Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 7 No. 1, April 2022, pp. 89 – 102

Gold Price Prediction Using Support Vector Regression

Yupie Kusumawati*[1], **Karis Widyatmoko**[2]
Universitas Dian Nuswantoro Semarang
E-mail : yupie@dsn.dinus.ac.id[1], karis.widyatmoko@dsn.dinus.ac.id[2]*
**Corresponding author*

Candra Irawan[3]
Universitas Dian Nuswantoro Semarang
E-mail : candra.irawan@dsn.dinus.ac.id[3]

Abstract - In this modern era, one of the businesses that continues to grow is investment. Gold has a more stable value. In Indonesia, there are futures exchange companies that offer gold investment with an online transaction system (E-Trade). The amount of demand and supply, the rate of inflation, economic conditions, and many more can affect the high and low prices of gold. Due to changes in the conditions above, the price of gold may increase, decrease, or remain constant every day. The price of gold that can go up and down causes the need for gold price predictions so that future gold trading investment prospects can be seen. In this final project, the accuracy of Support Vector Regression will be investigated to find out how accurate it is in predicting gold prices with High, Low, Open, Close, and Volume variables. Based on the calculation of the best RMSE in the study, it was found that the best RMSE was to use a Linear kernel with a C of 35 and using a Y variable dataset of 7.4615. The Support Vector Regression Algorithm can predict quite well, as evidenced by the acquisition of fairly good RMSE results. It is necessary to do a simulation of buying and selling gold based on the prediction results and comparing the advantages of the testing data and the actual data.

Keywords – Prediction, Time Series, Support Vector Regression

1. INTRODUCTION

In Indonesia, there are futures exchange companies that offer gold investment with an online transaction system (E-Trade). One of the gold investments that can provide high profits but also high risk is Gold Trading [1]. Online gold trading is online gold investment on a contract basis through intermediary traders whose physical gold is stored in London in the Bullion Association. The movement of gold in the physical market is something that must be considered in depth in investing in the gold futures exchange [1]. The amount of demand and supply, the rate of inflation, economic conditions, and many more can affect the high and low prices of gold. Due to changes in the conditions above, the price of gold may increase, decrease, or remain constant every day, for that reason the price of gold can be categorized as time-series data.

Prediction is a method that predicts something systematically and sequentially about things that can happen in the future according to past and existing information so that the level of error and error from what you want to know can be minimized [2]. Implementation of data mining can be used to look for patterns or a model that can predict something from existing data and from previous data within a certain time [3]. Predicting the price of gold is an important issue in investing, predicting the price of gold can help investors in investing in gold.

The known past time series model can predict future values between going up or down [4]. Time series prediction methods such as Support Vector Machine (SVM) [1], Artificial Neural Network (ANN) (Sari, 2018), Fuzzy Logic (FL) [5] have been widely used by researchers to make predictions.

The Support Vector Machine (SVM) as a leading concept in the field of pattern recognition that does not have overfitting problems and can choose automatic models was introduced in 1992 for the first time by Vapnik [1]. The prediction problem is one of the problems that can be solved by the Support Vector Machine (SVM) method of the many methods. Large and complex datasets are suitable to be handled by SVM [6]. The existence of a kernel function makes the Support Vector Machine able to handle non-linear problems, so that time-series forecasting can be used with this method [2]. SVM has the basic principle of classifying data that is linear, then by incorporating the concept of a kernel trick in a high-dimensional workspace, SVM is developed so that it can work on non-linear data [7]. The SVM method relies on kernel functions which makes it a nonparametric technique [8]. SVM tries to find the best hyperplane in the input space, in contrast to the Artificial Neural Network strategy which tries to find a dividing hyperplane between classes.

Research on the Support Vector Machine has been done previously by several researchers to predict the price of gold. In a study conducted in 2015, which was predicting the price of gold using the Support Vector Machine Algorithm by looking for the best variables to predict, variable A (Open, High, Low and Close) and variable B (Open, High, Low and Factory news) were variables that used. Variable A gets the best RMSE value so that variable A becomes the variable that has the best accuracy [1]. Support Vector Machine has advantages, namely linear and non-linear problems can be handled with this method [9], SVM also has a drawback, namely the kernel function and parameters used greatly affect SVM in the transition process for the accuracy of the resulting model.

The price of gold that can go up and down causes the need for gold price predictions so that the prospect of investing in gold trading in the future can be seen for investors so that it is useful. In this article, the accuracy of the Support Vector Regression will be examined to find out how accurate it is in predicting the price of gold with High, Low, Open, Close, and Volume variables.

2. RESEARCH METHOD

2.1. Prediction

The process of estimating or predicting sequentially and systematically about something that might happen in the future based on the information in the past and present that is owned, so that the level of error and error can be minimized is the notion of prediction [2]. Implementation of data mining can be used to look for patterns or a model that can predict something from existing data and from previous data within a certain time [3]. Prediction techniques can be divided into two, namely quantitative and qualitative according to the technique in predicting.

2.2. Machine Learning

Machine Learning (ML) is an Artificial Intelligent science that studies human nature in solving problems, this science is widely used by humans to automate [10]. Machine learning can analyze and find answers by looking at patterns from previous events. Machine learning, of course, cannot solve all problems. However, Machine Learning can solve complex problems easily and simply. Machine Learning is unconsciously often encountered in everyday life such as Disease Diagnosis, Spam Detection, Face Recognition, market analysis, NPCs in games.

