*Advances in Parallel Computing Algorithms, Tools and Paradigms*
*D.J. Hemanth et al. (Eds.)*

*© 2022 The authors and IOS Press.*
*This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).*
*doi:10.3233/APC220037*

# A Novel Approach for Correlation Analysis on FBProphet to Forecast Market Gold Rates with Linear Regression

Haari Kishann[a] and L.RamaParvathy[b,1]
[a] *Research Scholar, Dept. of CSE, Saveetha School of Engineering,*
[b] *Professor, Dept. of AI&DS, Saveetha School of Engineering,*
*SIMATS, Chennai, Tamilnadu, India.*

**Abstract:** This work aims to forecast gold prices for future dates using FBprophet and Linear Regression. For predicting the gold price using Linear Regression with a sample size of 140, FBprophet for time series analysis was suggested. The Dickey-Fuller test extracts seasonality (non-stationary) data and converts it to static data. The accuracy of FBProphet is 97.2 percent, compared to 85.6 percent for linear regression. Compared to linear regression, FBProphet tends to do substantially better than linear regression, with a significance level of ($p<0.05$). FBProphet can help predict the percentage of gold rate with greater precision.

**Keywords**. Gold price prediction, Novel FBprophet, Linear Regression, Time series analysis, Univariate Time Series, Machine learning.

## 1. Introduction

In today's financial markets, gold prices are highly volatile. FBprophet and Linear Regression were used to forecast gold prices in the future. The price can change regularly to understand the trends in the data and find patterns in the data. Time series analysis is a statistical method for dealing with time-series data used in this article. Time series data is information organized into a set of fixed time frames or intervals [1]. [2-4] implemented stock price prediction analysis using LSTM in Recurrent Neural Network and Ensemble Learning.

In [5] used analysis to determine the relationship between the gold price and various factors that influence it, such as the stock market, inflation, and interest rate. The best accuracy is found in the stacking performance model in [6]. Gold's price predictability is essential in various fields, including economics, trading, and investing. Many researchers used a decision tree, SVM, KNN, and linear regression to predict the gold price. The K-Nearest Neighbour algorithm is the only one that has shown satisfactory results. However, since our forecast is only based on one aspect, historical gold price movements, this kind of output is still acceptable [4, 7].The price of gold fluctuates regularly, making forecasting the next price difficult. While the gold price

[1]L. Rama Parvathy,Department of AI&DS, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Chennai, India. Email:ramaparvathyl.sse@saveetha.com

cannot be regulated, it can be indicated and predicted to make informed decisions in the future.

## 2. Materials and Methods

The study setting for the proposed work is the Saveetha School of Engineering. The number of groups identified for the study is two. Group 1 is FBprophet, and group 2 is Linear Regression. The FBprophet algorithm and Linear Regression algorithm were iterated ten times with a sample size of 140. The required samples for this analysis are done using G power calculation [8]. The minimum power of the research is fixed at 0.8, and the maximum accepted error is set at 0.5.

From 2009 to 2020, Figure 1 shows the price of gold over the last decade. Dates, price in USD, and price in INR are all part of the attribute information. Also, make sure the data is free of null values. Delete the null values after they've been preprocessed.

**Figure 1.** Gold price prediction dataset description

| | Date | Price USD per Oz | Price INR per Gm |
|---|---|---|---|
| 0 | 2009-01 | 857.726 | 2264.618110 |
| 1 | 2009-02 | 939.763 | 2481.216972 |
| 2 | 2009-03 | 925.989 | 2444.850056 |
| 3 | 2009-04 | 892.663 | 2356.860811 |
| 4 | 2009-05 | 926.855 | 2447.136519 |
| 5 | 2009-06 | 947.807 | 2502.455209 |
| 6 | 2009-07 | 934.272 | 2466.719314 |
| 7 | 2009-08 | 949.500 | 2506.925166 |
| 8 | 2009-09 | 996.443 | 2630.866807 |
| 9 | 2009-10 | 1043.511 | 2755.138480 |
| 10 | 2009-11 | 1126.119 | 2973.244930 |
| 11 | 2009-12 | 1135.012 | 2996.724746 |
| 12 | 2010-01 | 1119.575 | 2955.967081 |
| 13 | 2010-02 | 1095.800 | 2893.194942 |
| 14 | 2010-03 | 1115.554 | 2945.350602 |

