InComTech: Jurnal Telekomunikasi dan Komputer
vol.15, no.01, April 2025, 56-65
http://publikasi.mercubuana.ac.id/index.php/Incomtech
P-ISSN: 2085-4811 E-ISSN: 2579-6089

Gold Price Prediction using Linear Regression and Polynomial Regression Method

Ardi Azhari*, Naeli Umniati

Information System Management, Gunadarma University,
Jl. Margonda Raya 100, Depok 16424, Indonesia
*Email Penulis Koresponden: ardiazhari@gmail.com

Abstract
Gold is a precious metal with high value from mining prices and usage. His ability to purchase power and gold was often used as an investment tool in protecting value and wealth (protector of value and wealth). The high demand for gold from year to year is increasing yearly. So, it is essential to have some knowledge in predicting gold prices which aims to find out the opportunities for gold prices that will happen in future. The purpose of this study is to predict the price of gold to find out the price of gold in the coming year. Predictions are made using linear regression and polynomial regression methods, which are then compared to get the best model for the dataset. The dataset used 5703 data from gold prices from 2000-2022.

Keywords:
Gold prices;
Linear Regression;
Prediction;
Polynomial Regression;

Riwayat Artikel:
Received May 22, 2024
Revised Desember 15, 2024
Accepted April 09, 2025

DOI:
10.22441/incomtech.v15i1.27311

1. INTRODUCTION

Gold is currently in great demand from investors because it is considered profitable. The movement of gold investment tends to be more stable, and its value increases. Gold is a long-term investment, unlike investing in currency exchange or stocks [1]. In addition, gold is a form of liquid investment, meaning it can be accepted in any region or country and is easy to withdraw quickly. Investment can be defined as deferring present consumption for use in efficient production over a certain period. Meanwhile community investment activities will continuously increase economic activity and employment opportunities, national income, and community prosperity. This role comes from three essential functions of investment activities [2].

Investment can be interpreted as an activity that aims to develop assets to obtain profits in the future. One of them is a gold investment, according to the opinion of [3]. Investment is one way to gain profits in the future. Gold is a precious metal in great demand by investors because its price tends to rise, besides being easy to

liquidate or having an elevated level of liquidity. Investments in gold, whether in the form of gold coins, gold bars, or gold jewelry, have higher yields when compared to deposits. Moreover, they are much more stable when compared to stocks [4] [5]. According to financial experts, gold investment is second after businesses in the real sector that are doing well. Investment in gold is a prediction because the price of gold can change at any time [6]. These price changes force investors to be able to make decisions quickly and accurately. For that, we need a prediction of the price of gold. Several previous studies discuss gold price predictions, gold price predictions using the Average-Based Fuzzy Time Series method, which produces the best error value of 3.4216 % using 750 to 2500 training data. The amount of training data to predict with the Average-Based Fuzzy Time Series has no significant effect on prediction. The amount of training data that is too much is less effective for making predictions [7]. Univariate Convolutional Neural Network (CNN) can be used to predict time series data, in this case, namely the price of gold. Univariate CNN for predicting using predictive data suitable for predicting only a few values in the future. Proper optimization of univariate CNN models is needed to get reliable results, for example, Multi-Step CNN Models and Vector Output Models, which allow predictions several steps ahead [8]. Subsequent research uses a fourth-order polynomial regression in predicting gold price by [9] with the results R2 = 0.9217, MSE = 80549907, and RMSE = 8975 . Prediction is a process of systematically estimating something that is most likely to happen in the future based on past and present information that is owned so that the error (the difference between something that happened and the estimated result) can be minimized [10]. Predictions do not have to give a definite answer to what happened but try to find answers as close as possible to what will happen. Investment is a commitment to several funds or other resources made at this time to obtain several benefits in the future [11]. The term investment can be related to various kinds of activities. Investing funds in the real sector (land, gold, machinery, or buildings) or financial assets (deposits, stocks, or bonds) is standard.

