SEj
Vol. 8, N° 1, 2024

Journal of Sciences and Engineering

A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

Predictive Machine Learning models to estimate the price of gold [Modelos predictivos de Machine Learning para estimar el precio del oro]

Joela Noemi Sotelo Cenas*[a], Helin Julissa Gervacio Arteaga [a], Carmen Lizeth Carranza Rios[a].

[a] Facultad de Ingeniería, Universidad Nacional de Trujillo. Av. Juan Pablo II s/n –Ciudad Universitaria, Trujillo, Perú.
jsotelo@unitru.edu.pe

Received: 09 January 2024; Accepted: 01 February 2024; Published: 10 March 2024

Resumen

El propósito de este estudio fue determinar el algoritmo óptimo para estimar el precio del oro e identificar las variables más incidentes en su variación. Se usó una metodología de nivel exploratorio, enfoque cuantitativo y de diseño no experimental. Los resultados que se obtuvieron al realizar EDA muestran que las variables de mayor correlación con respecto al precio del oro son el costo de producción con 44 % y el S&P_500 con 30 %. Al validar los modelos dieron como resultado que el algoritmo de Gradient boosting tiene un R2 óptimo con 99.4 %, dicho valor justifica la importancia del modelo para poder estimar el precio del oro. Así mismo sin dejar de lado el algoritmo de Random Forest también muestra un R2 de 99.3 %. Así mismo se identificó que las variables de mayor incidencia son el Cost_prod con un 51.5 % y el USD_X con un 30.4 %. Finalmente se concluye que el uso de estos algoritmos como el Gradient boosting y el Random Forest pueden estimar el precio del oro tomando en cuenta las variables que inci-den en su variación.

Palabras clave: Modelos de regresión; Precio del oro; Gradient boosting; Random Forest.

Abstract

The purpose of this study was to determine the optimal algorithm to estimate the price of gold and identify the variables most incident to its variation. An exploratory level methodology, quantitative approach and non-experimental design was used. The results obtained when performing EDA show that the variables with the highest correlation with respect to the price of gold are the cost of production with 44% and the S&P_500 with 30%. When validating the models, the result was that the Gradient boosting algorithm has an optimal R2 of 99.4%, this value justifies the importance of the model in order to estimate the price of gold. Likewise, without leaving aside the Random Forest algorithm, it also shows an R2 of 99.3%. Likewise, it was identified that the variables with the highest incidence are Cost_prod with 51.5% and USD_X with 30.4%. Finally, it is concluded that the use of these algorithms such as Gradient boosting and Random Forest can estimate the price of gold taking into account the variables that affect its variation.

Keywords: Regression models; Gold price; Gradient boosting; Random Forest

Please cite this article as: Sotelo J.N., Gervacio H.J., Carranza C.L., Predictive Machine Learning models to estimate the price of gold, Journal of Sciences and Engineering, vol. 8, no 1, 2024, pp. 6-19.
https://doi.org/10.32829/sej.v8i1.204

SEj
Vol. 8, N° 1, 2024
Journal of Sciences and Engineering
A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

1. Introduction

The use of algorithms to estimate the price of Au has emerged as a strategic and advanced tool in the financial field (Ghule & Gadhave, 2022). In such a dynamic market and susceptible to various influences, the ability to anticipate fluctuations in the value of gold has become crucial for investors, analysts and traders (Makala & Li, 2021). The analysis focuses on the development of a machine learning algorithm designed to accurately estimate the price of gold, leveraging machine learning's ability to analyze complex patterns and trends in historical market data (Liang et al, 2022). The inherent volatility of financial markets, combined with the multitude of factors that impact the price of gold, presents a significant challenge to conventional analytical approaches (Cohen & Aiche, 2023). The application of machine learning algorithms offers an innovative solution by allowing the system to detect non-linear relationships and subtle patterns that could go unnoticed by traditional methods (Sević, J. & Stakić, A. 2022). The approach promises greater accuracy in predictions, providing investors with a competitive advantage by making informed and strategic decisions (Wagh et al, 2022).

Throughout this work, we will explore the methodology behind the selected machine learning algorithm, analyzing the selection of features, the development of the training and validation of the algorithm, as well as its ability to adapt to changes in market conditions (Khani et al., 2021). Additionally, we will examine the quality, relevance of the data used, as well as the interpretability of the resulting algorithm (Sadorsky, 2021).

In summary, the application of a machine learning model to forecast the price of gold represents a significant advance in the ability to anticipate market fluctuations (Tebin & James, 2022). This approach not only seeks to improve the accuracy of predictions, but also to offer a valuable tool for those seeking to make informed decisions in an increasingly complex and dynamic financial environment.

The purpose of this research was to determine the optimal algorithm to estimate the price of gold and identify the variables most incident to its variation.

2. Materials and Methods

2.1 Materials

The primary tools used for this study were the use of specific digital and programming tools to estimate the price of gold. For this, Maching Learning regression models were used, such as the Bosting Gradient algorithm and Ramdom Forest, complemented with additional analytical tools such as Microsoft Excel, Python. (Zúñiga, P. 2019).

