An Intelligent Gold Price Prediction Based on Automated Machine and k-fold Cross Validation Learning

Yakubu S. Baguda† and Hani Meateg Al-Jahdali††

Information Systems Department, Faculty of Computing and Information Technology,
King Abdulaziz University, Rabigh, Kingdom of Saudi Arabia.

Summary
The rapid change in gold price is an issue of concern in the global economy and financial markets. Gold has been used as a means for trading and transaction around the world for long period of time and it plays an integral role in monetary, business, commercial and financial activities. More importantly, it is used as economic measure for the global economy and will continue to play an important economic vital role - both locally and globally. There has been an explosive growth in demand for efficient and effective scheme to predict gold price due its volatility and fluctuation. Hence, there is need for the development of gold price prediction scheme to assist and support investors, marketers, and financial institutions in making effective economic and monetary decisions. This paper primarily proposed an intelligent based system for predicting and characterizing the gold market trend. The simulation result shows that the proposed intelligent gold price scheme has been able to predict the gold price with high accuracy and precision, and ultimately it has significantly reduced the prediction error when compared to baseline neural network (NN).

Key words:
Gold price, automated machine learning, prediction, neural network, cross validation

1. Introduction

Gold has been one of the precious metals used by the financial industry and market for business and commercial purposes. Gold has had huge impact on the global economy and financial market due to its scarcity and value as well. It has been used as an equivalent to the currency presently circulating within the financial and business institutions. In addition, the gold unique features such as resistance to corrosion, scarcity and value are the vital characteristics which make it a suitable for exchange in the financial and business industry. Furthermore, the gold supply and demand are caused by its uniqueness when compared to other precious metals that possess the same properties as gold. Therefore, it serves as the safe way to save investors' money in the event of a financial crisis or instability and volatility in the global markets. These factors have led to the increasing demand for more reliable and effective approaches to predict gold prices and its trend.

Notably, gold has long been recognized as a symbol of wealth and frontier less currency that can be easily exchanged among different monetary systems [1,2]. In recent decades, gold has gradually become a popular nonmonetary tool in the financial market, which is characterized by high-yield and high-risk. Gold price is partly regarded as reflection of investors' expectations and the world's economic trends. Therefore, gold price forecasting is a vital issue to the global economy. At the same time, it is noted that, during the financial crisis in 2008 and early 2009, the global gold price has increased by 6% on average, while many mineral prices have dropped by 40% approximately [3]. Ultimately, the gold price trend is entirely different when compared to other mineral commodities, and hence predicting gold price is more challenging due to the market volatility and rapid change in global economy.

More importantly, it is difficult to predict exactly the price of gold based on the historical data and information collected as a result of its price volatility and market dynamics. Many factors can lead to instability and volatility in gold price - this primarily included demand and supply, exchange rate, market and inflation. Developing an approach to effectively forecast on how the gold price trend change over a period is extremely important to the financial and stock market. Specifically, how the aforementioned factors can affect the gold price will require more concerted effort and support from financial institutions, investors and researchers to develop more effective strategies and scheme to predict the gold price with high precision and sophistication. The ability to explore the huge information and data collected by the financial and business institutions has witnessed an exponential growth in recent years. This is primarily due to significance of the information collected in designing and modeling systems in predicting future trend.

Additionally, there has been an alarming increase on the application of intelligent approach to help in forecasting the market trend based on the historical data collected. Machine and deep learning tools have been applied in exploring the hidden information in the financial dataset. Initially, statistical approaches have been used to predict the financial

Manuscript received April 5, 2021
Manuscript revised April 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.4.10

market. In [4] and [5], statistical schemes for predicting the stock market have been proposed, but their prediction accuracy and precision is less due to the small size of the data used. More importantly, the statistical approaches are efficient in predicting linear historical data and this is primarily their weakness. Consequently, the development of machine and deep learning has greatly assisted in solving the problem of predicting non-linear historical data. More effort and resources have been dedicated toward using specific in data and subsequently applying the selected features in predicting the future trend in banking and investment sectors. This has had more dramatic impact on the precision and accuracy of the prediction schemes.

In this paper, an intelligent gold price prediction scheme which uses time series learning has been developed for financial and monetary institutions. In contrast with the schemes, the proposed intelligent gold price prediction scheme which uses time series and machine learning to extract the variation in gold price over a period. Additionally, the intelligent learning forecast scheme determines the trend based on the historical information collected from the gold price dataset. Consequently, it helps in extracting and learning the features which are difficult to learn with other approaches. Additionally, this is accomplished with relative ease and high accuracy. In summary, this paper mainly contributes to the development of a time series-based learning scheme for efficient prediction of the daily gold prices by enhancing the prediction accuracy using automated learning.

The rest of the paper is organized as follows: Section 2 will cover the literature and related works. In section 3, the development of the intelligent gold price-based time series learning and forecasting algorithm has been covered. Numerical analysis and discussion for experimental simulation is presented in Section 4. Finally, the conclusions for the research are drawn in Section 5.

2. Related Works

Relentless efforts have been dedicated to prediction and forecasting prices and trends in various aspects – which included the application of time series, machine and deep learning, and hybrid systems in solving complex prediction problems. This section primarily deals with review of the related works used for the prediction of price trend of valuable metals. It will serve as the fundamental basis of understanding the application of prediction in determining the price and future trend of the financial market. Different approaches have been used to predict price, but there is need for high precision and accuracy in prediction. This has led to the development of different hybrid schemes to help improve the prediction accuracy in different fields.

