models for March-April 2015

## References:

1. C. Corti and R. Holliday, Gold. Boca Raton, FL: CRC Press, 2010.
2. A. Parisi, F. Parisi and D. DÃ-az, 'Forecasting gold price changes: Rolling and recursive neural network models', Journal of Multinational Financial Management, vol. 18, no. 5, pp. 477-487, 2008.
3. A. Escribano and C. Granger, 'Investigating the relationship between gold and silver prices', Journal of Forecasting, vol. 17, no. 2, pp. 81-107, 1998.
4. P. Achireko and G. Ansong, 'Stochastic model of mineral prices incorporating neural network and regression analysis', Mining Technology, vol. 109, no. 1, pp. 49-54, 2000.
5. Z. Ismail, A. Yahya and A. Shabri, 'Forecasting Gold Prices Using Multiple Linear Regression Method', American Journal of Applied Sciences, vol. 6, no. 8, pp. 1509-1514, 2009.
6. S. Hussein, M. Shah, M. Jalal and S. Abdullah, 'Gold price prediction using radial basis function neural network', 2011 Fourth International Conference on Modeling, Simulation and Applied Optimization, 2011.
7. Yuan, G. 'Study on Gold Price Forecasting Technique Based on Neural Network Optimized by GA with Projection Pursuit Algorithm', JCIT, vol. 7, no. 18, pp. 558-565, 2012.
8. P. Ping, S. Yaziz, M. Ahmad and N. Miswan, 'Forecasting Malaysian gold using a hybrid of ARIMA and GJR-GARCH models', ams, vol. 9, no. 29-32, pp. 1491-1501, 2015..
9. P. Sarangi and S. Dublish, 'Prediction of Gold Bullion Return Using Garch Family and Artificial Neural Network Models', Asian Journal of Research in Business Economics and Management, vol. 3, no. 10, pp. 217-230, 2013.
10. M. KangaraniFarahani and S. Mehralian, 'Comparison between

Artificial Neural Network and neuro-fuzzy for gold price prediction', 2013 13th Iranian Conference on Fuzzy Systems (IFSC), 2013.
11. B. Li, 'Research on WNN Modeling for Gold Price Forecasting Based on Improved Artificial Bee Colony Algorithm', Computational Intelligence and Neuroscience, vol. 2014, pp. 1-10, 2014.
12. C. Cortes and V. Vapnik, Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.
13. Herbrich, R., Graepel, T., Obermayer, K. (eds.): Large margin rank boundaries for ordinal regression. MIT-Press. 2000.
14. H. Yu, 'Selective sampling techniques for feedback-based data retrieval', Data Mining and Knowledge Discovery, vol. 22, no. 1-2, pp. 1-30, 2010
15. N. Sarikaya, K. Guney and C. Yildiz, 'Adaptive Neuro-Fuzzy Inference System For The Computation Of The Characteristic Impedance And The Effective Permittivity Of The Micro-Coplanar Strip Line', Progress In Electromagnetics Research B, vol. 6, pp. 225-237, 2008.
16. J. Jang, 'ANFIS: adaptive-network-based fuzzy inference system', IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665-685, 1993
17. Jang, J., Sun, C., Mizutani, E.: Neuro-fuzzy and soft computing. Prentice Hall, Upper Saddle River, NJ. 1997.

Table 2: Actual and Predicted Gold Prices for April 2015

| | Actual Value | SVM | ANFIS-GP | ANFIS-SC |
|---|---|---|---|---|
| 1/4/2015 | $1,552.06 | $1,551.96 | $1,552.38 | $1,550.58 |
| 2/4/2015 | $1,582.43 | $1,551.78 | $1,556.89 | $1,555.97 |
| 7/4/2015 | $1,594.28 | $1,594.18 | $1,594.56 | $1,593.38 |
| 8/4/2015 | $1,576.67 | $1,576.57 | $1,601.16 | $1,601.68 |
| 9/4/2015 | $1,556.28 | $1,581.99 | $1,572.64 | $1,574.32 |
| 10/4/2015 | $1,560.35 | $1,559.40 | $1,552.41 | $1,551.84 |
| 13/04/15 | $1,580.87 | $1,580.77 | $1,560.00 | $1,562.88 |
| 14/04/15 | $1,572.71 | $1,581.18 | $1,588.33 | $1,589.23 |
| 15/04/15 | $1,564.82 | $1,564.72 | $1,572.76 | $1,574.75 |
| 16/04/15 | $1,560.30 | $1,560.20 | $1,563.29 | $1,564.57 |
| 17/04/15 | $1,536.26 | $1,553.22 | $1,560.14 | $1,561.49 |
| 20/04/15 | $1,539.11 | $1,539.21 | $1,534.12 | $1,533.24 |
| 21/04/15 | $1,548.60 | $1,548.50 | $1,539.59 | $1,541.12 |
| 22/04/15 | $1,551.91 | $1,553.52 | $1,554.64 | $1,552.76 |
| 23/04/15 | $1,530.48 | $1,554.05 | $1,557.01 | $1,555.63 |
| 24/04/15 | $1,528.35 | $1,528.45 | $1,530.00 | $1,528.10 |
| 28/04/15 | $1,523.53 | $1,526.01 | $1,529.15 | $1,528.72 |
| 29/04/15 | $1,509.94 | $1,528.76 | $1,526.59 | $1,522.58 |
| 30/04/15 | $1,511.33 | $1,511.34 | $1,506.61 | $1,510.56 |