Gold Price Prediction Using Machine Learning

Dinesh Babu H [1], Kali Karthick Reddy[2], Veera Kishore Reddy E[3], Yaswanth G [4],
Dr.P.Rama [5]*

Department Of Computer Science and Engineering, Bharath Institute of Science & Technology affiliated to Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.

*Corresponding authors mail id: ramaponnuvarman@gmail.com

ABSTRACT

Among valuable investments, gold is regarded as timeless. The use and production of yellow metal are what largely drive the gold market in India. Both have a significant impact on the trade balance, unemployment, input to money market profits, and economic value creation. According to PwC report for According to the Precious Metals Board, jewelry has a direct economic impact. On India's economy was projected to be above $30 billion. Around 7% of the country's GDP and 15.71% of its total exports come from the gem and metal sector in all of India's exports of goods, reflects the importance and effect of gold. India has one of the largest gem and jewelry industries in the world, and it accounts for about 29% of global consumption. This industry accounted for 13.30% of all merchandise exports from the nation in FY 2014–15. There is concern that these exorbitant costs will continue to rise and then decline. The association between the price of gold and specific economic factors has been the subject of numerous research, nevertheless. Although it generally focused on projecting stocks rather than commodities, machine learning was widely used to predict financial variables. In this study, we proposed utilizing right progression near grow a guess representative prediction models for calculating the eventual medal's market value.

Keywords: Financial asset, Commodity, Precious Metal, Gross Domestic Product (GDP),

1. INTRODUCTION

Gold is a rare precious metal found in quartz veins and stream gravel, in its pure form. Gold is a commodity with a long history that dates back to Ancient Egypt and is very important to the world economy [1]. Although the metal is valued above and beyond its usefulness, supply, and demand factors influence gold's price. Some investors use gold as a safe-haven asset during recessions or periods of uncertainty, or as a hedge against inflation. High inflation historically has boosted gold's price because investors turn away from fiat money and towards the yellow metal. So, central banks' monetary strategy for containing inflation is crucial in determining gold's price [2]. As a tradable commodity, gold is denominated in the US dollars, which creates an inverse relationship with the greenback. When the US dollar rises against other currencies, gold becomes more expensive, this hurts demand. When USD falls, on the other hand, this boosts the gold's price as the metal becomes cheaper for overseas buyers. Gold is also used to produce jewelry, which is especially popular in China and India - some of the world's biggest buyers - for festivals and weddings [3]. The biggest gold importers in 2021 were Switzerland, India, the UK and China, according to Statista. Gold prices surged to near-record levels in February 2022 passing $2,000 pre-Oz mark as Russia invaded Ukraine and pessimism prevailed in the markets [4]. The bullishness, however, was short-lived as the commodity embarked on a downtrend, falling over 20% by September 2022. The fall was driven by strong US dollar and aggressive interest rate hikes by the US Fed. In late 2022 and the first weeks of 2023, however, the precious metal saw a trend reversal, enjoying a series of higher highs and higher lows. The gold's price rose by 14% since November 2022 (as of 8 February), supported by a less hawkish tone by the Fed's Jerome Powell [5]. Plus, the reopening of China's economy and hence stronger jewelry demand boosted the price at the start of 2023. Everyone's life comprises