The dataset that you want to test is collected and chooses a good method to use such as SVM, Linear Regression, Neural Network, Logistic Regression, and others which are then trained on the method used and then evaluates the method used and makes predictions is how the system works. Machine Learning [10].

2.3. K-fold Cross Validation

K-fold cross-validation can be used to survey the accuracy of a model [7]. This method randomizes the input attributes and repeats to test a system can get some new random input attributes. K-fold cross-validation is an approach for data training and data testing. This method defines K as n, the size of the dataset. The advantage of K-fold cross-validation is to use as much data as possible for training data. K-fold cross-validation also has the disadvantage that it requires a lot of calculations to repeat the procedure n times.

2.4. Support Vector Regression (SVR)

Support Vector Regression is part of the Support Vector Machine. SVR can be used for regression cases where the output is a continuous number or a real number. SVR can overcome overfitting which makes it able to produce good performance [11]. Overfitting is a condition where a model does not describe the main relationship between variables.

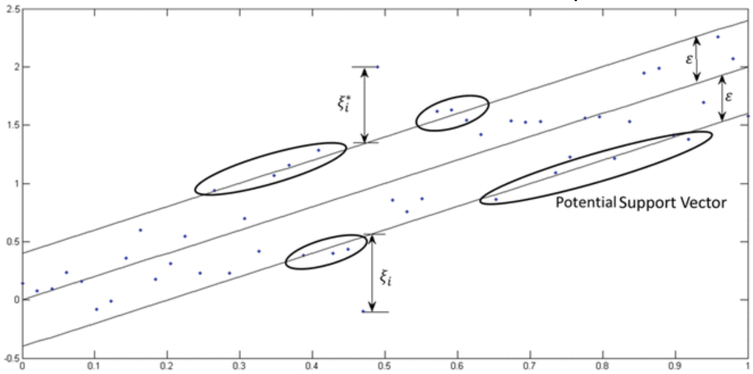


Figure 1. SVR illustration

Figure 1 describes an SVR illustration showing a hyperplane flanked by two boundaries + and a - boundary line. The distance between the hyperplane and the two boundary lines is indicated by . The circled data points are potential support, which means they are potential barriers so that data points can become one cluster by minimizing the value of .

2.5. Root Mean Square Error (RMSE)

Comparing the projected value with the original value or the studied value can provide an explanation of the accuracy of the contents / overall forecasting model carried out such as moving average and exponential smoothing [7]. RMSE has the following formula:

$$\text{RMSE} = \sqrt{\text{MSE}} \qquad (1)$$

MSE (Mean Squared Error) is the average number of squared forecasting errors [12], with the following formula:

$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_i - \acute{y}_i)^2 \qquad (2)$$

The smaller the RMSE value, the better the prediction accuracy level. Prediction has an accuracy value where this accuracy is determined from each accuracy method which has the smallest value [13]. Finding the best level of accuracy can be done by using several different prediction techniques and seeing the best result is the one with the smallest RMSE value.

3. RESULTS AND DISCUSSION

3.1. Data Collection Procedure

The data used in this study is time-series data in the form of datasets obtained from https://id.investing.com/commodities/gold-historical-data. The dataset used in this study is the gold futures price dataset from January 2013 to September 2014.

Table 1. Gold Price Example

Date	Close	Open	High	Low	Volume
02/01/2013	1687.90	1672.80	1693.80	1670.00	0.04
03/01/2013	1673.70	1686.10	1686.80	1662.00	0.14
04/01/2013	1648.10	1647.00	1658.30	1625.70	0.20
07/01/2013	1645.50	1656.50	1659.90	1643.80	0.05
08/01/2013	1661.50	1647.70	1659.50	1647.70	0.02
09/01/2013	1654.80	1658.60	1662.10	1651.20	0.01
10/01/2013	1677.30	1672.50	1673.80	1672.50	0.01
11/01/2013	1660.00	1673.10	1673.10	1654.20	0.15
14/01/2013	1668.90	1665.40	1673.80	1664.10	0.01
15/01/2013	1683.40	1671.60	1684.30	1671.60	0.02
16/01/2013	1682.70	1674.40	1680.50	1674.40	0.01
17/01/2013	1690.40	1676.40	1694.60	1671.00	0.08
18/01/2013	1686.60	1687.40	1692.00	1687.30	0.01
22/01/2013	1692.80	1690.30	1694.70	1687.80	0.02
23/01/2013	1686.30	1691.30	1693.70	1685.20	0.03
24/01/2013	1669.50	1680.80	1681.70	1668.00	0.01
25/01/2013	1656.40	1663.00	1663.00	1657.20	0.04
28/01/2013	1652.40	1651.60	1655.50	1651.60	0.08
29/01/2013	1660.70	1661.20	1661.50	1660.70	0.02
30/01/2013	1679.90	1662.40	1683.20	1661.80	44.89

Table 1 is an example of a gold futures price dataset in January 2013 that will be used.

3.1.1. Data Preparation

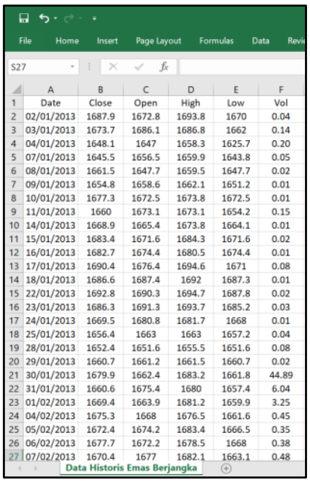
From the dataset that was downloaded and then opened with the Microsoft Excel application, the dataset was obtained in csv form. The dataset is still raw because it has not been checked and the data has not been preprocessed for further use in the Modeling stage. After the dataset is obtained, then the column separation is carried out so that the data can be seen easily.

