

Blacks can be found in almost all continents of the world. In some countries, they constitute a large section of the minority like in America. African Americans have contributed greatly to inventions. We will need to take a critical look into that aspect since this article only acknowledges such contributions and goes no further than that.

REFERENCES

1. “Discovery of the Oldest Adornments in the World”, (2007), Science-daily.com, Retrieved August 2012.
2. C.S Henshilwood, et al, (2011), “A 100,000-year-old ochre-processing workshop at Blombos Cave, South Africa, - PubMed”, Science, 334 (6053): 219 -22,
3. Ehret, Christopher (2002), “The Civilization of Africa”, Charlottesville: University of Virginia Press, pp. 22, ISBN 0-8139-2085-X.
4. “Civilizations in Africa: The Iron Age South of the Sahara”, (2007), Washington state University.
5. https://upload.wikimedia.org/wikipedia/commons/f/fl/Africa_map_regions.svg
6. Woods, “G. Science in Ancient Egypt” (1988).
7. Van Sertima, I “The Lost Sciences of Africa: An Overview.” Blacks in Science: Ancient and Modern, 7 – 26, (1983).
8. Durant Will, (1950), “The story of Civilization IV: The Age of Faith, Simon and Schuster, New York, pg. 330 – 31.
9. Asante, M. et al. “Great Zimbabwe: An Ancient African City-State.” Blacks in Science: Ancient and Modern, 84–91 (1983).

Zhang Weijia (China)

National Research Tomsk Polytechnic University, Tomsk

Scientific adviser: Gubin Evgeny Ivanovich, associate professor

SIMPLE MACHINE LEARNING MODEL ABOUT PREDICTING THE PRICE OF GOLD

Introduction

Gold is a symbol of wealth and luxury. Compared to currency, gold is more valuable and is largely immune to inflation. In today's volatile situation, gold has a high investment value.

The goal of this work is to explore the secrets of the price of gold.

To achieve the goal of this work, the following tasks were set:

1. Preparing and cleaning data.
2. Choose the suitable machine learning model to predict the price of gold.
3. Establishing a prediction model for gold prices.

To achieve the objectives and goals in this work, we use the pandas library for working with handle data. In addition, we also use linear model library for esablishing machine learning model and train test split library to split the data into train data and test data and matplotlib for plotting graphs. In this work we use the gold price data from London Bullion Market Association [1].

Data preparation

First, we checked the gold price dataset. Columns include, USD(AM), USD(PM), GBP (AM), GBP (PM),

EURO (AM), EURO (PM). These data represent the price of gold denominated in different currencies on the same date. We choose the price of gold denominated in US dollars and keep the date.

The US dollar was chosen because the US dollar is still the most widely used international currency, and in the data set, the dollar-denominated gold price data is more comprehensive.

We find that the price of gold denominated in US dollars is divided into USD(AM) and USD(PM). We choose USD (AM) as our analysis data, because USD (AM) contains a larger amount of data, and a larger amount of data can make the prediction model is more accurate.

We can notice that USD(AM) data has 1 Nan value. We replace that value by average of column.

Finally, we create the data set we need, which performed on the figure 1, and drawn the gold price chart from 1938 to 2020, which performed on the figure 2.

	Date	USD (AM)
0	2020-03-20	1504.45
1	2020-03-19	1480.70
2	2020-03-18	1506.00
3	2020-03-17	1472.35
4	2020-03-16	1504.65
...
13195	1968-01-08	35.14
13196	1968-01-05	35.14
13197	1968-01-04	35.14
13198	1968-01-03	35.16
13199	1968-01-02	35.18

Fig. 1. The table of gold price USD(AM)

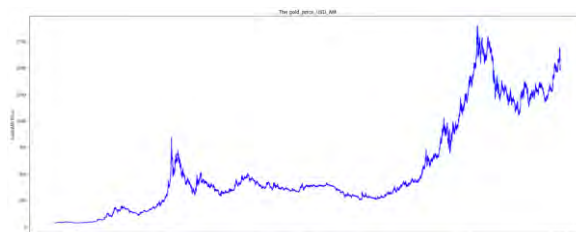


Fig. 2. The Graphics of gold price USD(AM)

Choose the suitable machine learning model

The price of gold is linear, as can be seen from figure 2, and this work predicts the price of gold, which is a numerical value, so linear regression is the most suitable machine learning model. Choosing the right model can help us solve the problem better.

Establishing a prediction model for gold prices

In order to establish a linear regression model for gold price prediction, the independent variable (X) and the dependent variable (Y) need to be determined. The price of gold on a certain day is naturally the dependent variable(Y), so what is the independent variable(X)?

The price of gold is a continuous line, as can be seen in Figure 2, so the current price of gold is related to the price of gold in the past, and the future price of gold can be predicted from the price of gold in the past.

Based on this idea Take the mean value of the gold price for the past three days(B_3) and the mean value of the gold price for the past nine days(B_9) as the dependent variable (X) , which performed on the figure 3.

