Rough Set Based Affinity Propagation Model for Prediction of Future Gold Price in Indian Scenario

Rudra Kalyan Nayak, N. Satya Teja, Ramamani Tripathy, G. Venkatesh, I. Sai Ram

Abstract: In international market, trading of metals has played a vital role. Metal cost might affect the nation’s economy. There are so many base metals available which have been utilized in world trading for construction and manufacture of goods. Among them gold, silver, platinum, palladium have been treated as precious metals which has economic values. Therefore today’s researchers have concentrated their investigation on metal prediction using diversified algorithms like Auto Regressive Integrated Moving Average (ARIMA), KNN (K-Nearest Neighbor), Artificial Neural Network (ANN) and Support Vector Machine (SVM) etc. In this paper our foremost objective is to predict gold price, so we put our research on this metal. In this work we have employed rough set based affinity propagation algorithm for predicting future gold price and we compared our proposed model with rough set and ARIMA model basing upon the performance measures such as root mean square error (RMSE) and mean absolute percentage error (MAPE). The experimental result shows that the proposed model outperforms rough set and ARIMA model.

Keywords: Rough Set, Affinity Propagation, Prediction, Gold Price.

I. INTRODUCTION

From early days metal price forecast has been played a significant role. Metal price might update periodically because it is fluctuate with market demand. Basically metals market can be categorized in to five parts such as base metals, Steel and Ferro-Alloys, Minor Metals, Platinum Group Metals and precious metals. Here we have concentrated our study on precious metal like gold. Gold is treated as very important asset for all human beings. According to global economy trading now a day’s gold plays major role and most of the investor have keen interest to invest their money on gold. Moreover it has been one of the principal commodities in both long run and short run. Economic value of gold may vary from other metals in terms of production and price. In case of fiscal crisis, numerous financial products dropped by approximately 40% whereas gold price increased by about 6% and this hypothesis shows the importance of gold in economic monetary market [1-5]. From ancient era, gold metal is preferred as sophisticated product for all Indian people. Because it has an Indian tradition that all populace must wear gold ornaments in every occasion and also gift gold items to others in their festivity. For that major investment can be done on gold in market. According to researcher it has proved that gold metal is safe during economic recession and other natural disaster [1-4]. From the literature survey it is reveals that gold price fluctuation might affect other commodities like crude oil, minerals, and foreign currencies. Therefore future gold price prediction has treating as difficult task for all scientists. A diversity of techniques has applied to forecast the gold price and all method have not getting that much of success. To realize the problems new and novel approaches are implemented such as Traditional Box Jenkins methods (ARIMA and SARIMA), Decision tree classifier, neural network and ARIMA, Decision tree and support vector regression (SVR), Fuzzy Linear Programming, Neural network with Genetic algorithm, Ensemble machine learning and naïve Bayes, and k-nearest neighbors [5-9]. The main objective of this paper is to build a future price list of gold in a long horizon that would assist all investors and policy maker who have an interest on investing their money on purchasing gold. Here we have proposed a hybrid model rough set based affinity algorithm and it was observed that our model performs well using the opening price of gold dataset according with prediction accuracy. Rest of the article is arranged as follows: section 2: Covers up the literature works. In section 3: we provide dataset description and mention about proposed methodology. Next in section 4: the methodology and experimental results are elaborated which can be used in prediction of opening gold price. Finally, in last section we presented the conclusion of our work.