Machine learning is a series of techniques that can assist in handling and predicting extensive data by representing these data with learning algorithms [12]. Machine learning is a computational method that uses experience to improve performance or make accurate predictions [13]. Here, knowledge refers to past information available to the learner, usually in electronic data, and can be analyzed. The regression model is a series of statistical processes used to estimate or predict the target or dependent variable based on the independent variables. Regression models have many variants, such as linear regression, ridge regression, stepwise regression, polynomial regression, etc. [14]. Regression is a statistical technique used for numerical analysis between input variable X and output variable Y, where the input variable is the independent variable, and the output variable is the dependent variable [15]. An equation can be considered linear if the relationship between the independent and dependent variables is linear. Otherwise, it is non-linear.

Several methods can be used in making predictions, namely Linear Regression and Polynomial Regression. The dataset is 5703 (2000-2022), sourced from the Kaggle data set. In this study, the authors used the Linear Regression and Polynomial Regression methods to forecast gold prices because the Linear Regression method produces a slight error of 3.79%. The results of the predictions

will be evaluated to determine which is the best by looking at the results of the coefficient of determination (R^2), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). The coefficient of determination can be interpreted as the proportion of the variance in the dependent variable that is predictable from the independent variables [16]. MSE is another performance metric for regression models, having a very intuitive interpretation in terms of relative error. MSE aids in the efficient convergence to minima for tiny mistakes as the gradient gradually decreases [17].

2. RESEARCH METHOD

The following is a schematic method used in this study.

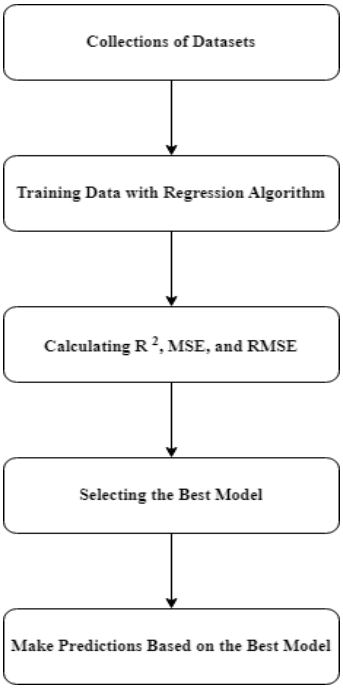

Figure 1. Research Methods Scheme

Figure 1. outlines the research method scheme conducted in this study. The first stage of this research is to collect or obtain data sets to make predictions. Then the data is imported or loaded into the application. This research aims to predict the price of gold with the regression method. Data training was conducted on the regression models used, namely linear and polynomial, to get reliable results. The training results will produce a predicted value of n data. The prediction results are then evaluated to determine the best model with the R^2, MSE, and RMSE methods. After the best model is obtained, the final stage is predicting the price of gold in the future.

The dataset used in this study is a collection of daily gold price data from January 4, 2000, to September 2, 2022. Data obtained from Daily gold Price Historical Data of the Kaggle site. The data was accessed in October 2022 at the link https://www.kaggle.com/datasets/psycon/daily-gold-price-historical-data.

The data set is a collection of daily gold prices from 2000-2022, totaling 5703. The data set contains the daily gold price in ounces in United States Dollars (USD). Data sets are stored using Microsoft Excel in .csv.

In this study, the regression process in Python applications, such as training data, predicting, and evaluating models, is conducted using the scikit-learn machine learning library.

The next step is to decide and create a regression model that will be used in forecasting. The models to be used in forecasting are linear regression models and polynomial regression models. The following is the model used:

1. Linear regression
2. Second-order polynomial regression
3. Third-order polynomial regression
4. Fourth-order polynomial regression

The following process is training and predicting after deciding and making the model in the program. Table 1. is a table used as a reference for predicting linear regression where X is a date, where the value of X for 1 February 2022 is 1, 2 February 2022 is 2, and so on, and Y is the price of gold.

Table 1. Reference data for predicting linear regression manually.