### *2.1 FBProphet*

FBProphet must be installed before the model can be trained [7]. Then use the Prophet kit, which is an open-source statistical forecasting algorithm developed by Facebook for users with no technical knowledge of statistics or statistical forecasting [1]. The Prophet employs a decomposable time series model that consists of three key components: trend, variance, and standard deviation. They are combined using equation (1), which can be expressed as the sum of the following factors:

$$(t) = g(t) + s(t) + h(t) + \epsilon t (1)$$

Where g(t) represents the overall growth trend, s(t) represents the annual seasonality, s(t) represents the weekly seasonality, and h(t) represents the holiday effects.The Prophet is frequently presented with a data frame that contains two columns: date and price. The format for the date (datestamp) column Year-Month-Date for dates and Year-Month-Date hours, minutes, and seconds for timestamps. The price column must be numerical and reflects.

***Pseudocode for FBProphet:***
**Input:** Past decades Gold rates
**Output:** Future predicted Gold rates

**Step 1**: Import the FBProphet and any required packages using Matplotlib and Seaborn, then import the dataset using pandas.

**Step 2** The dataset should then be plotted to determine whether or not stationarity exists. Then, fit the data to the model and tweak the parameters such as period, Fourier order, prior scale, and seasonalities.

**Step 3**: Then, to forecast the future price, produce future date stamps. When the time stamps are produced, use the stamps to estimate the time.

**Step 4**: Use matplotlib or seaborn to plot the forecast results.

**Step 5**: Examine the model's efficiency metrics.

*2.2 Linear Regression*

The sklearn module does not accept linear regression as a time series forecasting tool (Scikit library). Linear Regression is the fundamental paradigm for statistical analysis, and it explicitly agrees with univariate and multivariate time series data (one vs. one). Using equation (2), a future price forecast is determined using three new hyperparameters: the explanatory variable or independent variable, line slope, and intercept. Since linear regression is not a time series analysis, it does not conduct a seasonality search on the results.

$$Yi=\beta 0 + \beta 1Xi1 + \beta 2Xi2 + \ldots + \beta pXip + \epsilon i \quad (2)$$

Pseudocode for Linear Regression :

**Input:** Past decades Gold rates

**Output:** Future predicted Gold rates

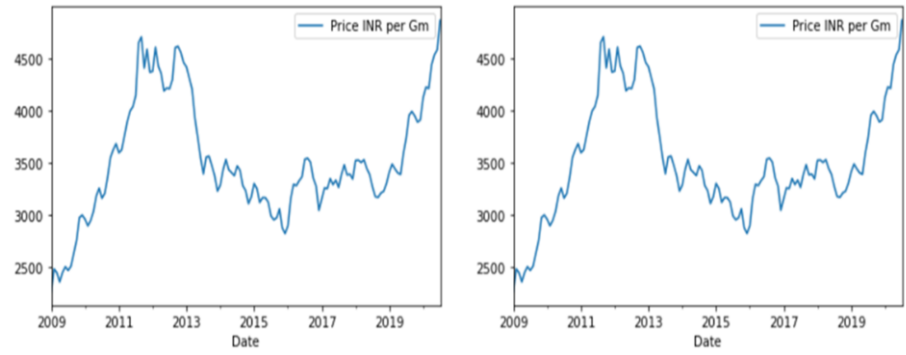
**Step 1**: Import the Linear Regression and necessary packages and import the dataset using pandas

**Step 2:** Then, divide the data collection into train and test data. And then submit the training data to Linear Regression, where the algorithm is trained on the data. After this, a test for the training's precision was performed.