II. RELATED WORKS

To resolve the movement of gold price, there are a group of dissimilar methods being applied by different authors to introduce their models. The authors BanhiGuha et al. [10] have implemented ARIMA time series model for forecasting future gold price. Here they depicted about gold’s importance and its price’s impact on global economy. Based on past data from November 2003 to January 2014 gold price was predicted and researcher concluded that their model was fitting well.
using three parameter such as Autoregressive (AR), integrated (I), Moving Average (MA) over 10 years dataset. The authors Khan et al. [11] in their paper, implemented Box-Jenkins, ARIMA model for predicting gold price. They used RMSE, mean absolute error, and mean absolute percentage error for calculating prediction accuracy. Authors took the dataset for prediction purpose from January 02, 2003 to March 1, 2012 and concluded that the results obtained from model were most suitable for gold prediction. Another authors Yaziz, Azizan et al. [12] have proposed a novel hybridized model for forecasting gold price using ARIMA-GRACH approaches. Here researchers elaborated the importance of hybrid model. Basically ARIMA model is well suited for forecasting the short period of time series data and also able to handle non stationary data. But it has limitation that it doesn’t have capability to handle the volatility and nonlinearity data series. So GARCH model was mixed with ARIMA method and revealed significant result among all. Again, Wongdhamma et al. [13] have explored their research on forecasting the currency based on wavelet analysis (WA). In this paper author compared ARIMA and ARIMAX-WD models using mean absolute percentage (MAPE) and adjusted R-squared in the out-of-sample data partition. The outcomes implied that the projected ARIMAX model based wavelet decomposed signals as explanatory variables (ARIMAX-WD) did fine in a long-standing with nonlinear time-series forecasting application. A novel multiple regression method has been implemented by Baker et al. [14] based on fundamentalist technique to decide the variables which was disturbing the gold price and model constructing for long run gold price forecasting. Commodity price index volatility and changes in the value of US dollars might be affecting on gold price. This paper recommended that gold price can be projected to go up there is a common raise in commodity prices. We know that day by day populations are growing and their requirements are also changing. For this purpose investigation on human need is increased and there are numerous fiscal factors present which could influence the movement of gold price. So many multiple linear regression techniques have been taken to forecast the gold price. Here researchers Ismail et al. [15] utilized multiple regression in stepwise manner to eliminate the correlation between variables. Durbin-Watson statistics is taken to make sure whether autocorrelation exist in the error terms. Error mean square MSE and mean square prediction error MSPR were used to evaluate the model accuracy.

III. MODEL DESCRIPTION

A. Data Set

Our study is based on secondary yearly data which is ranging from December 28th 1979 to August 30th 2019 and monthly, weekly and daily data ranging from January 4th 2010 to August 30th 2019 for Gold opening price prediction and the dataset is collected from [16]. We have taken three parameters of our data such as opening price, high value and low value in daily, weekly, monthly and yearly bases. Sample data of gold are revealed below in Table-I.

| Table-I: Sample monthly Gold Dataset |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Date       | Price | Open | High | Low | Vol | Change % |
| Aug 19    | 1.52  | 1,425.60 | 1,565.00 | 1,421.10 | 9.04M | 6.37% |
| Jul 19    | 1.43  | 1,418.00 | 1,467.00 | 1,396.40 | 1,70M | 0.89% |
| Jun 19    | 1.42  | 1,323.10 | 1,453.70 | 1,322.70 | 155.25 K | 7.75% |
| May 19    | 1.32  | 1,302.30 | 1,323.00 | 1,285.10 | 109.37 K | 1.46% |
| Apr 19    | 1.30  | 1,314.40 | 1,332.20 | 1,286.00 | 37.47K | -0.98% |
| Mar 19    | 1.31  | 1,339.10 | 1,348.40 | 1,306.70 | 38.17K | -1.83% |
| Feb 19    | 1.34  | 1,349.80 | 1,374.30 | 1,329.90 | 19.49K | -0.67% |
| Jan 19    | 1.35  | 1,319.90 | 1,354.60 | 1,308.00 | 28.98K | 2.86% |
| Dec 18    | 1.31  | 1,260.60 | 1,317.10 | 1,260.30 | 8.86K | 4.46% |

B. Model Discussion

According to economic market each time commodity products price may increase or decrease. Different financial souk products such as silver, gold, diamond, mines, minerals, crude oil, stock markets and currency price forecasting is an crucial aspect of all researchers. As market price is always fluctuating due to economic recession and other natural disaster so it is really a challenging facet to forecast the entire metal products in current days. Among all metals forecasting gold price is a very tough task for all scientists.