	A	B	C	D	E	F	G	H
1	Tanggal,"Terakhir","Pembukaan","Tertinggi","Terendah","Vol.","Perubahan%"							
2	30/09/2014,"1.366,70","1.214,80","1.218,20","1.204,00","0,68K","-0,51%"							
3	29/09/2014,"1.373,70","1.216,30","1.222,20","1.214,70","6,34K","0,24%"							
4	26/09/2014,"1.370,40","1.214,10","1.214,10","1.214,10","0,00K","-0,44%"							
5	25/09/2014,"1.376,40","1.209,50","1.216,10","1.209,50","0,01K","0,11%"							
6	24/09/2014,"1.374,90","1.216,80","1.216,80","1.216,80","0,00K","-0,17%"							
7	23/09/2014,"1.377,20","1.221,00","1.221,00","1.221,00","0,01K","0,22%"							
8	22/09/2014,"1.374,20","1.214,90","1.214,90","1.212,40","0,01K","0,08%"							
9	19/09/2014,"1.373,10","1.218,30","1.218,30","1.218,30","0,00K","-0,74%"							
10	18/09/2014,"1.383,40","1.216,20","1.220,40","1.216,20","0,01K","-0,62%"							
11	17/09/2014,"1.392,00","1.236,20","1.236,20","1.221,90","0,01K","-0,07%"							
12	16/09/2014,"1.393,00","1.238,30","1.238,30","1.235,90","0,00K","0,12%"							
13	15/09/2014,"1.391,30","1.226,70","1.235,00","1.226,70","0,02K","0,25%"							
14	12/09/2014,"1.387,80","1.227,00","1.227,00","1.227,00","0,05K","-0,54%"							
15	11/09/2014,"1.395,40","1.237,40","1.237,40","1.237,40","-","-0,47%"							
16	10/09/2014,"1.402,00","1.256,70","1.256,70","1.243,50","0,10K","-0,19%"							
17	09/09/2014,"1.404,60","1.246,80","1.246,80","1.246,80","-","-0,40%"							
18	08/09/2014,"1.410,20","1.268,40","1.268,40","1.262,20","0,00K","-0,91%"							
19	05/09/2014,"1.423,20","1.263,20","1.268,10","1.263,20","0,12K","0,06%"							
20	04/09/2014,"1.422,30","1.268,70","1.272,70","1.260,30","0,29K","-0,22%"							
21	03/09/2014,"1.425,50","1.266,30","1.271,10","1.263,10","0,28K","0,46%"							
22	02/09/2014,"1.419,00","1.284,50","1.286,40","1.262,50","0,28K","10,22%"							
23	01/09/2014,"1.287,40","1.287,60","1.290,90","1.285,90","-","-0,02%"							
24	31/08/2014,"1.287,70","1.288,50","1.289,00","1.286,80","-","-10,61%"							
25	29/08/2014,"1.440,60","1.289,50","1.289,70","1.284,80","0,12K","-0,22%"							
26	28/08/2014,"1.443,80","1.284,50","1.293,70","1.284,30","0,82K","0,48%"							
27	27/08/2014,"1.436,90","1.281,50","1.282,30","1.281,50","0,04K","-0,17%"							

(a) (b)

Figure 2. The dataset obtained is downloaded and opened in the form of .csv (a), the data has been separated by column

After the column separation is done, the dataset still has attributes that are not needed in the modeling process, namely the "Change%" attribute so that the Data Reduction process is carried out, namely by removing unnecessary attributes, the time sequence also shows the time from the newest to the oldest so it must be changed to time. oldest to newest and corrects the letters so that the numbers used are numeric. So you get:



	A	B	C	D	E	F
1	Date	Close	Open	High	Low	Vol
2	02/01/2013	1687.9	1672.8	1693.8	1670	0.04
3	03/01/2013	1673.7	1686.1	1686.8	1662	0.14
4	04/01/2013	1648.1	1647	1658.3	1625.7	0.20
5	07/01/2013	1645.5	1656.5	1659.9	1643.8	0.05
6	08/01/2013	1661.5	1647.7	1659.5	1647.7	0.02
7	09/01/2013	1654.8	1658.6	1662.1	1651.2	0.01
8	10/01/2013	1677.3	1672.5	1673.8	1672.5	0.01
9	11/01/2013	1660	1673.1	1673.1	1654.2	0.15
10	14/01/2013	1668.9	1665.4	1673.8	1664.1	0.01
11	15/01/2013	1683.4	1671.6	1684.3	1671.6	0.02
12	16/01/2013	1682.7	1674.4	1680.5	1674.4	0.01
13	17/01/2013	1690.4	1676.4	1694.6	1671	0.08
14	18/01/2013	1686.6	1687.4	1692	1687.3	0.01
15	22/01/2013	1692.8	1690.3	1694.7	1687.8	0.02
16	23/01/2013	1686.3	1691.3	1693.7	1685.2	0.03
17	24/01/2013	1669.5	1680.8	1681.7	1668	0.01
18	25/01/2013	1656.4	1663	1663	1657.2	0.04
19	28/01/2013	1652.4	1651.6	1655.5	1651.6	0.08
20	29/01/2013	1660.7	1661.2	1661.5	1660.7	0.02
21	30/01/2013	1679.9	1662.4	1683.2	1661.8	44.89
22	31/01/2013	1660.6	1675.4	1680	1657.4	6.04
23	01/02/2013	1669.4	1663.9	1681.2	1659.9	3.25
24	04/02/2013	1675.3	1668	1676.5	1661.6	0.45
25	05/02/2013	1672.4	1674.2	1683.4	1666.5	0.35
26	06/02/2013	1677.7	1672.2	1678.5	1668	0.38
27	07/02/2013	1670.4	1677	1682.1	1663.1	0.48