**Step 3:** Then, for forecasting future prices, generate future date stamps. When the time stamps are produced, use the stamps to estimate the time.

**Step 4:** Map the estimated data using matplotlib or seaborn. Examine the model's efficiency metrics.

To validate it, the training of the data is performed. To begin, load the dataset in any format, such as CSV, xlsx, or JSON. Then, imagine the data to look for seasonality. If it exists, delete the seasonality, make it stationary, and feed the training data to the model. The model is then learned to forecast future prices by creating timestamps (future dates). Then, for future forecasts, give this data to the model. Figure 2(a) and (b) depict the input data from 2009 to 2020, regardless of whether it is stationary or non-stationary, using the novel FBProphet and Linear Regression.




(a) (b)

Figure2(a) & (b). Stationary data or non-stationary data using FBProphet and Linear Regression respectively from 2009 to 2020 (X-Axis: Date Y-Axis: Price Per Gram)

### *2.3 Statistical Analysis*

IBM SPSS version 21 was used for the statistical analysis. It's a type of statistical software that's used to analyze data. For both the FBprophet and Linear Regression algorithms, different iterations with a sample size of 20 were performed, and the expected accuracy was recorded for each iteration toanalyze the accuracy. An independent and dependent variables sample T-test was conducted for accuracy. The SPSS software tool was used to calculate standard deviation and mean standard errors.

## 3. Results

**Table 1** shows the minimum and maximum gold price predictions using FBProphet from (2021-3-1) to (2021-7-1). Table 2 shows the forecast projection of gold prices using Linear Regression from (2021-3-1 to 2021-7-1). Figure 3(a) and 2(b) show the forecast values of the gold rate from 2009 to 2022 using FBProphet and Linear Regression.

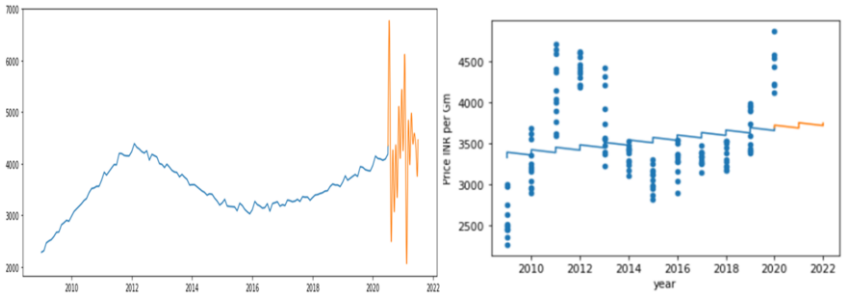
**Table 1.** Prediction of gold prices using FBProphet and Linear Regression from 2021-3-1 to 2021-7-1

| | FBProphet | | | Linear Regression |
|---|---|---|---|---|
| **DATE** | **Lower** | **Upper** | **Predict value** | **Forecast** |
| 2021-3-1 | 3706.718 | 4271.178 | 3982.61 | 3695.03 |
| 2021-4-1 | 4093.899 | 4666.83 | 4387.544 | 3700.91 |
| 2021-5-1 | 4082.85 | 4676.93 | 4382.475 | 3706.78 |
| 2021-6-1 | 4109.224 | 4709.17 | 4410.82 | 3712.66 |
| 2021-7-1 | 4146.521 | 4753.818 | 4459.740 | 3718.53 |

In Table 2, the accuracy table (FBProphet, Linear Regression), the accuracy of the FBProphet algorithm is roughly 95.25 percent and the accuracy of the Linear Regression algorithm is approximately 85.6 percent. The accuracy varies depending on the test size in decimals.

