Gold prices prediction:
Comparative study of multiple forecasting models

Udai Bhan Trivedi [1], Thakur Vats Singh Somvanshi[2], Suraj Prasad J [2]

[1] Pranveer Singh Institute of Technology, Kanpur,India
[2]Pranveer Singh Institute of Technology, Kanpur,India
[3]Liverpool John Moores University, UK

Corresponding Author: udaibhantrivedi@gmail.com

Abstract:

Gold is a rare and valuable metal as its price has attention globally. Gold price varies very frequently, almost daily and it is in high focus by the government, investors, and industrialists. Trends of Increasing in the Gold price indicate gold is also one of the better investment plans to yield better profits in future. Forecasting the gold price is a primary financial problem for a lot of people as it is considered an investment asset and gold is used as raw material in industrial products.Hence forecasting the gold price is much needed for the financial economy of any nation. Implementing an accurate forecasting technique to predict the gold price is Crucial. Better investments in gold can be made only when we predict the price meticulously. This research paper focuses on forecasting the gold price by implementing multiple forecasting models to identify which model performs better on the dataset to get better error measures.

Keywords: Time Series Prediction, ARMA, ARIMA, LSTM, MAPE, RMSE

1 Introduction

1.1 Background

Gold can be considered the most valuable and significant commodity in the market, as it aids in the financial development of both individuals and governments. The role played by gold is very important in the world economy, for many nations, it is the most important aspect of their monetary enterprise reserves, gold is seen as a rate-controlling tool and a strategic monetary resource (Li et al., 2021). Investor expectations, linked market movements, and political events all have an impact on the gold market, which is nonstationary and volatile by nature (Li et al., 2021). Gold Price forecasting is the technique of predicting long-term market patterns using historical data[18] [19].

Recently, economists and investigators have been looking into mathematical approaches for analysing financial data to make accurate divination. In today's environment, financial data processing and analysis are becoming increasingly vital. It is exceedingly complex and computationally demanding to investigate how financial data behaves. The pricing of various commodities is influenced by a variety of variables. Market fluctuations, economic policies, sicknesses, weather conditions, and other factors can all play a role. The financial sector is very complicated and nonlinear, with a large number of variables that influence predictions. As a result, financial information is never linear (Vidya and Hari, 2020)[27]. Deep learning techniques are being used to anticipate and model gold prices, which are carried out using LSTM CNN models to forecast the gold price (Vidya and Hari, 2020)[27]. Also,to estimate gold prices, the usage of the whale optimization technique with a multilayer perceptron neural network (Alameer et al., 2019)[2].One of the study projects tells us how to anticipate gold pricescan be done using the Hybrid VM-ICSS-BiGRUalgorithmic trading and estimating gold futures Prices (Li et al., 2021) [13]. (Tripathy, 2017) The autoregressive integrated moving average (ARIMA) model was used to forecast gold prices. [23]. To measure the model's prediction accuracy, he uses the mean absolute error (MAE), mean absolute percentage error (MAPE), and other forecasting performance metrics. Several factors impact the price of gold, making price movement unpredictable. These factors include, among others, the rate of expansion, demand and supply, and policy-driven issues. When the economy expands, the price of gold rises, and when there is a low inventory of any commodity, the price rises. Furthermore, when countries fear that the value of the dollar, which is the world's biggest currency, would collapse, the price of gold will rise as a result of the increased demand for gold. As a result of its significance, other writing has named it a place of refuge during monetary emergencies (Makala and Li, 2021)[16].

Our research study will be implementing multiple Forecasting models and compare their performance using error measures namely MAPE and RMSE for each of the models and find the best one to estimate the price.

2 Related Research

The list of the paper which have been reviewed based on problems under consideration, the purpose of study, pre-processing of data, Used Algorithm, Evaluation Matrix, summary and remarks is given in table 1.

Table1. Table on Related Works

Sl No	Year	Title	Author(s)	Dataset	Problem(s)	Purpose	Preprocessing	Algorithm s	Evaluation	Summary	Remarks
1	2020	Gold Price Prediction and Modelling using Deep Learning Techniques	(Vidya and Hari, 2020) [27]	Data taken from World Gold Council for year 1987 to44013	Gol pricing nonlinearity.	To forecast gold price using LSTM	Data is scaled Between Minimum and Maximum Value (0-1)	Long Short-term Memory Network s (LSTM)	Root mean square error (RMSE)	It gave good RMSE value of about 7.385.	Performed better than traditional Forecasting models.
2	2015	Gold Price Prediction Using Type-2 Neuro- Fuzzy Modeling and ARIMA	(Modeling et al., 2015)	Gold Price historical data	When time factor is included in dataset there will uncertainty in results which might occur in future	To predict accuracy in predicting price of gold.	1)Type-2 TSK2) Fuzzy Rules 3) Crisp Set 4) Hybrid Learning Algorithm	type-2 neuro-fuzzy modeling and ARIMA (Auto regressive Integrated Moving Average)	Root mean square error (RMSE)Mean Absolute Percentage error (MAPE)Mean Absolute Error (MAE)	Algorithm implemented in research gave better results.	Accuracy is achieved with implement ed technique.
3	2017	Forecasting Gold Price with Auto Regressive Integrated Moving Average Model	(Tripathy, 2017) [23]	Price of gold from July 1990 to February 2015	Forecasting models forecasting is inaccurate	To gain more accuracy using ARIMA (Auto regressive Integrated Moving Average)	Check the stationarity of data.	Box- Jenkins' ARIMA (Auto regressive Integrated Moving Average)	Provides good results for the error measures used.	This result says ARIMA (0,1, 1) is best model among all.	Wavelet analysis can be implemented ed.
4	2018	Using Classification Techniques to Predict Gold Price Movement Wedad	(Al-Dhuraibi and Ali, 2018) [1]	Weekly basis data collected from the year January 1st 2007 to December 24th	Unpredictable investment decisions on gold	To check if price of gold inclines or declines	Data processig is done using "Rapid miner"	Decision Tree, Support Vector Machin e and other method	Precision, Recall	Performance of K- Nearest Neighbour is acceptable.	We have to select the models based on correlation.
5	2020	Deep belief network for gold price forecasting Pinyi	(Zhang and Ci, 2020) [9]	Price of gold from January 1984 to December 2019,	Multi -factor influence on price of gold.	To Implement deep belief network (DBN)model to forecast Gold Price	Its carried out by well-known methods.	Unique ML Models	MAPE: Mean absolute percentage error MAE: Mean absolute error RMSE: Root mean square error	The proposed models give good results.	New model Deep belief model can be used for gold price forecasting.

