

DEPLOYMENT OF RANDOM FOREST ALGORITHM TO PROPHECIZE THE EXACT PRICE OF GOLD

Utsab Ray (Department of Biomedical Engineering, JIS College of Engineering, Kalyani, West Bengal),

Debatwa Banik (Department of Computer Application, JIS College of Engineering, Kalyani, West Bengal),

Sourajit Sarkar (Department of Computer Application, JIS College of Engineering, Kalyani, West Bengal),

Karabi Ganguly (Department of Biomedical Engineering, JIS College of Engineering, Kalyani, West Bengal)*

Karabi Ganguly* : Corresponding Author

Abstract: Historically, gold, unlike other payment channels, was employed to support trading acquisitions worldwide. We forecast future gold prices based on twenty-two market variables using a machine learning technique. In order to analyze these data, one machine learning approach, random forest regression, was applied. Numerous states have kept and increased their gold reserves while being progressive and rich. In reality, central banks throughout the world keep precious metals like gold on hand to ensure foreign debt service in addition to stabilizing inflation. The primary goal of this research is to anticipate the increase and fall in routine gold rates, which will assist investors in deciding whether to purchase or sell gold. Logistics forecasting is critical to the fiscal performance of an organization. The secondary market is derived by examining the dataset including the previous year's gold price. There is concern that these high prices will not be sustainable and will fall despite the fact that several research has been conducted to examine the relationship between the exchange rate volatility and various economic factors. We employed machine learning through python to forecast financial indicators.

Keywords: Gold, machine learning, random forest, python, gold price, ETF, dataset

- 1. Introduction:** In addition to other forms of payment, gold was utilized to finance global commerce transactions. Various states kept and increased their gold holdings and were regarded as prosperous and progressive. Our project will let investors and controlling banks determine whether to commit to this commodity [1]. The commodity is known as gold in this country. Several global companies and people have also made investments in gold reserves. Big speculators have also been drawn to this palladium and have invested large sums in it. Using machine learning techniques, we forecast future gold prices based on 22 market indicators. The results suggest that we can properly estimate daily gold rates [2]. The immediate exchange price at which a substance is acquired or sold for immediate payment and transportation is known as the spot price. It differs from the perspective sum, which is the premium at which two parties agree to

trade at a later period. Gold spot prices are set twice daily by the gold market's supply and demand. A fractional shift in the price of gold can result in large profits or losses for these speculators as well as the government's banks. Forecasting the daily increase and fall in gold rates can assist investors in determining whether to purchase (or sell) the asset. Investing in gold has evolved over time, either through traditional methods such as purchasing jewellery, or through modern methods such as purchasing gold coins and nuggets. Historically, gold was utilized as a kind of currency in several parts of the world, including the United States[3]. In recent years, gold has retained its value and has been used to gauge a country's monetary strength. Big investors have also been intrigued in this product and have put large sums into it. Recently, emerging world economies such as China, Russia, et India have been significant consumers of gold, whilst the United States, South Africa, & Australia have been large sellers of this commodity. The value of gold is affected by both Chinese & Indian traditional occasions. More money is being spent on the buying of these things at this time[4]. Small investors recognize this product as a safe investment option rather than alternative investment options that include inherent investment risks. The aforementioned countries' internal monetary conditions have a significant impact on gold spot rates. Gold is another advantage that is being considered a wonderful investment avenue by many investors due to its expanding value and consequently the region of usage[15]. Investors' demand for gold as a safe haven will rise as a result of their pessimistic outlook for the developed exchange markets and thus the capital markets. Gold is also regarded as "the plus of ultimate instance," implying that it is the plus that investors put their trust in when the developed global financial markets are unable to deliver thrilling profitability. As a result, investors use gold as a tool for hedging against swings in other markets[5].