493

savings then investments now nearly method. Investing is the process of using present resources with the expectation of a future profit [6]. Inside this financial system, an asset is anything that is purchased for long-term wealth building as opposed to current consumption. A financial Stock is an investment in a financial asset with the hope that it may one day provide income or otherwise be traded at a gain. India has a higher income level that is disposable and a wide variety of buying prospects due to its rapid economic growth [7]. Stocks, savings, commodities, and real estate are a few of the investment options available to investors. Their respective risk and return characteristics vary from one another [8]. Another asset that many investors find appealing is gold, because of its rising value and wide range of applications. Investors are increasingly choosing metal as a protection asset due to their negative predictions for the status of the oversea financial sector and the capital markets [9]. In addition, gold is regarded as "the asset of last resort," or the asset on which investors rely when the developed world's financial markets are unable to deliver the desired profitability. As a result, it might be claimed that investors view gold as a hedge against changes in other markets. Since gold is a precious metal, its price should depend on supply and demand just like it does for other items. The supply of gold has been built up over centuries, but because it can be stored, this year's production has little impact on its prices [10]. Both as a financial asset and as a commodity, gold is used. Compared to long-lived assets like stocks or bonds, gold has fewer characteristics of a commodity [11]. Inflation rates, currency fluctuations, and political unrest are just a few of the many interconnected factors that affect the price of gold. More investors are turning to gold as a desirable investment as a result of the rising price of gold and the volatility and decline in prices of other markets, including the capital and real estate markets [12, 13]. However, as of late, the price of gold is also experiencing tremendous volatility, making gold investments riskier. There is concern over whether these high prices will remain this way and when they will start to decline. Even though the fact that numerous studies have examined the relationship between the price of gold and various economic factors. To understand the dynamic consequences of these interactions, but it is still thought that a study that reveals the influence and impact of different macroeconomic factors on the price of gold in the current environment will be useful [14, 15]. As a result, the purpose of this essay is to investigate the connection between the price of gold and particular economic and market variables. Knowing this relationship can help investors, fund managers, and portfolio managers make better market-related investment decisions, in addition to monetary policymakers [16]. In addition, this study analyses these data using three machine learning algorithms: linear regression, random forest regression, and gradient boosting regression [17]. We can determine the accuracy of these three approaches under various circumstances by comparing them. This essay is organized with a part on literature reviews, then sections on data and technique, findings and discussion, and finally a section on conclusion [18, 19].

2. METHODOLOGY

2.1 Dataset Collection

For such an exercise or any follow-ups, we will want the data points of several number of instruments for the preceding ten years. We may import data using a variety of sites, both commercial (like Reuters and Bloomberg) and unpaid. They were really helpful and straightforward because they allocate resources to a wide range of asset classes, including shares, debt, precious metals, and commodities. I've created an inventory of the instruments that require data import. The list includes the trading icons and their descriptions. We import the file and create separate lists for the respective marks and names.

2.2 Pre-Processing

Emphasized that to order to forecast future returns on Gold, we will use the lag time of the mentioned instruments' returns. Now, we continue by computing the historical returns over the short and long terms for each instrument individually. The basic premise behind it is that there is a higher possibility of portfolio rebalancing if a certain asset has either outperformed or underperformed, which would affect future returns on other asset classes. For

494

instance, asset managers could wish to book gains, Prepare yourself for one share price collapse by moving some funds to, say, precious metals. If stock markets (let's say the S&P500) have produced fantastic returns over the preceding six months. The figure below illustrates the connection between it value of the metal and the S&P 500 index throughout time.

2.3 Feature Extraction

The setup function in PyCaret is the initial step in every modeling activity. The aim classify the information inside the collection are the only required variables here. All fundamental and essential data transformations, such as removing IDs, categorical components from One-Hot Encoding, and outlier detection restoration, take place automatically in the background. Moreover, PyCaret offers more than 20 also before the choices. In this instance, we would stick to the core arrangement and test out different pre-processing techniques in the subsequent trials.

2.4 Model Training

The tune-model () program uses an uneven method look for the collection of settings that will best optimize the selected statistics. (In this example, Require) and displays the again checked value for the modified model. Each method in PyCaret has a predefined grid. A string-passed abbreviation of an estimator is required instead of a trained model, which is not accepted. Linear regression will be tuned.

2.5 Testing Model

As we can see above, linear regression's accuracy considerably improved after tweaking, well outpacing both it and boost, which did not improve. This may be a result of the grid search method being randomized. They may get better after a certain extremely high number of iterations. A base Extra Tree model would also be developed, as its initial performance (before tuning) was quite comparable to tuned linear regression.