Sl No	Year	Title	Author(s)	Dataset	Problem(s)	Purpose	Preprocessing	Algorithms	Evaluation	Summary	Remarks
6	2021	Structural analysis and forecast of gold price returns	(Chai et al., 2021) [8]	Daily data of each variable of gold price	Due to complicated market prices of gold.	This work provides dynamic relation of price and factors.	To Eliminate non-overlapping data	STL- ETS, neural network and Bayesian structural time series model	Forecast error variance decomposition	This paper explains positive impact on crude oil.	Proposed model gives better results.
7	2008	Forecasting gold price changes: Rolling and recursive neural network models	(Parisi et al., 2008)	Price observation s from 10th January 2000 to 05th April 2005 is used	Dynamic Gold price changes and non-linearity of data	To implement and use rolling neural networks.	In sample set and out sample set	Rolling and recursive neural network models	Predictive capacity Rolling operation Recursive operation	Used model provides good results which can be relied on.	dynamic neural networks can be used in any economic scene
8	2019	Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm	(Alameer et al., 2019) [2]	360monthly observa tions 1987 September to 2017 August	Gold price fluctuations plays	accurately forecasting long- term monthly gold price fluctuations	Data splitting is done.	whale optimization algorithm (WOA) perceptron neural network (NN)	Different and multiple machine learning methods used	Study will be proving promising outcomes.	Fold price volatility can be done using predict vars.
9	2021	Chaotic behaviour in gold, silver, copper and bitcoin prices	(Bildirici and Sonustun, 2021) [6]	Data set of February 2012 to May 29,2020	How covinf-19 influenced the price of Gold.	To investigate the behaviour of multiple items.	Econometric methodology	Switching and multiplayer type models.	Performance model was analysed.	It came up with which is best commodity to be invested in.	Suggestion of best investment is extracted.
10	2021	A New Hybrid VMD- ICSS-BiGRUAppro ach for Gold Futures Price Forecasting and Algorithmic Trading	(Li et al., 2021) [13]	Raw data of year 2000 of January to January 2019	Complex and nonstationary data.	Market movements of gold ix captured.	Data set splitting and normalization.	Variatio nal mode decomp osition (VMD)-iterated cumulative sums of squares (ICSS)-bidirecti onal gated recurrent unit (BiGRU)	Generic error measures are taken care and calculated.	The models and methods implement ed gives good results.	Strategy proposed is long only strategy and we also need short term strategy.
11	2019	Gold and Diamond Price Prediction Using Enhanced Ensemble Learning	(Pandey et al., 2019) [19]	Previous data of the product.	Variation in price of gold market.	To analyze and examine the patterns of previous close prices.	Remove noise and blank entries	linear regression on and random forest	Mean, Best and worst is calculated for preciseness.	Multiple models used in this paper gave better results.	Models were efficient and gave better outputs.

3 Gold prises predictive Models

The research for analysing the effectiveness of multiple forecasting models on gold price data set involves major key processes Such as collection of datasets, data pre-processing, building multiple forecasting models, plotting the outcome of every forecasting model, calculating the error measures, and also comparing the effectiveness of all the forecasting models. The relevant dataset consists of the daily gold price over the last five years. Data set pre-processing has been doneas per the feasibility of each forecasting model. Each forecasting model's error matrix of MAPE and RMSE must be calculated, and the effectiveness of all forecasting models compared in the conclusion.

3.1 The Simple Auto-Regressive Model (AR)

The basic autoregressive model anticipates the future observation (dependent variable)based onpast observations. Autoregressive is a time series model that predicts values for the next time step by using data from the previous time step as input to the regression equation. This is a very simple idea and can lead to accurate predictions of many time series problems. The lag order parameter of the model is denoted by the letter 'p.' The greatest number of lags required to create the 'p' numbers of past data points needed to anticipate future data points is called lag order.

Figure1. Partial autocorrelation function

Choose p as the longest latency at which partial autocorrelation is the highest. In figure 1, the lag values of 1, 2, 4, and 12 have a high level of confidence. i.e., a significant amount of control over future observations. Create an equation for an autoregression model $\boldsymbol{y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{y}_{t-1} + \boldsymbol{\beta}_2 \boldsymbol{y}_{t-2} + \boldsymbol{\beta}_4 \boldsymbol{y}_{t-4} + \boldsymbol{\beta}_{12} \boldsymbol{y}_{t-12}$

1, 2, 4, and 12 are the historical values that have considerable significance. As a result, the independent variables y_{t-1}, y_{t-2}, y_{t-4}, and y_{t-12},which are historical observations, have been used to forecast the regression model's dependent variable y_t.

3.2 Moving Average Model (MA)

The Moving Average (MA) is a phrase used to denote the average of two or more moving averages. The current value is linearly reliant on the present and past error terms in a moving average process or moving average model. Like white noise, the error terms are thought to be mutually independent and dispersed frequently. In a regression-like model, prior forecast mistakes are used to model future projections. In this model, the window size parameter 'q' is employed to calculate the linear combination of errors.

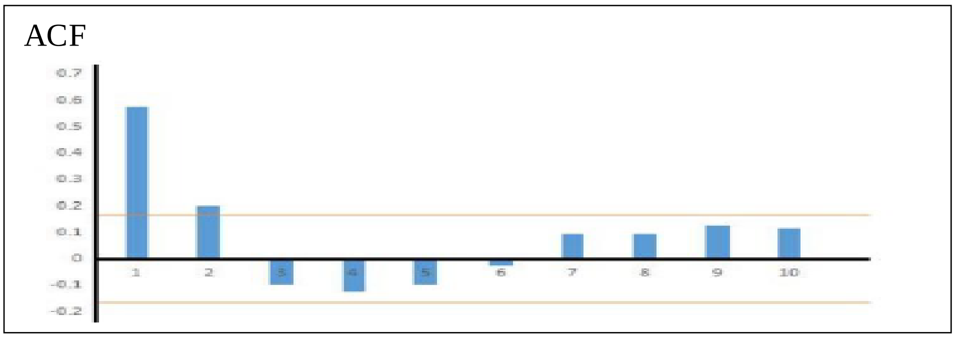


Figure 2. Autocorrelation function

The value of q is calculated asthe maximum lag beyond which autocorrelation dies to exist. In the above figure2. The value of lag 1 and lag 2 the autocorrelation value is above the significance level[5]. So, value of q will be 2 as it is thegreatest lag after which autocorrelation ceases to exist.Moving Average model equations for above figure $\hat{y} = \mu + \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2}$

3.2 Auto Regressive Moving Average (ARMA)

ARMA is a model for predicting in which the methods of autoregression (AR) analysis and moving average (MA) are both used on well-behaved time-series data. In ARMA, the time series is thought to be stationary, and when it varies, it does so evenly around a certain period. A period of series which shows the qualities of an AR(p), as well as MA(q) cycle, will be reproduced utilizing an ARMA (p,q) model.