1.1 Gold ETF: A security is now an agreement fund (ETF) that tracks the price of physical gold. These are gold-focused defensive investment instruments that invest in gold bars. Each element of the fund is worth one gram of gold and is backed by genuine gold of the highest quality. Gold ETFs combine simplicity with equity ownership as well as the option of trading in cash. Gold ETFs are listed and traded as securities on the National Stock Exchange of India (NSE) and the Bombay Stock Exchange Ltd. (BSE). Gold ETFs, like any other corporate stock, trade on the cash section of the BSE and NSE and may be bought and transferred at market prices. Purchasing Gold ETFs entails purchasing gold in electronic form. Gold ETFs may be bought and sold in the same way that stocks can. When you reclaim a Gold ETF, you obtain the cash equivalent rather than physical gold. Gold ETF investment is done through a vaporized portfolio (Demat) and dealer, making it an extremely simple way to trade directly in gold.

1.2 Price Forecasting: Price forecasting is the process of determining the amount of a commodity/product/service by considering many aspects such as its presence, pricing, forecasts, the costs of other items (such as gasoline), numerous manufacturer offers, and so on. Price forecasting may be seen in consumer-facing travel services designed to increase client loyalty and involvement, such as Railway line or Hopper. During the same time span, certain firms might still benefit from

information regarding possible expenses. Entrepreneurs must decide the best time to acquire a commodity to adjust the costs of services or products that need a material (timber, chocolate, gold) or assess the investment worth of fixed assets.

1.3 Machine Learning: Machine learning is indeed a data analytics technique that automates the creation of analysis methods. It is a subfield of artificial intelligence based on the concept that computers can learn via data, spot trends, and make judgments with little or no user intercession. Machine learning aims to create systems capable of spotting trends in data, learning through it without human intervention, and clear reprogramming. To answer the price prediction problem, data scientists must first comprehend what data to utilize to train machine learning models, which is why succinct analytics are required.

2. Literature Review: Buyers have been paying close attention in recent years to investments in the gold market due to the possible future rewards. Gold is the only asset that preserves its value throughout political and economic turmoil. Gold values are frequently intimately related to other resources. Because of the unpredictability of market risk, investors utilize future gold price projections as an alarm mechanism. As a result, detailed forecasting of gold prices is required to foresee market patterns.

Gold is the only metal that preserves its value throughout political and economic changes[14]. The gold value unit of measurement is typically closely related to various resources. Due to unpredictability in the market, investors' alarm mechanism is future gold value forecast. As a result, detailed gold price forecasting is required to forecast market patterns[6]. a number of machine-learning approaches for gold forecasting applications Over the last decade, the square measure has been highlighted. The relationship between the price per ounce and the prices of several commodities, particularly petroleum, has also been thoroughly researched. However, the findings of this research on the square measure were found to be contradictory. a lot of studies on the factors determining gold value, as well as different approaches for discovering these relationships[6]. A number of machine-learning approaches for gold forecasting applications Over the last decade, the square measure has been highlighted. The relationship between price per ounce and the prices of several commodities, particularly petroleum, has also been thoroughly researched. However, the findings of this research on the square measure were found to be contradictory. a lot of studies on the factors determining gold value, as well as different approaches for discovering these relationships[7].

3. Related Work: Xiaohui Yang [15] stated that despite increases in 2016 and 2017, the world gold price has been in a slump since 2013. The volatility of gold prices will have a significant impact on people's, businesses, and nations' investment decisions. This inquiry focuses on the World Gold Council's figure for platinum expenses from July 2013 to June 2018 and means to assess and investigate daily gold costs of USD in the first half of July 2018 on the basis of the ARIMA model. This analysis also employs AC, PAC, AIC, and BIC to assess model precision. Exact findings show that ARIMA (3, 1, and 2) seems to be the best method for predicting the USD gold price.

In [16] Mrs. B. Kishori forecasted that Gold is a precious metal that is valuable as a financial resource, ornament, and investment option. Its significant rising costs pique the interest of financial experts as a venture option. In any case, the price of gold fluctuates. It varies consistently for a variety of reasons. The purpose of this research is to calculate the gold value using the ARIMA model. It employs memorable information for judging.

R Hafezi in [17] figured on the number of price fluctuations is a major source of concern in financial markets. As a result, financial professionals must develop a precise and robust determining choice model. Lead representatives use gold as a budget planning switch because gold has shown an outstanding capacity to calm swelling variations. As a result, greater information regarding future gold price patterns will aid in making clear decisions. This research aims to present a clever model based on artificial neural networks (ANNs) to forecast future gold prices. To make ANN fit for the following variations, the suggested intelligent system includes a meta-heuristic computation termed BAT calculation.