Machine and deep learning based approaches can efficiently handle the challenges anticipated by statistical approaches and yield better performance in terms of precision and prediction. However, [6] proposed a stock price trend prediction model based on dual features and attention in order to predict direction and variation in stock price. This has had more dramatic impact on the precision and accuracy of the stock prediction schemes. A trend prediction model based on features and attention mechanism was meant to predict the direction and changes in stock prices.

Much research efforts focused on extracting information of specific point and this has led to the development of different techniques which will ultimately address the need. Encoding and decoding techniques have been reported in [7, 8, 9]. In [10], it proposed an approach which uses feature correlation and particle swarm optimization to achieve high performance. A hybrid feature selection-based algorithm which employ genetic algorithm is reported in [11]. Both [12] and [13] have utilized segmentation in order to improve the precision of prediction of stock data. Other approaches use machine and deep learning in modeling the financial time series data such that the non-linearity issue can be appropriately handled. It has been reported in [14] that the stock price can be predicted using wavelet transform, neural network and fuzzy logic with the precision needed. Deep learning scheme with multi-level data abstraction have been applied in financial field [15]-[18]. In [19], it proposed an approach using Long Short-Term Memory (LSTM) to enhance the prediction accuracy of stock market.

An asset pricing model has been developed to establish the relationship between return and risk in capital investment [20]. Subsequently, [21] added the size and value of risk to the approach in [20] to effectively determine the market risk. With the development of deep learning approaches, complicated features in large data can learn with relative ease and high precision [22]. This has greater role to play in financial and economic sector where there is huge demand for extracting and exploring the large data and information collected. It is obvious that the Autoregressive Integrated Moving Average model (ARIMA) has been one of the conventional machines learning widely used for prediction in financial and banking [43]. For instance, [23] have used ARIMA model in predicting stock market. Moreover, it has been used in [24] for predicting and analyzing the gold price market. However, time consumption and inability to process huge data have been a major key bottleneck in ARIMA.

The fact that the deep learning-based approaches can be capture the necessary features from large data set when compared machine learning [25]. It has been applicable in different fields as indicated in [26][27]. In [28], the proposed deep learning approach has been able to predict electricity theft after it have had learn from the electricity consumption dataset. There has been report from [29]-[31] that attention-based schemes were used to tackle different problems.

A classical time series for analyzing stationary and non-stationary conditions was developed by [32] has been used in conjunction with recurrent neural network to process the time series data. In addition, [33] has proposed LSTM network approach for determining the trend of large-scale financial market and price prediction as well. In [34]-[37], strategies for extracting the patterns for financial series data have been proposed. More importantly, state frequency memory recurrent network was proposed in order to determine the pattern of the market data for both long and short time period.

In [38], an approach using support vector machine (SVM) was developed to assist in predicting the direction of stock exchange market. A combination of SVM and multilayer perceptron (MLP) has been used in predicting the daily direction and price of stock in [39]. In addition, [40] uses feed forward neural network to predict the mid-price direction based on the data. Furthermore, it has been proved that based on the approach developed in [41] that using different data set size to train the neural network model can lead to degradation in the prediction accuracy.

3. Intelligent Gold Price Prediction Learning Algorithm

This section presents the main approach used for predicting the gold price based on the intelligent learning scheme. In addition, the model used for predicting the gold prices over period is proposed and discussed. Figure 1 shows the typical automated learning model for predicting gold price. The basic idea used in developing the model is the fact that the trend of the gold prices represents some certain patterns over time. Hence, these patterns need to be fully exploited and critically analyze to efficiently determine the gold price trend.

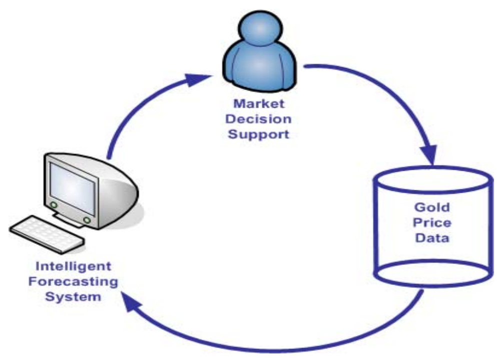

Fig 1. An Intelligent Gold Price Prediction Model

The proposed approach has been developed primarily to predict the daily gold prices to support investor and financial marketers in making effective decision. Figure 1 shows the general conceptual diagram for the gold price prediction model. The gold price prediction model consists of the following:

1. **Data Preprocessing.** The dataset used for the experimentation was analyzed, evaluated and thoroughly processed as part of the preliminary preparations before embarking on the usage of the dataset. This helps in handling the massive gold price data efficiently when it comes to the implementation of the proposed algorithm.
2. **Intelligent Information Extraction.** Extracting intelligence information from the available data is important and necessary for the advancement of new approaches and improvement of the existing techniques. The complexity of the information obtained requires thorough analysis of the data and many features need to be taken into consideration to achieve effective decision. As it has been shown in figure 1, the intelligent gold prediction scheme extracts the information from the gold price dataset and the algorithm learns the change in gold price over the period. The output of the algorithm can be used by investors and marketers in making decision about issues related to gold price and its trend prediction.
3. **Intelligent Prediction Mechanism.** In order to ultimately determine the gold prices under different condition, we deploy the mechanism of machine and k-fold cross validation learning to predict the trend of gold based on the captured information from the gold price dataset. The deviation in the gold price was tracked over the period, and machine learning mechanism utilizes the knowledge acquired to analyze the information and decide the output for any given scenario. The detail description about the intelligent gold price scheme has been presented in Section 3.1.