The ARMA (1,1) when p=1 and q=1 the equation $\hat{y} = \beta_0 + \beta_1 y_{t-1} + \phi_1\varepsilon_{t-1}$the anticipated value is y.

To figure out what the 'p' and 'q' are

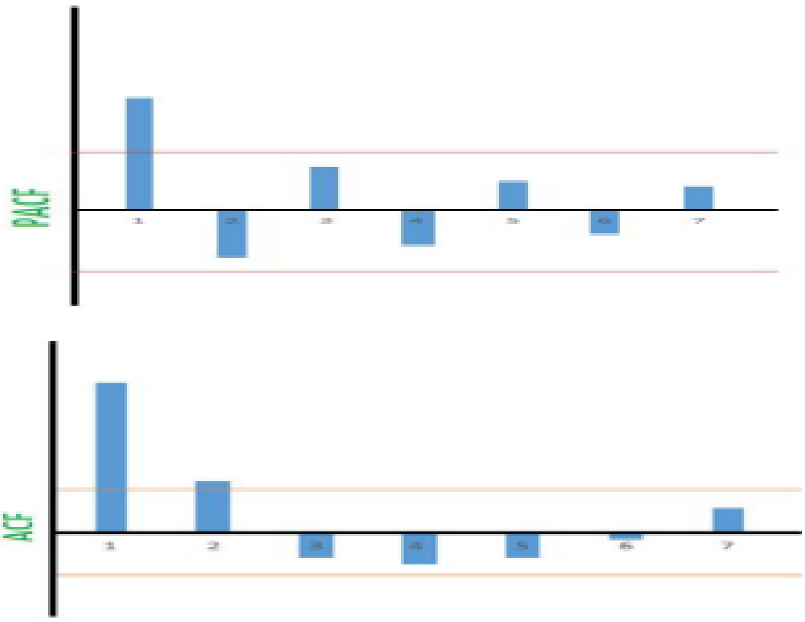


Figure 3. Autocorrelation function and Partial autocorrelation function

In the above figure 3 plot for PACF and ACF, p = 1is inertness with the greatest incomplete autocorrelation and q = 2 as the maximum lag beyond which autocorrelation is no longer visible in the ACF plot[5].

3.4 Auto-Regressive Integrated Moving Average (ARIMA)

It's a statistical and economic model for calculating events across time. The model is used to interpret historical data or forecast future data in a series. To stabilize the time series, it is differentiated, and the differentiated series is constructed as a linear regression with one or more past observations of previous prediction errors[28].
There are 3 considerations in the ARIMA model.
‘d’: degree of differencing, ‘q’: Number of previous error terms,
‘p’: Highest lag
‘t’ eliminates the pattern (non-stationarity) before re-coordinating it into the first dataset.
To change the data into a stationary time-series box cox transformation and differencing at first is used. A research paper uses box cox before the model is generated and then differencing is handled by the model itself (i.e., the trend component).

- $Z_t$is first-order differencing in time series.
- Determine the parameters 'p, d, and q'.
- 'd’: To make the original time series stationary, choose d as the order of difference. The stationarity tests ADF and KPSS determine whether or not this differenced stationary series.
- Plot the ACF and PACF of the first request differenced time series for 'p' and 'q'. Likewise, with the first Auto-Regressive Models, decide the value of 'p' and 'q'.
- The first time series gauge is recuperated in the ARIMA model's last stage.
  For Example,ARIMA (1,1,1)
  $z_t = \phi_1 z_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$ Where $z_t = y_{t-1} - y_t$

3.5 Seasonal Auto-Regressive Integrated Moving Average (SARIMA)

This model differs from an ARIMA model based on seasonal patterns by one step. Seasonal influences may be found in many time series data sets. For example, the temperature in this season will almost probably have a strong correlation with the temperature in the same season the previous year[14].

- With Benefits of ARIMA, SARIMA has a seasonality component. To make the time series stationary, it is differed.
- Future observations are modelled as a linear regression of previous data and forecast errors. SARIMA's seasonal components
- Perform seasonal differencing on time series, and model future seasonality as a linear regression of past seasonality observations and past seasonality forecast errors.
  - The parameters 'p', 'd', 'q', and 'P', 'D', 'Q':

- Elements that aren't seasonal: p: order of trend autoregression • d: order of trend difference • q: order of trend moving average
- Seasonal ingredients: • D: Seasonal difference order • Q: Seasonal moving average order P: Seasonal autoregressive order
  $z_t = \beta z_{t-1} + \phi\epsilon_{t-4} + \epsilon_t$ Where $z_t = y_t - y_{t-4}$

3.6 SARIMAX Model(Seasonal Auto-Regressive Integrated Moving Average with exogenous variables)

The ARIMA model has been upgraded to SARIMAX. SARIMAX, like SARIMA and Auto ARIMA, is a seasonal equivalent model. It's also capable of dealing with external influences. This aspect of the model sets it apart from others. Seasonal elements are not available. A linear regression of prior observations and forecast errors is used to model future data.Time series are differentiated to make them stationary[14]. The direct regression of chronicled perceptions and estimated mistakes from past seasons is utilized to characterize model irregularity. The model uses seasonal differencing to make series fixed.An external variable is used in this case. Future observations of the model as a linear regression of an external variable

$z_t = \beta z_{2t-1} + \phi\epsilon_{t-4} + \alpha x_t$ Where $z_t = y_t - y_{t-4}$

The PACF plots are used to calculate ‘p’ which is non-seasonal. The ACF plots are used to locate ‘q’ which is non-seasonal. To identify the value of 'd,' use stationarity tests. To find the best seasonal P, D, and Q parameter values, use grid search[22].

3.7 Univariate LSTM

The Long Short-Term Memory engineering is based on a Recurrent Neural Network (RNN)-design utilized in regular language handling and time series anticipating (Webpage/Towards DS, n.d.)[4].