2.6 Performance Evaluation

Model diagnostics on trained models should be carried out. For seeing the collection of graphs and other diagnostics, we'll utilize PyCaret's evaluate-model () function. To give a variety of system informative charts & pattern rules requires a trained model. For both of our top models, we would do model diagnostics.
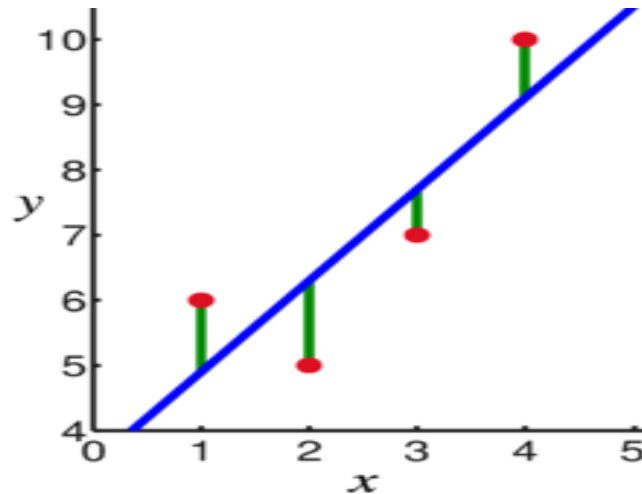
2.7 Prediction

Once we've saved our model, we'd like to forecast fresh data as it comes in. It provides us with the closing prices for each instrument, but to order to apply the model, we must again compile the fresh data. The processes will be the same as those we took to prepare the training data, with the exception that we will import the most recent data and avoid establishing labels (we can't because we don't know future prices). The data should be imported and shaped using the following code chuck to make it ready for prediction.

2.8 ALGORITHM USED
2.8.1 Linear Regression

To estimate the quantity of a variable analysis of direct worsening is used depending on the other variable's value. The parameter is the one you're exasperating to calculation. An "independent variable" is the one you're spending to prediction the cost depending on the extra component. This type of analysis derives the linear model's solutions using one or more variables that may most accurately predict the results for the dependent factor. By perfectly matching a line or ground, the difference between anticipated and real sample points has lessened when employing linear regression. The "least squares" approach is used for the straightforward linear purpose to take tools

495

in graph 1 to find the best line for a collection of data sets. The quantity for X (the reliant on moveable) is then roughly calculated using Y. (the own variable)

Graph 1. Discovering the best-fit line for a set of paired data.

It is possible to create predictions using the simple mathematical method employed in linear regression models. The application of linear regression can be advantageous in several number of business and academic fields. In business as well as the biological, behavioural, environmental, and social sciences, linear regression is often used. With the use of linear regression models, new ways to think can be finally produced with scientific rigor and dependability. As linear retrogression is a statistical method that has been used for a very long period, its characteristics are well recognized and can be learned pretty quickly.

2.9 ARCHITECTURE DIAGRAM:

In Figure 1, it illustrates the two primary data mining techniques, which are classification and prediction. To examine the data and learn more about the unexplored data, we employ these two strategies. Hence, Fig.1 depicts the outcome of using linear regression while changing these 2 elements' values over the test dataset, on the validation set.
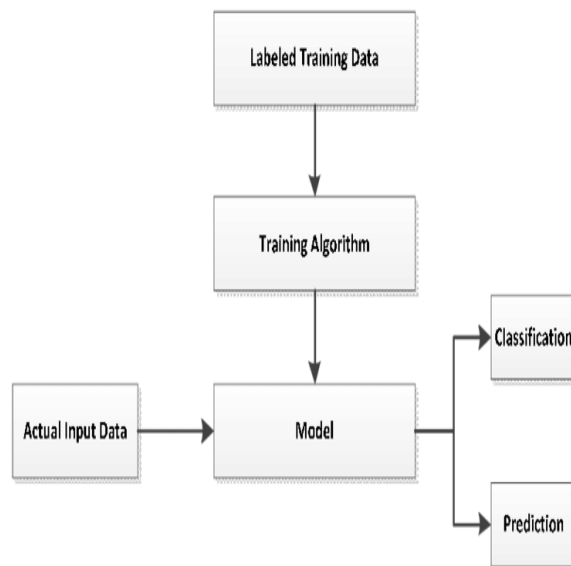
496

Fig 1: Mining of data using training algorithm

Predicting the category of things for whom the class label is unknown is the aim of classification. The process of classification entails creating a good model that adequately describes the various data classes or concepts. Simply put, the classification may be seen as the process of categorizing newly acquired data based on our present or past ideas and the facts we already know. A prediction could be thought of as an event that could happen in the future. Similar to how we utilize the info we possess to make presumptions regarding the future when making forecasts, we are using the info we already possess to figure out or anticipate the absent or inaccessible information for a query point. A continuous value is produced in prediction.