2.2 Procedure

The process began with the collection and preparation of data on the price of gold. Likewise, the database and EDA were organized and cleaned, in this way work began with five algorithms such as Decision Tree that is used from historical data to estimate their future values through a regression technique. The Linear Regression algorithm was used to predict the value of gold based on independent variables. The SRV algorithm is used to fit the training data by identifying support vectors and creating a hyperplane that maximizes the separation margin for gold price prediction. The Random Forest algorithm is used to predict the price of gold by combining multiple decision trees and averaging the individual predictions and finally the Gradient Bosting algorithm is based on training multiple weak models sequentially, where each model focuses on correcting

SEj
Vol. 8, N° 1, 2024

Journal of Sciences and Engineering

A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

the errors. errors of the previous model, in which the optimal ones to estimate the price of gold were Random Forest and Gradiente Bosting. (Rodríguez, A. 2021).

2.3 Methods

A quantitative approach methodology and statistical data analysis were used in a descriptive manner, a review and interpretation of information from different documents on Machine Learning regression algorithms was carried out to determine the price of gold (Talavera, F. 2020).

2.4 Population and smaple

Population.

The population of this study is made up of a database of the variables that affect the price of gold, such as the USD price index, world gold production, cost of gold production, gold price index, consumption world gold, oil price, gold price, for more detail of the aforementioned variables are shown in Table 1. Machine Learning algorithms can help analyze these factors and estimate the price of gold in the future (Villada, F. 2018).

Table 1. Definition of study variables

Variables	Units	Variable type
Dollar Price Index (USD_X)	Und	Independent
World Gold Production	t	Independent
Gold production cost (cost_prod)	US$/troy ounce	Independent
S&P_500 Index	Und	Independent
World Gold Consumption	t	Independent
Oil price (Prec_petr)	U$S/barrel	Independent
Gold price (Prec_au)	US$/troy ounce	Dependent

Sample

The method used for this research is through systematic probabilistic sampling, because the sample was taken at intervals from 01/04/2021 to 09/30/2023, there was a total of 730 data for each variable, for greater detail it is shown the first 5 and the last 5 data in Table 2, which will be used to estimate the price of Au (Reales. L. at al 2022).

Table 2. Database visualization

Date	USD _X	Production	Cost_prod	S&P_500	Consumption	Pric_petr	Pric_Au
4/01/2021	89.82	832.13	1,070	3,700.65	863.38	47.62	1,946.6
5/01/2021	89.39	832.13	1,070	3,726.86	863.38	49.93	1,954.4
6/01/2021	89.48	832.13	1,070	3,748.14	863.38	50.63	1,908.6
7/01/2021	89.77	832.13	1,070	3,803.79	863.38	50.83	1,913.6
8/01/2021	90.05	832.13	1,070	3,824.68	863.38	52.24	1,835.4
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
26/09/2023	105.9	971.06	1,349	4,273.53	812.35	90.39	1,919.8
27/09/2023	106.4	971.06	1,349	4,274.51	812.35	93.68	1,890.9
28/09/2023	105.9	971.06	1,349	4,299.70	812.35	91.71	1,878.6
29/09/2023	105.8	971.06	1,349	4,288.05	812.35	90.79	1,866.1
30/09/2023	105.8	971.06	1,349	4,288.05	812.35	90.79	1,866.1

SEj
Vol. 8, N° 1, 2024
Journal of Sciences and Engineering
A publication of CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

3. Results

3.1 Exploratory data analysis (EDA)

EDA is a critical phase in initially exploring and understanding data before performing more advanced analyzes or statistical modeling (Courtney, 2021). Its importance lies in providing a deep and preliminary understanding of a data set (Nicodemo & Satorra, 2022).
The statistical overview of the research shows statistical parameters such as amount of data, standard deviation, mean, quartile representation as shown in Table 3. There is a very high standard deviation in the S&P_500, this reflects a larger dispersion of its data compared to the other independent variables.

There is a standard deviation with a value of 92.39 in the price of gold, this value will allow us to have a considerable correlation with the independent variables.

Table 3. Description of the statistical parameters

Statistical parameters	USD _X	Production	Cost_prod	S&P_500	Consumption	Pric_petr	Pric_Au
Count	711	711	711	711	711	711	711
Mean	99.38	905.16	1,213.57	4,197.41	819.02	79.19	1,840.08
Std	6.34	45.69	92.01	281.99	104.9	13.65	92.09
Min	89.39	832.13	1,070	3,577.03	684.94	47.62	1,630.9
25%	92.96	860.17	1,109	3,958.67	735.96	70.23	1,777.6
50%	101.12	912.7	1,232	4,183.18	777.78	77.85	1,825.1
75%	104.09	946.71	1,298	4,436.27	869.89	86.94	1,921.6
Max	114.05	971.06	1,349	4,796.56	1,044.34	114.67	2,055.7

The Box Plot for each study variable is displayed in Figure 1, for which 19 outlier values were obtained in the oil price variable. These data were eliminated to avoid future errors in the use of the different algorithms. It should be noted that by eliminating these outliers the data is no longer 730 data but rather 711 data distributed in each column where the independent variables can be evidenced.