A sort of artificial recurrent neural network (RNN), the Long Short-Term Memory (LSTM), may be used to estimate asset value based on previous data. It's intended to address the issue of long-term addiction while also preventing the gradient from disappearing. Because they keep an internal state to monitor the data that has already been shown, LSTMs are good for modelling sequence data. Since LSTMs include feedback connections, time series and natural language processing are two popular uses. It can handle full data sequences as well as individual data points [26] [15].

figure 4shows, that many memory blocks comprise the LSTM. The two states are kept in the next block. Cell state (information is stored and loaded) and concealed state (sends and overwrites information from the previous state). A technique called gates is used by LSTMs to learn. These gates can learn to keep or reject information in the sequence. Thus, the LSTM has three gates: input, forgetting, and output. [12] [15].

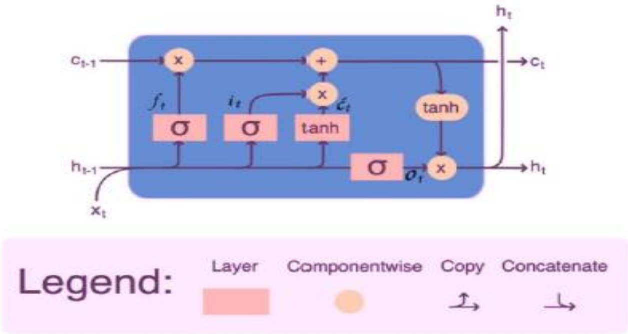


Fig 4. LSTM Architecture Flow Diagram

$f_t = \sigma_g(W_f x_t + U_f c_{t-1} + b_f)$
$i_t = \sigma_g(W_i x_t + U_i c_{t-1} + b_i)$
$o_t = \sigma_g(W_o x_t + U_o c_{t-1} + b_o)$
$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + b_c)$
$h_t = o_t \circ \sigma_h(c_t)$
f_t= Forget gate, i_t= Input gate, o_t= Output gate, c_t= Cell state, h_t= Hidden state

The LSTM tackles a critical repetitive neural organization issue: a short memory. The LSTM utilizes a progression of 'gates,' each with its own RNN, to keep, neglect or disregard information focuses given a probabilistic model. Problems with explosive and disappearing gradients can also be solved with LSTMs [24] [25].

Each forecast puts a small amount of uncertainty onthe model. Toprevent exploding gradients, values are squished using sigmoid and before gate entrance and the output modelneeds tanh activation functions (Webpage/Towards DS, n.d.)[4].

4 Error Measures (Evaluation Metrics)

The research paper measures MAPE and RMSE values for all estimating models and compared them.

4.1 Mean Forecast Error (MFE)

Essentially take away the genuine upsides of the reliant variable, i.e., 'y,' from the expected upsides of 'y,' to show up in this major technique[10][1]. $MFE = \frac{1}{n}\sum_{i=1}^{n}(y_{actual} - \hat{y}_{forecast})$

4.2 Mean Absolute Error (MAE)

This calculation is more evident as this considers the difference between the actual and projected values for which absolute values, whereas overestimated values and underestimated values are cancelled out by MFE[7].

$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y = (y_{actual} - \hat{y}_{forecast})|$

4.3 Mean Absolute Percentage Error (MAPE)

The difficulty with MAE is that you can compare incorrect values of it to any parameters. MAPE calculates the mean absolute error (MAE) as a percentage of the actual values of 'y' to indicate how well the forecast performed based on real data[10][11].

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

4.4 Mean Squared Error (MSE)

Mean squared error and mean absolute error both have the same goal: to capture absolute deviations without cancelling out negative and positive variances. MSE square the error numbers and adds them together and takes the average.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y_{actual} - \hat{y}_{forecast} \right)^2$$

4.5 Root Mean Squared Error (RMSE)

RMSE is calculated by taking the square root of the MSE value, because the MSE error term is squared and not in the same dimension as the goal variable 'y'[10][11].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_{actual} - \hat{y}_{forecast} \right)^2}$$

5 Research methodology

The research for analysing the effectiveness of multiple forecasting models on gold price data set involves major key process Such as collection of datasets, data pre-processing, building multiple forecasting models, plotting the results of every forecasting model, calculating the error measures, and also comparing the effectiveness of the estimating models.
The research methodology flow chart is as shown below,

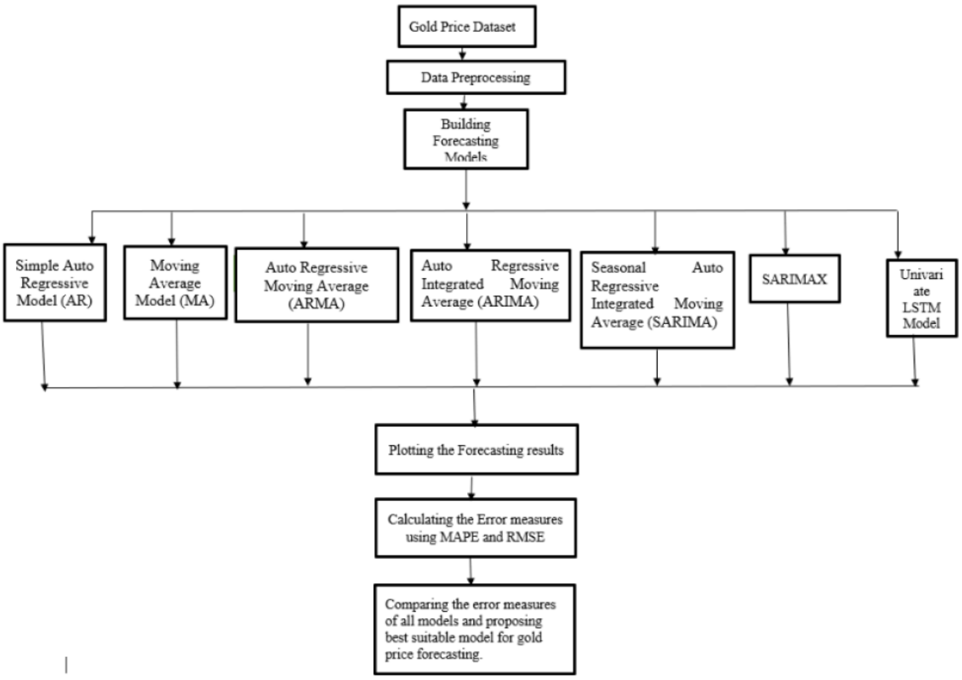


Figure 5. Flow Chart of Research Process

5.1 Selecting Dataset

The major and initial step of any forecasting model is to collect the valid and relevant data set. As we are forecasting the gold price dataset has been collected from Kaggle, the data set of gold price for the period of 01-01-2014to 29-12- 2021[3]. The dataset is pre-processed for time series forecasting. Data set looks as shown below.