No.	X	Y	X^2	Y^2	XY
1	1	1798.9	1	3236041.21	1798.9
2	2	1800.8	4	3242880.64	3601.6
3	3	1807.4	9	3266694.76	5422.2
4	4	1805.4	16	3259469.16	7221.6
5	5	1808	25	3268864	9040
6	6	1821.5	36	3317862.25	10929
7	7	1827	49	3337929	12789
8	8	1834	64	3363556	14672
9	9	1827	81	3337929	16443
10	10	1862.2	100	3467788.84	18622
11	11	1873.8	121	3511126.44	20611.8
12	12	1855.5	144	3442880.25	22266
13	13	1872	169	3504384	24336
14	14	1901.4	196	3615321.96	26619.6
15	15	1903.5	225	3623312.25	28552.5
16	16	1901.2	256	3614561.44	30419.2
17	17	1911.9	289	3655361.61	32502.3
18	18	1906.5	324	3634742.25	34317
19	19	1921	361	3690241	36499
Total	190	35239	2470	65390946.06	356662.7

Table 1 is the result of calculating the values of X, Y, X^2, Y^2, and XY linear regression, which will be a reference in determining the importance of a and b to predict the gold price.

$$a = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{n (\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

After the training results and data predictions for each model have been obtained, the following process calculates the errors for these models. Calculating R^2, MSE, and RMSE is done using the scikit-learn machine learning library.

$$R^2 = 1 - \frac{\sum_{1=1}^{n}(y_i - f(\hat{y}_i))^2}{\sum_{i=1}^{n}(y_i - \tilde{y})^2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{Y}_i)^2$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y - \hat{Y}_i)^2}{N}}$$

After getting the R^2, MSE, and MRSE values for each regression model, the next step is determining which model is the best from the data set used. To determine which model is the best, what needs to be considered are the rules of each method used. For $R^{2,}$ the best results are those close to the value 1. So, the more relative to the value 1, the better the relationship between the dependent variable (Y) and the independent variable (X). Whereas for MSE and RMSE, the best results are those with the lowest scores. So, the smaller the MSE and RMSE values, the closer the actual value.

It can be concluded that the best model is prediction using the 3rd-order polynomial regression method. The resulting data has an R^2 value close to 1 and the lowest MSE and RMSE values.

After getting the best model, the last step is to make predictions based on the 4th-order polynomial model regression method. The formula used with 3rd-order and replacing the value of X with the date to be predicted.

3. RESULTS AND DISCUSSION

Gold price prediction using the linear and polynomial regression methods uses a dataset totaling 6703 from 2000 – 2022. This study's prediction and training data were made using the Python programming language, where predicting and executing the linear method uses scikit-learn machine learning library tools.

Data is generated from each regression model after training and prediction using existing datasets.

Table 2. Training results and predictions for each model

No.	Date	Linear	Order 2	Order 3	Order 4	Actual Value
1	1/3/2022	1912.5074	1819.0824	1742.2399	1769.157102	1830.1
2	1/4/2022	1911.7148	1821.6248	1750.2732	1773.985632	1801.1
3	1/5/2022	1910.9222	1824.1272	1758.1017	1778.782662	1815.2
4	1/6/2022	1910.1296	1826.5896	1765.7272	1783.545312	1810.6
5	1/7/2022	1909,337	1829012	1773.1515	1788.27075	1791
6	1/10/2022	1908.5444	1831.3944	1780.3764	1792.956192	1796.4
7	1/11/2022	1907.7518	1833.7368	1787.4037	1797.598902	1801.3
8	1/12/2022	1906.9592	1836.0392	1794.2352	1802.196192	1821.3

9	1/13/2022	1906.1666	1838.3016	1800.8727	1806.745422	1825.7
10	1/14/2022	1905.374	1840524	1807.318	1811.244	1822.1
11	1/17/2022	1904.5814	1842.7064	1813.5729	1815.689382	1818.7
12	1/19/2022	1903.7888	1844.8488	1819.6392	1820.079072	1812.9
13	1/20/2022	1902.9962	1846.9512	1825.5187	1824.410622	1840.1
14	1/21/2022	1902.2036	1849.0136	1831.2132	1828.681632	1840.6
15	1/24/2022	1901,411	1851036	1836.7245	1832.88975	1835.2
16	1/25/2022	1900.6184	1853.0184	1842.0544	1837.032672	1843.6
17	1/26/2022	1899.8258	1854.9608	1847.2047	1841.108142	1848
18	1/27/2022	1899.0332	1856.8632	1852.1772	1845.113952	1819
19	1/28/2022	1898.2406	1858.7256	1856.9737	1849.047942	1799.6
20	1/31/2022	1897.448	1860548	1861596	1852,908	1792.8

Table 2 shows these models' training result data and gold price predictions. So, if Table 2 is projected and matched with each model, it will produce a curve like Figure 2.

