

# The Prediction of Gold Price Movement by Comparing Naive Bayes, Support Vector Machine, and K-NN

Yahya Suryana<sup>1\*</sup>, Tjong Wan Sen<sup>2</sup>

<sup>1,2</sup>:informatics Engineering Study Program, Faculty of Computer, President University Email:  
[yahya.surryana@student.president.ac.id](mailto:yahya.surryana@student.president.ac.id), [wansen@president.ac.id](mailto:wansen@president.ac.id)

**Abstract** – Gold is a yellow precious metal that can be forged so it is easy to form with various forms of jewelry such as pendants, earrings, rings, bracelets and others, gold has a high value. Gold itself is an exchange rate used in ancient times before the existence of money as it is today. Gold also can be used as an investment that is profitable for the investor and it has less risks. Investment is a form of fund management to give benefit by putting fund in allocation that is predicted will give additional benefits. Prediction of gold price movements or predictions of gold price in gold stock investment, this research uses 3 (three) algorithms that will be implemented in analysis and increase accuracy, in the discussion or research that was made using the Naïve Bayes algorithm, Support Vector Machine and K-Nearest Neighbor, the dataset is obtained from the website, namely [www.finance.yahoo.com](http://www.finance.yahoo.com) the data was then tested using Rapid miner tools so that the average value of the Support Vector Machine algorithm with an accuracy rate of 57.59%, precision 58,73% and recall 51,78%. The next is the Naïve Bayes algorithm so that it is known to have an accuracy rate of 55.59%, precision 54.55% and recall 51.70%. Based on the comparison of the three algorithms, it is known that the one with the best accuracy, precision, and recall is the K-NN algorithm with 61.90% accuracy, 60.98% precision, and 60.35% recall. Furthermore, the results of testing the K-Nearest Neighbor algorithm have good results compared to the 3 (three) other algorithm tests and the Naïve Bayes algorithm testing has a low level of accuracy, namely 55.59%, precision 54.55% and recall 51.70%. The research uses 3 algorithms, namely naive bayes, K-nearest neighbor and Support Vector Machine, because the three algorithms are well-established algorithms to be applied to research, especially in time series gold price research and are very good, especially for classification.

**Keywords** – Data Mining, Naïve Bayes, KNN, Support Vector Machine.

## I. INTRODUCTION

Gold is one of the most malleable and highly malleable yellow precious metals. Gold is the exchange rate used before the existence of money as it is today therefore the risk impact of investing in gold is very small, this yellow precious metal has two types, namely, gold for investment only, gold for jewelry only such as necklaces, rings, bracelets, earrings and others, gold is investment in gold stocks or gold futures. Investment process contains risk and uncertainty. The investment that everyone can do is a gold investment, so this so that gold becomes an investment that is in great demand and becomes a prima donna among the upper class, upper or lower middle, but basically in investing in gold this is when price fluctuations occur every day, every month or even every year, this fluctuation risk is called time series risk, where the price is always going up and down. In order not to happen and avoid risk, in this study a strong prediction is made about the price of gold by using time series data so that gold investors can know when to invest and when to resell so that they can provide benefits for gold investors according to the plans that have been made, This gold price forecast is made so that investors get profits and according to the plans made in this research. This gold precious metal investment is more profitable, However, investors must know that there are several factors that influence when the gold exchange rate occurs Gold is one of the most valuable commodities and is traded in various parts of the world, such as one of the leading countries,

namely Saudi Arabia, with the most investors reaching 70%. Gold is synonymous with a symbol of luxury and glory, but gold is vulnerable and heavily influenced by various economic indicators such as interest rate, inflation, and Bruto Domestic Product (Produk Domestik Bruto). Certain information is utilized its knowledge. One of the approach that can be used to analyze a group. Several research methods have been carried out to predict gold prices, one of which is the Prediction of Gold Price Movements in Gold Stock Investments By Comparing the Naïve Bayes Algorithm, KNN and Support Vector Machine[1]. The Naïve Bayes algorithm has been used to predict gold prices as was done by Mohammad Guntur, Yulius Santony and Yuhandri in 2018 with the title Gold Price Prediction Using The Naïve Bayes Method In Investing To Minimize Risk[2] with the results of his research, namely predicting the price of gold which can help make decisions in determining whether to sell or give gold to predict the price of gold for the next 14 days The data used for testing are 16 data and an accuracy of 75% [3] is obtained, but Naïve Bayes has many deficiencies in classification problems, classification so that further research is needed to improve accuracy. In this study, several tests were carried out on the price of gold by adding a dataset so as to get high accuracy Based on the brief discussion of the above problems, predictions are made by increasing accuracy of the gold price prediction, This research uses gold price data with of gold time series data obtained from a time series gold data website, namely [www.finance.yahoo.com](http://www.finance.yahoo.com) and processing gold data so that