3. RESULT AND DISCUSSION

The impact of the United States money on the determination of unrefined petroleum and medal pricing on the global shop is covered by writers in the Current System. Moreover, they examine the impact of the US currency between 1996 and 2009 on Securities trading, underwriting, Federal Reserve prices, hyperinflation, and the state of the economy. They take into account significant global events that might have influenced US dollar exchange rates. In addition to current happenings, they are also taking various past occurrences into account. Also, they heavily depend on political events. In the proposed work, the user will utilize the search engine to inquire about the Future gold price by providing economic factors. These parameters are preprocessed to make it simpler to locate the parameters keyword that aids in making rapid predictions of the gold price throughout the subsequent phase. Where all computation is postponed until classification and the function is just locally approximated. The best feature of the current system has been determined to be this one. In this project, we are forecasting the price of gold using a linear regression algorithm. The benefit is improved accuracy and simple gold price prediction utilizing economic factors.
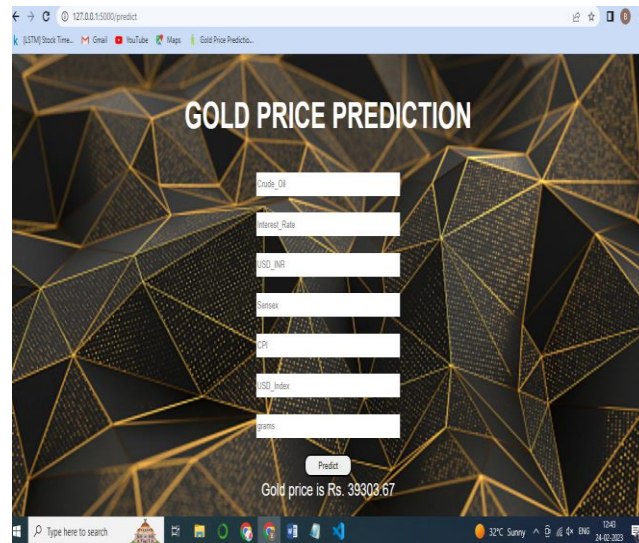
Fig.2 prediction of gold price

Python 3.8 was used to implement the study's experiments. Implementations of deep learning have been made using Keras, and some loss evaluations have been made using Scikit-learn. Time series analysis can forecast data for the next 1, 2, and 30 days in both a single and a multi-step manner. This study has varied tweaked the history and the range of the data to infer from both single and multi-step forecasts. To order to fine-tune the 1 and 2-day forecasts, 5, 9, 15, 22, and 30 history days must be used. For the 30-day prediction models, 45 and 60 historical days must be used. The best model checkpoints were used for each model's 400-epoch training period. Using the suggested linear regression to successfully create the model, we then found the medal value outcomes for the exercise and check collections. To order to fine-tune the 1 and 2-day forecasts, 5, 9, 15, 22, and 30 history days must be used. For the 30-day prediction models, 45 and 60 historical days must be used. The best model checkpoints were used for each model's 400-epoch training period. Using the suggested linear regression to successfully create the model, we then predicted the gold price outcomes for the training and test sets. There are several tuning strictures for linear regression. We tested the most of them, and the two that had the biggest effects on the performance were the amount of films and wisdom level. The four figures match the four training set iterations. Similar to Fig. 2, Fig. 3 shows how LR performs as its ridge parameter is changed. The outcomes are positive. Fig. 5(d) demonstrates that there is just a 19 percent root mean squared error. When training with as few as 920 days of data (i.e., 45–75% of the total data set). Given that the average price of gold in the test data is over $1200, this is a very small inaccuracy (Fig. 2). Although the performance of LR is less than that of linear regression, the difference is not statistically significant. The advantage of LR over linear regression is that it requires less training time. Although it is interesting to observe that both classifiers perform well when using smaller training data (i.e., 45-75 percent exercise set variety), it is typical for ML algorithms to perform better when using greater training data. The reasoning behind this phenomenon is that even though the fact that gold rates have changed dramatically in recent years, the larger training set (i.e., the 0-75% training data version) has records of gold rates dating way back to 2005. For instance, the average price of gold was slightly over 400 in 2005 whereas it was over 1200 in 2016. So, using too much history (going way back to 2005) tends to make the classifiers perform worse. The use of solely recent history is far more advantageous. Just a little more than 2.5 years of information before the data in the test set make up 45-75% of training data. Moreover, Fig. 5(a)-(d) sheds some light on how linear regression functions. When using a smaller training set, linear regression with