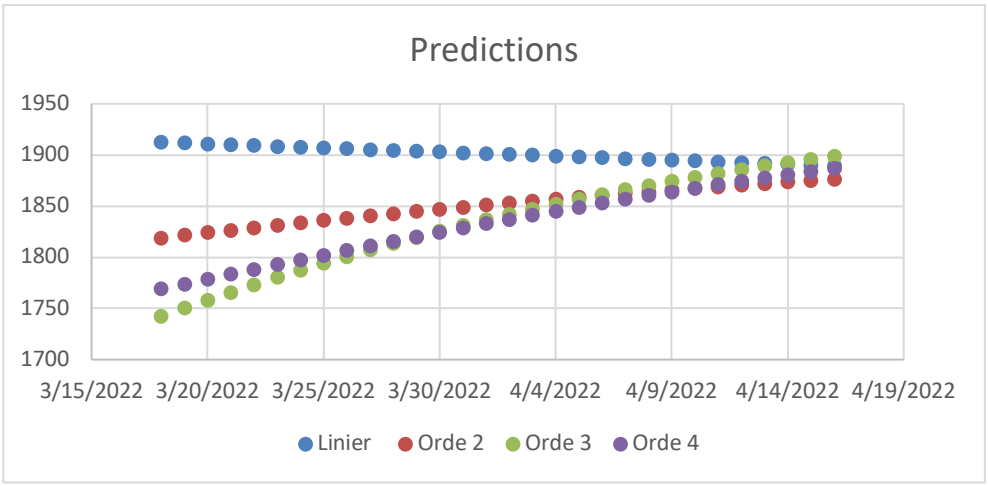

Figure 2 (a). The prediction result model with the regression method the model of each regression

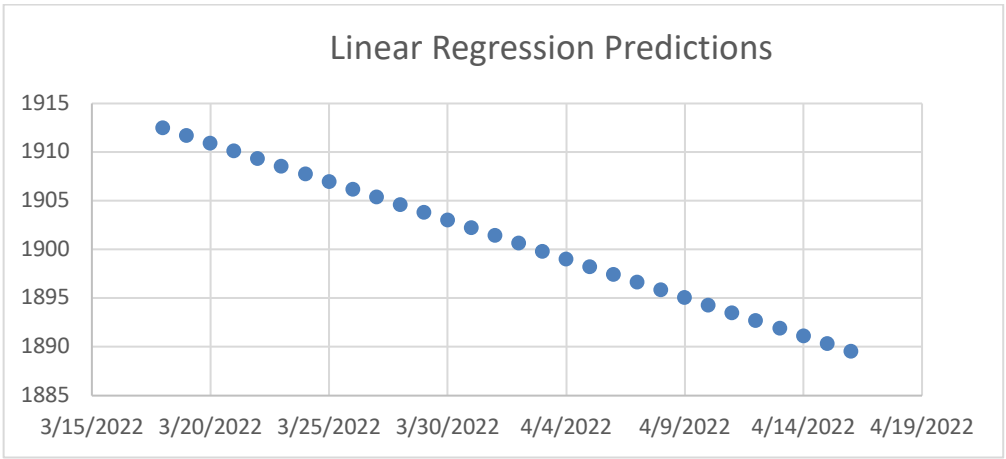

Figure 2 (b). The prediction result model with the regression method the linear regression model