the results of gold price predictions can be known. Based on the explanation described above, this research was conducted so that compare Naïve Bayes[4], Support Vector Machine[5], and K-Nearest Neighbor algorithm[6],[8],[9],[10].

## II. RESEARCH METHODOLOGY

In this study, analysis and methodology were carried out on the gold time series datafacilites research and able to run systemmatically fulfill the purpose as expected therefore steps in research research stage is made and will be run as follows

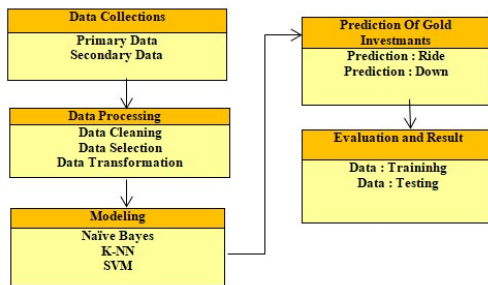


Figure 1 Research Stages

### 2.1. Data collection

At the stage of data collection, research is carried out using 2 (two) references to the types of data obtained, including:

#### 1. Primary Data

The data taken in this study is data from a website [www.finance.yohoo.com](http://www.finance.yohoo.com), namely the price of gold time series for 5 (five) years, from 2014 to 2019, this data is used as study material for learning and training in research. next.

At this stage, namely taking existing references, namely looking for references in books or journals and also on papers that are similar to the research being conducted.

### 2.2 Data Processing

At this stage, data processing is carried out by pre-processing the data to be processed. The data obtained in this study are time series gold price data taken from a website [www.finance.yahoo.com](http://www.finance.yahoo.com) as an initial stage, the processing is carried out in 3 stages, the first stage is the elimination of some noise data, then the second stage namely dividing into training data and testing data, 90% for training and 10% for testing data. The third stage of processing is carried out using rapidminer tools to find accuracy, precision and recall on rapidminer tools.

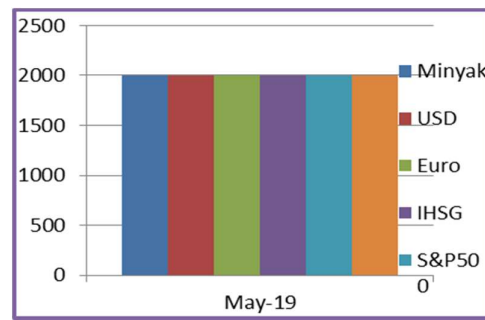


Figure 2 Dataset Graph

At the data cleaning stage, it aims to clean up inconsistent data empty values or commonly called empty tuples, duplicate data and correct data errors, the data repair process is carried out manually, with the help of spreadsheet tools

### 2.4 Data Selection

Data selection is the process of selecting data from existing operational data before entering the data and information mining stage. At this stage, the following steps will be carried out:

1. The data sample is taken randomly with the attribute parameters on the data [www.finance.yahoo.com](http://www.finance.yahoo.com). The document that has the largest amount of data to be used as a dataset and ensures that the selected data is suitable for use in the modeling process.
2. After viewing this dataset, you will get the amount of testing data.
3. Choose the attributes that will be used and analyzed, because in the initial data there are some unneeded attributes such as attributes.

Table 1 Data and Attribute Used

No	Attribute						
	Date	Oil	USD	Euro	IHSG	S&P500	Gold
1	31/05/19	53.5	14385	15,964.	6209	2,752.0	1,30
2	30/05/19	56.5	14385	16,027	6209	2,788.8	1,28
3	14/05/10	58.8	14478	16,125	5907	2,840.2	1,27

### 2.5 Transformation Data

The data transformation stage is the process of changing the initial data format into a standard data format for the process of reading data using the algorithms in the programs or tools used.