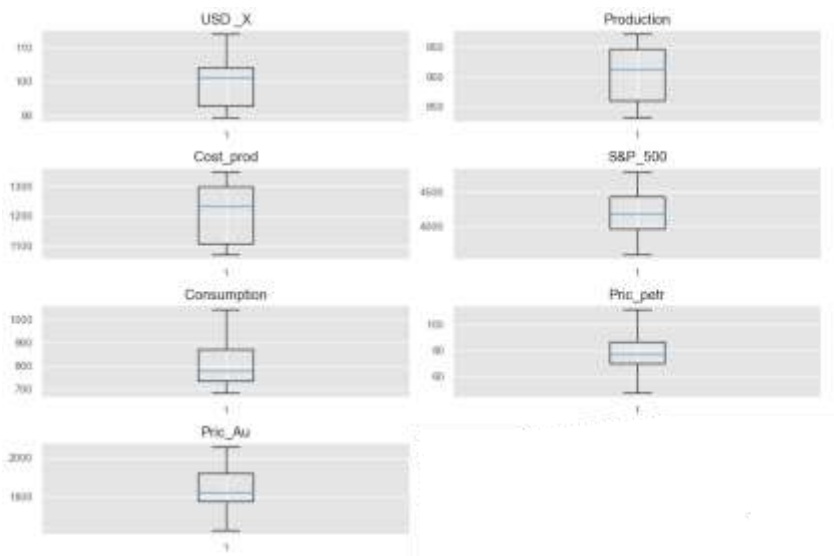


Figure 1. Box Plot of each of the variables

SEj
Vol. 8, N° 1, 2024

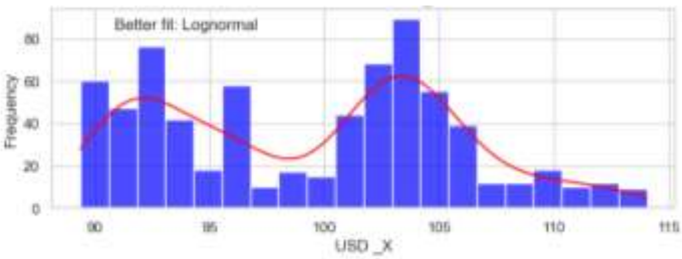
Journal of Sciences and Engineering

A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

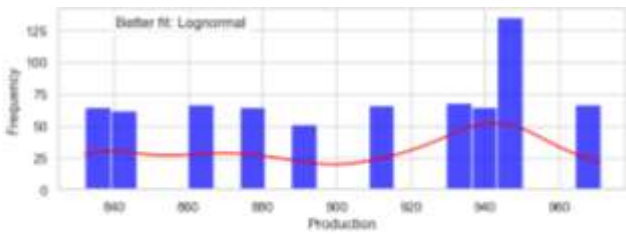
ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

Histograms provide a valuable graphical representation that makes it easy to identify patterns, trends, and symmetries, as well as other crucial aspects of data distribution. In addition, they are effective tools to visualize dispersion and concentration (Bakker et al, 2019).

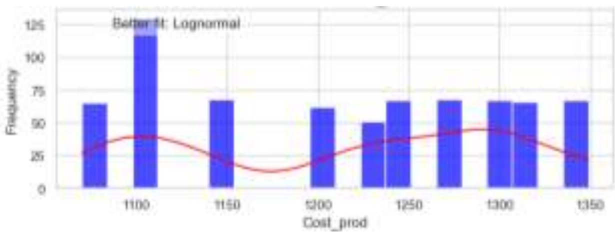
Figure 2 shows a distribution of the data through the histograms of the 6 variables analyzed, as well as the type of function, for greater detail each one will be explained: (a) a not so defined lognormal adjustment is evident in the variable USD_X, (b) the production variable and the production cost variable (c) show a very irregular distribution of their data and also have a log normal fit, (d) the S&P_500 variable shows a normal fit which indicates that their data have symmetry with each other, (e) the Consumption variable shows a gamma adjustment due to the behavior of the variable with positive asymmetry, (f) the oil price variable and the Au price variable (g) show a better gamma adjustment compared to the others. Like Gulandi, S. & Toscani, G. (2018), in their research they mention that the log normal distribution is adequate for each category of data, while the fit is worse for data sets where a wide range is combined. of adjacent data categories for evaluation. The appearance of a lognormal distribution can be fully justified on the basis of modeling.



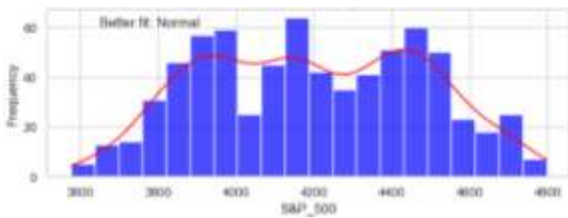
(a) Distribution of USD_X.



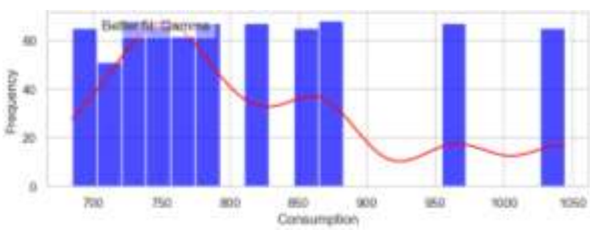
(b) Production Distribution.