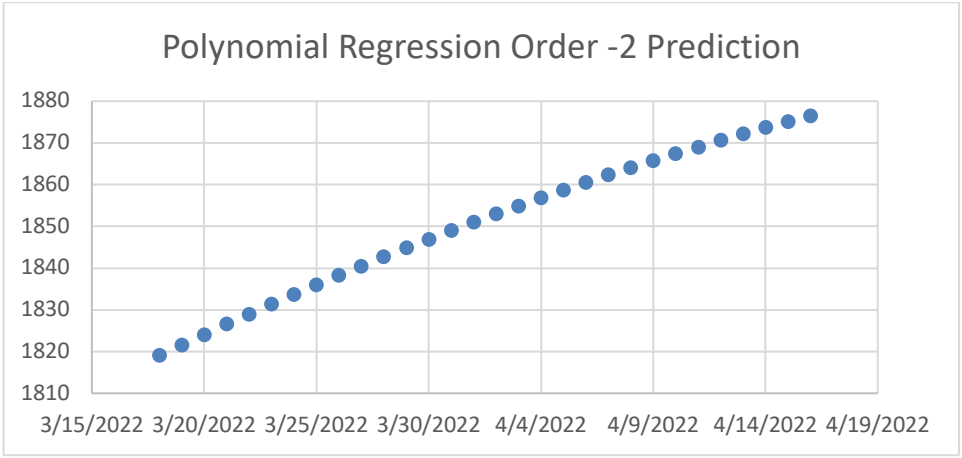

Figure 2 (c). The prediction result model with the regression method the 2nd-order polynomial regression model

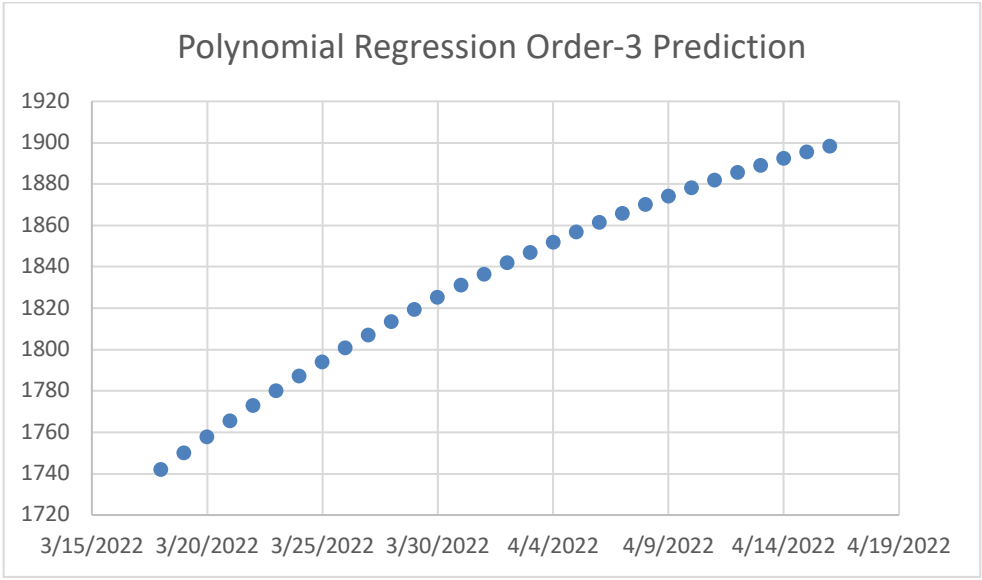

Figure 2 (d). The prediction result model with the regression method the 3rd-order polynomial regression model