In financial markets, the quantity of price variations is a major cause of concern. As a response, financial experts must create a precise and strong decision-making model. Because gold has demonstrated an exceptional ability to control swelling variations, top executives utilize it as a budgetary switch. As a result, having more information about future spot price patterns will help you make better selections. The goal of this study is to offer a sophisticated model for forecasting future gold prices that uses artificial neural networks (ANNs). The proposed intelligent system contains a meta-heuristic computation known as BAT calculation to make ANN suitable for the following changes.

Shian-Chang Huang and Cheng-Feng Wu [18] demonstrated on oil which is an important vitality item. The nonlinearity and does it's of its constituents make determining oil costs difficult. Nonetheless, oil prices are inextricably linked to global financial markets and economic situations, providing us with enough information to forecast them. Conventional models are basic and parametric, and they are not very good at forecasting oil prices. This assessment established a new process to handle these difficulties. Deep (or multi-leveled) different bit learn (DMKL) was used to forecast the oil value having to contact. Traditional insights and AI tactics typically involve shallow models; nonetheless, they can't adequately speak to mind-boggling, compositional, and diverse leveled information highlights. This explains why conventional tactics fail to consider oil value aspects. This investigation aimed to address this issue by combining deep learning and multiple segment machines with data from the oil, gold, and financial system. DMKL is capable of manipulating a variety of data sources. It may correctly distinguish significant data while also selecting a different information portrayal. The DMKL parts were transplanted in a synchronized non-cyclic diagram (DAG), which is a substantial model capable of speaking to complicated and creative information highlights. This provided a robust foundation for extracting the essential highlights of oil value aspects. Our innovative framework outperformed conventional models and drastically reduced estimating errors by using actual data for observational testing.

4. Aims and Objectives of the work: In this study we aim to present an adaptive, flexible, and scalable predictive model by using the characteristics of provided computationally sophisticated neural network simulations to improve the training cognitive development and accelerate convergence. With marginal median square error, the proposed research has the maximum possibility of obtaining higher project rate prediction precision for the studied gold price projection in the market scenario.

In general, this effort is done to recommend appropriate predictor models for efficiently forecasting the deemed gold there in the business world using datasets deemed from their separate databases from prior years.

The goal of this presentation is to correctly estimate the future adjusted current value of Gold ETF hereafter for a set length of time. Supervised Machine Learning was used in this work.

Following are the objectives:

- This study is based on the usefulness of the presented machine learning algorithms, which have shown their efficiency in predicting gold prices with a higher predictive rate.
- To investigate various variables on which the price of gold will be dependent, as well as various Machine Learning Algorithms and Techniques that may be used for value prediction.
- To employ the most suitable Machine Learning Approach
- To disassemble and verify the obtained results

5. Methodology:

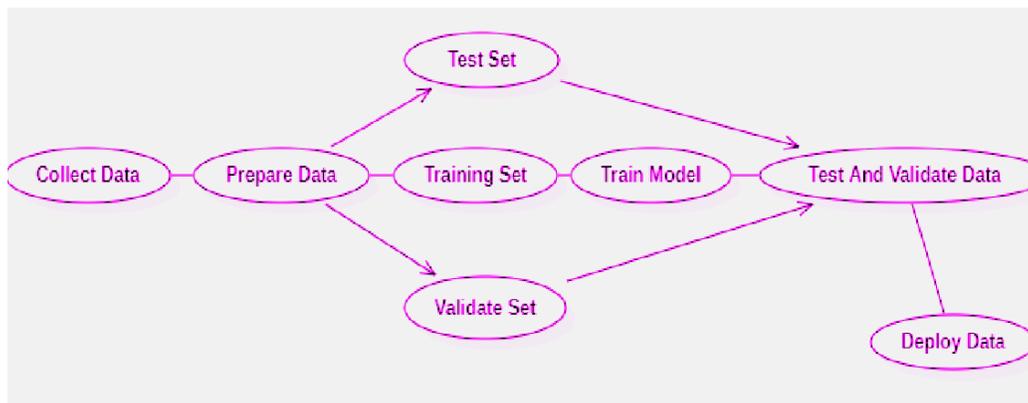


Figure (i): Methodology for price prediction

Figure (i) illustrates the methodology undergone which underlies the data collection from the dataset.