3.1 Intelligent Gold Price Prediction Algorithm

The flow chart for the intelligent gold price forecasting scheme is presented in figure 2 below. The steps in the aforementioned sections can be further described in more detail using the flow chart in figure 2.

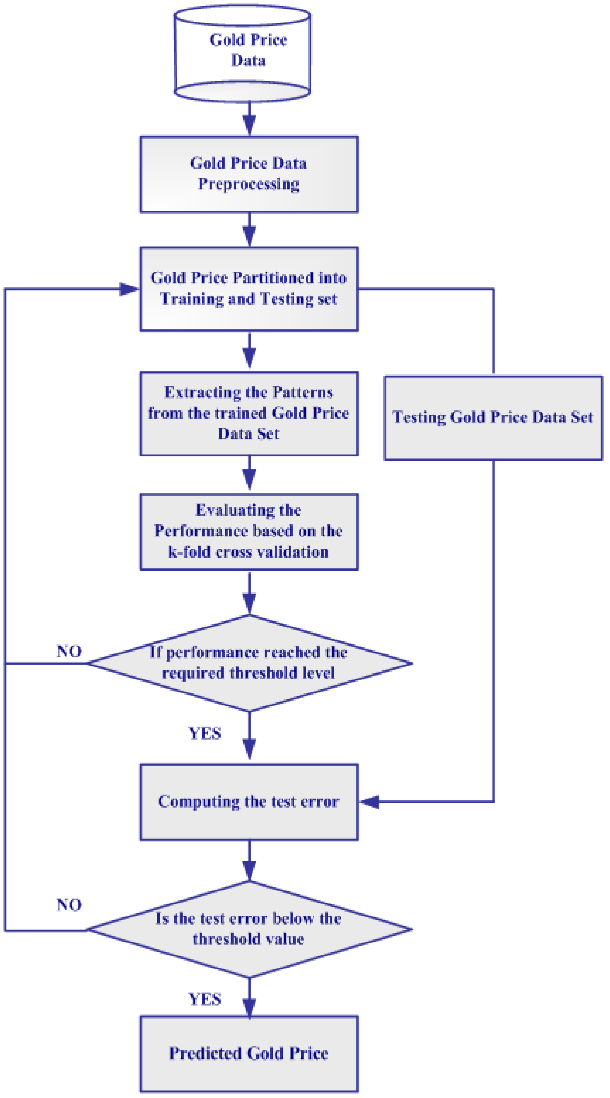

Fig 2. Intelligent Gold Price Prediction Algorithm Flow Chart

3.1.1 Initialization step

The trend for the gold price time series data can be presented as X = (x_1, x_2,…………x_n) which represent the gold price over period of time. It is assumed that $x_i \in X$ which can be greater or less than the average gold price at a particular time t. Massive daily gold price data was collected over the period t. To determine the exact trend of the time series data X, the difference between the daily consecutive gold prices is computed to determine their respective relationship. This tracks the variation and changes in the gold prices over the period to be examined. Therefore, the relationship can be represented diagrammatically as follows:

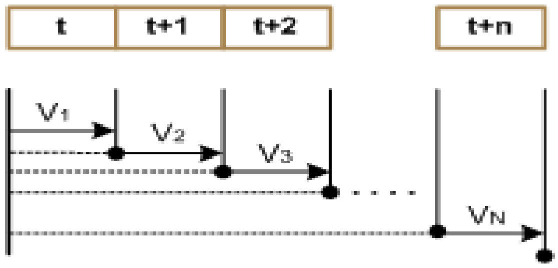

Fig 3. Representation of Gold Price Deviation Variation

where V represents the variations over period t. This can be presented in vector form as V = (v_1, v_2 ….v_n). Each of the values can be computed $V_i(t) = P_t - P_{t-1}$. The difference in V is track throughout the gold price time series data, it essentially helps toward ensuring that the intelligent scheme learns the deviations over the period. More importantly, it identifies the trend direction which has direct correlation with the time series dataset represented by X. V(t) is determine over the range V_0 to V_N.

Interestingly, the data involved in the experimentation has clearly shown that it has non-linear pattern representation over the period it has been examined. This is mainly due to the volatility and instability in gold as many factors can affect its price. To improve the non-linearity and to efficiently handle rapid variations in gold prices, a sigmoid activation function was used developing and evaluating the proposed scheme.

To improve the precision and accuracy of the proposed scheme, a *k*-fold cross validation has been used. The data have been partitioned randomly into *k* folds f_1, f_2, …… f_k. Each fold is approximately of equal to the another. Both the training and testing process were repeated k times to ultimately enhance the precision and accuracy. Initially, one of the folds is used as a test data and other folds used as training data. More precisely, each sample has been used for the same number time in training and subsequently used only once for testing. In essence, the proficiency of the proposed is enhanced with the introduction of the k-fold cross validation. The performance achievement has been demonstrated in section 4.4.

In addition, it is particularly important to note that the time series, machine and k-fold cross validation learning based algorithm uses primarily sigmoid function to map the features and eventually make the decision. This is because the time series gold price data is one dimensional and there is need to adapt a sigmoid function which has similar properties and it ultimately yields the needed non-linearity. In order to accomplish and increase the non-linearity behavior within the scheme, the sigmoid activation function which been described in equation (1) below is employed:

$$F(\Delta) = \frac{1}{1 + e^{\Delta}} \qquad (1)$$

where Δ is the deviation in gold price over a period t. It is important to note that the features extracted were normalized using the mathematical expression in equation (2) and it is presented as follows:

$$X_{\Delta} = \frac{X_i - X_m}{\sigma} \qquad (2)$$

where X_{Δ} is the parameter to be normalize. X_i represents the gold price values and X_m is the mean value of the average gold price. σ is the standard deviation of the gold price data.

