



# PREDICTION OF GOLD PRICE MOVEMENT CONSIDERING THE NUMBER OF INFECTED WITH THE COVID 19

Jovana Stokanović Sević,  
Ana Jovancai Stakić\*

Singidunum University,  
Belgrade, Serbia

## Abstract:

The beginning of the health and economic crisis caused by the appearance of the new Covid-19 virus showed us that, in the perception of investors, gold is still highly valued as a guardian of value. This paper aims to test several models and select the best one for predicting the price of gold on the world market for the next day, in five and ten days, taking into account the number of cases and deaths from the Covid-19 virus. We believe that predictions with Covid-19 parameters give more accurate results than predictions that take only historical gold prices as information. These predictions can help decision-makers whether, at what point, and in what amount, it is best to invest in gold and gold-related financial instruments, relative to the projected price of gold from the model. The paper tests models called Decision tree, K-nearest neighbours, Linear regression model, and Support vector machine based on the information on gold prices and the number of cases and deaths from the Covid-19 virus. It will be seen in the paper that even models with only information on the price of gold give quite reliable predictions, but in unstable times like this, models that take into account the instability factor give more accurate predictions. The research is aiming to determine the optimal amount of information on which the models will "learn" to give the most accurate possible result. This work's data processing and models are done in the Python programming language.

## Keywords:

Gold price, COVID 19, Decision tree, K-nearest neighbours, Linear regression model, Support vector machine.

## INTRODUCTION

Throughout the history and development of economic systems, various goods and materials have played the role of money. At the time of barter, goods were exchanged for goods, later it was possible to exchange goods of greater value for more goods of lesser value. Eventually, a system was established where one type of goods became the general equivalent of payment, accepted by all participants. For a large part of history, the function of money was performed by gold. That gold was in the form of coins of precisely determined quality, shape and weight, and in return, it was assigned a value. Of course, bigger, heavier coins, with higher purity had a higher value and thus in return, they could get larger quantities of goods and services.

## Correspondence:

Ana Jovancai Stakic

## e-mail:

ajovancaistakic@singidunum.ac.rs



Gold has always been considered a custodian of value and its basic function is to preserve purchasing power in times of great uncertainty. In the last few years, we have been experiencing a period of great uncertainty regarding the health situation, which further leads to uncertainty in the financial, market of goods and services, as well as in all other fields. A new virus called SARS-CoV-2 has appeared in 2019 in China as an infectious disease caused by severe acute respiratory syndrome. The World Health Organization (WHO) has declared the Coronavirus outbreak of 2019/20 a pandemic and public health threat of international importance. Evidence of local disease transmission has been found in several countries, in all six WHO regions.

Although the death rate from the virus was quite low at the very beginning, it is increasing over time, and the number of patients is growing. A pandemic was soon declared and thus began instability in the financial, but also in all other markets. This is best seen through the significant jump in the price of gold. Yousef and Shehadeh [1] explain this situation quite well, where they prove that there is a correlation between the number of patients and the jump in the value of the price of gold. That is why the research problem in this paper is the prediction of the price of gold in conditions of uncertainty, such as the appearance of the COVID-19 virus pandemic. There are many methods in the literature that make predictions of the price of gold based on historical data. These methods, although effective, lose precision when major market disruptions occur. Therefore, the topic of this paper is to prove that such models have less accuracy than models that include historical data on the cause of market disruptions, which in this case is a pandemic of the COVID-19 virus. This paper starts from the assumption that investors need to be able to follow the trend of gold prices, and this paper relatively accurately (over 90% accuracy) follows models that give predictions for tomorrow, five days, and ten days in the future. In this way, investors can determine at what point and in what amount they can invest and thus get the most value. Also, this kind of research helps those who operate in short-term markets like the Forex market where things change quickly. Such analyses can help them make the greatest possible profit in the shortest possible time. Predictions are of great importance for financial decision-makers. The data within the models used in this paper can easily be replaced by data on any other financial instrument in the event of any similar crisis that will almost certainly occur in the future. We believe that this is the greatest contribution of this paper.

## 2. BACKGROUND AND LITERATURE REVIEW

Concerning the importance of the gold price in the overall economic environment, predicting the price of gold is very significant, and different studies and models have been used for this purpose. In some eminent research, classical econometric methods were used for this prediction [2] [3]. Different techniques for the gold price prediction, were used and although various models give very good results, the ARIMA (autoregressive integrated moving average) model is the most precise of all traditional statistical models [4] [5] and it is good to use a sliding dataset for the prediction [6]. By comparing the models on the same data set for profit prediction it can be concluded that the choice of the dataset is very important and that parameters unrelated to gold can help in a better prediction [7].