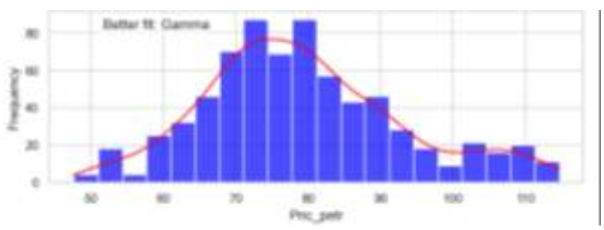
(c) Distribution of Cost_prod.



(d) Distribution of S&P_500.



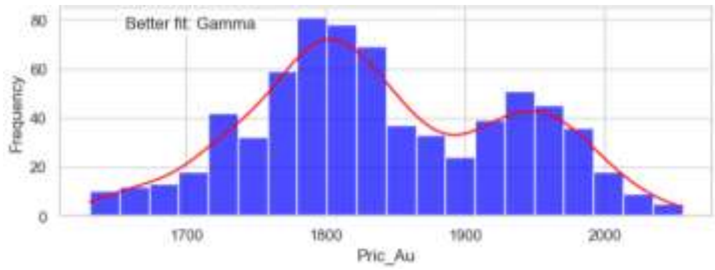
(e) Distribution of Consumption



(f) Distribution of Pric_petr.

SEj
Vol. 8, N° 1, 2024
Journal of Sciences and Engineering
A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204



(g) Distribution of Pric_Au.

Figure 2. Representation of histograms for each variable

Figure 3 shows the correlation between variables that provide a preliminary scope of the possible relationships that exist with the price of gold. In this case, the variable that shows the greatest correlation with the price of gold is the production cost variables with 44% and the S&P_500 variable with 30%. It is evident that the production cost variable in relation to the dollar price index has an 83% correlation. Like Castillo, O. (2022), in his research he developed predictive regression models of various data where he revealed that inflation in the United States had a 78% impact and the S&P 500 index had a 82% impact.

Figure 1. Correlation Matrix

SEj
Vol. 8, N° 1, 2024
Journal of Sciences and Engineering
A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

3.2 Training predictive algorithms to estimate the price of gold

The hyperparameters used for this research with respect to each algorithm are shown in Table 4, Table 5, Table 6, Table 7, Table 8, these hyperparameters are determined using GridSearchCV, GridSearchCV is an optimization technique that seeks the best combination of hyperparameters in a predefined set for an artificial intelligence algorithm, based on its performance during cross-validation (Laura-Ochoa, L. (2020).

Linear regression modeling is a statistical and machine learning method for modeling variables that depend on each other and one or more independent variables. Its purpose is to estimate numerical data focused on the relationship between the variables (Espinosa, Z & Javier, J. 2020).

Table 4. Optimal hyperparameters for linear regression

Type	Predefined	Optimal
fit_intercept	True, False	True

The decision tree algorithm is an automatic training method that focuses on structuring the tree in order for classification and regression, dividing the data set into smaller subsets based on certain characteristics (Sandoval, L 2019).

Table 5. Optimal hyperparameters for the Decision Tree

Type	Predefined	Optimal
max_depth	None, 5, 10, 15, 20	15
min_samples_leaf	1, 2, 4	1
min_samples_split	2, 5, 10	5
splitter	Best, Random	Random

The Random Forest algorithm focuses on several decision trees to develop predictions. Every tree "votes" for the optimal prediction, which reduces overfitting in precision, compared to a single decision tree (Cánovas et al 2018).

Table 6. Optimal hyperparameters for Random forest

Type	Predefined	Optimal
n_estimators	50, 100, 200	200
max_depth	None, 10, 20, 30	None
min_samples_split	2, 5, 10	2
min_samples_leaf	1, 2, 4	1
bootstrap	True, False	True

The SVR algorithm is based on the concept of support vectors to find the best fit line that maximizes the margin of separation between the value points and the regression line. This helps to estimate numerical values by finding an optimal hyperplane in a high-dimensional space (González et al 2018).

Table 7. Optimal hyperparameters for SVR

Type	Predefined	Optimal
C	0.1, 1, 10, 100	100
kernel	linear, rbf', poly	rbf
gamma	Auto, scale, 0.01, 0.1, 1	scale
coef0	-1, 0, 1	-1
degree	2, 3, 4	2

SEj
Vol. 8, N° 1, 2024
Journal of Sciences and Engineering
A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

The Gradient boosting algorithm is a machine learning that builds a predictive model in the form of a set of decision trees (Campillo et al., 2018).

Table 8. Optimal hyperparameters for Gradient boosting

Type	Predefined	Optimal
n_estimators	50, 100, 200	100
max_depth	3, 4, 5, 6	6
learning_rate	0.01, 0.05, 0.1	0.05
subsample	0.8, 0.9, 1.0	0.8
min_samples_split	2, 3, 4	2
min_samples_leaf	1, 2, 3	2

By applying the regression algorithms, models were generated with their respective hyperparameters, subsequently a test was carried out, as an example with 5 tests of different dates, in order to corroborate the prediction of the gold price, for this data from the independent variables considered in this study. The results can be viewed in Table 9, where the real value vs. the estimated value and the error of the 5 algorithms used in this study are specified. Like Espinoza, (2020) in his research mentions that the trained Random Forest algorithm reached an error of 40% for the OOB estimate, which gives us a precision of 60%.