### 2.6 Modelling

The modeling in this study was carried out using data mining classification techniques for the Naïve Bayes algorithm, Support Vector Machine, and K-Nearest Neighbor. This technique was chosen because it is a commonly used method in data mining research to classify or recognize new data that has never been studied, especially in predicting gold price movements or predicting gold prices in gold stock investments. The The algorithm that will be used for the analysis of this research is the Naïve Bayes algorithm, Support Vector Machine,



and K-Nearest Neighbor [11],[12],[13],14]. These 3 (three) algorithms are algorithms that have been established and is widely implemented in classification techniques. In addition, this algorithm has advantages, namely in the form of good accuracy in handling a processed dataset

### 2.7 Testing and Validation of Reseach

Method testing is carried out with the aim of knowing the results of the analyzed calculations and measuring the methods and algorithms used whether they function properly or not. The testing process uses the rapidminer tool and sees whether the data is in accordance with the results obtained through the tool. While the validation of the methods and algorithms of Naïve Bayes[15],[16],[17], Support Vector Machine[18],[19],[20], and K-Nearest Neighbor[21],[22] is done by measuring the results of accuracy, percision and recall and can be calculated using the Confusion Matrix as follows:

### 2.8 Proposed Method

The method of this research uses 3 (three) algorithms, namely K-NN, Naive Bayes and Support Vector Machine to train data accuracy in time series gold price prediction assisted by rapid miner software.

### 2.9 K-fold Cross Validation

Using K-Fold Cross-Validation for statistical analysis that will be generalized to independent data sets (Suyanto, 2019). This technique is mainly used to make model predictions and estimate how accurate a predictive model will be when run in practice. The purpose of defining to test the model phase is, data validation, to limit problems such as the occurrence of such To provide insight into how the model will generalize independently of the dataset (i.e., unknown dataset, for example from a real problem), the following is an example of a data iteration table on K-Fold Cross Validation.

Pengujian Ke	Dataset									
1	■									
2		■								
3			■							
4				■						
5					■					
6						■				
7							■			
8								■		
9									■	
10										■
	Data Testing	■	Data Training							

Figure 3 Testing With 10 K-fold Cross Validation

#### 1. KNN

The accuracy value is calculated by adding up the correct data, namely positive (True Positive) plus true negative (True Negative) divided by the number of correct data, namely true positive (True Positive), true negative

(True Negative) and added by the false data which is positive ( False Positive), Negative (False Negative).

Table 2 Confusion Matrix (K-NN) Accuracy Calculation Formula

Score Prediction	True Value	
	Calculation K-NN	
	True	False
True	TP 11	FP 3
False	FN 0	TN 0

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \cdot 100 \quad (1) \\
 &= \frac{11 + 0}{11 + 0 + 3 + 0} * 100\% \\
 &= \frac{11}{14} * 100\% \\
 &= 0,7857142857 * 100\% \\
 &= 78,57\%
 \end{aligned}$$

The precision value is calculated by dividing the

number of true positive data (True Positive) divided by the number of true positive data (True Positive) and false positive data (False Positive).

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} \cdot 100\% \quad (2) \\
 &= \frac{11}{11 + 3} * 100\% \\
 &= \frac{11}{14} * 100\% \\
 &= 0,7857142857 * 100\% \\
 &= 78,57\%
 \end{aligned}$$

The recall value is calculated by dividing the correct

data, which is true positive (True Positive) and divided by the number of correct data, namely true positive (True Positive) and false data, namely false negative (False Negative).

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} * 100\% \quad (3) \\
 &= \frac{11}{11 + 0} * 100\% \\
 &= \frac{11}{11} * 100\% \\
 &= 1 * 100\%
 \end{aligned}$$



$$= 100\%$$

The accuracy value is calculated by adding up the correct data, namely true positive (True Positive) plus the

Tabel 3 Confusion Matrix SVM Precision Calculation SVM

true negative value (True Negative) also divided by the number of correct data, namely true positive (True Positive), true negative (True Negative) and added to the data False is false positive (False Positive), and the data is divided by false negative (False Negative).