498

two layers of neurons works best. On the other hand, a linear model with 5 nerve layers works best with a larger test dataset. This suggests that linear regression with five layers over fits the 45-75% exercise agreed for a smaller training set, and that its comparative act increased by the size of the training set. Equally the scope of the exercise usual was improved, the under fitting in the linear regression with two layers was also reduced. Yet, due to its consistency of performance, linear regression by 3 coatings top spasms the facts.

4. CONCLUSION

Precious metals have always ranked among the most essential sectors. Central banks need to keeps treasuries to order to preserve the current state of the world market. Several important corporations and investors also make substantial investments in gold. Even if it is challenging, being able to predict the gold's value would help investors and national banks achieve wiser purchasing and selling choices, boosting their income. Now this work, we active machine learning procedures near properly prediction medal values. Our analysis, which is the largest to date, takes into account a variety of economic data from different nations and businesses. The store prices of significant medal dealing then production businesses and Russian interest rates have never before been utilized effectively as a predictor of gold rates. Contrarily, we demonstrate that the US economy has less of an impact on gold rates than the stock value of a significant corporation. In the future, we want to use deep learning and ensemble learning to enhance our outcomes. A bigger dataset than the one used in this research will be used in future work to raise the standard of the output and raise the reliability and effectiveness of such metal value forecasts.

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] Sami, I. and Junejo, K.N., 2017, "Predicting Future Gold Rates using Machine Learning Approach," Int. J. Adv. Comput. Sci. Appl., 8(12), pp. 92–99, doi: 10.14569/IJACSA.2017.081213.

[12] Livieris, I.E., Pintelas, E., and Pintelas, P., 2020, "A CNN–LSTM Model for Gold Price Time-Series Forecasting," Neural Comput. Appl., 32(23), pp. 17351–17360, doi: 10.1007/s00521-020-04867-x.

[13] Vidal, A. and Kristjanpoller, W., 2020, "Gold Volatility Prediction using a CNN-LSTM Approach," Expert Syst. Appl., 157, pp. 113481, doi: 10.1016/j.eswa.2020.113481.

[14] Shahid, F., Zameer, A., and Muneeb, M., 2020, "Predictions for COVID-19 with deep learning models of LSTM, GRU and BiLSTM," Chaos, Solitons & Fractals, 140, pp. 110212, doi: 10.1016/j.chaos.2020.110212.

[15] Phi, M., 2018, "Illustrated Guide to LSTM's and GRU's: A step by step explanation," towards data science. https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-sa-step-by-step-explanation-44e9eb85bf21 (accessed Apr. 01, 2021).

[16] Hansun, S., Charles, V., Gherman, T., Subanar, and Indrati, C.R., 2021, "A Tuned Holt-Winters White-Box Model for COVID-19 Prediction," Int. J. Manag. Decis. Make. 20(3), pp. 241-262, doi: 10.1504/IJMDM.2021.10034422.

[17] Hansun, S., Charles, V., Indrati, C.R., and Subanar, 2019, "Revisiting the Holt-Winters' Additive Method for Better Forecasting," Int. J. Enterp. Inf. Syst., 15(2), pp. 43–57, doi: 10.4018/IJEIS.2019040103.

[18] Yahoo! finance, 2021, "Gold". https://finance.yahoo.com/quote/GC%3DF?p=GC%3DF (accessed May 15, 2021).

[19] Moreno, J.J.M., Pol, A.P., Abad, A.S., and Blasco, B.C., 2013, "Using the R-MAPE Index as a Resistant Measure of Forecast Accuracy," Psicothema, 25(4), pp. 500–506, [Online]. Available: http://www.psicothema.com/pdf/4144.pdf.