Table 9. Results after applying the regression algorithms

Date	Real	Linear regression		Decision tree		Random forest		SVR		Gradient boosting	
		Estimated	Error	Estimated	Error	Estimated	Error	Estimated	Error	Estimated	Error
4/01/21	1946.6	1817.2	129.4	1938.2	8.4	1933.9	12.6	1925.2	21.4	1944.3	2.3
11/05/21	1836.1	1826.1	9.96	1837.3	-1.2	1844.6	-8.5	1854.0	-17.9	1842.5	-6.4
1/07/21	1758.4	1791.8	-33.4	1759.8	-1.4	1759.0	-0.6	1755.8	2.6	1761.3	-2.9
1/10/21	1800.1	1905.7	-105.6	1794.7	5.45	1802.7	-2.6	1757.7	42.4	1796.9	3.2
3/01/22	1923.7	1890.8	32.8	1925.6	-1.9	1925.1	-1.4	1948.3	-24.6	1926.1	-2.4

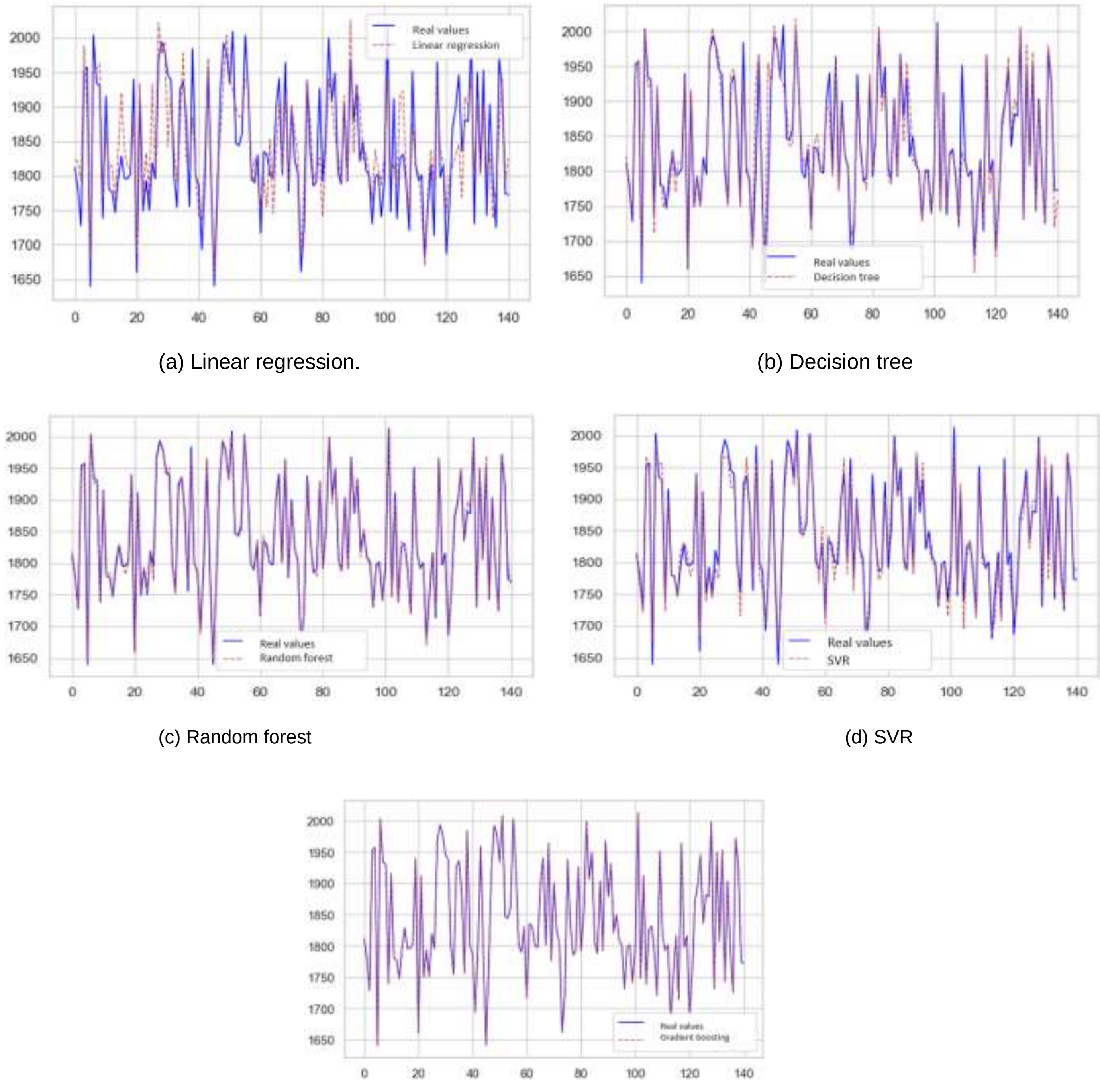
Figure 4 shows information about all the algorithms with their respective predictions. This was done in order to find the optimal model to estimate the price of gold; The results obtained for each algorithm are detailed below: (a) the linear regression algorithm indicates an unacceptable prediction, (b) the Decision Tree algorithm has an acceptable prediction but has pronounced highs, (c) the Random algorithm Forest is optimal and can be used without any problem, (d) the SVR algorithm can be used but we must not leave aside the hyperparameters that make it work, (e) the Gradient boosting algorithm is the one that has the best behavior to estimate the price of gold due to its low margin of error.