![Proposed model of gold price prediction](image)

Figure 1 in the above depicts about the working procedure of our manuscript. The objective of our paper is to predict future gold opening price which is useful for all investors and stakeholders. In first step we have collected gold data from [16] according with their parameters as daily, weekly and monthly, yearly. In next step we have used smoothing technique for smoothing the raw data. Basically it is a technique which is used in time series data to eliminate the fine-grained variation between time steps. The wish of smoothing is to eradicate noise and well again expose the signal of the underlying causal
processes. In next step we have applied min-max approach on smoothing data for normalization. Normally normalization is a procedure of maintaining database without delicacy among data and also enhancing data integrity. In the next immediate step we have explored our model rough set based affinity propagation for prediction of gold opening price including daily, weekly, monthly and yearly in separate manner. Finally we have calculated prediction accuracy using performance measures.

IV. METHODOLOGY AND EXPERIMENTAL PROCEDURE

A. Rough Set Algorithm

In every minutes of the day a large number of information are available in the market. According to the vast improvement in technology all research issues are solving slowly but still some tasks are there which are difficult to resolve in recent days. In current trends artificial neural network, data mining, soft computing and various statistical approaches are utilized in various domains to solve the complex problems. Among all rough set theory is also treated as a pivotal approach which deals with vagueness and uncertainty emphasized in decision making. In the year 1982 rough set theory [17-20] was introduced by scientist Z. Pawlak. This theory depicts about lower range approximation, upper range approximation and boundary region of data. Rough set theory is discussed as $B = \{U, m \cup D, W, Z\}$, leading which $U$ refers as universe of each and every one non-void set of entity, $m$ and $D$ are treated as non-void predeterminet set of characteristic and feature whichever assure $m \cup D = N$, $F$ denotes to the domain of all attributes like $F = U_{\in E_{m}}F$, wherever $F_{m}$ is represent as finite number of $f_{m}$, $Z$ is also called as function for all characteristic like $Z = U_{\in E_{m}}Z_{c}$, everywhere $Z_{c}$ is a sum function $Z_{c} : U \in F_{m}$. All compartment of characteristic $E \subseteq N$ might be combined with an indiscernibility relation $Z(E)$expressed as (1).

Affinity Propagation Approach:

Input

1. Initialize availabilitys $= 0$, $\forall j$, $m$, $a(j, m) = 0$
2. Update responsibility and availability to converge until

$$r(j, m) = \sum_{i | a(i, m) > a(j, m)} [a(i, m) + a(j, m)]$$

$$\forall j, m : a(j, m) = \begin{cases} \min_{i \in E_{m}}(0, r(j, m)) & \text{if } j \in \text{ forming } j \\ \max_{i \in E_{m}}(0, r(i, m)) & \text{if } i \notin \text{ forming } j \end{cases}$$

$$Z(E) = \{ \{o, p\} \in U \times U | \forall g \in E, \exists \{o, p\} \in Z(g) \}$$

Rough set theory creates an equivalent relation whichever includes three dissimilar types of mapping among objects such as relation may be reflexive, symmetric and transitive. For each and every object $o \in U$, all substances alike to $o$ create equivalence class of $o$, which is symbolized by $\alpha_{E} = \{ p | p \in U, (o, p) \in Z(E) \}$. All unique equivalence classes on the basis of $E$ where it composes a separation persuaded by $E$ and this type of equivalence classes are known as elementary sets in this hypothesis.

The pivotal element of this theory is basically based on two sets, where it might be given a clear view of elementary sets, to approximate an indescribable subset $O$ of Universal set. Here these two dissimilar sets are represented as vital thought in case of rough hypothesis and are depicted as the lower and upper approximations of $O$, correspondingly.

Lower boundary substance $E(O)$ holds every part of objects whichever might be surely classed as objects of $O$ on the basis of the set of attributes $E$. In case of upper boundary $\overline{E}(O)$ is known as the objects substance where it might be probably classified as objects of $O$. In twosome cases $[E(O), \overline{E}(O)]$ is represented by scientist Z. Pawlak. The conception of optimistic, pessimistic and boundary regions are described as in equation (4), (5) and (6), respectively,

$$POS_{E}(O) = \overline{E}(O)$$

$$NEG_{E}(O) = \overline{E}(O)$$

$$DN_{E}(O) = \overline{E}(O) - E(O)$$

The main motto of optimistic area $POS_{E}(O)$ is same as the lower boundary set. The pessimistic part $NEG_{E}(O)$ is the set of all objects which certainly not belong to $O$ based on $E$. Finally the border region is the set of all objects whichever are hard to be classified to $O$ or to the complement of $O$.