Figure 3. Dataset after reduction

Figure 3 is a display of the dataset after Data Reduction has been performed. This dataset has 452 data with attributes to be used are Open, High, Low, Close, and Volume. The dataset is ready to be carried out at the next stage.

3.1.2. Data Cleaning

The existing dataset still has Missing Value. The missing value in the Volume attribute occurs because the missing value line date is a national holiday date so that there are no gold buying and selling transactions and no recorded transaction volume. Therefore, data that has a missing value will be deleted.

423	28/08/2014	1443.8	1284.5	1293.7	1284.3	0.82
424	29/08/2014	1440.6	1289.5	1289.7	1284.8	0.12
425	31/08/2014	1287.7	1288.5	1289	1286.8	-
426	01/09/2014	1287.4	1287.6	1290.9	1285.9	-
427	02/09/2014	1419	1284.5	1286.4	1262.5	0.28
428	03/09/2014	1425.5	1266.3	1271.1	1263.1	0.28
429	04/09/2014	1422.3	1268.7	1272.7	1260.3	0.29
430	05/09/2014	1423.2	1263.2	1268.1	1263.2	0.12

(a)

167	28/08/2013	1419	1417.3	1428	1415.1	0.02
168	29/08/2013	1412.9	1417.4	1417.4	1404.5	0.38
169	30/08/2013	1396.1	1407.2	1409.8	1392.3	0.10
170	03/09/2013	1412	1393	1414.4	1382.3	0.14
171	04/09/2013	1389.9	1413	1413	1387.1	0.10
172	05/09/2013	1373.1	1393.5	1397.5	1367.1	0.04

(b)

Figure 4. Dataset that still has Missing Value in Volume attribute (a), Dataset after Cleaning process (b)

After the Data Cleaning process was carried out, it was found that the amount of existing data amounted to 434 data, which means this dataset has 18 data that have missing values.

3.1.3. Data Split

After doing Data Cleaning, the dataset is then separated into training data and testing data with a percentage of 70% training data and 30% testing data. The separation process is done manually by taking the last 30% of data as testing data and the remaining 70% as training data and then saving it in .xlsx format. After the separation process, the amount of training data obtained is 303 data and testing data is 131 data.

3.1.4. Data Transformation

Training data and testing data have been created and then normalized. Before normalization is carried out, feature selection is carried out, namely for X variables, namely Open, High, Low, Close with labels/targets Close and Y variables being Open, High, Low, Close, and Volume with labels/targets Close. Normalization is intended to balance the value between existing features so that no feature dominates. In normalizing this data, use MinMaxScaler or range transformation with a minimum range of 0.0 and a maximum range of 1.

3.2. Modelling

This modeling stage uses the Support Vector Regression algorithm and cross validation for the training process on the training data. The preprocessed training data will be carried out 10-fold cross validation using linear SVR, RBF, and Polynomial kernels with various parameters to be tested.

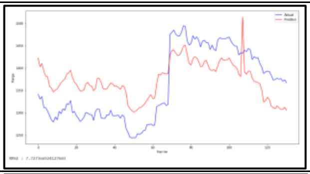
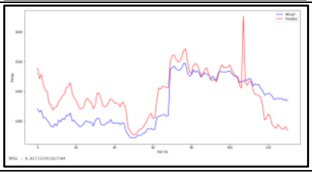
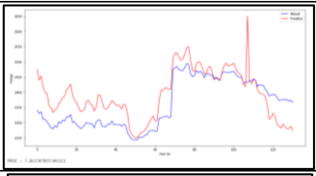
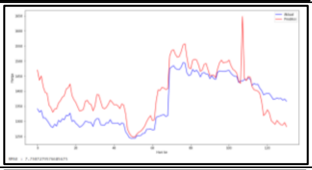
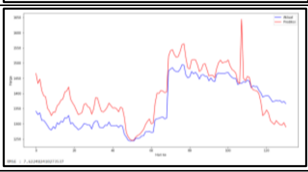
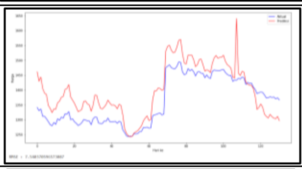
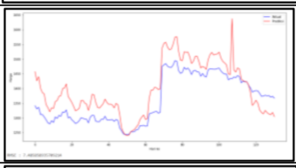
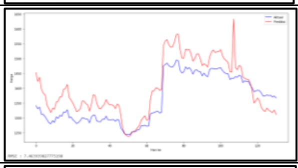
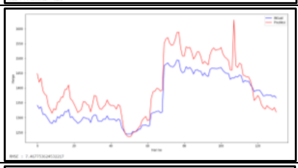
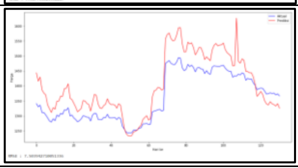
3.3. Evaluation

The evaluation will see how the model that has been made can predict the testing data and see the results of the RMSE. Experiments carried out using training data as many as 303 data that were preprocessed with the attributes of the X variable, namely Open, High, Low, Close with the Close label and the Y variable, namely Open, High, Low, Close, and Volume with the Close label and cross validation experiments with testing linear kernel, RBF, and Polynomial with the parameters that have been described in the Modeling stage are then used to predict the testing data.

