International Journal of Scientific Modern Research and Technology
(Volume: 2, Issue: 1, Number: 1)
ISSN: 2582-8150

Study of Gold Price Prediction using Machine Learning with Ant Colony Optimization

Karishma Rai[1] and Avinash Sharma[2]
[1]PG Scholar, [2]Head and Associate Professor
[1,2]Dept. of CSE, MITS, Bhopal, India

Abstract- Gold is a metallic instrument and liquid asset too. It has been used as a symbol for purity, value, royalty. This precious metal catches huge attraction towards all stages of people since it is a good investment option. Due to imperfections of expectation exactness, legitimization of gold forecast isn't distinguished reasonably. Regression strategy is excessively intricate, it might produce high blunder rate. Prediction precision may strife to gold forecast in the middle of two periods. Supervised Learning is one of the progressing redesigns of guileless Regression. Regression handles the issue of self-governance by averaging all models created by regular one dependence estimator and is suitable for continuous learning. Regression creates great results stood out from ordinary models. Prediction precision is improve subsequently legitimization of gold forecast is precisely. Regression model develop with CNN-ACO (Convolution Neural Network with Ant Colony Optimization) strategy, result is anything but difficult to upgrade, it create low mistake rate. Prediction precision improve up to 98.76%.

Keywords: Metallic Instrument, Regression, Prediction, Convolution Neural Network, Ant Colony Optimization, Precision.

I. INTRODUCTION

Reserve funds and Investments structure a basic piece of everybody's life. Ventures allude to the work of present assets with a target of acquiring an ideal profit for it in future. In a financial sense, a speculation can be considered as the acquisition of advantages that are not devoured today yet are utilized later on to make riches. In money, a venture is acquisition of a fiscal resource with the possibility that the advantage will give pay later on or will later be sold at a more significant expense for a benefit. The Indian economy being one of the quickest developing on the planet has brought about higher discretionary cash flow level and a plenty of venture roads.

There are various speculation roads accessible for speculators, which incorporates stocks, stores, items and land. Every one of them contrasts as far as hazard and bring attributes back. Gold is another advantage which is being considered as an alluring speculation road by numerous speculators because of its expanding esteem and the zone of utilization. Financial specialist inclination for gold as a defensive resource increments because of their negative desires concerning the circumstance in the created unfamiliar trade markets and the capital markets [1]. Gold is likewise viewed as "the benefit of definite occasion" for example is the benefit financial specialists depend on, when the created world capital markets are not fit to give alluring profit ability [2]. In this manner it tends to be said that financial specialists consider gold to be an apparatus to support against the changes in different markets. Gold is a valuable metal, so like some other products, gold's cost ought to rely upon gracefully and request. In any case, since gold is storable and the gracefully is collected over hundreds of years, the current year's creation has little impact on its costs. Gold is utilized both as an item and as a monetary resource. Gold acts less like a product than enduring resources, for example, stocks or bonds. Cost of gold relies upon a horde of interrelated factors, including swelling rates, cash variance and political turmoil[3]. This raising estimation of gold combined with the volatilities and fall in costs of different markets like capital markets and land markets has pulled in an ever increasing number of speculators towards gold as an alluring venture. Yet, generally cost of gold is additionally seeing high

instability and interests in gold are going to be less secure. There is a dread with respect to whether these significant expenses are maintainable and when the costs would switch.

II. PREVIOUS WORK

There are a few papers which have been contemplated and alluded on my work.

Although, 2016 and 2017 have risen, the worldwide gold expense has been unhappy since 2013. The eccentrics of gold costs will significantly influence the endeavor selections of individuals, attempts and countries. This examination bases on the figure of gold expenses from July 2013 to June 2018 as demonstrated by the World Gold Council, and intends to check and look at step by step gold expense of USD in the central portion of the time of July 2018 through the establishment of ARIMA model. This assessment in like manner uses AC, PAC, AIC, BIC to assess the accuracy of models. Precise outcomes show that ARIMA (3, 1, 2) is the best model to expect the gold expense of USD. The check aftereffects of ARIMA Model are principal for people to grasp the capability of gold expenses and make unfathomable endeavor choices. (Xiaohui Yang; 2019)