In order to ensure fairness and to make sure that the gold price used in prediction experimentation is within the acceptable range, we consider the average of the USD (AM) and USD (PM) gold prices. The average gold price is observed for the morning and evening gold prices for any day. Also, the average gold price is computed over period t. Therefore, the average gold price can be computed mathematically using equation (3) as follows:

$$AVG(t) = \frac{USD_{AM}(t) + USD_{PM}(t)}{2} \qquad (3)$$

Additionally, the variation of the gold price from one day to another is computed to track the difference as the time elapse. Hence, the variation (Δ) can be expressed using equation (4) as follows:

$$\Delta_n = \frac{AVG_n - AVG_{n-1}}{AVG_{n-1}} \qquad (4)$$

Δ_n represent the variation of the USD gold price between two consecutive days. While AVG_n represent average gold price in a particular day n. Also, AVG_{n-1} represents the previous gold price in the subsequent day *n-1*.

4. Numerical Experiments and Performance Analysis

This section primarily focuses on the details of dataset preprocessing, experimentation and analysis of the result obtained. Several experiments were conducted to evaluate the performance of the proposed neural network with K-fold cross validation to enhance the prediction accuracy. In addition, the comparison between the proposed model and baseline neural network model is presented in section 4.4. Furthermore, the performance of the developed scheme has further been investigated and compared in order to determine the impact of different parameters on the proposed scheme performance.

4.1 Data Description & Analysis

In this section, the description about the gold price data set used has been covered. This is mainly to ensure that the series of embedded facts about the data used for the experimentation is obtained through observation, searching and extraction. These facts when fully exploited can be used in making effective and efficient decision. The gold price data was collected from [42].

The gold prices are represented in US dollar, Euro and Great Britain Pound (GBP). Each of the gold price time series features numerical variables contained AM (morning) and PM (evening) gold price values. In our experimental work, US dollar gold prices have been used in the simulation experimentation in order to ensure consistency and uniformity. Figure 4 shows the original structure of the US dollar gold price dataset. We collected the gold price time series data for each day, and the data covers the period from April 1, 1968, to October 23, 2017.

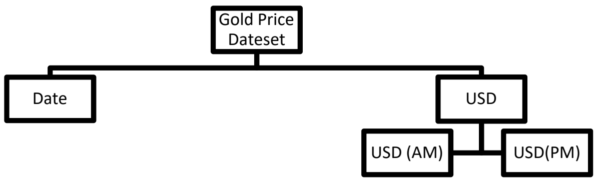

Fig 4. Block diagram representation of the original dataset

The AM & PM gold price of the collected data can be represented graphically as shown in Figure 5. As it can be notice from the graph that there has been exponential growth in gold price from 1968 to 2017. The gold trend over the period under examination clearly indicated that the investment into gold market has profitable and brighter future when compared to other financial markets. The develop scheme will enhance the investor's ability when it comes to buying and selling gold. In addition, the gold prices trend can be evaluated thoroughly and predicted with high precision and accuracy.

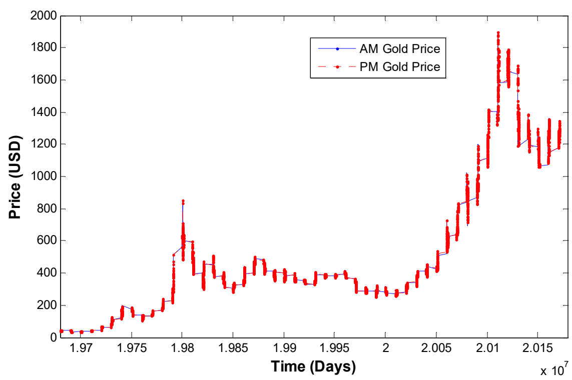

Fig 5. AM & PM Gold prices from 1968 to 2017

As it can be seen from figure 5, the gold price has been increasing exponentially (1968–2017) over the period under examination. At the beginning of 1980, there was spike in gold prices and after which the price remained stable up to the year 2000. After that, the gold price continues to grow exponentially from 2000 to 2012 and it starts to depreciate up to 2017. This clearly shows how the gold market changes very rapidly within the past 7 decades. More importantly, this has justified the dramatic need to develop more effective and high precision approach to determine the gold price trend and volatility. The average gold price eventually has similar trend as in figure 5 since both the AM and PM gold prices have the same graphical pattern.

To track the variation alone would not be enough in designing more effective solution for predicting the direction of the gold prices over time, but there is need to conduct some basic analysis in order to acquire the knowledge necessary for efficient feature selection. Hence, the statistical analysis of the raw gold price USD dataset is presented in Table 1. It is obvious that normalization is important in most of the learning schemes; the extracted features are normalized based on the average computed mean and standard deviation of the gold price dataset.