In some prominent research, the artificial neural network model has been used for modeling the gold price and compared with the traditional statistical model of ARIMA. The three performance measures, the coefficient of determination ( $R^2$ ), root mean squared error (RMSE), and mean absolute error (MAE), are utilized to evaluate the performances of different models developed. The results show that the ANN model outperforms the ARIMA model, in terms of different performance criteria during the training and validation phases [8] [9].

Machine learning has often been applied to the prediction of financial variables, but usually with a focus on stock prediction rather than commodities [10]. The application of machine learning in trading with financial instruments has shown very good results [11] as well as predicting gold price direction [12] especially by using a decision tree algorithm and support vector [13].

## 3. DATA

The time frame of the data set is dictated by the COVID-19 data set frame. Virus data and gold data are merged into one data set. For the models to be able to predict such a union of data, all N/A values have been dropped, and data on the value of gold will start at the beginning of 2020, which is from the first of January. Also, the stock exchanges are closed from Friday from 4 pm to Monday at 8 am, and there is no weekend information.



A set of data containing information about gold was used from the Python library Yahoo! Finance. The following values were taken directly for prediction:

- ◆ Close - the value indicates the final price of gold on the stock exchange for a given day.
- ◆ High - the highest daily price of gold.
- ◆ Low - the lowest daily price of gold.
- ◆ Volume - daily turnover, number of transactions for a given day.

Based on the close value, two columns were added, which represent the average price of gold in the previous three and nine days. This way of creating a model is based on the idea of Shah and Pachanekar [14]. This way of processing data is called moving average, it is good for predictions of any period and it is very easy to present graphically. When we take into account 3 days, this method is called simple moving average and is used as a significant indicator among brokers because it gives equal importance to all three prices and thus shows the price trend.