This study's data was gathered from several sources between December 2011 and September 2016 (kaggle.com). Oil worth, NYSE, Standard and Poor's (S&P) 500 index, US Bonds rates (10 years), and EuroUSD market prices were gathered[8]. Data from several sovereign central banks and five massive corporations that have invested massive sums in gold have also been gathered. The analysis also includes the

value of precious metals over this time period. Table I shows the internet sites from which this data was obtained[9].

We employ two-milliliter models, specifically Random Forest and linear regression. Random forest is a Controlled Machine Learning Approach that is commonly utilized in classifying and predicting problems[13]. They include, in combination with the source and destination nodes, one may be more hidden nodes of neurons that attempt to learn non-linear prediction errors that divide completely different groups of information[10].

For anticipating continuous-valued attributes, a regression can be utilized. We employ the RapidMiner tool's application of LR and ANN[11]. Each model is optimized based on the RMSE efficiency live.

This information is then collected, selected, planned, preprocessed, and converted by the specialists. After this stage is completed, the specialists begin developing prognosticative models[12].

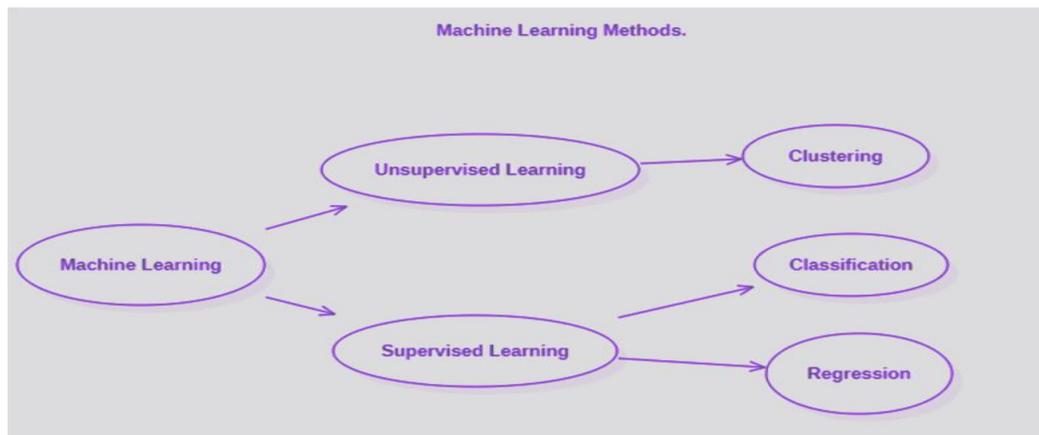


Figure (ii): Machine Learning Methods

Machine learning algorithms are frequently classed as either supervised or unsupervised.

Data Understanding following this work pipeline pertains to the information you have about facts, the demands it will meet, its content, and placement.

When developing machine learning research, we do not always come across clean and prepared data. As a result, we apply the data pre-processing task for this.

Data Pre-processing entails the following steps:

- Obtaining the dataset
- Adding libraries
- Dataset importation
- Identifying Missing Data Encoding Categorical Information

- dividing the dataset into testing and training sets
- Scaling of features

EDA is predominantly a significant method of analyzing data , trends, and patterns, or to validate assumptions using statistical summaries and graphical representations.

Statistical measures mainly fore lies mean, median, mode, and standard deviation. It enables to examination and visualizes data in order to uncover previously hidden patterns. The scientific study requires statistical methodologies. In fact, statistical approaches dominate scientific research since they involve planning, designing, data collection, analysis, meaningful interpretation, and publishing of study findings.

The set of data on which actual training happens is referred to as training data. By fine-tuning the network after each epoch, validation split aids in optimizing model performance. The test set advises us about the model's final accuracy when the training phase is completed.

