**Gold Rate Historical Data Analysis**

S.Vasudevan1

1Department of Mathematics, School of Applied Sciences, REVA University, Bangalore-560064, Karnataka, India.

Email : svasudevanidvr@gmail.com

**Abstract**

The most popular investment option in India is gold, and it is also one of the biggest consumers of the metal, accounting for over 25% of global demand. Particularly in times of economic and financial crisis, gold is viewed as a secure investment. Gold is a tool used by investors to protect their portfolio investments. In addition to fluctuations in interest and exchange rates, government policies, and other factors, gold prices fluctuate due to monetary and political unpredictability.

In this case, the arising of the gold was rapidly increasing. Analysing the gold price according to the previous state with respect to year wise.

**Keywords**: Excel, Dataset, Python, Machine Learning Algorithms

**Introduction**

In 2020, the coronavirus epidemic completely changed the gold market. Supplies were impacted by the temporary shutdown of refineries and mines, while demand was harmed by social isolation, curfews, and lockdowns. When other investment possibilities appeared uncertain during the pandemic, demand increased.

The COVID-19 pandemic helped to start the year off on a bright note, but the price of gold displayed a changing tendency throughout 2020. Gold became more in demand as a result of investors using the precious metal as a haven, which also raised the price. When the price of gold somewhat dropped at the end of 2020, the equity market began to rebound after the pandemic.

**Preliminaries**

India's gold prices in 2022 have been erratic. The cost of gold has increased considerably since 2021. The price of the yellow metal rose by around Rs. 3,000 over the first six months of the year, an increase of nearly 6.5 percent. The war between Russia and Ukraine, the rise in US Fed interest rates, and inflation have all contributed to rising gold prices. Since the year's commencement, the equities market has fallen due to the rise in gold demand.

**Dataset**

It's crucial to remember that gold prices fluctuate during the year, and the sum mentioned here represents the average price for that period. Except maybe a few low points in between certain years. The gold price trend in the table shows that it has traditionally been on the rise, supporting the claim that gold is a secure investment over the long term.

|  |  |  |
| --- | --- | --- |
|  | **Year** | Price |
| **0** | 1964 | 63.25 |
| **1** | 1965 | 71.75 |
| **2** | 1966 | 83.75 |
| **3** | 1967 | 102.5 |
| **4** | 1968 | 162 |

Table: 1 Head of the Dataset.

**Cor-relation Analysis**

We conducted a correlation analysis between two attributes we obtained are strongly associated to gold price. The outcome of the correlation analysis is displayed in Fig. 3. It offers some insightful information r = 0.82. where, the Year and Price should be Perfectly Corelated. While comparing with other it strongly and highly corelated with respect to the dataset.

To Find the co-relation anlysis, we used scatted plot, heatmap etc…This should be Positively corelated. It also be shown in the scatter plot analysis.

|  |  |  |
| --- | --- | --- |
|  | **Year** | **Price** |
| **year** | 1 | 0.82 |
| **Price** | 0.82 | 1 |

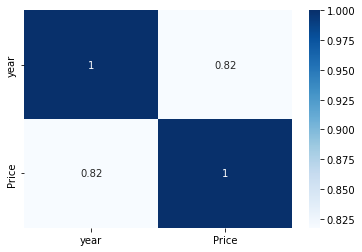
Table: 2 Relation Analysis of Year & Price

Figure 1 Co-relation Analysis on heat map

Pair plots are used to determine the most distinct clusters or the best combination of features to describe a connection between two variables.

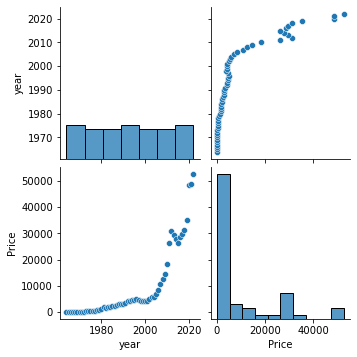
****

Figure 2 Pair plot Analysis

**Machine Learning Models**

As is well known, machines are capable of learning from data in order to gain experience and make future predictions. Supervised, Unsupervised, and Reinforcement learning were the three divisions. To carry out the particular duty, the categories were divided into various models. Given that the dataset is supervised (labelled). The rest is an independent variable, with the year being the dependent. Here, we train and evaluate the dataset for estimating price with respect to the year in the future.

**Regressor Analysis.**

1. **Linear Regressor Analysis**

A variable's value can be predicted using linear regression analysis based on the value of another variable. Regression uses independent variables to model a goal prediction value. It is mostly used to determine how variables and forecasting relate to one another. Regression models vary according to the number of independent variables they use and the type of relationship they take into account between the dependent and independent variables.

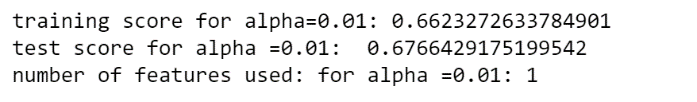
As the thing, the mean square error = 58396481.115

The mean absolute error = 6043.826

And the Root mean squared error = 7641.75

1. **Regularization**

we Introduced the Regularization analysis to overcome. In order to minimize the adjusted loss function and prevent overfitting or underfitting.