Table 2. Structure of the Gold Price dataset (Webpage/Kaggle, n.d.)

Date	Price	Open	High	Low	Volume	Chg%
1/1/2014	29542	29435	29598	29340	2930	0.25
1/2/2014	29975	29678	30050	29678	3140	1.47
1/3/2014	29727	30031	30125	29539	3050	-0.83
1/4/2014	29279	29279	29279	29279	0	-1.51
1/6/2014	29119	29300	29395	29051	24380	-0.55
1/7/2014	28959	29130	29195	28912	18710	-0.55
1/8/2014	28934	28916	29029	28820	18140	-0.09
1/9/2014	28997	28990	29053	28865	15130	0.22
1/10/2014	29169	29030	29198	28960	15810	0.59

Dataset mentioned above has a few sample entries of all the columns collected. This data set has the following columns, Date,Price. Open, High, Low, Volume and chg.%...For us, as we are implementing univariate estimating models our major focus will be on Date columns and target column Price (Webpage/Kaggle, n.d.). So, using this we can observe the change in gold price every day and implement the estimating models.

5.2 Data Pre-processing

Any kind of dataset must be pre-processed before inputting intothe model. Following are the few data pre-processing procedures which must be taken care of for auto-regressive models.

In an autoregressive model, the regression technique is used to define a time series problem. To construct autoregressive models, we expect future observations using a linear combination of previous observations. For example, we need to predict the future forecasting variable y_t , to do so we need previous samples of y_t such as y_{t-1}, y_{t-2}, and so on. We need to build autoregressive models on the following assumptions,

- Stationarity
- Auto Correlation

Assuming a period series is steady, factual properties like mean, difference, and covariance will stay reliable over the long haul, regardless of when you take a gander at it. Since the series will be fixed, factual properties like mean, fluctuation, and covariance should be something similar, requiring the utilization of fleeting information.

5.3 Stationarity Tests

There are two conventional tests for stationarity given theory testing. A regular way of recognizing whether or not a period series is fixed is to utilize the unit root test. Assuming the presence of a unit pull for a series can't be invalidated, it is considered non-fixed

5.4 The KwiatkowskiPhillipsSchmidtShin Test(KPSS Test)

- When the p-value is larger than 0.05, the null hypothesis (H0) states that the series will remain stationary.
- When the p-value is less than or equal to 0.05, the Alternative Hypothesis (H1) suggests that the series is not stationary.

5.5 The ADF (Augmented Dickey-Fuller) Test

- P value>0.05 value for Null Hypothesis (H0) which denotes series is notstationary
- P value<=0.05 value for Alternate Hypothesis (H1) which denotes theseries is stationary

5.6 Changing a Non-Stationary Time-Series to a Stationary

For implementing the auto-regressive models, time series must be stationary, and thus, for a non-stationary time series, we need to first convert it into a stationary one.Two tools can be used to make a stationary series out of a non-stationary series, namely Differencing and Transformation.

In a periodor series, the strategy of differencing is utilized to eliminate the pattern (making the mean consistent). In differencing, you register the distinctions between back-to-back perceptions, as the name infers... It's equivalent to differencing an inclined line to zero of the slants. By settling the mean of a period series and decreasing variances in the level of the time series, differencing wipes out (or limits) patterns and irregularity[20] [21].

One more strategy to acquaint stationarity is to make the fluctuation consistent. For making a non-fixed series fixed, there are other change techniques, yet we'll zero in on the Box-Cox transformation here.

It is based on mathematical equations.

$\boldsymbol{y(\lambda)}=\{(\boldsymbol{y}\text{^}\boldsymbol{\lambda}-\mathbf{1})/\boldsymbol{\lambda}, \boldsymbol{if\lambda}=\mathbf{0}$

$\boldsymbol{y(\lambda)}=\{\boldsymbol{logyif}\ ,\boldsymbol{\lambda}\neq\mathbf{0}$

where y(λ) will be changed time series and y(λ) going to be the first-time series The Box-Cox change is done by choosing an ideal worth of - 5 to 5 that limits the fluctuation of the changeover information.

5.7 ACF and PACF

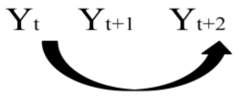
The connection between perceptions y_t at time t and y_{t-k} at time k time stretch before t is caught via autocorrelation. To put it another way, autocorrelation lets us know how a variable is impacted by its own slacked values. Here, we checked out two different autocorrelation measures:

1.The Autocorrelation function 2.The Partial autocorrelation functions

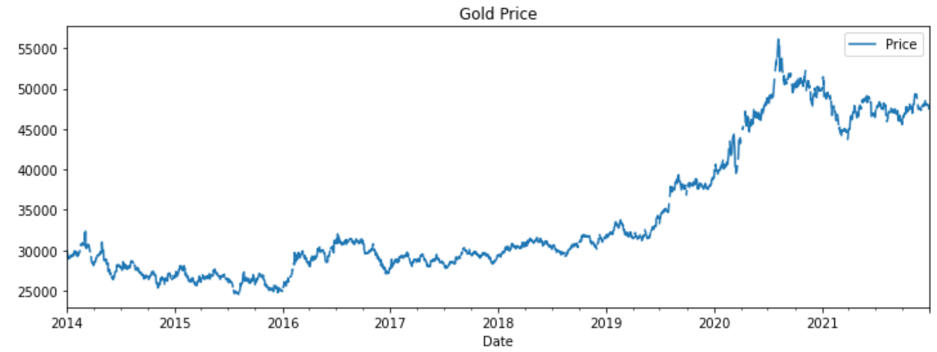
Autocorrelation work portrays how a solitary perception and its slacked values are connected. It helps you in concluding which observational slack generally affects it.

Both immediate and circuitous connections between factors are caught by the autocorrelation work.