B. Affinity Propagation Algorithm

We know that clustering is treated under unsupervised learning approach which is used for arranging data into meaningful groupings. Data clustering is based on similarity among data points and it is a vital part in computer science domain. Among all clustering algorithms affinity propagation (AP) algorithm [21-26] is one which has been used in diversified fields. In the year 2007 scientist Frey and Dueck has proposed the affinity algorithm. This algorithm is an exemplar-based clustering method which has demonstrated high-quality performance on an extensive variety of datasets. Affinity propagation is totally a data-driven approach where it separates the data points to dissimilar clusters and classifies exemplars amid them at the same time allowing for every one data points as probable exemplars and swapping messages amongst data point until a high-quality set of exemplars and clusters come out. On the other hand, the original AP approach presumes features are numeric valued, that signifies that the methodology could not have progression features of categorical values or miscellaneous type of values.

C. Experimental Analysis

In this work, for prediction we have utilized hybrid approach rough set based affinity propagation algorithm. Here we have done our experiment using Intel i3 with 4GB Hard disk and windows 7 operating system and complete code is written using python.

3. In essence, our analysis is
Rough Set Based Affinity Propagation Model for Prediction of Future Gold Price in Indian Scenario

organized in 5 phases. Phase 1: Dataset retrieval: In first phase we have retrieved gold data from [16] based on yearly, monthly, weekly and daily basis. Generally each and every research analysis is depending on correct data. So it is very important step.

Phase 2: Dataset smoothing: With the environmental point of views raw data are always noise. So first we smooth the data using technical indicators. We opt some of the particular technical indicators such as: Williams %R, simple moving average (SMA), true strength index (TSI) and relative strength index (RSI) [27-29] given in equations (9), (10), (11) and (12) respectively.

\[
\%R = \frac{\text{high}_M \text{ days} - \text{close}_{\text{today}}}{\text{high}_M \text{ days} - \text{low}_M \text{ days}} \times -100
\]  

(9)

\[
SMA_j = \frac{1}{2m+1} \left( MA(j + M) + MA(j + M - 1) + \cdots + MAj - M \right)
\]  

(10)

\[
TSI(c_0, a, b) = 100 \times \frac{EMA(EMA(m)a, b)}{EMA(EMA(lm)a, b)}
\]  

(11)

\[
RSI = 100 - \frac{100}{1 + EMA(0.25a, b)}
\]  

(12)

Phase 3: Scaling of dataset: In next step we have applied min-max scaling technology on smoothing data which is mentioned in equation (13).

\[
\tilde{r}_{1m} = \frac{n_{\text{max}} - r_{\text{min,m}}}{n_{\text{max,m}} - n_{\text{min,m}}}
\]  

(13)

Where, \(n_{1m}\) denotes \(m^{th}\) attribute value of \(l^{th}\) feature \(n_l\), designed for the dataset, \(m^{th}\) maximum value is \(n_{\text{max,m}}\) plus maximum value is \(n_{\text{max,m}}\) in addition to normalized price of \(l^{th}\) day is \(\tilde{r}_l\). Phase 4: Implementation of rough set based affinity propagation model: After receiving whole correct structure data we implemented our novel hybrid approach rough set based affinity propagation algorithm for exact forecasting of the original data. The rough set method first takes the normalized opening gold price to explore the hidden patterns in that by making them transformed into lower and upper approximations and boundary regions. Afterward affinity propagation model receives similarities within data points and preferences of data point’s appropriateness for being an exemplar which are denoted by a matrix. Preferences are present in the prime diagonal. In each loop messages are passed computing responsibilities and availabilities of exemplar \(k\) and point \(i\). The algorithm carries on until it converges that is we terminate the process when variation in values is less than or equal to the threshold value. As a result we are getting the promising accuracy in predicting opening gold price. Figure 3 which is given below shows the actual yearly opening price of gold and Figure 4 describes the comparison graph of actual versus predicted graph of yearly opening price of gold.