Table 1: Average USD gold price statistic

Gold US Dollar Price	
Parameter	**Value**
Mean	500.33
Median	375.24
Std. Deviation	420.20
Minimum	34.76
Maximum	1895.75

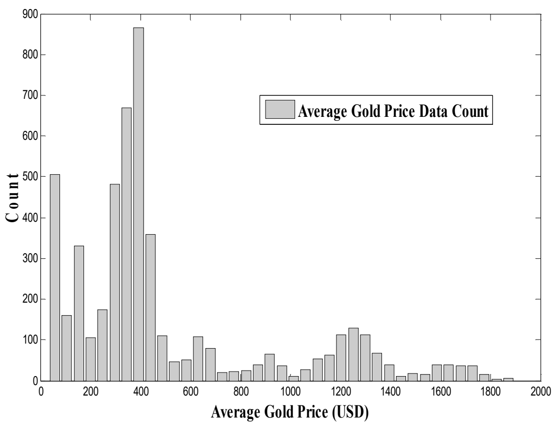

Fig 6. Average Gold price data count

Figure 6 shows the average data count for the gold/ price dataset. It clearly indicated that the gold price has consistently been stable within the price range between 300 to 400 US dollars for long period of time. This can be seen from figure 5 which indicated that the gold price was reasonably constant within the period from 1980 to 2005. The data count for the average gold price between 500 to 2000 US dollars is relatively small when compared to 0 to 500 US dollars count.

In addition, we computed the variation in gold price as it changes from one day to another to keep track of its volatility from 1968 to 2017. Figure 7 shows the gold price variations over the period under study. The deviations show that the price in gold is volatile as the price can be affected by various factors mentioned the previous section. More importantly, the graph in figure 7 shows the rise and fall in gold price over the period. The gold price deviation was computed based on equation (4) and represented graphically in figure 7 shown below.

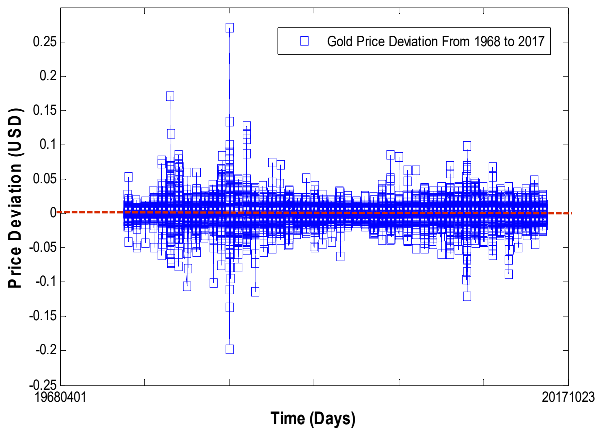

Fig 7. Gold price deviation from 1968 to 2017

Furthermore, it is particularly important to note that the average value has been considered in our experimentation. Figure 7 shows the price variations in time for the gold price data set. The gold price for each day in the time series is tract based on equation (5) shown below:

$$\Delta_t = \begin{cases} 0 & AVG_{t-1} \leq AVG_t \\ 1 & AVG_{t-1} > AVG_t \end{cases} \tag{5}$$

where Δ_t represents whether there is increase or decrease in gold price from t-1 to t. This eventually tracks the gold price variation over the period time under investigation.

Additionally, the K-fold cross-validation was used in the development of the proposed scheme. It primarily divides the data into k-folds and eventually iterates the folds. With the application of cross-validation, it effectively ensures better training and testing of the dataset. Subsequently, it leads to high precision and prediction of the gold price. In our simulation, the value of k has been set to 10 (k=10) and different scenarios were run to observe the performance of the developed algorithm. More importantly, the cross validation ensures more credible feature selection amongst different possible combinations.

4.2. Experimental Setup

To effectively evaluate and analyze the performance of the developed intelligent gold price-based time series learning and forecasting scheme, the learning capability of the scheme has been explored by acquiring useful information and predict the gold price trend over period of time based on the gold price time series data. An extensive experimentation on the gold price dataset was conducted and compared with another method. The scheme has been implemented in MATLAB environment and executed by 2.7GHz processor, 8 GB RAM and 64-bit operating system. Different test and experimentation were conducted to examine the capability of the developed scheme. The impact of various parameters on the developed scheme have been investigated and studied to critically examine the proposed scheme behavior under different conditions.

4.3 Performance Metrics

In order to effectively train the proposed scheme, the input vectors, output, weight and bias are adjusted to minimize the error between the computed and actual achievable output – this eventually leads to fast convergence. The performance evaluation function used in our simulation is described by equation (6) which is minimized until the convergence has been accomplished. The error function can be represented as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i) = \frac{1}{N}\sum_{i=1}^{N}(y_i - d_i) \quad (6)$$

N represents the number of input samples which eventually yields the same number of outputs. The actual achievable is represented by d_i while y_i represents the output computed using equation (7). In addition, the root means square error (RMSE) is considered as the metric used to evaluate the performance of the developed intelligent gold price prediction based on time series, machine and k-fold cross validation learning algorithm, and to juxtaposed proposed scheme with other similar schemes as well. The RMSE has been considered primarily to measure how close the predicted value to the actual value. The main goal is to significantly reduce the RMSE in order to achieve the desired outcome. More importantly, the mean absolute error (MAE) has been used to determine the difference between the variable and the average absolute error. Therefore, the *RMSE* and *MAE* can be express mathematically as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - d_i)^2} \quad (7)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|(y_i - d_i)| \quad (8)$$

It is important to note that the closer the RMSE and MAE value to 0 indicated better chance for forecasting the gold price value. Therefore, the RMSE value is reduce in each iteration until when the targeted threshold is accomplished or reached. Similarly, when the RMSE remains constant, it indicated that the convergence point has been reach and it depends on the condition set up to attain the convergence.