3.3.1. SVR Linear Kernel Prediction with Variable X Dataset

The first experiment is to use a linear kernel on SVR with a dataset of X variables to predict the Close price and evaluate it with visualization and RMSE results. Experiment details for each kernel parameter can be seen in the following results:

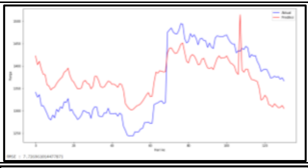
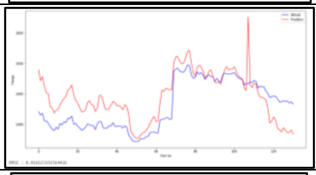
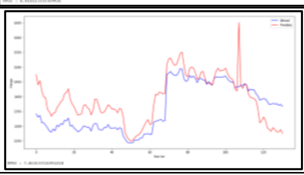
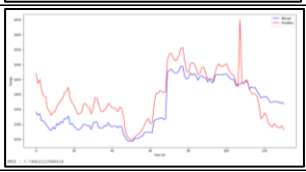
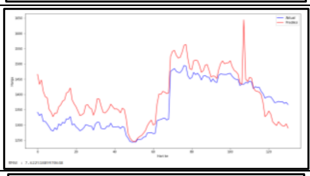
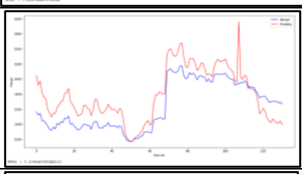
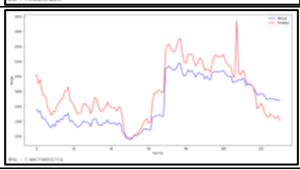
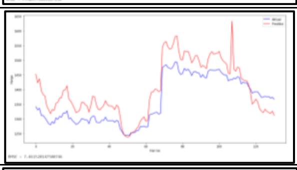
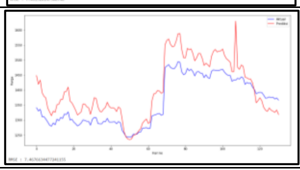
Table 2. Variable X prediction graph using Linear Kernel

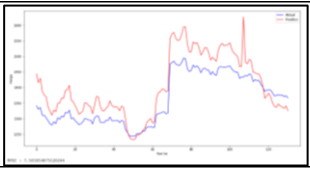
Prediction Result Chart	Graphic Description
	[Prediction of variable X using linear kernel with parameter C = 1] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 1 produces an RMSE of 7.727366924127603.
	[Prediction of variable X using linear kernel with parameter C = 5] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 5 produces an RMSE of 8.017355993267504.
	[Prediction of variable X using linear kernel with parameter C = 10] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 10 produces an RMSE of 7.863307893346111.
	[Prediction of variable X using linear kernel with parameter C = 15] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 15 produces an RMSE of 7.7307279576685675.
	[Prediction of variable X using linear kernel with parameter C = 20] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 20 produces an RMSE of 7.622482410273537.
	[Prediction of variable X using linear kernel with parameter C = 25] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 25 produces an RMSE of 7.540370596573087.
	[Prediction of variable X using linear kernel with parameter C = 30] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 30 produces an RMSE of 7.485858935709214.
	[Prediction of variable X using linear kernel with parameter C = 35] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 35 produces an RMSE of 7.463939827775198.
	[Prediction of variable X using linear kernel with parameter C = 40] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 40 produces an RMSE of 7.467753624532217.
	[Prediction of variable X using linear kernel with parameter C = 45] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 45 produces an RMSE of 7.503942710051336.

3.3.2. SVR Linear Kernel Prediction with Variable Y Dataset

The next experiment is to use a linear kernel on the SVR with the Y variable dataset to predict the Close price and evaluate it with visualization and RMSE results. Experiment details for each kernel parameter can be seen in the following results:

Table 3. Variable Y prediction graph using Linear Kernel

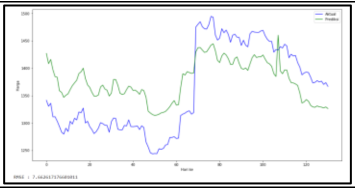
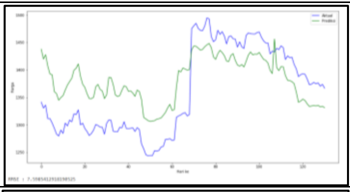
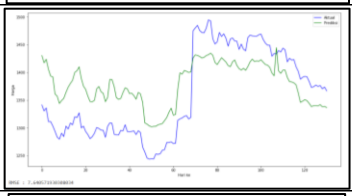
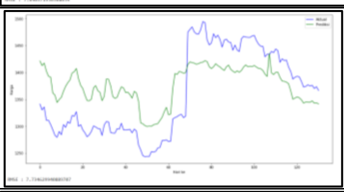
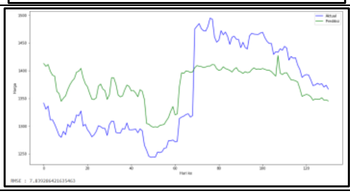
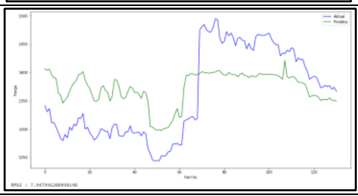
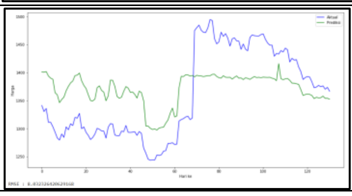
Prediction Result Chart	Graphic Description
	[Prediction of variable Y using linear kernel with parameter C = 1] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 1 produces an RMSE of 7.726961014477871.
	[Prediction of variable Y using linear kernel with parameter C = 5] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 5 produces an RMSE of 8.016623192564426.
	[Prediction of variable Y using linear kernel with parameter C = 10] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 10 produces an RMSE of 7.863633516992218.
	[Prediction of Y variable using linear kernel with parameter C = 15] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 15 produces an RMSE of 7.730621127044618.
	[Prediction of variable Y using linear kernel with parameter C = 20] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 20 produces an RMSE of 7.62251889970648.
	[Prediction of variable Y using linear kernel with parameter C = 25] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 25 produces an RMSE of 7.53960359504133.
	[Prediction of variable Y using linear kernel with parameter C = 30] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 30 produces an RMSE of 7.484779009117156.
	[Prediction of variable Y using linear kernel with parameter C = 35] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 35 produces an RMSE of 7.461520147500746.
	[Prediction of variable Y using linear kernel with parameter C = 40] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 40 produces an RMSE of 7.4676634477241155.

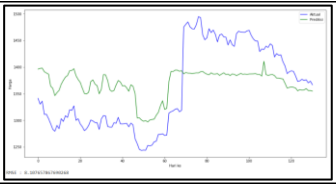
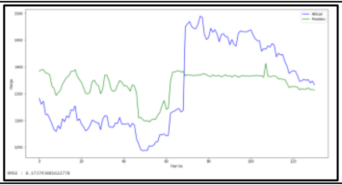
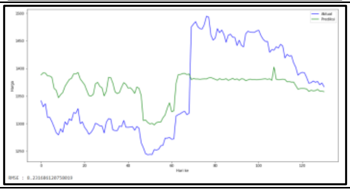
	[Prediction of variable Y using linear kernel with parameter C = 45] Visualization of the predicted Close price (red) and the actual Close price (blue). A linear kernel with parameter C = 45 produces an RMSE of 7.503854075620244.

3.3.3. SVR RBF Kernel Prediction with Variable X Dataset

The third experiment is to use the RBF kernel on SVR with a dataset of X variables to predict the Close price and evaluate it with visualization and RMSE results. Experiment details for each kernel parameter can be seen in the following results:

Table 4. Variable X prediction graph using Kernel RBF

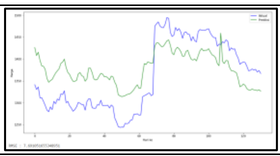
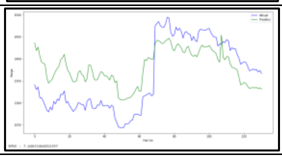
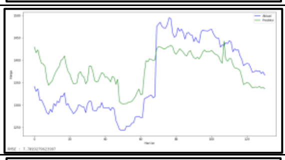
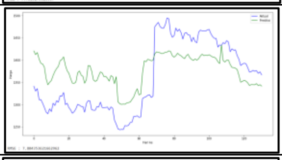
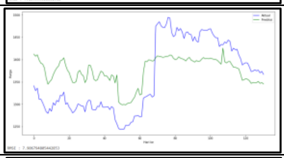
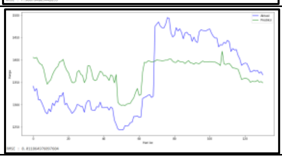
Prediction Result Chart	Graphic Description
	[Prediction of variable X using RBF kernel with parameter = 1] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 1 produces RMSE of 7.662617176601011.
	[Prediction of variable X using RBF kernel with parameter = 2] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 2 produces RMSE of 7.5985412918190525.
	[Prediction of variable X using RBF kernel with parameter = 3] visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 3 produces RMSE of 7.640571930308034 .
	[Prediction of variable X using RBF kernel with parameter = 4] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 4 produces RMSE of 7.734629940889707.
	[Prediction of variable X using RBF kernel with parameter = 5] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 5 produces RMSE of 7.839286421635463.
	[Prediction of variable X using RBF kernel with parameter = 6] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 6 produces RMSE of 7.947996280498148.
	[Prediction of variable X using RBF kernel with parameter = 7] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 7 produces RMSE of 8.032326420629168.

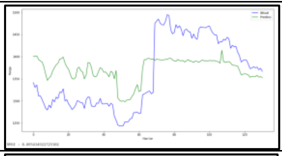
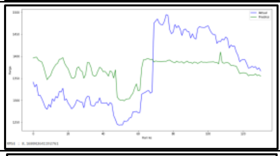
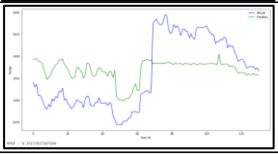
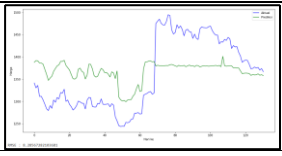
	[Prediction of variable X using RBF kernel with parameter = 8] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 8 produces RMSE of 8.107657867690268.
	[Prediction of variable X using RBF kernel with parameter = 9] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 9 produces RMSE of 8.173793881622778.
	[Prediction of variable X using RBF kernel with parameter = 10] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 10 produces RMSE of 8.231686120758019.