SEj
Vol. 8, N° 1, 2024

Journal of Sciences and Engineering

A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204



(a) Linear regression.

(b) Decision tree

(c) Random forest

(d) SVR

(e) Gradient boosting

Figure 4. Graphical representation of actual values vs estimated values of each algorithm

SEj
Vol. 8, N° 1, 2024

Journal of Sciences and Engineering

A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

3.3 Validation of predictive models to estimate the price of gold

Table 10 and Table 11 display the validation metrics data for each of the algorithms in training and testing. The Gradient boosting algorithm has the optimal coefficient of determination (R2) with 99.4% in training, this value justifies the importance of the model to be able to estimate the price of gold. Likewise, without leaving aside the Random Forest algorithm, it also shows an R2 of 99.3% during training, which indicates that either of these two algorithms can be used to estimate the price of gold. According to Castillo, O. (2022), in his study he carried out the validation through various algorithms, the Gradient Boosting Regressor algorithm being optimal, where he obtained R2 = 0.51, MAE = 0.81, RMSE = 0.81.

Table 10. Validation metrics values for each algorithm in training

Algorithms	MSE	RMSE	R2	Variance	Error max	MAE
Linear regression	2699.52	51.95	0.68	0.68	130.46	42.67
Decision tree	159.69	12.64	0.98	0.98	76.20	8.32
Random forest	58.82	7.67	0.993	0.993	36.02	5.64
SVR	551.28	23.48	0.94	0.93	43.29	19.79
Gradient boosting	62.24	7.90	0.993	0.993	24.76	6.13

Tabla 1. Valores de las métricas de validación para cada algoritmo en el testeo

Algorithms	MSE	RMSE	R2	Variance	Error max	MAE
Linear regression	2870.52	53.56	0.67	0.67	117.06	44.95
Decision tree	576.51	24.01	0.93	0.94	79.30	18.53
Random forest	377.65	19.43	0.957	0.958	64.40	15.38
SVR	603.41	24.56	0.93	0.93	77.05	19.62
Gradient boosting	364.48	19.09	0.958	0.959	59.70	15.20

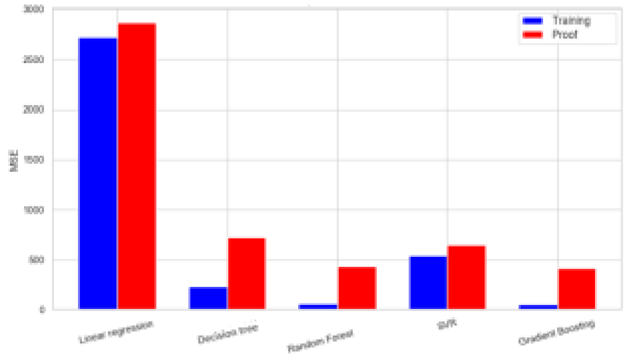
For a greater interpretation of the validation metrics of the algorithms, Figure 5 is shown where it is evident that the aforementioned algorithms excel in each metric.

SEj
Vol. 8, N° 1, 2024

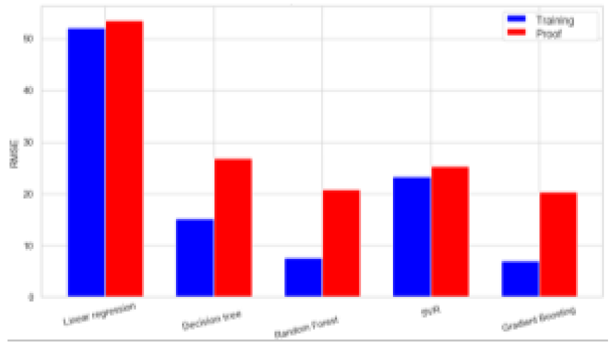
Journal of Sciences and Engineering

A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

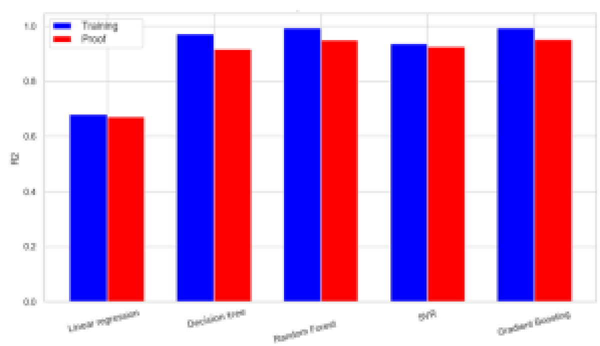
ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204



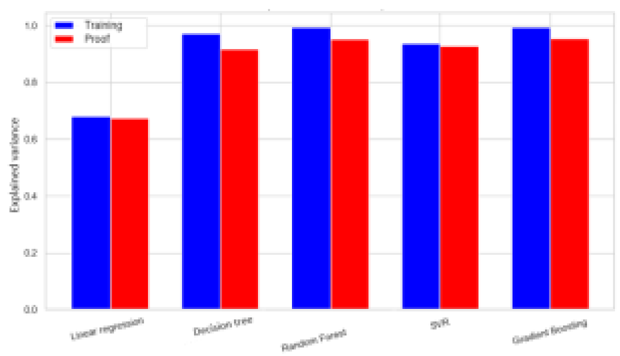
(a) Comparison of MSE.