**Table 2.** Accuracy Table (FBProphet, Linear Regression)

| S. No. | Test Size | FBProphet Algorithm | Linear Regression Algorithm |
|---|---|---|---|
| 1 | 0.1 | 95.25 | 64.86 |
| 2 | 0.2 | 95.68 | 77.15 |
| 3 | 0.3 | 96.34 | 82.01 |
| 4 | 0.4 | 97.2 | 80.76 |
| 5 | 0.5 | 95.39 | 80.40 |
| 6 | 0.6 | 94.98 | 82.69 |
| 7 | 0.7 | 95.25 | 84.63 |
| 8 | 0.8 | 95.51 | 84.96 |
| 9 | 0.9 | 95.43 | 84.96 |
| 10 | 1.0 | 95.44 | 85.68 |




**(a)** **(b)**

**Figure 3(a) and (b).** Prediction of Gold rate using FBProphet and Linear Regression respectively

For the FBProphet and Linear Regression algorithms, separate iterations with a sample size of 20 were performed, and the expected accuracy was recorded for each iteration toanalyze the accuracy. Table 3 shows the T-test for Independent Samples findings with a 95% interval of confidence and a margin of 0.05 significance. (FB Prophet tends to do slightly better than linear regression with a value of p =0.000, mean-variance of 0.24931, and standard deviation of 39.47).

**Table 3.** Independent Samples Levene's T-test results

| | F | Sig | t | df | Sig (2tailed) | Mean Diff. | Std. Error Difference | Lower | Upper |
|---|---|---|---|---|---|---|---|---|---|
| Equal Variances assured | 161.6 | 0.0 | 0.006 | 276 | .995 | .249 | 39.4 | -77.46 | 77.95 |
| Equal Variances not assured | | | 0.006 | 151.6 | .995 | .249 | 39.47 | -77.46 | 78.24 |

Table 4 displays the Group Statistics Results. The T-test for FBProphet Standard Error Mean for FBProphet is 38.53, which is better than the Linear Regression of 38.58. Figure 4 compares the accuracy percentage of FBProphet (97.2 %) and Linear Regression (85.6%). In Figure 4, a T-test was used to assess the precision of two

algorithms. The FBProphet model achieved 95.25% accuracy, while Linear Regression achieved 85.6% accuracy. As opposed to other algorithms, the proposed FBProphet technique outperformed Linear Regression.

**Table 4.** Group Statistics Results

| | Clusters | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| **Predicted** | FB Prophet | 139 | 3514.94 | 454.26 | 38.53 |
| | Linear Regression | 139 | 3514.69 | 101.23 | 38.58 |
| **Actual** | FB Prophet | 139 | 3514.69 | 564.52 | 47.88 |
| | Linear Regression | 1`39 | 3514.69 | 564.51 | 47.88 |

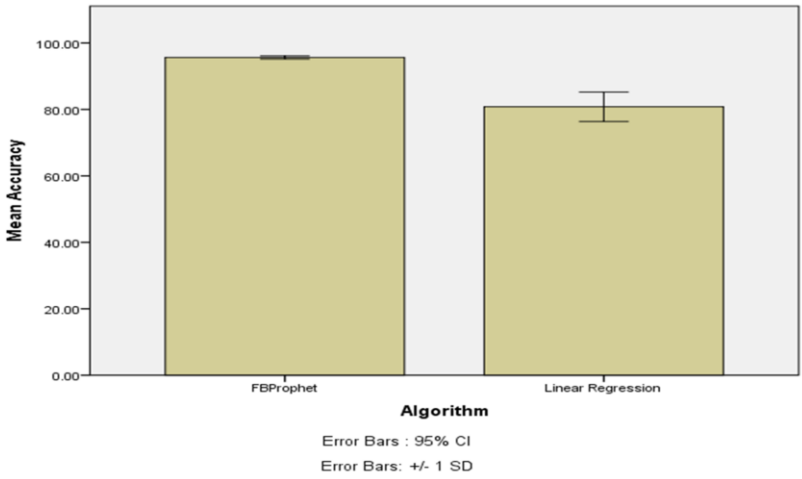



**Figure 4.** Evaluation of mean accuracy of FBProphet and Linear Regression.

## 4 Discussion

In this experiment, it is detected that FBProphet has 97.2% better performance than Linear Regression has 85.6%. Compared to the prediction graph and the results from the SPSS, the FBProphet is better.