Phase 5: Performance Measurement: "Equations (14) and (15) represent Root Mean squared error (RMSE) and mean absolute percentage error (MAPE) for measuring our proposed model’s efficiency.

\[
\text{RMSE} = \sqrt{\frac{1}{D} \sum_{j=1}^{D} (a_j - \hat{a}_j)^2}
\]  

(14)

\[
\text{MAPE} = \frac{1}{D} \sum_{j=1}^{D} \left| \frac{a_j - \hat{a}_j}{a_j} \right| \times 100
\]  

(15)

Where, \(D\) is known as the sum of testing data, \(a_j\) and \(\hat{a}_j\) is the required and forecasted outputs correspondingly.

D. Simulated Result Analysis

The testing error values of proposed rough set based affinity propagation model is compared to rough set and recently worked ARIMA model [30] which are depicted below in table-II.
Table-II: Performance comparison of Rough set based Affinity Propagation model for one day, one week, one month and one year ahead opening price prediction with Rough set and ARIMA method.

<table>
<thead>
<tr>
<th>Time Horizon</th>
<th>Rough Set</th>
<th>ARIMA</th>
<th>Rough Set based Affinity Propagation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAPE</td>
<td>RMSE</td>
</tr>
<tr>
<td>1 Day</td>
<td>1.545</td>
<td>0.964</td>
<td>0.99</td>
</tr>
<tr>
<td>1 Week</td>
<td>1.85</td>
<td>1.296</td>
<td>1.222</td>
</tr>
<tr>
<td>1 Month</td>
<td>2.851</td>
<td>2.124</td>
<td>2.32</td>
</tr>
<tr>
<td>1 Year</td>
<td>4.7</td>
<td>3.985</td>
<td>4.56</td>
</tr>
</tbody>
</table>

The prediction accuracy of opening price of gold resulted by proposed model and two aforesaid models are analyzed here. The RMSE of 1 day prediction by proposed model is 0.57 where as rough set based and ARIMA models generate 1.545 and 0.99 respectively. Similarly MAPE of 1 day prediction by proposed model is 0.647 where as rough set based and ARIMA models generate 0.964 and 0.84 respectively. Test error values for weekly, monthly and yearly predictions are mentioned in table II above.

V. CONCLUSION

In this work we have taken a novel hybridized approach that is rough set based affinity propagation model to forecast the gold opening price correctly. The beauty of rough set theory to handle uncertainty in data and subsequent capability of affinity propagation yielded better resultants in advance. The prediction results were compared with rough set based model and ARIMA model. The prediction performance measures such as RMSE and MAPE are evaluated and compared for all the models. The proposed model has shown the higher prediction accuracy as compared to the aforesaid models for the gold dataset in Indian scenario.

REFERENCES


AUTHORS PROFILE

Dr. Rudra Kalyan Nayak is presently working as Associate Professor in the Department of Computer Science & Engineering at Koneru Lakshmaiah Education Foundation (Deemed to be University), Green Fields, Vaddeswarum, Andhra Pradesh, India. He has got his M.Tech in Information Technology and Ph.D in Computer Science & Engineering from Sikkim O’N & Sciences Publication.
Rough Set Based Affinity Propagation Model for Prediction of Future Gold Price in Indian Scenario

Dr. Ramamani Tripathy is presently working as Assistant Professor of department of Master of Computer Application at United School of Business Management in Bhubaneswar, Odisha, India. She has 10+ years of teaching, mentoring research and academic experience at USBM. She has a Ph.D in Computer Science & Engineering (CS&E) from SOA University, Odisha. She is professional member of ISTE and ACM. Her research interest lies in the field of Bioinformatics, Financial Engineering and Data mining.

G. Venkatesh is currently graduating in Department of Computer Science & Engineering at Koneru Lakshmaiah Education Foundation (Deemed to be University), Green Fields, Vaddeswaram, Andhra Pradesh, India. His research interest lies in Operating Systems, Real Time Systems, Data Base Management Systems.

I. Sai Ram is currently graduating in Department of Computer Science & Engineering at Koneru Lakshmaiah Education Foundation (Deemed to be University), Green Fields, Vaddeswaram, Andhra Pradesh, India. His research interest lies in Operating Systems, Real Time Systems, Data Base Management Systems.