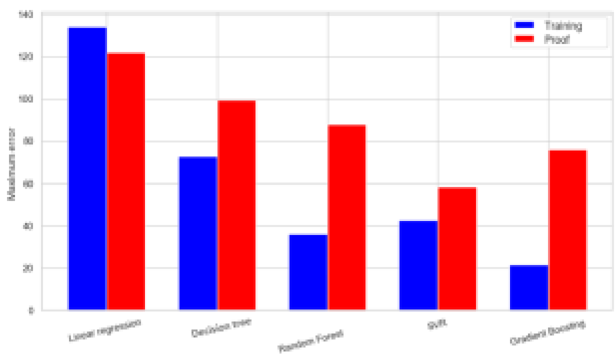
(b) RMSE comparison



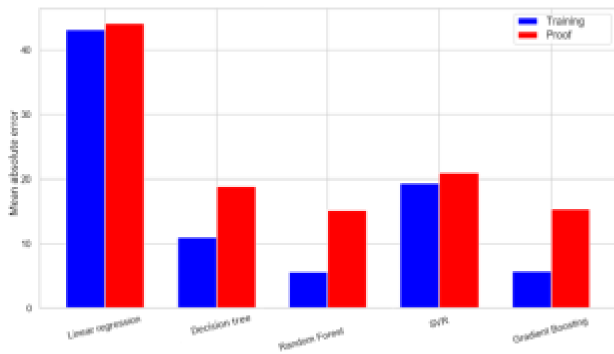
(c) Comparison of R^2.



(d) Comparison of Explained Variance



(e) Maximum error



(f) Mean absolute error comparison

Figure 5. Graphical representation of validation metrics

SEj
Vol. 8, N° 1, 2024
Journal of Sciences and Engineering
A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

3.4 Identification of the most representative variables

To identify the independent variables that have the greatest impact on the gold price forecast, optimal algorithms such as Gradient boosting and Random forest were used. Table 12 shows that the variables with the highest incidence are the Cost_prod with 51.5% and the USD_X with 30.4% and the variable with the lowest incidence is the world consumption of gold with 0.2%. Diaz et al. (2020) in their study mentions that they used the Random forest algorithm that determined the most incident variable was the production cost of 38.3% and a USD price of 24.8% of gold, while the variable with the least incidence is world consumption, which was a con 0.21% in gold price performance.

Table 12. Percent incidence of each variable

Variables	Random forest	Gradient boosting
USD _X	30.4 %	31.4%
Production	2.3 %	2.7 %
Cost_prod	51.5 %	50.5 %
S&P_500	4.7 %	3.9 %
Consumption	0.2 %	0.6 %
Pric_petr	10.9 %	10.9 %

4. Conclusions

In this study, a meticulous analysis of 711 data was developed, focusing on the use of regression algorithms to estimate the price of gold, recognizing its strategic importance in the financial field. The use of algorithms such as Gradient Boosting and Random Forest, supported by analytical and programming tools, allowed us to identify the key variables to estimate the price of gold. The results highlight the effectiveness of the Gradient Boosting and Random Forest algorithms, with determination coefficients greater than 99% during training. Validation reveals superior performance compared to other algorithms, underlined by metrics such as R2, MSE, and RMSE. The identification of variables highlights the significant incidence of Cost_prod and USD_X in the prediction. Finally, the successful application of regression algorithms, especially Gradient Boosting and Random Forest, offers a valuable tool to anticipate gold market fluctuations, highlighting the importance of specific variables in this process.

5. Acknowledgments

Special thanks to Engineer Juan Vega, for his support and guidance in the development of this research.

References

Bakker A., Boels, L., Dooren W., Drijvers P. (2019). Conceptual difficulties when interpreting histograms: A review. Educational Research Review, 28. https://doi.org/10.1016/j.edurev.2019.100291

Cánovas F., Alonso F., Gomariz F., & Oñate F. (2018). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. Computers & Geosciences, 103, 1-11. https://doi.org/10.1016/j.cageo.2017.02.012

Campillo J. Martínez Julián & Carmona Pedro. (2018). Análisis de la utilidad del algoritmo Gradient Boosting Machine (GBM) en la predicción del fracaso empresarial. Spanish Journal

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

of Finance and Accounting / Revista Española de Financiación y Contabilidad, 47, 1-26. doi.org/10.1080/02102412.2018.1442039

Castillo O. (2022). Desarrollo de modelos predictivos de regresión en la industria minera mediante el uso de algoritmo de machine Learning. https://orcid.org/0000-0001-6996-923x.