Model deployment has been the procedure that involves putting a fully functional machine learning algorithm into operations so that it can make data-driven predictions.

6. Results and Discussions:

[]	Date	SPX	GLD	USO	SLV	EUR/USD
0	1/2/2008	1447.160034	84.860001	78.470001	15.180	1.471692
1	1/3/2008	1447.160034	85.570000	78.370003	15.285	1.474491
2	1/4/2008	1411.630005	85.129997	77.309998	15.167	1.475492
3	1/7/2008	1416.180054	84.769997	75.500000	15.053	1.468299
4	1/8/2008	1390.189941	86.779999	76.059998	15.590	1.557099

Figure (iii): Dataset illustration

Figure (iii) signifies a crucial step in data processing where the dataset is explored and studied for further processing.

```
# getting some basic informations about the data
gold_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2290 entries, 0 to 2289
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Date        2290 non-null   object
 1   SPX         2290 non-null   float64
 2   GLD         2290 non-null   float64
 3   USO         2290 non-null   float64
 4   SLV         2290 non-null   float64
 5   EUR/USD     2290 non-null   float64
dtypes: float64(5), object(1)
memory usage: 107.5+ KB
```

Figure (iv): Data understanding

From Figure (iv) it is clear that machine learning analysis of the data uses algorithms to continually develop itself over, but good data is required for these models to function properly and involves accessing and studying data through the use of tables and graphics.

```
# getting the statistical measures of the data
gold_data.describe()
```

	SPX	GLD	USO	SLV	EUR/USD
count	2290.000000	2290.000000	2290.000000	2290.000000	2290.000000
mean	1654.315776	122.732875	31.842221	20.084997	1.283653
std	519.111540	23.283346	19.523517	7.092566	0.131547
min	676.530029	70.000000	7.960000	8.850000	1.039047
25%	1239.874969	109.725000	14.380000	15.570000	1.171313
50%	1551.434998	120.580002	33.869999	17.268500	1.303296
75%	2073.010070	132.840004	37.827501	22.882499	1.369971
max	2872.870117	184.589996	117.480003	47.259998	1.598798