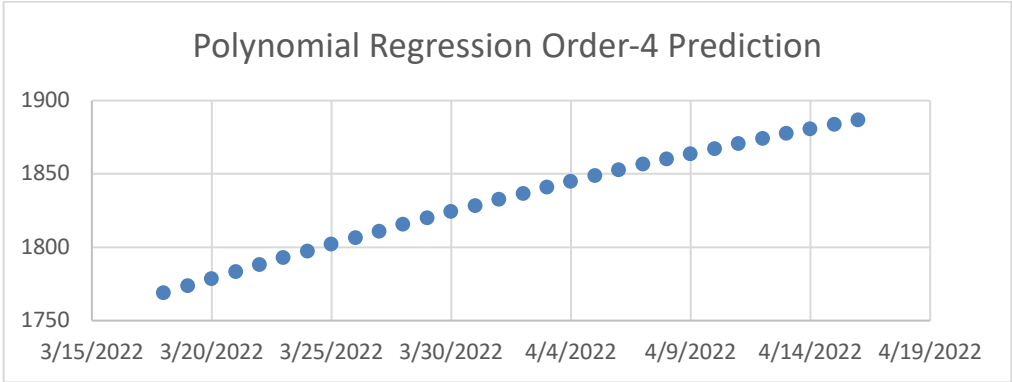

Figure 2 (e). The prediction result model with the regression method the 3rd- order polynomial regression model 4

Figure 2 (a), (b), (c), (d), and (e) are the results of gold price prediction using the linear regression method and order 2, 3, and 4 polynomials, with the y-axis being the price of gold on the x-axis date.

To get optimal predictions, it is necessary to evaluate each model. The evaluation used the R^2, MSE, and RMSE methods.

Table 3. The results of R^2, MSE, and RMSE

	Linear	**Order 2**	**Order 3**	**Order 4**
R^2	0.94	0.97	0.9567	0.9596
MSE	4218.146	2411.259	4654,499	15181.03
RMSE	64.94725	49.10458	68.22389	123.2113

Based on the evaluation results of Table 3, the second-order regression model has an R^2 value that is closest to number 1 compared to the other regressions, which is 0.97. The 2nd Order Polynomial Regression also has the lowest MSE and RMSE values compared to the other regressions, namely 2411.259 and 49.10458. So, the 4^{th}-order polynomial regression is the best regression model to predict the price of gold.

R^2, MSE, and RMSE methods, it can be concluded that for the dataset used to predict gold prices, the best model is using a 2^{nd}-order polynomial regression method.

Based on the evaluation results of Table 3, which concludes that the second-order polynomial regression model is the best, the final stage of this research is to predict the price of gold with this model. The gold price that will be expected is the price of gold starting September 2022. Table 4. is the result of predictions using the 2^{nd}-order polynomial regression method.

Table 4. Gold Price Prediction Results for September 2022.

No.	**Date**	**Gold price**
1	9/1/2022	1689.2232
2	9/2/2022	1685.0856
3	9/3/2022	1680908
4	9/4/2022	1676.6904
5	9/5/2022	1672.4328
6	9/6/2022	1668.1352
7	9/7/2022	1663.7976
8	9/8/2022	1659.42
9	9/9/2022	1655.0024
10	9/10/2022	1650.5448
11	9/11/2022	1646.0472
12	9/12/2022	1641.5096
13	9/13/2022	1636,932
14	9/14/2022	1632.3144
15	9/15/2022	1627.6568
16	9/16/2022	1622.9592
17	9/17/2022	1618.2216
18	9/18/2022	1613,444
19	9/19/2022	1608.6264
20	9/20/2022	1603.7688
21	9/21/2022	1598.8712
22	9/22/2022	1593.9336
23	9/23/2022	1588,956
24	9/24/2022	1583.9384

25	9/25/2022	1578.8808
26	9/26/2022	1573.7832
27	9/27/2022	1568.6456
28	9/28/2022	1563,468
29	9/29/2022	1558.2504
30	9/30/2022	1552.9928

Table 4 shows the results of the gold price prediction in September 2022.

4. CONCLUSION

Based on the gold price prediction experiment using a dataset with a total of 6703 data, it can be concluded that the best model for predicting the price of gold is using the 2nd-order polynomial regression method. Based on the evaluation results, this can be completed by comparing the prediction results with the R^2, MSE, and RMSE with respective values $R^2 = 0.97$, MSE = 2411.259, and RMSE = 49.10458.

Using the second-order polynomial regression method, a prediction is also made to determine the future price of gold. The forecast is made to find out the price of gold for September 2022. The cost of gold shows that in September 2022, the price will fall.