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (4) \\ &= \frac{11 + 0}{11 + 0 + 3 + 0} * 100\% \\ &= \frac{11}{14} * 100\% \\ &= 0,7857142857 * 100\% \\ &= 78,57\% \end{aligned}$$

The precision value is calculated by dividing the number of true positive data (True Positive) divided by the number of true positive data (True Positive) and false positive data (False Positive)

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} * 100\% \quad (5) \\ &= \frac{11}{11 + 3} * 100\% \\ &= \frac{11}{14} * 100\% \\ &= 0,7857142857 * 100\% \\ &= 78,57\% \end{aligned}$$

The recall value will be calculated by dividing the correct data, namely true positive (True Positive) then dividing by the correct number of data, namely true positive (True Positive) and incorrect data, namely false negative (False Negative).

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} * 100\% \quad (6) \\ &= \frac{11}{11 + 0} * 100\% \\ &= \frac{11}{11} * 100\% \\ &= 1 * 100\% \\ &= 100\% \end{aligned}$$

The accuracy value is calculated by adding up the correct data which is positive (True Positive) plus the negative value (True Negative) divided by the number of correct data which is positive (True Positive), Negative (True Negative) and added by false data which is positive ( False Positive), Negative (False Negative).

Table 4 Confusion Matrix Accuracy Calculation Naïve bayes

Score Prediction	True Value	
	True	False
	Calculation Naïve Bayes	

Score Prediction	True Value	
	Calculation SVM	
	True	False
True	TP 7	FP 3
False	FN 4	TN 0

	True	False
True	TP 7	FP 3
False	FN 4	TN 0

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (7) \\ &= \frac{7 + 0}{7 + 0 + 3 + 4} * 100\% \\ &= \frac{7}{14} * 100\% \\ &= 0,5 * 100\% \\ &= 50\% \end{aligned}$$

The precision value is calculated by dividing the number of true positive data (True Positive) divided by the number of true positive data (True Positive) and false positive data (False Positive).

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} * 100\% \quad (8) \\ &= \frac{7}{7 + 3} * 100\% \\ &= \frac{7}{10} * 100\% \\ &= 0,7 * 100\% \\ &= 70\% \end{aligned}$$

The recall value is calculated by dividing the correct data which is positive (True Positive) with the sum of the correct data which is positive (True Positive) and the incorrect data which is negative (False)..



$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} * 100\% \\
 &= \frac{7}{7 + 4} * 100\% \\
 &= \frac{7}{11} * 100\% \\
 &= 0,636363664 * 100\% \\
 &= 63.64\%
 \end{aligned}
 \tag{9}$$

From the sample data as many as 14 data, then the results from these data state the level of accuracy, recall and persicion of the K-NN, SVM and Naïve Bayes algorithms. The following are the results of the accuracy, recall and persicion values

Table 5 Results of accuracy, recall and precision values

### III. RESULTS AND DISCUSSION

In this study using the SVM, Naïve Bayes, K-NN algorithm which will be tested so that it will get the results of accuracy, precision, and recall values as well as predictions that can be used in making decisions when investing in gold. The source of data as an object in this study is gold price data over a period of years The data used in this study consists of attributes or variables such as, Oil, USD, Euro, IHSG, S&P500, Gold, so that it can find out a result of knowledge in predicting gold data and in this test using 2000 data and the data is divided into two 90% training data and 10% testing data, so that it can produce models or values of accuracy, precision and recall obtained from the three algorithms to test gold data, the data used is divided randomly or randomly into subsets, namely t1, t2, t3, ... , t10 with the same data size.

From the 200 testing data and 1800 training data, data modeling will be formed to be tested by rapidminer in finding results so that they can find out the accuracy, precision and recall results. and compare the 3 algorithms that have good results or have high levels of accuracy, precision and recall

from the test results can produce information and knowledge in predicting gold investment. The following is Table 6 of the modeling process in testing gold data.