This article relies upon an examination prompted grasp the association between gold expense and picked factors influencing it, to be explicit monetary trade, crude petrol esteem, rupee dollar change scale, expanding and credit cost. Month to month esteem data for the time frame January 2000 to December 2018 was used for the assessment. The data was furthermore part into two periods, period I from January 2000 to October 2011 during which the gold worth shows a raising example and period II from November 2011 to December 2018 where the gold expense is demonstrating a level example. Three AI estimations, direct backslide, discretionary forest backslide and tendency boosting backslide were used in looking at these data. It is found that the association between the variables is strong during the time frame I and delicate during period II. While these models show strong match with data during period I, the wellbeing isn't worthy during the time frame II. While sporadic woodlands backslide is found to have better conjecture accuracy for the entire time period, point boosting backslide is found to give better precision for the double cross edges taken autonomously. (Manjula K. A., Karthikeyan P; 2019)

Gold is metal which is critical as monetary asset, decorations, Investment decision. As adventure decision it grabs the interest of monetary experts by its high uplifting costs. Regardless, the gold expense isn't consistent It fluctuates reliably due to various reasons. This paper is planned to calculate the gold worth using ARIMA model. For guaging it uses critical data. (Mrs.B. Kishori 1, V. Preethi; 2018)

III. PROBLEM IDENTIFICATION

The issue conspicuous verification in existing work is according to the accompanying:

- Due to blemishes of assumption precision, legitimization of gold conjecture isn't recognized sensibly.
- Regression system is unreasonably complex, it may create high screw up rate.
- Prediction exactness may hardship to gold estimate in two periods.

IV. OBJECTIVES

The objections dependent on issue unmistakable verification in existing work are according to the accompanying:

- To examine various factors on which gold expense will depends and distinctive Machine Learning Algorithm, Techniques may be used for esteem assumption.
- To apply a best fitting Machine Learning systems.
- To separate and check the gained results

V. METHODOLOGY

Step 1: Convolution process: utilize a trainable channel Fx, deconvolution of the gold price, at that point include a predisposition bx, we can get convolution layer Cx.

Step 2: Update inclination through sigmoid capacity: refresh the estimation of predisposition through straight sigmoid capacity.

 ISSN: 2582-8150

$$f(x)=\begin{cases}0 & \text{if } x \le x_{min} \\ mx+b & \text{if } x_{max} > x > x_{min} \\ 1 & \text{if } x \ge x_{max}\end{cases}$$

Step 3: Optimize through Ant Colony Optimization: The main goal of ACO is to create scientific systems for optimization, and not to create precise technique of nature. The basic scheme of ACO is as below:

3.1 Generate ants and states in image area.

3.2 Select next side probabilistically according to the attractive and visibility.

3.3 Probability is calculate as follows

$$\Pr = \frac{r(e).\eta(e)}{\sum_{available,edges,e'} \tau(e').\eta(e')}$$

3.4 Every ant maintains a register of infeasible transforms for that repetition.

3.5 Modify attractive of a side as per to the number of ants that pass through

The value of pheromone is update as follows

$$\tau(e) := \begin{cases}(1-\rho)\cdot\tau(e), & \text{if edge is not traversed} \\ (1-\rho)\cdot\tau(e) + \text{new pheromone}, & \text{if edge is traversed}\end{cases}$$

Where argument $0 \le \rho \le 1$ is known as evaporation tempo, Basically Pheromones is equal to max-term storage of an ant colony and following conditions is satisfied.

ρ small → min evaporation → slow evaporation

ρ large → max evaporation → fast evaporation

Step 4: Sampling process: n highlights of every area through pooling steps, turn into an element, and after that by scalar weighting Wx + 1 weighted, include predisposition bx + 1, and after that by an actuation work, create a thin n times include outline Sx + 1. Then conclude the gold prediction.

IV. RESULTS AND ANALYSIS

IMPORTING LIBRARIES

Step1:

LinearRegression is a linear regression machine-learning library

```
from sklearn.linear_model importLinearRegression
```

Step2:

pandas and numpy are used to manipulate data

```
import pandas as pd
import numpy as np
```

NumPy stands for 'Python Numeric' or 'Python Numeric.' It is a Python open source module that delivers easy mathematical computation on arrays and matrices. Since arrays and matrices are an integral part of the Machine Learning ecosystem, NumPy completes the Python Machine Learning Ecosystem along with Machine Learning modules such as Scikit-learn, Pandas, Matplotlib, TensorFlow, etc. NumPy offers the basic multidimensional array-oriented features of computing optimized for high-level mathematical functions and scientific computing.