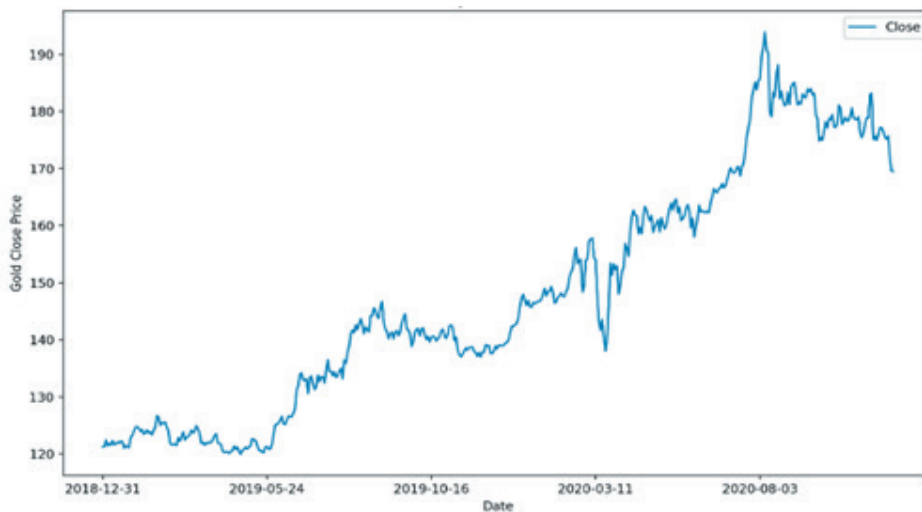


Figure 1 – Gold price from the beginning of 2020

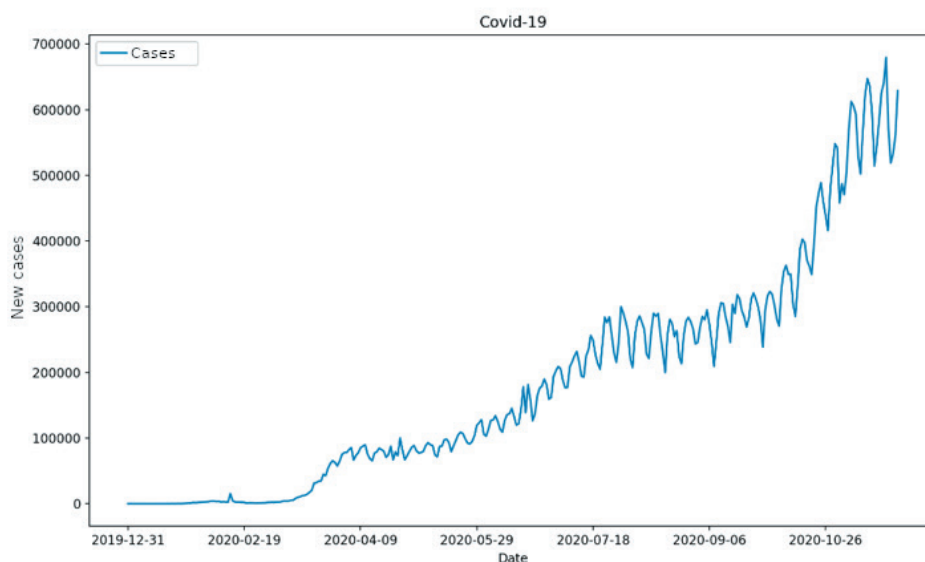


Figure 2 – The number of people infected with the COVID-19 virus in 2020

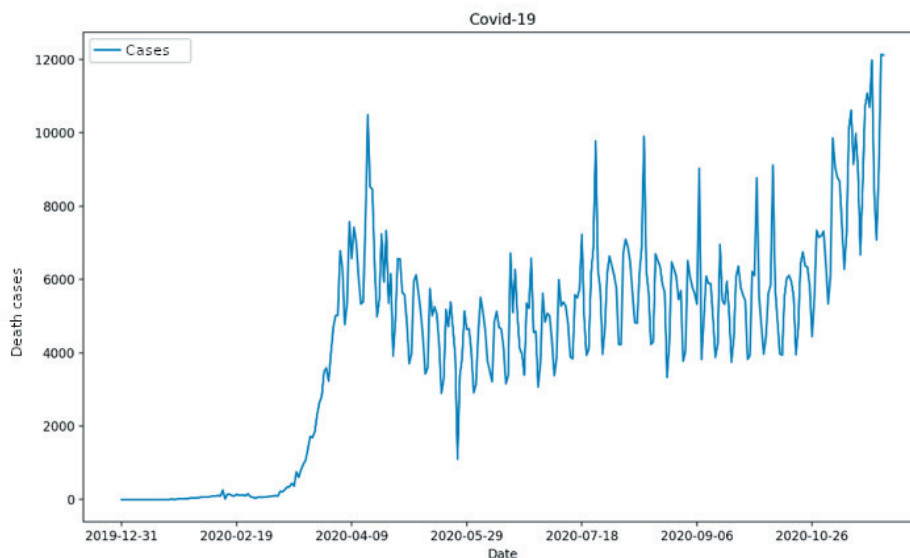


Figure 3 – Number of deaths from Covid-19 virus in 2020

#### 4. MODELS USED FOR THE RESEARCH

The paper uses four popular models of machine learning that apply the following algorithms: linear regression, the decision tree, K-nearest neighbour, A support vector machine.

**Linear regression** tries to show the values of the dependent variable in the most accurate way possible about the independent variable with a linear function. This is a common way of predicting the value of financial instruments, especially if the values are inert.

**The decision tree** is one of the best and most commonly used classification algorithms because, in addition to offering high prediction accuracy and clarity, it also easily maps nonlinear relationships. This algorithm easily solves regression and classification problems.

**K-nearest neighbour (kNN)** can be used for classification and regression problems. It is a model that classifies data points based on the points that are most similar to it. It uses test data to "learn how to guess" what to classify in an unclassified point. More specifically, one might wish to weight the evidence of a neighbor close to an unclassified observation more heavily than the evidence of another neighbor who is at a greater distance from the unclassified observation [15].

**A support vector machine (SVM)** is a model reminiscent of a more advanced version of linear regression. This model presents data as points in space that it classifies into two categories between which there is a gap. The SVM efficiently constructs linear or nonlinear classifica-

tion boundaries and is able to yield a sparse solution through the so-called support vectors, that is, through those observations that are either not perfectly classified or are on the classification boundary [16]. For the model not to be a linear regression model, the so-called kernel trick is used, which implies observing individual zones, and not the whole set.

#### 5. RESULTS

This research intends to predict the price of gold for one, five, and ten days in advance by using the models of linear regression, decision tree, K-nearest neighbour, and support vector machine. We attempted to see whether one model gives relatively accurate predictions of these prices and which. The algorithms did not change in the testing itself, but the intention is to get the best result through changes in the variables. To begin with, it was important to find the model that makes the most accurate prediction for the next day, and later on to predict the gold price five and ten days in advance.