In [5], gold price prediction using Ensemble Techniques was predicted for three periods (2000-2018, 2000-2011, and 2011-2018) and the accuracy for three periods was (95,97,62%). It is not stable. In [6], the prediction of gold and silver stock prices using ensemble methods was used to predict the accuracy for gold at 85% and silver at 65%. Using classification techniques to predict gold price movement has an accuracy of 75% [4, 7]. Gold and diamond price prediction using Enhanced Ensemble learning has an accuracy of 95%.

In paper [9], GA-ROSELM, developed in the gold price prediction experiment using ARIMA, SVM, BP, ELM, and OS-ELM, has an accuracy of 13.1%, 22.4%, 53.87%, 57.84%, and 37.72%, respectively. In paper [10], ARIMA, and SVM (Support Vector Machine) models are implemented to predict the gold rate using the performance measurement tools of RMSE and MAPE with an accuracy of 36.2%.

From these two papers, it is evident that the FBProphet has improved accuracy related to ARIMA. As a result, real-time or online data can be used to forecast the future.

## 5. Conclusion

According to the findings of this report, FBProphet outperforms linear regression in forecasting gold rates. Gold rate forecasting is performed with 97.2 percent accuracy in this proposed work using FBProphet.

## References

[1] Zainal NA, Mustaffa Z. A literature review on gold price predictive techniques. *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*. Epub ahead of print 2015. DOI: 10.1109/icsecs.2015.7333120.

[2] Christina C, Umbara RF. Gold price prediction using type-2 neuro-fuzzy modeling and ARIMA. *2015 3rd International Conference on Information and Communication Technology (ICoICT)*. Epub ahead of print 2015. DOI: 10.1109/icoict.2015.7231435.

[3] Hussein SFM, Shah MBN, Jalal MRA, et al. Gold price prediction using radial basis function neural network. *2011 Fourth International Conference on Modeling, Simulation and Applied Optimization*. Epub ahead of print 2011. DOI: 10.1109/icmsao.2011.5775457.

[4] Pandey AC, Misra S, Saxena M. Gold and Diamond Price Prediction Using Enhanced Ensemble Learning. *2019 Twelfth International Conference on Contemporary Computing (IC3)*. Epub ahead of print 2019. DOI: 10.1109/ic3.2019.8844910.

[5] Manjula KA, Karthikeyan P. Gold Price Prediction using Ensemble based Machine Learning Techniques. *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. Epub ahead of print 2019. DOI: 10.1109/icoei.2019.8862557.

[6] Mahato PK, Attar V. Prediction of gold and silver stock price using ensemble models. *2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*. Epub ahead of print 2014. DOI: 10.1109/icaetr.2014.7012821.

[7] Al-Dhuraibi WA, Ali J. Using classification techniques to predict gold price movement. *2018 4th International Conference on Computer and Technology Applications (ICCTA)*. Epub ahead of print 2018. DOI: 10.1109/cata.2018.8398669.

[8] Kane SP, Phar, BCPS. Sample Size Calculator, https://clincalc.com/stats/samplesize.aspx (accessed 23 March 2021).

[9] Weng F, Chen Y, Wang Z, et al. Gold price forecasting research based on an improved online extreme learning machine algorithm. *Journal of Ambient Intelligence and Humanized Computing* 2020; 11: 4101–4111.

[10] Makala D, Li Z. Prediction of gold price with ARIMA and SVM. *Journal of Physics: Conference Series* 2021; 1767: 012022.