	B_3	B_9
9	1652.183333	1556.983333
10	1665.500000	1576.111111
11	1673.666667	1599.033333
12	1670.350000	1614.750000
13	1659.750000	1633.911111
...
13195	35.153333	35.177778
13196	35.143333	35.172222
13197	35.140000	35.165556
13198	35.140000	35.158889
13199	35.146667	35.155556

Fig. 3. The table of dependent variable (X)

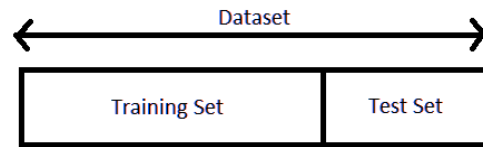


Fig. 4. Train Set and Test Set split

Next, need to split the training data set and the test data set. By training on a subset of data, and testing on a different subset of data that the learning algorithm has never seen, ensure that the learner regression model is actually finding real patterns in the data and not just memorizing it[2].

We usually split the data around 20%-80% between testing and training stages[3], which performed on the figure 4. In this work, we split a dataset into a training data (80%) and test data (20%).

Train the linear regression model through the training data, and get the final model expression as: Gold(AM) Price = 1.18 * Mean value Of the gold price for the past 3 days - 0.18 * Mean value Of the gold price for the past 9 days + 0.02.

Use the dependent variable (X) in the test data set which performed on the figure 5, and the linear model created from the training data set to predict the gold price. Compare the predicted gold price and the real gold price with a graphic as on the figure 6.

	Date	USD (AM)	B_3	B_9
6905	1992-11-24	333.95	334.466667	334.900000
11442	1974-12-13	178.40	186.500000	189.666667
9167	1983-12-13	389.75	386.533333	380.822222
13139	1968-04-10	37.60	38.050000	38.038889
321	2018-12-12	1244.75	1240.900000	1251.672222
...
1098	2015-11-17	1080.80	1076.250000	1070.344444
5438	1998-09-15	290.35	289.216667	290.777778
11466	1974-11-11	182.50	182.833333	185.227778
7099	1992-02-19	352.35	351.766667	351.133333
5749	1997-06-24	338.65	337.966667	332.138889

Fig. 5. The table of test data set

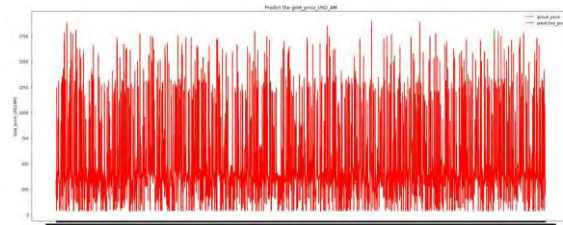


Fig. 6. The Comparison between

The predicted gold price and the real gold price

From Figure 5 and Figure 6, we can see that the data is irregularly arranged, because the test set and training set are divided randomly, which is obviously not easy to observe. Rearrange the test dataset by date, and a new comparison graph is obtained as shown in figure 7.

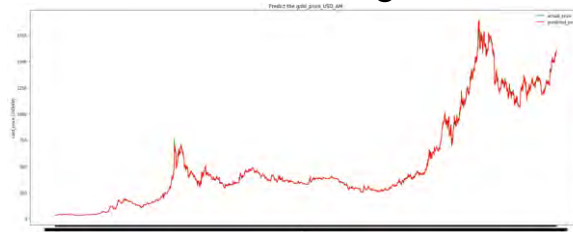


Fig. 7. The Comparison of gold price after reordering

From figure 7, we can observe that the red line (predicted price) and green line (actual price) overlap to a high degree, which proves that the obtained gold price prediction model is very accurate.

Finally, the accuracy of the model was checked, and the accuracy of the model obtained through the program was 99.95%, which proved that the gold prediction model was very successful.

Conclusion

Through this project, we can draw the following conclusions.

1. Carefully process the data, which directly affects the accuracy of the project.
2. Understanding the problem and choosing the right machine learning model can greatly improve the success rate of model building.
3. Understand the relationship between different columns of data in the data set. Sometimes the data set includes independent variables (X) and dependent variables (Y). Sometimes there are no independent variables (X) in the data set. At this time, how to deal with the dependent variable and how to derive the dependent variable from the data set is very important.

In this project, we accurately grasped the above three factors and obtained a gold price prediction model with an accuracy of 99.95%, which proved that the gold price is based on the past gold price and is predictable.

REFERENCES

1. Quandl/ 52 years of Gold price // URL: <https://www.quandl.com/data/LBMA-London-Bullion-Market-Association> – (accessed 20.03.2020).

2. In machine learning, what's the purpose of splitting data up into test sets and training sets? // URL: <https://www.quora.com/In-machine-learning-what-s-the-purpose-of-splitting-data-up-into-test-sets-and-training-sets>
3. Train and Test Set in Python Machine Learning – How to Split // URL: <https://data-flair.training/blogs/train-test-set-in-python-ml/>