Figure (v): Computation of statistical measures of dataset

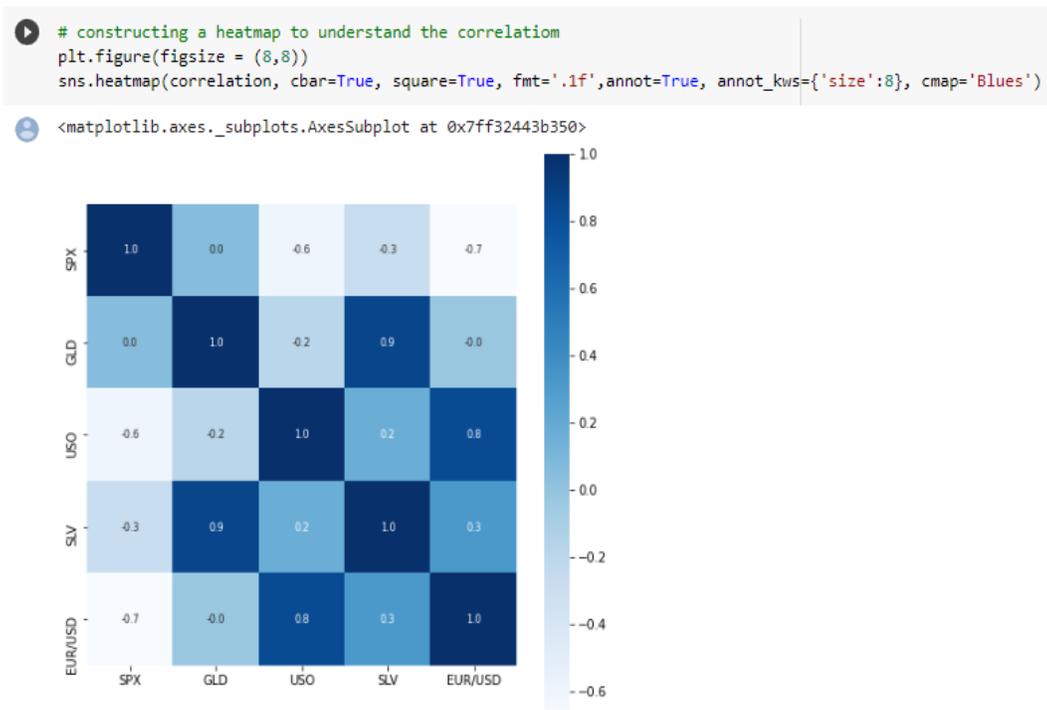


Figure (vi): Heatmap construction

In Figure (vi) a heat map is constructed to represent these values to show the degree of association between variables. It aids in the discovery of features that are optimal for the development of Machine Learning methods. The correlation matrix is converted into color labeling via the heat map.

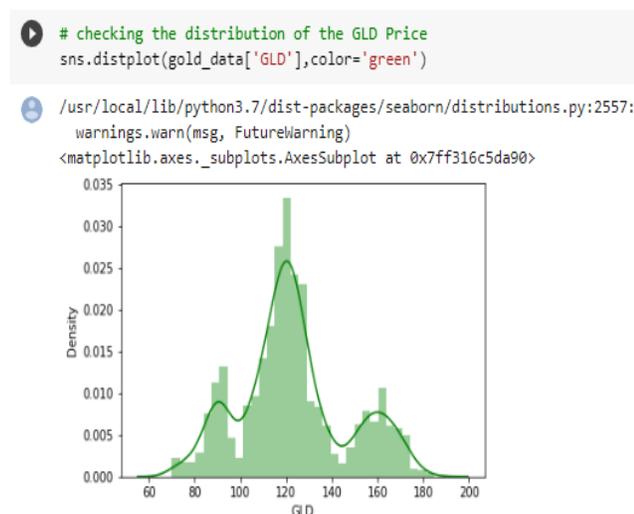


Figure (vii): Distribution representation of price in graphical format

```
[ ] # R squared error
error_score = metrics.r2_score(Y_test, test_data_prediction)
print("R squared error : ", error_score)
```

R squared error : 0.9887338861925125

Figure (viii): Calculation of r squared error

In Figure (viii) r squared error is calculated to represent how near the regression line is to the actual data values (i.e. the projected values shown). The R squared number ranges between 0 and 1, with 0 indicating that the model does not match the given data and 1 indicating that the model fits the data correctly.

Compare the Actual Values and Predicted Values in a Plot

```
[ ] Y_test = list(Y_test)

plt.plot(Y_test, color='blue', label = 'Actual Value')
plt.plot(test_data_prediction, color='green', label='Predicted Value')
plt.title('Actual Price vs Predicted Price')
plt.xlabel('Number of values')
plt.ylabel('GLD Price')
plt.legend()
plt.show()
```

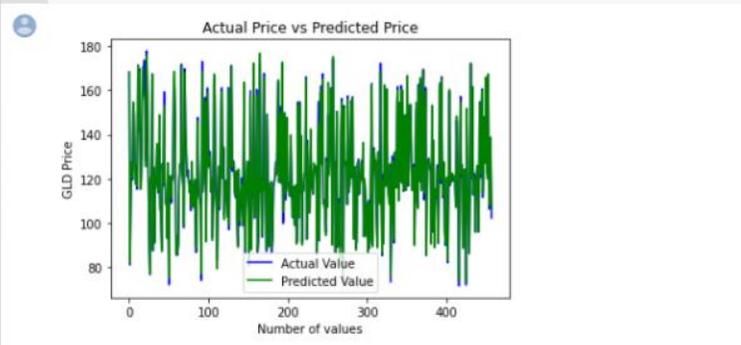


Figure (ix): Price prediction

Figure (ix) gives a graphical result of the predicted price against the actual gold price.

7. Conclusion: This research used machine learning techniques to accurately predict gold prices and when to sell and buy them. This study was conducted to clarify gold ETF stock picks using machine learning in Python. The following inferences can be drawn as per work:

(i) It is discovered that machine learning techniques combined with Random Forest analysis are quite beneficial in predicting gold prices.