3.3.4. SVR RBF Kernel Prediction with Variable Y Dataset

The next experiment is to use the RBF kernel on the SVR with the Y variable dataset to predict the Close price and evaluate it with visualization and RMSE results. Experiment details for each kernel parameter can be seen in the following results:

Table 5. Variable Y prediction graph using Kernel RBF

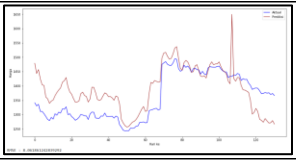
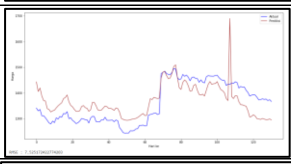
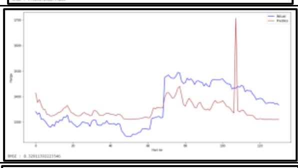
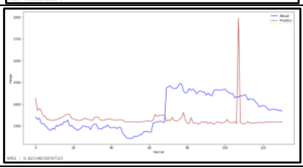
Prediction Result Chart	Graphic Description
	[Prediction of Y variable using RBF kernel with parameter = 1] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 1 produces RMSE of 7.691051655348951.
	[Prediction of Y variable using RBF kernel with parameter = 2] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 2 produces RMSE of 7.64655464852397.
	[Prediction of Y variable using RBF kernel with parameter = 3] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 3 produces RMSE of 7.7019279623997.
	[Prediction of Y variable using RBF kernel with parameter = 4] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 4 produces RMSE of 7.804353621662962.
	[Prediction of Y variable using RBF kernel with parameter = 5] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 5 produces RMSE of 7.906794805442853.
	[Prediction of Y variable using RBF kernel with parameter = 6] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 6 produces RMSE of 8.011864976097604 .

	[Prediction of Y variable using RBF kernel with parameter = 7] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 7 produces RMSE of 8.095434922725302.
	[Prediction of Y variable using RBF kernel with parameter = 8] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 8 produces RMSE of 8.168042641192761.
	[Prediction of Y variable using RBF kernel with parameter = 9] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 9 produces RMSE of 8.23137617247104.
	[Prediction of Y variable using RBF kernel with parameter = 10] Visualization of the predicted Close price (green) and the actual Close price (blue). RBF kernel with parameter = 10 produces RMSE of 8.28567282103681.

3.3.5. SVR Polynomial Kernel Prediction with Variable X Dataset

The fifth experiment is to use the Polynomial kernel on SVR with a dataset of X variables to predict the Close price and evaluate it with visualization and RMSE results. Experiment details for each kernel parameter can be seen in the following results:

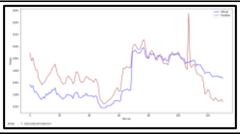
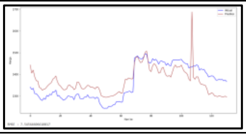
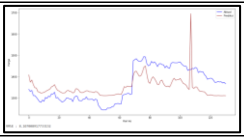
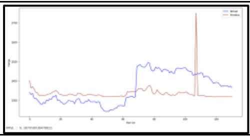
Table 6. Variable X prediction graph using Kernel Polynomial

Prediction Result Chart	Graphic Description
	[Prediction of variable X using Polynomial kernel with parameter d = 1] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 1 produces RMSE of 8.061861242839292.
	[Prediction of variable X using Polynomial kernel with parameter d = 2] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 2 produces RMSE of 7.525172422774203.
	[Prediction of variable X using Polynomial kernel with parameter d = 3] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 3 produces RMSE of 8.32911392223546 .
	[Prediction of variable X using Polynomial kernel with parameter d = 4] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 4 produces RMSE of 9.823348330747315.

3.3.6. SVR Polynomial Kernel Prediction with Variable Y Dataset

The last experiment is to use Polynomial kernel on SVR with Y variable dataset to predict Close price and evaluate it with visualization and RMSE results. Experiment details for each kernel parameter can be seen in the following results:

Table 7. Variable Y prediction graph using Kernel Polynomial

Grafik	Keterangan
	[Prediction of Y variable using Polynomial kernel with parameter d = 1] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 1 produces RMSE of 7.992198245584397.
	[Prediction of Y variable using Polynomial kernel with parameter d = 2] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 2 produces RMSE of 7.54564606610817.
	[Prediction of Y variable using Polynomial kernel with parameter d = 3] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 3 produces RMSE of 8.107008917733152.
	[Prediction of Y variable using Polynomial kernel with parameter d = 4] Visualization of the predicted Close price (brown) and the actual Close price (blue). Polynomial kernel with parameter d = 4 produces RMSE of 9.187958938478111.

3.4. Detailed Linear Kernel, RBF and Polynomial Prediction Results

Table 8. Detailed Linear Kernel

RMSE Kernel Linear		
C	Variable X	Variable Y
1	7.727366924127603	7.726961014477871
5	8.017355993267504	8.016623192564426
10	7.863307893346111	7.863633516992218
15	7.7307279576685675	7.730621127044618
20	7.622482410273537	7.62251889970648
25	7.540370596573087	7.53960359504133
30	7.485858935709214	7.484779009117156
35	7.463939827775198	7.461520147500746
40	7.467753624532217	7.46766344772411555
45	7.503942710051336	7.503854075620244

Table 8 shows the overall results of the RMSE SVR Linear kernel with its parameters. Parameter C of 35 shows the best RMSE results for the X variable dataset and Y variable dataset.