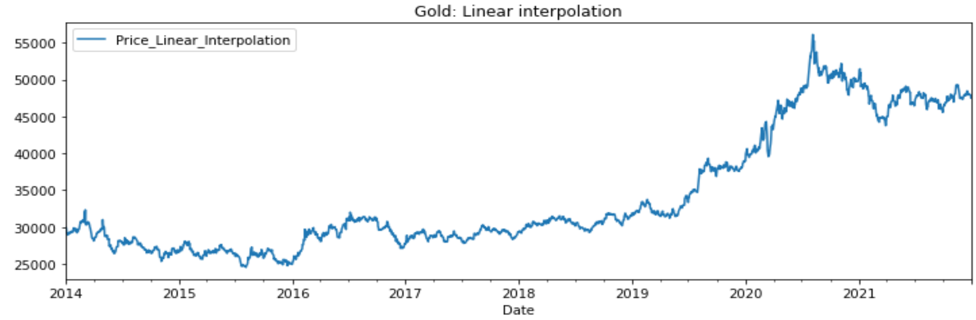
For y_t , y_{t+1} , y_{t+2} : Autocorrelation function captures both direct and indirect relationships with its lagged values. Here, the big arrow on the bottom indicates the direct relationship that is captured between y_t , y_{t+2} The indirect relationship between y_t and y_{t+2} through y_{t+1} is likewise captured by the autocorrelation function. In other words, y_t will have a relationship with y_{t+1}, and y_{t+1} will have a relationship with y_{t+2}. The indirect relationship reflected by the Autocorrelation function is the transitive correlation that flows through y_{t+1}. As a result, you can't use ACF to separate out just the direct relationship. Another metric is known as the Partial Autocorrelation Function, or PACF, which is used to capture just direct correlations.



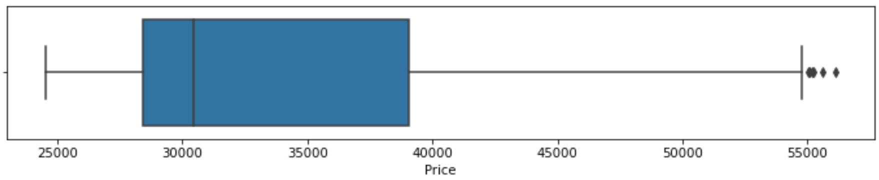
6 Performance Analysis of forecasting models on Gold Price Prediction



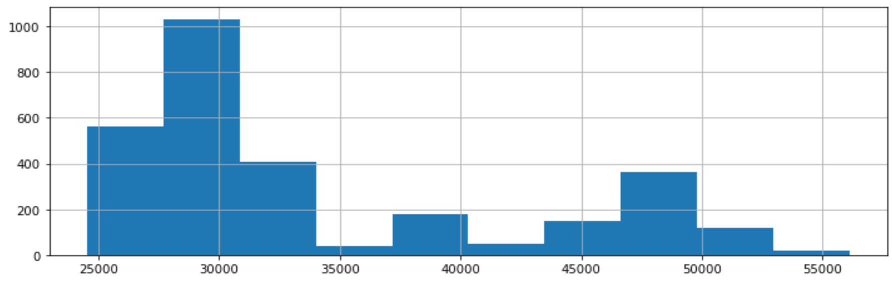
6.1 LinearInterpolation to Fill the missing Dates



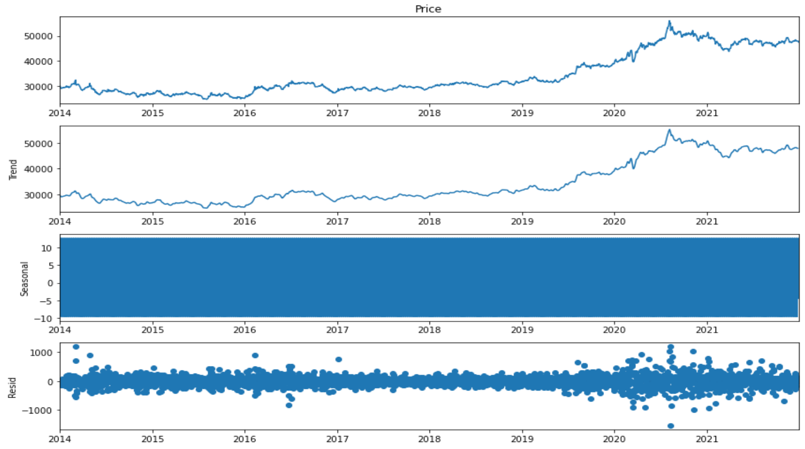
6.2 Box Plot to check outliers



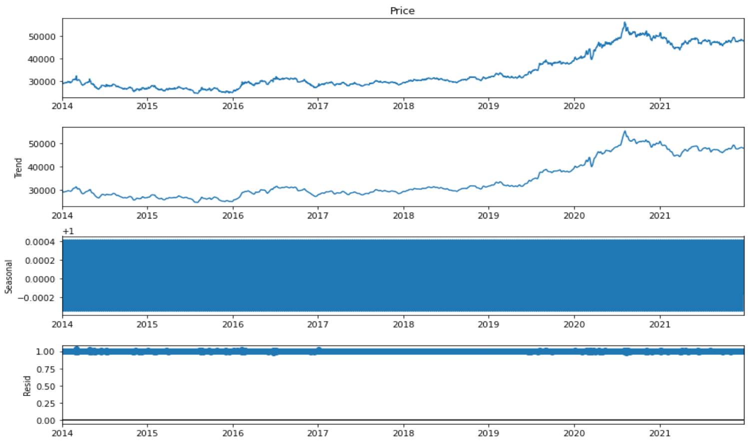
6.3 Histogram Plot



6.4.1 Time-Series Decomposition: Additive seasonal decomposition



6.4.2 Time-Series Decomposition: Multiplicative seasonal decomposition



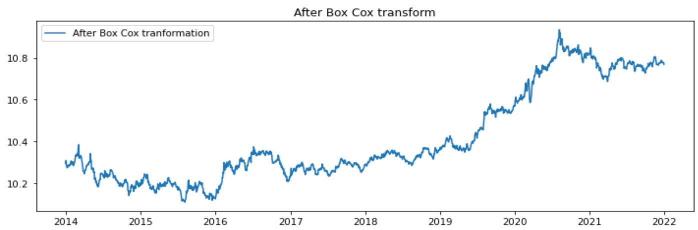
6.5 Augmented Dickey-Fuller (ADF) test

ADF Statistic: -0.183439, Critical Values @ 0.05: -2.86, p-value: 0.940467

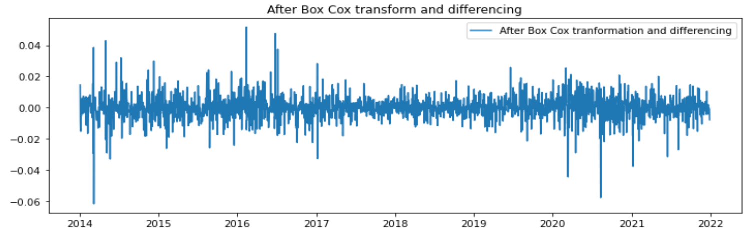
6.6 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