4.4 Performance Comparison

In this section, the performance of the proposed scheme has been investigated. The developed scheme is compared with baseline neural network in terms of error and prediction precision accuracy. By considering the error and prediction accuracy as metrics to measure the performance of the proposed scheme because they have widely used in determining the proficiency of prediction algorithms. Initially, we consider the error comparison between the baseline NN and the k-fold cross validation neural network (kfCV-NN). Both approaches have been tested to measure the performance in terms of output errors. The primary objective is to reduce output error to minimum in order to achieve better prediction accuracy and precision of the gold prices. This is because of the non-linearity of the gold price data. Hence, the performance of the proposed kfCV-NN reduced the error to minimum. Based on the simulation experimentation, kfCV-NN proved to be more effective since it converges more rapidly compared to the base line NN. As can be seen from figure 8, the kfCV-NN has been deployed to enhance the accuracy and stability of the developed scheme. The kfCV-NN scheme has been able to reduce the training error significantly when compared to the baseline NN as presented in Table 2. Both the RMSE and MAE for kfCV-NN clearly shows relatively low training error. Thus, it is an indication that the prediction accuracy of the proposed scheme outperformed the accuracy of the base line NN.

Table 2: RMSE & MAE comparisons

	kfCV-NN		Base line NN	
Iterations	**RMSE**	**MAE**	**RMSE**	**MAE**
10	0.0757	0.0412	0.213	0.395
20	0.0411	0.0274	0.087	0.064
30	0.0286	0.0203	0.081	0.059

Additionally, the RMSE testing error for the proposed scheme decreases rapidly to minimum as the number of iterations increases as it can be clearly seen in figure 8. To effectively and efficiently measured the percentage reduction in RMSE, the number of iteration was varied from 10 to 30 at the interval of 10. Under each scenario, the RMSE was computed based on equation (7) for both the kfCV-NN and baseline schemes. The average error for the two schemes have been computed and compared. The RMSE testing error has been significantly reduced by 53.6% when compared to the baseline NN.

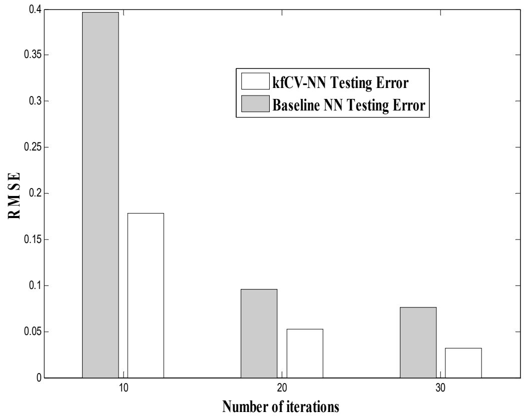

Fig 8. kfCV-NN Vs Baseline NN Gold price prediction RMSE comparison

Furthermore, it can be seen from figure 9 that the MAE testing error for the proposed scheme is relatively low when compared to the baseline NN as the number of iterations increases. The number of iteration was varied from 10 to 30 at the interval of 10, and under each scenario, the mean absolute error was computed based on equation (8) for both the kfCV-NN and baseline schemes. The average error for the two schemes have been computed and compared. The MAE error has been reduced by 57.6% compared to the baseline NN.

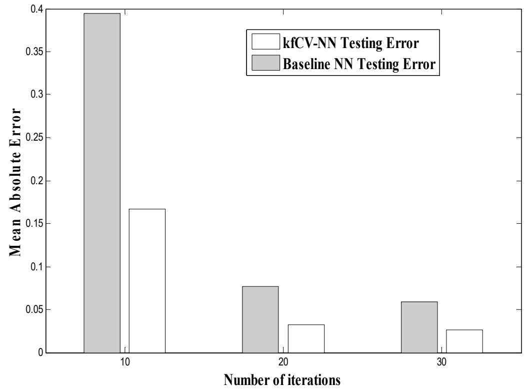

Fig 9. kfCV-NN Vs Baseline NN Gold price prediction MAE comparison

Finally, the precision and accuracy of the proposed was measured and compared the predicted values of the baseline NN through the simulation experimentation. Both the two approached were compared to exact gold price values to measure the prediction precision proficiency of each of the approaches compared to the exact gold price. The prediction result obtained from both the kfCV-NN and baseline NN have been compared and presented in Figure 9.

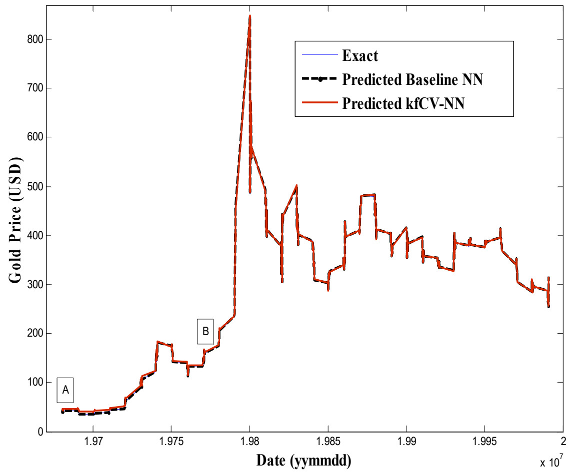

Fig 10. Gold Price Prediction Accuracy Comparison

As it can be notice from figure 10 which shows the prediction test result. The gold prices predicted by kfCV-NN scheme seem to be close to the exact gold price values over the period of time under observation. The baseline NN lags behind the kfCV-NN and exact values from point **A** up to point **B**. Immediately after point **B**, all three values seem to be close to one another – this clearly shows uniformity in gold price. Based on the results achieved, it is obvious that kfCV-NN approach outperform the baseline NN in terms of prediction accuracy. The k-fold cross validation has been used to improve the performance of the baseline NN, and consequently it ensures more stability and better proficiency while predicting the gold prices. In summary, the proposed kfCV-NN scheme has successfully been able to improve the prediction accuracy by 4% when compared to the baseline NN.