Cohen G., & Aiche A. (2023). Forecasting gold price using machine learning methodologies. Chaos, Solitons & Fractals, 175. https://doi.org/10.1016/j.chaos.2023.114079

Courtney M. (2021). Exploratory Data Analysis in Schools: A Logic Model to Guide Implementation. Internatioal Journal of Education Policy & Leadership, 17(4). https://doi.org/10.22230/ijepl.2021v17n4a1041

Diaz H., Ramírez S., & Salazar F. (2020). Los determinantes de las variaciones en el rendimiento del oro. La revisión por pares es responsabilidad. 65(2), 1-28. http://dx.doi.org/10.22201/fca.24488410e.2018.1973

Espinosa Z & Javier J. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. Ingeniería, investigación y tecnología, 21(3), 00002. https://doi.org/10.22201/fi.25940732e.2020.21.3.022

Espinoza S. (2020). Predicción de postulantes que cometerán fraude interno con Algoritmo de aprendizaje supervisado.

Ghule & Gadhave (2022). Gold Price Prediction using Machine Learning. International Journal of Scientific Research in Engineering and Management (IJSREM), 06. DOI: 10.55041/IJSREM15027.

González R. S. L., Gómez I., Pastrana J. L., & Hernández A. (2018). Algoritmos de clasificación y redes neuronales en la observación automática de registros. Cuadernos de Psicología del Deporte, 15(1), 31-40. https://dx.doi.org/10.4321/S1578-84232015000100003

Gulandi S., & Toscani G. (2018). Human behavior and lognormal distribution. A kinetic description. Department of Mathematics of the University of Pavia and IMATI CNR, Italy. Vol. 1.

Khani M., Vahidnia S., Abbasi A. (2021). A Deep Learning-Based Method for Forecasting Gold Price with Respect to Pandemics. SN Computer Science. https://doi.org/10.1007/s42979-021-00724-3.

Liang Y., Lin Y., & Lu Q. (2022). Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. Expert Systems with Applications, 15. https://doi.org/10.1016/j.eswa.2022.117847.

Makala D., & Li Z. (2021). Prediction of gold price with ARIMA and SVM. Journal of Physics: Conference Series. doi:10.1088/1742-6596/1767/1/012022.

Nicodemo C., & Satorra A. (2022). Exploratory data analysis on large data sets: The example of salary variation in Spanish Social Security Data. Article Reuse Guidelines, 25. https://doi.org/10.1177/2340944420957335.

Ochoa L. (2020). Evaluation of Classification Algorithms using Cross Validation. LACCEI International Multi-Conference for Engineering, Education, and Technology. http://dx.doi.org/10.18687/LACCEI2019.1.1.471.

Reales Chacón L., Robalino Morales G. E., Peñafiel Luna A. C., Cárdenas Medina J. H., Cantuña-Vallejo P. F. (2022). El Muestreo Intencional No Probabilístico como herramienta de la investigación científica en carreras de Ciencias de la Salud. Revista Universidad y Sociedad, 14(S5), 681-691. https://orcid.org/0000-0002-4242-3429.

Rodríguez A. (2021). Implementation of Artificial Intelligence algorithms in the prediction of new knowledge through constructivist teaching. Vol. 14(No. 3). doi: https://orcid.org/0000-0002-1238-0106.

Sadorsky P. (2021). Predicting Gold and Silver Price Direction Using Tree-Based Classifiers. Journal of risk and financial management, 14(5). https://doi.org/10.3390/jrfm14050198.

Sandoval L. (2019). Machine Learning algorithms for analysis and data prediction. 2017 IEEE 37th Central America and Panama Convention (Concapan xxxvii), 1-5. Managua, Nicaragua. IEEE. http://doi.org/10.1109/Concapan.2017.8278511.

SEj
Vol. 8, N° 1, 2024
Journal of Sciences and Engineering
A publication of
CINCADER
Centre of Research and Training for Regional Development
Online at www.journals.cincader.org

ISSN 2523-9503
DOI: https://doi.org/10.32829/sej.v8i1.204

Sević J., & Stakić A. (2022). Prediction of gold price movement considering the number of infected with the covid 19. The European journal of applied economics, 19(2). DOI: DOI 10.5937/EJAE19-39258.

Talavera F. J. (2020). Fundamentos Metodológicos de la Investigación: El Génesis del Nuevo Conocimiento. vol. 5(núm. 16). doi: https://doi.org/10.29394/Scientific.issn.2542-2987.2020.5.16.5.99-119

Tebin J. y James A. (2022). Gold Price Prediction using Machine Learning. Proceedings of the National Conference on Emerging Computer Applications (NCECA), 4. DOI: 10.5281/zenodo.6874188.

Villada F. (2018). Artificial Neural Networks applied to the Prediction of the Gold Price. vol.27(no.5). doi: http://dx.doi.org/10.4067/S0718-07642016000500016

Wagh A., Shetty S., Soman A., Maste D. (2022). Gold Price Prediction System. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 45(98). https://doi.org/10.22214/ijraset.2022.41623.

Zúñiga P. I. (2019). Metodología de la investigación científica: guía práctica. Vol. 7 Núm. 4(Núm. 4). doi: https://orcid.org/0009-0001-9418-8707