Table 9. Detailed RBF Kernel

RMSE Kernel RBF		
γ	Variable X	Variable Y
1	7.662617176601011	7.691051655348951
2	7.5985412918190525	7.64655464852397
3	7.640571930308034	7.7019279623997
4	7.734629940889707	7.804353621662962
5	7.839286421635463	7.906794805442853
6	7.947996280498148	8.011864976097604
7	8.032326420629168	8.095434922725302
8	8.107657867690268	8.168042641192761

9	8.173793881622778	8.23137617247104
10	8.231686120758019	8.28567282103681

Table 9 shows the overall results of the RMSE SVR kernel RBF with its parameters. The parameter of 2 shows the best RMSE results for the variable X dataset of 7.5985412918190525 and the Y variable dataset of 7.64655464852397.

Table 10. Detailed Polynomial Kernel

RMSE Kernel Polynomial		
d	Variable X	Variable Y
1	8.061861242839292	7.992198245584397
2	7.525172422774203	7.54564606610817
3	8.32911392223546	8.107008917733152
4	9.823348330747315	9.187958938478811

Table 10 shows the overall results of the RMSE SVR Polynomial kernel with its parameters. Parameter d of 2 shows the best RMSE results for the variable X dataset of 7.525172422774203 and the Y variable dataset of 7.54564606610817.

4. CONCLUSION

Based on the results of research and discussion conducted previously, the following conclusions can be obtained:

1. Support Vector Regression can be used to predict gold prices. The selection of features as well as the kernel and its parameters greatly affect the prediction results.
2. The best RMSE accuracy results for the Linear kernel are using a dataset of variables Y (Open, High, Low, Close, and Volume) of 7.461520147500746 with C of 35.
3. The best RMSE calculation result for the RBF kernel is to use the X(Open, High, Low, Close) variable dataset of 7.5985412918190525 with of 2.
4. The best RMSE result for the Polynomial kernel is to use the variable X(Open, High, Low, Close) dataset of 7.525172422774203 with d of 2.
5. Based on the best RMSE for each kernel, the best RMSE is to use a Linear kernel with a C of 35 and a Y variable dataset.

REFERENCES

If references are Journal:

[1] Dwi, N. (2015). Penerapan Algoritma Support Vector Machine untuk Prediksi Harga Emas. Jurnal Informatika Upgris, 1(1), 10–19. https://doi.org/10.26877/jiu.v1i1 Juni.805

[2] Kusumodestoni, R. H., & Sarwido, Komparasi Support Vector Machines (SVM) dan Neural Network Untuk Mengetahui Tingkat Akurasi Prediksi Tertinggi Harga Saham, 2017.

[3] Guntur, S. R., Kumar, N. S. S., Hegde, M., & Dirsala, V, R, In Vitro Studies of the Antimicrobial and Free-Radical Scavenging Potentials of Silver Nanoparticles Biosynthesized From the Extract of Desmostachya bipinnata, 2018.

[4] Yang, Y. and Liu, 1999, â€œProceedings of ACM SIGIR Conference on Research and Development in Information Retrieval â€œ, (SIGIR'99, pp 42--49).

[5] 28. Kim S, Jhong JH, Lee J, Koo JY. Meta-analytic support vector machine for integrating multiple omics data. BioData Min. 2017;10(1):2.

[6] Enri, U. (2018). Optimasi Parameter Support Vector Machines Untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika Serikat. Jurnal Gerbang, 8(1), 12–16.

[7] Pudjianto, E., & Supriyanto, C. (2015). Prediksi harga komoditas emas dan batubara di pasar dunia dengan algoritma support vector machine. 11(April), 13–23.

[8] Sari, E. P. (2016). Model Prediksi Harga Saham Media Sosial Berdasarkan Algoritma Svm Yang Dioptimasikan Dengan Pso. None, 12(2), 161–170.

[9] Panggabean, V., Nababan, E., & Bu, F. (2013). Analisis Fundamental Dan Analisis Teknikal Pada Investasi Trading Emas Online Dengan Value At Risk. Saintia Matematika, 1(4), 369–382.

[10] Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. October.

[11] Yasin, H., Prahutama, A., & Utami, T. W. (2014). Prediksi Harga Saham Menggunakan Support Vector Regression Dengan Algoritma Grid Search. Media Statistika, 7(1), 29–35. https://doi.org/10.14710/medstat.7.1.29-35

[12] Sari, Y. (2018). Prediksi Harga Emas Menggunakan Metode Neural Network Backropagation Algoritma Conjugate Gradient. Jurnal ELTIKOM, 1(2), 64–70. https://doi.org/10.31961/eltikom.v1i2.21

[13] Azam, D. F., Ratnawati, D. E., & Adikara, P. P. (2018). Prediksi Harga Emas Batang Menggunakan Feed Forward Neural Network Dengan Algoritme Genetika. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, 2(8), 2317–2322.

[14] Zulfikar, W. B., & Lukman, N. (2016). Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata. Jurnal Online Informatika, 1(2), 82–86. https://doi.org/10.15575/join.v1i2.33