KPSS Statistic: 8.086909, Critical Values @ 0.05: 0.46, p-value: 0.010000

6.7 Box-Cox transformation to make variance constant



6.8 Differencing to remove the trend



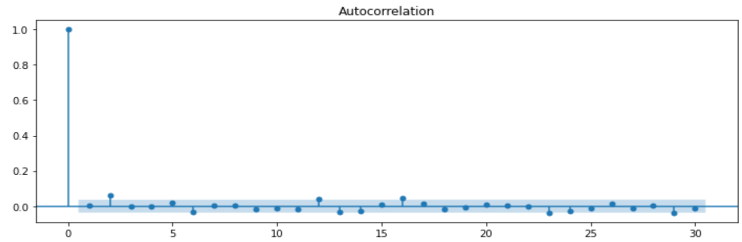
6.9 Augmented Dickey-Fuller (ADF) test

ADF Statistic: -35.858365, Critical Values @ 0.05: -2.86, p-value: 0.000000

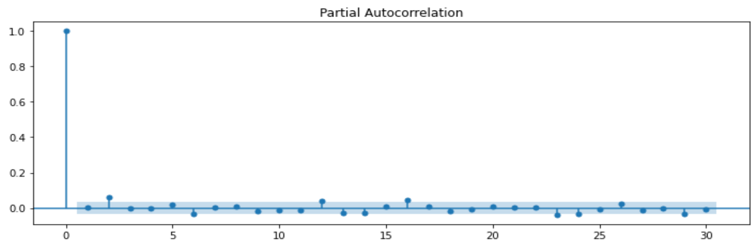
6.10 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test

KPSS Statistic: 0.209949, Critical Values @ 0.05: 0.46, p-value: 0.100000

6.11 Autocorrelation function (ACF)

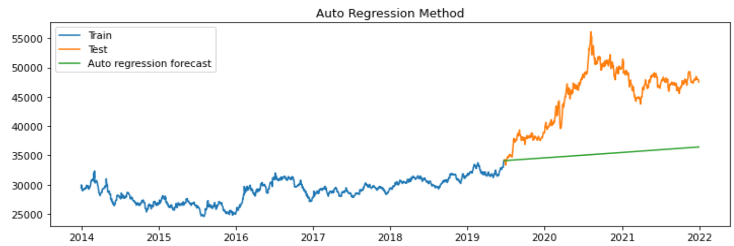


6.12 Partial autocorrelation function (PACF)

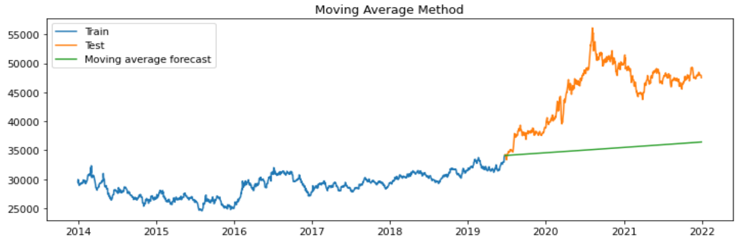


7 Build and evaluate time series forecast

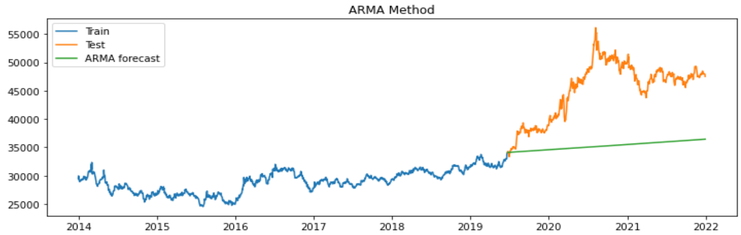
7.1 Auto-Regressive Method



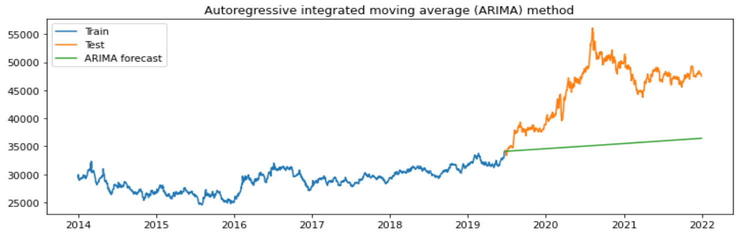
7.2 Moving Average Method



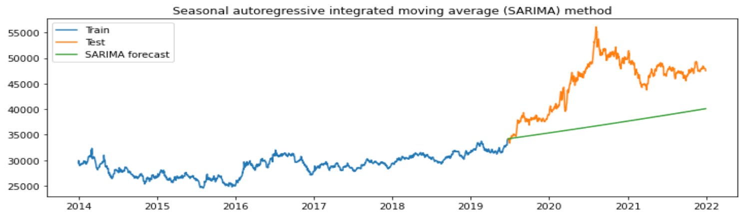
7.3 Auto-Regressive Moving Average Method



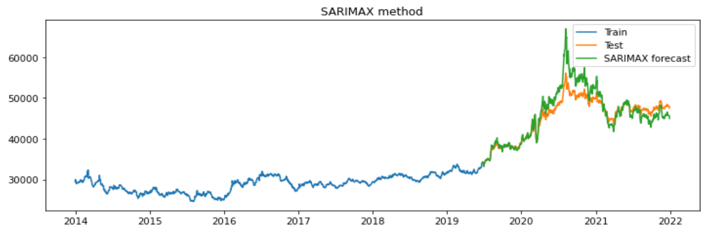
7.4 Autoregressive Integrated Moving Average Method



7.5 Seasonal Autoregressive Integrated Moving Average Method



7.6 SARIMAX



7.7 LSTM

The LSTM Model have very less RMSE and MAPE error so the predicted value is the same as the actual value of the gold price

Table 3. Various Time Series Model and Corresponding RMSE and MAPE Values

	Method	RMSE	MAPE
1	Autoregressive (AR) method	10942.77	21.18
2	Moving Average (MA) method	10940.91	21.17
3	Autoregressive moving average (ARMA) method	10931.27	21.15
4	Autoregressive integrated moving average (ARIMA)	10940.91	21.17
5	Seasonal autoregressive integrated moving average(SARIMA)	9059.90	17.29
6	SARIMAX method	2466.56	3.59
7	LSTM	26.59	0.015

8 Conclusion

After analysis of the output of time series forecasting and its corresponding Error matrix outputmentioned in table7.1. this can be easily verified that the Recurrent Neural Network (RNN) based Long Short-Term Memory (LSTM) Model-design gives the best forecast on test data with RMSE error of 26.59 and MAPE Error of 0.015. Out of the traditional Time Series Model SARIMAX is best with MAPE Error of 2466.56 and RMSE 3.59. So, it is recommended that LSTM Model will be used for gold price prediction purposes.