Sweet Spot

Overfitting

Bias

Underfitting

Variance

When our machine learning model is unable to recognise the data's underlying trend, underfitting occurs (i.e, high Bias). When a deep neural network or machine learning model performs noticeably better on training data than on test data, it is said to be overfit (i.e high Variance). So the model is not learning and flexible in the case of over and under fitting. This should be called as Dropout Layer. Our motive is to find midway or sweet spot (Balanced spot). To overcome that, by introducing the Ridge and Lasso Regression.

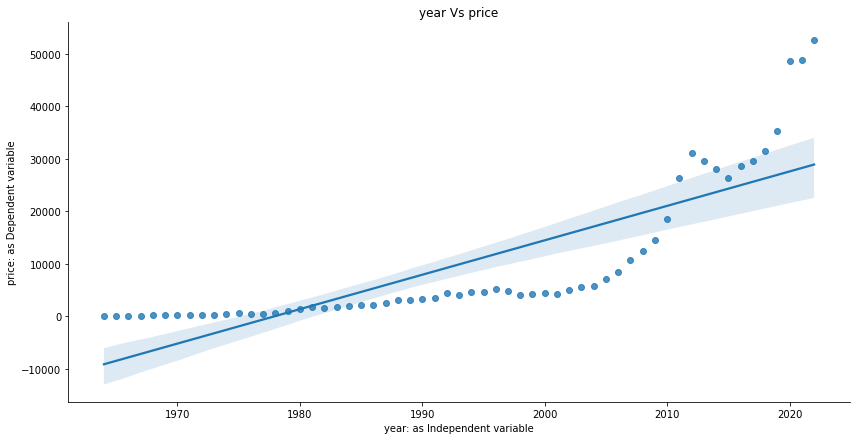
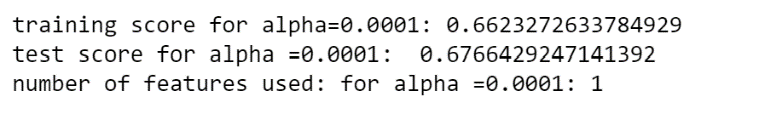


Figure 3 Regression Analysis

1. Ridge & Lasso

By the usage of Lasso Regression, the learning rate as zero. So the analysis of the gold still in the low accuracy of 0.67 that shown in Fig: 4. As I increasing the learning rate or alpha value as 0.01,0.0001 and so on. Still the accuracy is on 0.67 and 0.68 that can be shown in the Fig:5, Fig: 6. To improve my machine we moved to Decision tree Regressor.

Figure 4

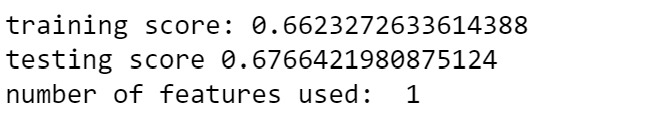
Figure 5

Figure 6

1. Decision tree Regressor

Decision tree regression trains a model in the form of a tree to predict data in the future and generate useful continuous output by observing the properties of an item. Continuous output denotes non-discrete output or results. But the dropout layer occurs to improve the accuracy again we introduce the Regularization concept like ensemble techniques.

1. Essemble Techniques

Methods that build several models and then combine them to deliver better outcomes. Bagging, Boosting and Stacking are the three ways. I am implementing Boosting concept, that split into three categories.

1. Ada Boost
2. Gradient Boost
3. XG-Boost.

As we can perform in Ada Boost and XGBoost Regressor, but for my convenience and to achieve a better result, I perform on Gradient Boost Regressor. Below, I go into more information regarding gradient and xgboost regressor.

1. **Gradient Boost**

A versatile non-parametric statistical learning method for regression and classification is gradient boosting. As the data set was randomly mixed for 30% of testing and 70% of training. 99 percent of the data that we gave for training and testing the machine were effectively learned from. Accuracy was attained satisfactorily.

Step 1:

Initialize the model with constant value

Step 2:



Step 3:

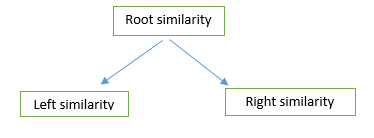
Calculate:

Step 4:

1. XG Boost

It provides the basic model () based on the average of the independent attributes. Determine the residual (y-). Build a decision tree (sequently according to features, i.e., it may be regressor or categorical)

1. To find. Similarity weight (root and terminal node)
2. Calculate Gain (split-wise)

= Similarly of left + similarly of right –root similarity

1. Similarly, split-wise can be done in a n-number of ways.
2. Compare the split which is higher

Calculate Residual , repeat the procedure until we find the output.

1. we want to prune(cut the tree)
2. we want to not prune

As a result Fig: 7, the Gradient Boosting Classifier by default uses a loss of ‘ls’, a learning rate of 0.1, a maximum depth of 3, a number of estimators of 500, and a random state of 0 and so on

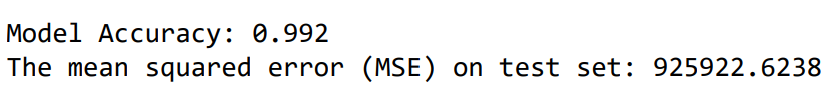


Figure 7

Conclusion

As per the dataset, the machine well trained with the help of algorithms successfully and the future predictions also done well .