5. Conclusions

In this paper, an intelligent gold price prediction scheme has been proposed to effectively determine the gold price and its trend. The developed approach primarily uses the gold price time series data, cross validation and machine learning approach to analyzed and predict the gold price market volatility and fluctuation at any particular given time. The neural network has been primarily included to learn the gold price trend and predict the future prices based on the knowledge acquired from the gold price time series data. Subsequently, the cross validation has significantly improved the prediction accuracy and precision of the proposed scheme. More importantly, the combination of the proposed method and neural network has resulted in better performance which reduced the overall error and increases the prediction accuracy. This improves the capability of the proposed scheme in predicting the gold price effectively with high precision based on the features detected and patterns learnt observed in the dataset. Our future work will primarily focus on the development of hybrid approach to reduce the error further which eventually leads to high prediction capability and convergence. Consequently, this will enhance the overall system performance and allows the proposed scheme to be used for highly dimensional and complex financial dataset.

Acknowledgments

The authors would like to thank all those who contributed toward making this research successful. Also, we would like to thanks the reviewers for their insightful and valuable comment. This work is supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Saudi Arabia, under grant D-64-830-1437. The authors are very grateful to the DSR for their technical and financial support throughout the period of the project.

References

[1] Y. Long, “Visibility graph network analysis of gold price time series. Physica A. 2013;392(19):3374-3384.

[2] A. Parisi, F. Parisi, and D. Diaz, “forecasting gold price changes: rolling and recursive neural network models. Journal of Multinational Financial Management. 2008;18(5):477-487.

[3] S. Shafiee, and E. Topal, “An overview of global gold market and gold price forecasting. Resources Policy. 2010;35(3):178-189.

[4] S. M. Idrees, M. A. Alam, P. Agarwal, "A prediction approach for stock market volatility based on time series data", *IEEE Access*, vol. 7, pp. 17287-17298, 2019.

[5] M. Khashei, M. Bijari, "An artificial neural network (p d q) model for timeseries forecasting", *Expert Syst. Appl.*, vol. 37, no. 1, pp. 479-489, Jan. 2010.

[6] Y. Chen, W. Lin, and J. Z. Wang “A Dual-Attention-Based Stock Price Trend Prediction Model With Dual Features,” IEEE Access, Vol. 7, 2019

[7] I. Sutskever, O. Vinyals, Q. V. Le, "Sequence to sequence learning with neural networks", *Proc. Adv. Neural Inf. Process. Syst.*, pp. 3104-3112, 2014.

[8] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, `Multimodal deep learning," in Proc. 28th Int. Conf. Mach. Learn., 2011, pp. 689-696.

[9] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, A. L. I. Oliveira, "Computational intelligence and financial markets: A survey and future directions", *Expert. Syst. Appl.*, vol. 55, pp. 194-211, Aug. 2016.

[10] P. Moradi and M. Gholampour, ``A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," Appl.Soft. Comput., vol. 43, pp. 117_130, Jun. 2016.

[11] H. Dong, T. Li, R. Ding, and J. Sun, ``A novel hybrid genetic algorithm with granular information for feature selection and optimization," Appl.Soft. Comput., vol. 65, pp. 33_46, Apr. 2018.

[12] P. C. Chang, C. Y. Fan, and C. H. Liu, ``Integrating a piecewise linear representation method and a neural network model for stock trading points prediction," IEEE Trans. Syst., Man, C (Appl. Rev.), vol. 39, no. 1, pp. 80_92, Jan. 2009.

[13] L. Zhao and L.Wang, ``Price trend prediction of stock market using outlier data mining algorithm," in Proc. IEEE 5th Int. Conf. Big Data Cloud Comput., Aug. 2015, pp. 93_98.

[14] S. K. Chandar, ``Fusion model of wavelet transform and adaptive neuro fuzzy inference system for stock market prediction," in Journal of Ambient Intelligence and Humanized Computing. Berlin, Germany: Springer, 2019.

[15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ``BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, arXiv:1810.04805, [Online]. Available: https://arxiv.org/abs/1810.04805

[16] X. Ding, Y. Zhang, T. Liu, and J. Duan, ``Deep learning for event-driven stock prediction," in Proc. 24th Int. Joint Conf. Artif. Intell., Jun. 2015, pp. 2327_2333.

[17] Y. Qin, D. Song, H. Chen,W. Cheng, G. Jiang, and G.W. Cottrell, ``A dualstage attention-based recurrent neural network for time series prediction," in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 2627_2633.

[18] D. M. Nelson, A. C. Pereira, and R. A. de Oliveira, ``Stock market's price movement prediction with LSTM neural networks," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), May 2017, pp. 1419_1426.

[19] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, ``An innovative neural network approach for stock market prediction," in The Journal of Super-computing. Springer, 2018. doi: 10.1007/s1122.

[20] F. Black, M. C. Jensen, M. Scholes, "The capital asset pricing model: Some empirical tests", Stud. Theory Capital Markets, vol. 81, no. 3, pp. 79-121, 1972.