References

1. Al-Dhuraibi, W.A. and Ali, J., (2018) Using classification techniques to predict gold price movement. 2018 4th International Conference on Computer and Technology Applications, ICCTA 2018, pp.127–130.
2. Alameer, ZElaziz, M.A., Ewees, A.A., Ye, H. and Jianhua, Z., (2019) Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. Resources Policy, [online] 6January, pp.250–260. Available at: https://doi.org/10.1016/j.resourpol.2019.02.014.
3. Anon (n.d.) Webpage/Kaggle. [online] Available at: https://www.kaggle.com/nisargchodavadiya/daily-gold-price-20152021-time-series.
4. Anon (n.d.) Webpage/TowardsDS. [online] Available at: https://towardsdatascience.com/lstm-framework-for-univariate-time-series-prediction-d9e7252699e.

5. Anon (n.d.) Webpage/Upgrad. [online] Available at: https://learn.upgrad.com/course/982/segment/7761/50153/144280/777534 .
6. Bildirici, M.E. and Sonustun, B., (2021) Chaotic behavior in gold, silver, copper and bitcoin prices. Resources Policy, [online] Volume 74 , December 2021, p.102386. Available at: https://doi.org/10.1016/j.resourpol.2021.102386.
7. Bilgin, M.H., Gozgor, G., Lau, C.K.M. and Sheng, X., (2018) The effects of uncertainty measures on the price of gold. International Review of Financial Analysis, 58February, pp.1–7.
8. Chai, J., Zhao, C., Hu, Y. and Zhang, Z.G., (2021) Structural analysis and forecast of gold price returns. Journal of Management Science and Engineering, 62, pp.135–145.
9. Chen, L. and Zhang, X., (2019) Gold Price Forecasting Based on Projection Pursuit and Neural Network. Journal of Physics: Conference Series, 11686.
10. E, J., Ye, J. and Jin, H., (2019) A novel hybrid model on the prediction of time series and its application for the gold price analysis and forecasting.
11. Fischer, T. and Krauss, C., (2018) Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, [online] 2702, pp.654–669. Available at: https://doi.org/10.1016/j.ejor.2017.11.054.
12. He, Z., Zhou, J., Dai, H.N. and Wang, H., (2019) Gold price forecast based on LSTM-CNN model. Proceedings - IEEE 17th International Conference on Dependable, Autonomic and Secure Computing, IEEE 17th International Conference on Pervasive Intelligence and Computing, IEEE 5th International Conference on Cloud and Big Data Computing, 4th Cyber Scienc, pp.1046–1053.
13. Li, Y., Wang, S., Wei, Y. and Zhu, Q., (2021) A New Hybrid VMD-ICSS-BiGRU Approach for Gold Futures Price Forecasting and Algorithmic Trading. IEEE Transactions on Computational Social Systems, 86, pp.1357–1368.
14. Liu, Z., Zhu, Z., Gao, J. and Xu, C., (2021) Forecast Methods for Time Series Data: A Survey. IEEE Access, 9, pp.91896–91912.
15. Livieris, I.E., Pintelas, E. and Pintelas, P., (2020) A CNN–LSTM model for gold price time-series forecasting. Neural Computing and Applications, [online] 3223, pp.17351–17360. Available at: https://doi.org/10.1007/s00521-020-04867-x.
16. Makala, D. and Li, Z., (2021) Prediction of gold price with ARIMA and SVM. Journal of Physics: Conference Series, 17671.
17. Pandey, A.C., Misra, S. and Saxena, M., (2019) Gold and Diamond Price Prediction Using Enhanced Ensemble Learning. 2019 12th International Conference on Contemporary Computing, IC3 2019, pp.2019–2022.
18. aza, S.A., Shah, N. and Shahbaz, M., (2018) Does Economic Policy Uncertainty Influence Gold Prices? Evidence from a Nonparametric Causality-in-quantiles Approach. Resources Policy, [online] 57December 2017, pp.61–68. Available at: https://doi.org/10.1016/j.resourpol.2018.01.007.
19. Risse, M., (2019) Combining wavelet decomposition with machine learning to forecast gold returns. International Journal of Forecasting, [online] 352, pp.601–615. Available at: https://doi.org/10.1016/j.ijforecast.2018.11.008.
20. Sezer, O.B., Gudelek, M.U. and Ozbayoglu, A.M., (2020) Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing Journal, [online] 90, p.106181. Available at: https://doi.org/10.1016/j.asoc.2020.106181.
21. Shafiee, S. and Topal, E., (2010) An overview of global gold market and gold price forecasting. Resources Policy, [online] 353, pp.178–189. Available at: http://dx.doi.org/10.1016/j.resourpol.2010.05.004.

22. Shankar, P.S. and Reddy, M.K., (2021) Forecasting Gold Prices in India using Time series and Deep Learning Algorithms. International Journal of Engineering and Advanced Technology, 105, pp.21–27.
23. Tripathy, N., (2017) Forecasting Gold Price with Auto-Regressive Integrated Moving Average Model. International Journal of Economics and Financial Issues, 74, pp.324–329.
24. Ul, I. and Nazir, K., (2017) Predicting Future Gold Rates using Machine Learning Approach. International Journal of Advanced Computer Science and Applications, 812, pp.92–99.
25. Verma, S., Thampi, G.T. and Rao, M., (2020) ANN-based method for improving gold price forecasting accuracy through modified gradient descent methods. IAES International Journal of Artificial Intelligence, 91, pp.46–57.
26. Vidal, A. and Kristjanpoller, W., (2020) Gold volatility prediction using a CNN-LSTM approach. Expert Systems with Applications, 157.
27. Vidya, G.S. and Hari, V.S., (2020) Gold Price Prediction and Modelling using Deep Learning Techniques. 2020 IEEE Recent Advances in Intelligent Computational Systems, RAICS 2020, pp.28–31.
28. Yaziz, S.R., Zakaria, R. and Suhartono, (2019) ARIMA and Symmetric GARCH-type Models in Forecasting Malaysia Gold Price. Journal of Physics: Conference Series, 13661.