Due to the specific nature of machine learning models, testing and result representation are adequately adjusted, with multiple repeating independent runs taking place during testing, and results showing statistical results of multiple iterations. Additionally, various parameter settings were tested in search of optimal performance.



	Linear regression		Decision tree		K-nearest neighbor		Supp. Vector machine	
	Gold	Gold and Covid-19	Gold	Gold and Covid-19	Gold	Gold and Covid-19	Gold	Gold and Covid-19
Explained_variance_score	-0.6391	-0.6295	0.9616	0.9627	0.9386	0.9387	0.9307	0.9306
Max_error	143.1799	143.1799	10.7200	10.7200	13.4921	13.4722	11.3136	11.3128
Mean_absolute_error	6.6431	6.5635	1.8054	1.8054	2.4035	2.4021	2.5986	2.5994
Mean_squared_error	274.9183	273.2479	6.4203	6.2792	10.7402	10.7300	12.1284	12.1423
Mean_squared_log_error	0.1156	0.1155	0.0002	0.0002	0.0003	0.0003	0.0004	0.0004
Median_absolute_error	2.8622	2.8233	1.3300	1.2699	1.8060	1.8060	2.0386	2.0307
r2_score	<b>0.6412</b>	-0.6313	0.9616	<b>0.9625</b>	0.9358	0.9359	0.9275	0.9275

Table 1 - Results of gold price prediction models with the model with the greatest accuracy

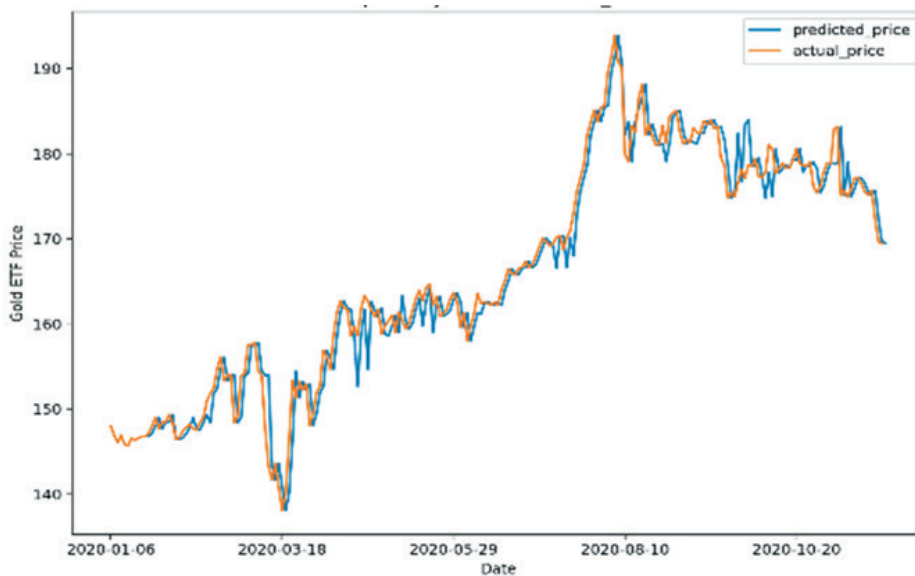


Figure 4 – Graphic view of most accurate model, decision tree with gold and Covid data

The results tell us that the best model for predicting the price of gold for tomorrow is the decision tree model, which, in addition to the historical price of gold, also contains data on the number of cases and deaths from the Covid-19 virus with 96.25% prediction accuracy. The model is trained on seven-day data, taking into account the historical price of gold and the average number of Coronavirus cases and deaths in the last seven days and that is the only model shown in this paper. The second-best model is still the decision tree, which contains only information on historical gold prices. This iteration is the first that, in the case of the support vector machine model, managed to equalize the accuracy of the prediction in the case of data with and without information about Covid.

Additionally, the same model is giving us the most accurate results for predictions five and ten days in advance with an accuracy of more than 95%.

## 6. CONCLUSION

Linear regression tries to show the values of the dependent variable as accurately as possible in relation to the independent variable with a linear function. Many research papers deal with the prediction of the price of gold with the help of linear regression, which can be explained by the relatively stable price of gold in the past, but due to the instability caused by the Covid 19 virus, this model does not give nearly accurate predictions. In fact, of all the four models tested, this model can be



called the least applicable for this type of prediction. We attribute this to the fact that the data do not follow any linear trend and are too "scattered" for this type of prediction to give meaningful results.