(ii) The model's R-square is calculated to be 98 percent. R-squared is often between 0% and 100%. A score close to 100% implies that the supplied model adequately explains the Gold ETF prices.

(iii) The suggested Random forest method-machine learning outperforms traditional and existing forecasting models, according to the results.

References:

- [1] Xiaohui Yang, "The Prediction of Gold Price Using ARIMA Model", 2nd International Conference on Social Science, Public Health and Education 2019.
- [2] Manjula K. A., Karthikeyan P, "Gold Price Prediction using Ensemble-based Machine Learning Techniques", Third International Conference on Trends in Electronics and Informatics, 2019.
- [3] Mrs. B. Kishori 1, V. Preethi, "Gold Price forecasting using ARIMA Model", International Journal of Research, 2018.
- [4] R. Hafezi*, A. N. Akhavan, "Forecasting Gold Price Changes: Application of an Equipped Artificial Neural Network", AUT Journal of Modeling and Simulation, 2018.
- [5] Xiaohui Yang, "The Prediction of Gold Price Using ARIMA Model", 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018).
- [6] Shian-Chang Huang and Cheng-Feng Wu, Energy Commodity Price Forecasting with Deep Multiple Kernel Learning, MDPI Journal, 2018.
- [7] Wedad Ahmed Al-Dhuraibi and Jauhar Ali, "Using Classification Techniques to Predict Gold Price Movement", 4th International Conference on Computer and Technology Applications, 2018
- [8] Iftikharul Sami and KhurumNazirJunejo, "Predicting Future Gold Rates using Machine Learning Approach", International Journal of Advanced Computer Science and Applications, 2017.
- [9] NalinipravaTripathy, "Forecasting Gold Price with Auto Regressive Integrated Moving Average Model", International Journal of Economics and Financial Issues, 2017.

- [10] K. R SekarManav Srinivasan, K. S. Ravichandran, and J. Sethuraman, "Gold Price Estimation Using A Multi-Variable Model", International Conference on Networks & Advances in Computational Technologies, 2017.
- [11] Sima P. Patil, Prof. V. M. Vasava, Prof. G. M. Poddar, " Gold Market Analyzer using Selection based Algorithm", International Journal of Advanced Engineering Research and Science, 2016.
- [12] S. Kumar Chandra, M. Sumathi and S. N. Sivanadam, "Forecasting Gold Prices Based on Extreme Learning Machine", International Journal of Computers Communications & Control, 2016. International Journal of Trend in Scientific Research and Development (IJTSRD) @ www.ijtsrd.com eISSN: 2456-6470 @ IJTSRD | Unique Paper ID – IJTSRD33143 | Volume – 4 | Issue – 5 | July-August 2020 Page 1456
- [13] NurulAsyikin Zainal and ZurianiMustaffa, "Developing A Gold Price Predictive Analysis Using Grey Wolf Optimizer", 2016 IEEE Student Conference on Research and Development, 2016.
- [14] Hossein Mombeini and AbdolrezaYazdaniChamzini, "Modeling Gold Price via Artificial Neural Network", Journal of Economics, Business and Management, 2015.
- [15] ZurianiMustaffa and NurulAsyikin Zainal, "A Literature Review On Gold Price Predictive Techniques", 4th International Conference on Software Engineering and Computer Systems (ICSECS), 2015.
- [16] Xiaohui Yang, "The Prediction of Gold Price Using ARIMA Model", 2nd International Conference on Social Science, Public Health and Education 2019.
- [17] Manjula K. A., Karthikeyan P, "Gold Price Prediction using Ensemble-based Machine Learning Techniques", Third International Conference on Trends in Electronics and Informatics, 2019.
- [18] Mrs. B. Kishori 1, V. Preethi, "Gold Price forecasting using ARIMA Model", International Journal of Research, 2018. [4] R. Hafezi*, A. N. Akhavan, "Forecasting Gold Price Changes: Application of an Equipped Artificial Neural Network", AUT Journal of Modeling and Simulation, 2018.
- [19] Xiaohui Yang, "The Prediction of Gold Price Using ARIMA Model", 2nd International Conference on Social Science, Public Health and Education (SSPHE 2018).