REFERENCES

[1] L. H. Rifqi, L. A. Mufiroh, and A. Z. Nihayah, "Analysis of gold investment with installment scheme," *J. Islam. Econ. Manag. Bus.*, vol. 3, no. 2, pp. 137–152, 2021, doi: 10.21580/jiemb.2021.3.2.11699.

[2] S. Setiyono, P. Priyagus, and S. Hudayah, "Pengaruh investasi swasta dan pendidikan serta belanja pembangunan terhadap pertumbuhan ekonomi dan kemiskinan di kabupaten kutai kartanegara," *J. Ilmu Ekon. Mulawarman*, vol. Vol 4, no. 3, 2019.

[3] R. H. Sugiastuti, V. Friseyla, and R. Pramesti, "How Financial Literacy and Investment Knowledge Influence Gold Investment Decisions," *PROFIT J. Adm. Bisnis*, vol. 18, no. 2, pp. 257–272, 2024, [Online]. Available: https://profit.ub.ac.id/index.php/profit/article/view/2734.

[4] J. Budiman, "PENDEKATAN PERILAKU KEUANGAN TERHADAP KEPUTUSAN INVESTASI EMAS," *Profit*, vol. 14, pp. 63–68, Jan. 2020, doi: 10.21776/ub.profit.2020.014.01.7.

[5] P. Thombare and R. Bhati, "ANALYSIS OF RELATIONSHIP BETWEEN GOLD PRICE AND STOCK MARKET PRICE IN INDIA ANALYSIS OF RELATIONSHIP BETWEEN GOLD PRICE AND STOCK MARKET PRICE IN INDIA," 2024.

[6] R. Anggara, "Investing In Gold In Times Of Global Crisis," *Int. Asia Law Money Laund.*, vol. 1, no. 4, pp. 228–233, 2022, doi: 10.59712/iaml.v1i4.43.

[7] M. R. I. Hariwijaya, M. T. Furqon, and C. Dewi, "Prediksi Harga Emas Dengan Menggunakan Metode Average-Based Fuzzy Time Series," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4 SE-, pp. 1258–1264, Jul. 2020, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7190.

[8] V. Nastiti, I. Halimi, and Y. Azhar, "Prediksi Harga Emas Menggunakan Univariate Convolutional Neural Network," *Repositor*, vol. 1, Dec. 2019, doi: 10.22219/repositor.v1i2.612.

[9] S. Ma'arif and S. Sukamulja, "ANALISIS STABILITAS EMAS SEBAGAI NILAI TUKAR DAN PENGARUH VARIABEL MAKRO EKONOMI TERHADAP PERGERAKAN HARGA EMAS DUNIA," Universitas Gadjah Mada, 2014.

[10] E. P. K. Orpa, E. F. Ripanti, and T. Tursina, "Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, p. 272,

Oct. 2019, doi: 10.26418/justin.v7i4.33163.
[11] F. Heeb, J. F. Kölbel, F. Paetzold, and S. Zeisberger, "Do Investors Care about Impact?," *Rev. Financ. Stud.*, vol. 36, no. 5, pp. 1737–1787, Apr. 2023, doi: 10.1093/rfs/hhac066.
[12] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.
[13] S. Brightwood and F. Gen, "The Evolution of Machine Learning Algorithms and their Impact on Software Development," 2024.
[14] G. Pandey, P. Chaudhary, R. Gupta, and S. Pal, "SEIR and Regression Model based COVID-19 outbreak predictions in India." 2020, [Online]. Available: https://arxiv.org/abs/2004.00958.
[15] I. Maryati, T. S. Sumartini, and D. Sofyan, "Experiences of Pearson formula in analysis regression," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 3, p. 032088, Mar. 2021, doi: 10.1088/1757-899X/1098/3/032088.
[16] G. Romeo, "Chapter 13 - Data analysis for business and economics," G. B. T.-E. of N. M. E. with E. Romeo, Ed. Academic Press, 2020, pp. 695–761.
[17] A. Jadon, A. Patil, and S. Jadon, "A Comprehensive Survey of Regression-Based Loss Functions for Time Series Forecasting BT - Data Management, Analytics and Innovation," 2024, pp. 117–147.