As with NumPy, **Pandas** is one of data science's most commonly used python libraries. It offers high-performance, easy-to-use frameworks, and methods for data analysis. Unlike the NumPy library that provides objects for multi-dimensional arrays, Pandas provides 2d table object named Data frame in memory. It's like a spreadsheet with names in columns and marks in rows. Hence, pandas can have many additional functionalities with 2d tables, such as building pivot tables, computing columns based on other columns, and plotting graphs.

Step 3:

The Matplotlib and seaborn are used for plotting graphs

```
import matplotlib.pyplot as plot
plot.style.use ('seaborn-darkgrid')
```

The Matplotlib and Seaborn are two of Python's best visualization tools. Essentially Seaborn library is based on Matplotlib. Below is a comparison of the two in detail:

Matplotlib: Matplotlib is used mainly for simple plotting. Using Matplotlib visualization usually consists of bars, pies, lines, scatter plots etc.

Seaborn: On the other hand, Seaborn offers a number of forms of visualization. This requires fewer code and has preset templates that are quickly of interest. This is specialized in analysis of information which is used where one needs to outline details in visualizations and even display the data distribution.

Step 4:

Yahoo Finance is used in data processing

Python module to get stock data from Yahoo! Finance

$ pip install yahoo-finance

```
import yfinance as yf
```

Yahoo! Finance is a part of the Yahoo! media property networking. Including stock quotes, press releases, financial reports and original content, it provides financial analysis, statistics and commentary. It also provides several personal financial software resources online.

Then we read daily Gold ETF price data from the past 15 years and store it in Df. We delete columns that are not significant, and use dropna ( ) function to drop NaN values. Then we're calculating a near price for the Gold ETF.

```
# Read data
Df = yf.download ('GLD', '2005-01-01', '2020-08-01', auto_adjust=True)
# only keep close columns
Df = Df[['Close']]
# Drop rows with missing values
Df= Df.dropna ()
# Plot the closing price of GLD
Df.Close.plot (figsize= (12, 8), color='g')
plot.ylabel ("Gold ETF Prices")
plot.title ("Gold ETF Price Series")
plot.show ()
```

Output:

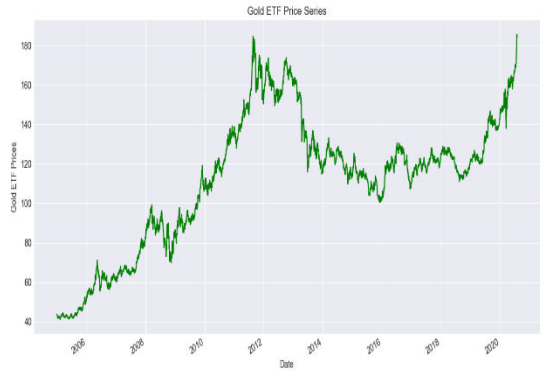

Figure 1: Variation of Gold price

Defining Explanatory and Dependent variable

An explanatory variable is a variable controlled the next day to determine the value of the price of the Gold ETF. Basically, they are the features we want to use to predict the price for the Gold ETF. In this strategy, the explanatory variables are the moving averages of the past 5 days and 10 days. Using the dropna () function, we drop the NaN values, and store the function variables in X. Nevertheless, we add more variables to X by which we think that are useful in predicting the Gold ETF's prices. Such variables may be technical indicators, the price of another ETF such as Gold Miners ETF (GDX) or Oil ETF (USO), or US economic data.

```
# defining explanatory variables
Df ['S_5'] = Df ['Close'].rolling (window=5).mean ()
Df ['S_10'] = Df ['Close'].rolling (window=10).mean ()
Df ['next_day_price'] = Df ['Close'].shift (-1)
Df = Df.dropna ()
X = Df [['S_5', 'S_10']]
```

The dependent variable is equally dependent on the explanatory variables values. Simply put, what we are trying to predict is the Gold ETF price. We store price for the Gold ETF in y.

```
# Define dependent variable
y = Df ['next_day_price']
```

Splitting Data into train and test data set and Creating Linear Regression model

Here, we split the predictors and the data output into train and test data in this stage. Training data is used by combining the input with expected output to construct the linear regression model. The test data are used to determine how well the model was equipped.