The k-nearest neighbour model shows the highest accuracy of the prediction in the case of predictions for the next five days, and only on the basis of historical data on the price of gold. The maximum error of the model is always quite similar for both the model with historical gold prices and the model with the Covid virus.

The support vector machine model in almost every iteration shows more precise results when the prediction is made only on the basis of the historical price of gold. Similar to the k-nearest neighbour model, the accuracy of this model is almost always over 80%. The maximum error of this model is almost the same regardless of what data we take into account, and very often it is smaller than the model that has higher levels of accuracy.

In each iteration and in almost every case, except one, the accuracy of the decision tree model is the highest compared to other tested models. The highest accuracy, of 96.25%, is given in the model trained on a seven-day data set, with information on the seven-day average number of patients and deaths from the Covid-19 virus where the prediction was made for the next day. It is believed that this level of precision is very successful in conditions of great imbalance.

Based on all the above, it is believed that the hypothesis that information on the number of infected and infected with the Covid-19 virus helps to create models with greater accuracy in determining the price of gold than those with only historical values of gold prices has been proven.

Future research could focus on adjusting the parameters of algorithms for creating models and testing other models. It is also believed that such models can be applied not only to the situation with gold and the Covid-19 virus but also to determine the price of any financial instrument at any time in a crisis if appropriate quantifiers of the crisis situation are inserted. Given that health crises are becoming more frequent in the last few decades, it is believed that this work will have wide applications in the future.

## 7. REFERENCES

- [1] I. Yousef, E. Shehadeh "The impact of COVID-19 on Gold Price Volatility," *International Journal of Economics and Business Administration*, vol. VIII, no. 4, pp. 353-364, 2020.
- [2] S. Shafiee, E. Topal, "An Overview of Global Gold Market and Gold Price Forecasting," *Resources Policy*, no. 35, p. 178-189, 2010.
- [3] G. Aye, R. Gupta, S. Hammoudeh, W. Joong Kim, "Forecasting the Price of Gold Using Dynamic Model Averaging," *International Review of Financial Analysis*, vol. 41, p. 257-66, 2015.
- [4] A. Azzuti, *Forecasting Gold Price A Comparative Study*, 2016.
- [5] X. Yang, "The prediction of gold price using ARIMA model," in *2nd International Conference on Social Science, Public Health and Education (SSPHE 2018)*, Atlantis Press, 2019.
- [6] J. Brownlee, "Machine Learning Mastery," 2020. [Online]. Available: <https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>. [Accessed 12 6 2020].
- [7] M. Riazuddin, "Towards Data Science," [Online]. Available: <https://towardsdatascience.com/machine-learning-to-predict-gold-price-returns-4bdb0506b132>. [Accessed 20 4 2020].
- [8] H. Mombeini, A. Chamzini, "Modeling Gold Price via Artificial Neural Network," *Journal of Economics, Business and Management*, vol. 3, pp. 699-703., 2015.
- [9] U. Hong, N. Majid, "Comparison of ARIMA model and artificial neural network in forecasting gold price," *Journal of Quality Measurement and Analysis*, vol. 2, no. 17, pp. 31-39, 2021.
- [10] P. Megan, "Stanford," 2013. [Online]. Available: <http://cs229.stanford.edu/proj2013/Potoski-PredictingGoldPrices.pdf>. [Accessed 12 1 2022].
- [11] J. Test, M. Broker, *Machine learning for algorithm trading: Master as a PRO applied artificial intelligence and Python for predict systematic strategies for option and stocks.*, Independently published, 2020.
- [12] S. Perry, "Predicting Gold and Silver Price Direction Using Tree-Based Classifiers," *Journal of Risk and Financial Manag.*, vol. 5, no. 14, p. 198, 2021.
- [13] D. Navin, "Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector," *International Journal of Science and Research*, vol. 4, no. 3, pp. 2026-2030, 2015.



- [14] I. Shah, R. Panchekar, "Quantisti," [Online]. Available: <https://blog.quantinsti.com/gold-price-prediction-using-machine-learning-python/>. [Accessed 8 7 2020].
- [15] S. A. Dudani, "The Distance-Weighted k-Nearest-Neighbor Rule," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, no. 4, pp. 325 - 327, 1976.
- [16] G. Van den Burg, P. Groenen, "GenSVM: A Generalized Multiclass Support Vector Machine," *Journal of Machine Learning Research*, vol. 17, no. 224, pp. 1-42, 2016.