[21] E. F. Fama, K. R. French, "Common risk factors in the returns on stocks and bonds", J. Financial Econ., vol. 33, no. 1, pp. 3-56, 1993,

[22] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Commun. ACM, vol. 60, no. 2, pp. 2012, 2012.

[23] R. C. Chung, W. Ip, S. Chan, "An ARIMA-intervention analysis model for the financial crisis in China’s manufacturing industry", Int. J. Eng. Bus. Manage., vol. 1, pp. 5, Mar. 2009.

[24] B. Guha, G. Bandyopadhyay, "Gold price forecasting using ARIMA model", J. Adv. Manage. Sci., vol. 4, no. 2, pp. 117-121, Mar. 2016.

[25] A. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton, M. A. Picheny, "Deep belief networks using discriminative features for phone recognition", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 5060-5063, May 2011.

[26] J. Zhou, Y. Lu, H.-N. Dai, H. Wang, H. Xiao, "Sentiment analysis of chinese microblog based on stacked bidirectional LSTM", IEEE Access, vol. 7, pp. 38856-38866, 2019.
[27] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids", IEEE Trans. Ind. Informat., vol. 14, no. 4, pp. 1606-1615, Apr. 2018.
[28] T. Kim, H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data", PLoS ONE, vol. 14, no. 2, pp. 1-23, Feb. 2019.
[29] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, "Recurrent models of visual attention", Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS), vol. 2, pp. 2204-2212, 2014
[30] R. He, W. S. Lee, H. T. Ng, D. Dahlmeier, "An unsupervised neural attention model for aspect extraction", Proc. 55th Annu. Meeting Assoc. Comput. Linguistics Long Papers, vol. 1, pp. 388-397, Jul. 2017
[31] N. Si, H. Wang, W. Li, Y. Shan, "Chinese POS tagging with attention-based long short-term memory network" in Advances in Intelligent Systems and Interactive Applications, Cham, Switzerland:Springer, pp. 586-592, Jun. 2018.
[32] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. Hoboken, NJ, USA: Wiley, 2015.
[33] T. Fischer and C. Krauss, ``Deep learning with long short-term memory networks for _nancial market predictions," Eur. J. Oper. Res., vol. 120, no. 2, 2017.
[34] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, ``Modeling long-and short-term temporal patterns with deep neural networks," in Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2018.
[35] S. Jeon, B. Hong, and V. Chang, ``Pattern graph tracking-based stock price prediction using big data," Future Gener. Comput. Syst., vol. 80, 2018.
[36] X. Xu, J. Zhang, and M. Small, ``Superfamily phenomena and motifs of networks induced from time series," Proc. Nat. Acad. Sci., vol. 105, no. 50, 2008.
[37] L. Zhang, C. Aggarwal, and G.-J. Qi, ``Stock price prediction via discov-ering multi-frequency trading patterns," in Proc. 23rd ACM SIGKDD Int.Conf. Knowl. Discovery Data Mining, 2017.
[38] W. Huang, Y. Nakamori, and S.-Y. Wang, ``Forecasting stock market movement direction with support vector machine," Comput. Oper. Res.,vol. 32, no. 10, 2005.
[39] K. J. Kim, ``Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, nos. 1_2, 2003.
[40] D. T. Tran, J. Kanniainen, M. Gabbouj, and A. Iosi_dis. (2019). Datadriven neural architecture learning for _nancial time-series forecasting." [Online]. Available: https://arxiv.org/abs/1903.06751
[41] S. Walczak, ``An empirical analysis of data requirements for financial forecasting with neural networks," J. Manage. Inf. Syst., vol. 17, no. 4, 2001.
[42] 26. Gold price dataset (1968-2017). [online] Available: https://www.kaggle.com/gwhittington/lbma-gold-price-19682017 (Accessed December 2018)
[43] M. Almasarweh, S. Alwadi, "ARIMA model in predicting banking stock market data", Modern Appl. Sci., vol. 12, no. 11, pp. 309, 2018.

Yakubu S. Baguda (M’05–SM’18) was born in Kano, Nigeria. He received the B.Eng in Electrical Engineering from Bayero University, Nigeria, in 2001. He received an MSc in Information Systems from the Robert Gordon University, Aberdeen, U.K. in 2003 and PhD degree in Electrical Engineering with distinction from the University of Technology, Malaysia, in 2010. From 2011 to 2012, he was a postdoctoral researcher at the UTM-MIMOS Centre of Telecommunication Excellence, Malaysia. He is currently an Assistant Professor with the Faculty of Computing & Information Technology, King Abdulaziz University, Saudi Arabia. He is the author of more than 30 journal and conference articles. His research interests include game theory, e-healthcare system, cross layer design and optimization, energy-efficient schemes, system modeling & prediction, intelligent decision systems, cognitive radio systems, resource allocation & management, network optimization and performance analysis, bioinspired optimization algorithms, wireless multimedia communication and networking. Dr. Yakubu received the chartered engineer (CEng) award from the UK Engineering Council in 2017. He is currently a senior member of both IEEE and ACM.

Hani Al-Jahdali received his BSc in Computer Science from King Abdulaziz University, Saudi Arabia in 2005. He received an MSc in Information Technology (2009) and PhD degree in Computer Science from the University of Glasgow, Scotland, United Kingdom in 2015. From 2005 to 2007, he was working at the Saudi Electricity Company as a Budget and System analyst. Since 2011, he has been with the Information systems department, faculty of Computing & Information Technology, King Abdulaziz University, Saudi Arabia. His research interests include Information Security, Human Computer Interaction, Graphical Password Authentication and Robotics.