First 75% of the data were used for training purposes and remaining data for research.

X _train & y_train Sample Testing

X_test & y_test are collections of test data

```
# splitting the data into train and test dataset
t = .75
t = int (t*len (Df))
# Train dataset
X_train = X [: t]
y_train = y [: t]
# Test dataset
X_test = X [t:]
y_test = y [t:]
```

Now we're building a model of linear regression. So what exactly is linear regression? If we seek to capture a statistical equation between 'x' and 'y' variables that describes the observed values of 'y' in terms of the observed values of 'x' by running a line through a scatter plot, then this is called a linear regression analysis.

 ISSN: 2582-8150

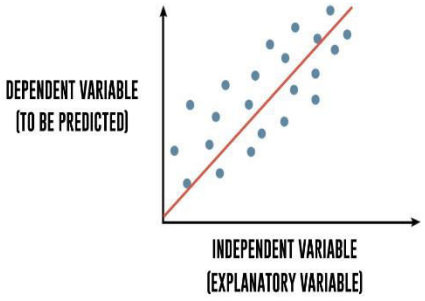

Figure 2: Regression Anaysis

To further break it down, regression in terms of independent variables describes the variance in a dependent variable. The dependent variable - 'y' is the one we want to predict. The independent variables - 'x' are the explanatory variables we use to evaluate the variable based upon. The following equation for regression describes this relation:

$$y = (m_1 \times X_1) + (m_2 \times X_2) + C$$

$$\boldsymbol{Gold\ ETF\ Price} = (\boldsymbol{m_1} \times \boldsymbol{5\ day\ moving\ average}) + (\boldsymbol{m_2} \times \boldsymbol{10\ day\ moving\ average}) + \boldsymbol{C}$$

```
# creating a linear regression model
linear = LinearRegression().fit(X_train, y_train)
print ("Linear Regression model")
print ("Gold ETF Price (y) = %.f * 5 Days Moving Average (x1) \
+ %.f * 10 Days Moving Average (x2) \
+ %.f (constant)" % (linear.coef_[0], linear.coef_[1], linear.intercept_))
```

Output:

Gold ETF Price (y) = 1.4523852334619429260698098 * 5 Days Moving Average (x1) + -0.4546956945100898428435698 * 10 Days Moving Average (x2) + 0.30034053112238723315385 87 (constant)

Predicting the gold ETF prices

Now it's time to verify if the model in the test dataset is working. We forecast prices for the Gold ETF using the linear model generated with the train dataset. For the given explanatory variable X the predict method finds the Gold ETF price (y).

```
# predicting the Gold ETF prices
predicted_price = linear.predict(X_test)
predicted_price = pd.DataFrame(
predicted_price,index=y_test.index,
columns=['price'])
predicted_price.plot(figsize=(12, 8))
y_test.plot()
plot.legend(['predicted_price', 'actual_price'])
plot.ylabel("Gold ETF Price")
plot.show()
```

Output:

Figure 3: The graph above shows the expected and actual Gold ETF price.

Calculation of coefficient of determination or R squared

The determination coefficient, or R squared, is a statistical measure of how close the data pairs are to their matched regression line within a range. This measure ranges from 0 to 1, indicating the extent of predictability of the dependent variable within a data set. The R square of 0 means the independent variable cannot predict the dependent variable, while a R square of 1 means it can be predicted without error.

```
# Coefficient of determination or R squared
rsquare_score = linear.score(X[t:], y[t:])*100
float("{0:.2f}".format(rsquare_score))
```

Output: 98.76%

As we can see, the model's R-square is 98.76 percent. R-squared is usually 0 to 100 per cent. A score close to 100 per cent indicates that the Gold ETF prices are well explained by the given model.

Plotting of Cumulative return and use of this model to predict daily moves

Here we quantify this strategy's cumulative returns for analysis of its results. The steps needed to measure the combined returns are:

i. Generate a percentage change in gold price per day
ii. Build a buy-trading signal represented by "1" when the expected price of the next day

is more than the predicted price of the current day. No other position is taken

iii. Calculate the returns of the plan by comparing with the trading signal the increase in percentage per day.

iv. Ultimately, we'll plot the graph of cumulative returns

```
gold = pd.DataFrame()
gold['price'] = Df[t:]['Close']
gold['predicted_price_next_day'] = predicted_price
gold['actual_price_next_day'] = y_test
gold['gold_returns'] = gold['price'].pct_change().shift(-1)
gold['signal']=np.where(gold.predicted_price_next_day.shift(1)<gold.predicted_price_next_day,1,0)
gold['strategy_returns'] = gold.signal * gold['gold_returns']
((gold['strategy_returns']+1).cumprod()).plot(figsize=(12,10),color='r')
plot.ylabel('Cumulative Returns')
plot.show()
```

Output:

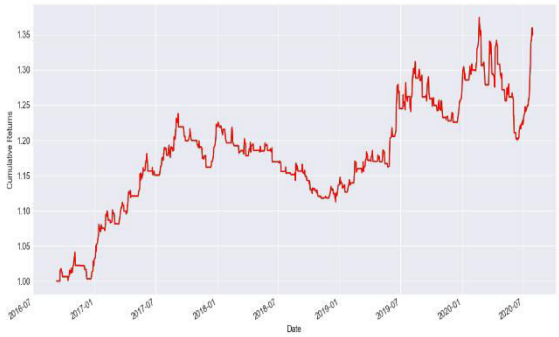

Figure 3: The graph above shows the expected and actual Gold ETF price.

Date	Open	High	Low	Signal	s_5	s_10	Predicted Gold Price
29-07-2020	184.009995	186.14	182.39	**Buy**	183.703	177.409	181.456
30-07-2020	183.440002	184.33	182.13	**Buy**	184.213	178.924	181.931
31-07-2020	184.509995	185.75	184.17	**Buy**	184.773	180.534	181.693

V. CONCLUSION

The new forecasting model implemented on the dataset of gold prices, and the results showed that the prediction model outperforms other benchmark models such as ARIMA, ANN, and ANFIS. The investigated ANFIS model has been modified and optimized recently by Yazdani-Chamzini et al. Thus, it can be claimed that CNN-ACO model based on regression analysis is an accurate model that can be used for gold price forecasting and also other financial markets.

The outcomes on proposed model is as follows:

(1) Prediction accuracy is improve hence justification of gold prediction is accurately.

The Sharpe ratio which Nobel laureate William F. Sharpe introduced in 1966 is a measure for calculating risk-adjusted return. The Sharpe ratio is the average return per unit of uncertainty gained over risk-free cost.

We also calculate the Sharpe ratio using:

```
'Sharpe Ratio %.25f' % (gold['strategy_returns'].mean()/gold['strategy_returns'].std() (252**0.5))
```

Output:

'Sharpe Ratio 0.89945'

We may now use the following application to forecast the gold prices and send a trading signal as to whether we can purchase GLD or not.

```
data = yf.download('GLD', '2005-01-01', '2020-08-01', auto_adjust=True)
data['S_5'] = data['Close'].rolling(window=5).mean()
data['S_10'] = data['Close'].rolling(window=10).mean()
data = data.dropna()
data['predicted_gold_price'] = linear.predict(data[['S_5', 'S_10']])
data['signal']=np.where(data.predicted_gold_price.shift(1)<data.predicted_gold_price,"Buy","No Position")
data.tail(3)
```

Output:

(2) Regression model build up with CNN-ACO method, result is easy to optimize, it generate low error rate.
(3) Prediction accuracy improve upto 98.76%.
As in future, it will be interesting to improvise the prediction capabilities by hybridizing Amoeba Optimization algorithm with others such as Linear Regression method or Neural Network to predict the gold price.

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] Iftikhar ul Sami and Khurum Nazir Junejo, "Predicting Future Gold Rates using Machine Learning Approach", International Journal of Advanced Computer Science and Applications, 2017.
[9] Naliniprava Tripathy, "Forecasting Gold Price with Auto Regressive Integrated Moving Average Model", International Journal of Economics and Financial Issues, 2017.
[10] K.R Sekar Manav Srinivasan, K.S.Ravichandran and J.Sethuraman, "Gold Price Estimation Using A Multi Variable Model", International Conference on Networks & Advances in Computational Technologies, 2017.