Tabel 6. 10 x test

No	Testing					
	10 x test					
Test 1	200	200	200	200	200	200
Test 2	200	200	200	200	200	200
Test 3	200	200	200	200	200	200
Test 4	200	200	200	200	200	200
Test 5	200	200	200	200	200	200
Test 6	200	200	200	200	200	200
Test 7	200	200	200	200	200	200
Test 8	200	200	200	200	200	200
Test 9	200	200	200	200	200	200
Test 10	200	200	200	200	200	200

Based on table 7 it is shown that the p value used by the testing process. Here are the steps for testing the data with 10 tests.

Table 7 Sharing of Training Data and Testing Data

No	Data Training and Data Testing	
	Data Training	Data Testing
Test 1	1800	200
Test 2	1800	200
Test 3	1800	200
Test 4	1800	200
Test 5	1800	200
Test 6	1800	200
Test 7	1800	200
Test 8	1800	200
Test 9	1800	200
Test 10	1800	200

The data used is divided into 10 parts, namely p1, p2, ..., p10 and training data, = (1, 2, 3, 4, 5, 6, 7, ..., 10) is used as testing data and others as training data.

Accuracy results are calculated for each test (test-1,

test-2, test-3, test-4, test-5, test-6, test-7, test-8, test-9, test-10) then calculate the average level of accuracy from

No	Result			
	Algorithm	Accuracy	Precision	Recall
1	K-NN	78.57%	78.57%	100%
2	SVM	78.57%	78.57%	100%
3	Naïve Bayes	50%	70%	63.64%

all tests to get the level of accuracy, precision and recall of the overall data

The data is saved in excel workbook format which is then converted into a data frame with the read excel command.

Table 8 Sample Data Testing

No	Attribute						
	Date	Oil	USD	Euro	IHSG	S&P500	Gold
1	31/05/19	53.5	14385	15,964.	6209	2,752.0	1,30
2	30/05/19	56.5	14385	16,027	6209	2,788.8	1,28
3	29/05/19	591	144	16,028	6104	2,783.0	1,28
4	28/05/19	591	143	16,043	6033	2,808.3	1,27
5	27/05/19	591	143	16,103	6098	2,802.1	1,28
6	26/05/19	587	143	16,129	6098	2,808.3	1,28
7	25/05/19	586	143	16,126	6098	2,802.1	1,28
8	24/05/19	586	144	16,126	6057	2,826.0	1,28
9	23/05/19	579	145	16,167	6032	2,822.2	1,28
10	22/05/19	614	144	16,199	5939	2,856.2	1,27
11	21/05/19	629	144	16,160	5951	2,864.3	1,27
12	20/05/19	631	144	16,147	5907	2,8402	1,27
13	19/05/19	627	144	16,129	5907	2,840.2	1,27
14	14/05/10	58.8	14478	16,125	5907	2,840.2	1,27

The training and testing data will be processed using the SVM, Naïve Bayes and K-NN algorithms, the data is tested with the rapidminer tools, this study will evaluate the classification results of the data, the real data that have been tested, so that the results of accuracy, recall and precision can be seen on tools rapidminer determining data result of powder coating production. Here is the



overall process of data testing by using rapid miner. From the testing result with 10 testings randomly has generates highest accuracy level, it is K-NN algorithm.

Table 9 Accuraci Value Result

No	SVM		Naïve Bayes		K-NN	
	Accuracy		Accuracy		Accuracy	
	Test	Value	Test	Value	Test	Value
1	Test 1	58.50%	Test 1	60.00%	Test 1	63.00%
2	Test 2	59.50%	Test 2	60.00%	Test 2	63.00%
3	Test 3	60.50%	Test 3	59.50%	Test 3	58.50%
4	Test 4	53.50%	Test 4	48.50%	Test 4	62.00%
5	Test 5	55.50%	Test 5	57.00%	Test 5	59.00%
6	Test 6	54.50%	Test 6	51.00%	Test 6	66.00%
7	Test 7	55.00%	Test 7	55.50%	Test 7	62.00%
8	Test 8	52.00%	Test 8	43.50%	Test 8	53.50%
9	Test 9	64.50%	Test 9	55.00%	Test 9	62.00%
10	Test 10	66.00%	Test 10	67.50%	Test 10	70.00%

From the test results by conducting 10 random tests, the highest accuracy level is the K-NN algorithm.

Table 10 Precision Value Result

No	Algorithm					
	SVM		Naïve Bayes		KNN	
	Test	Precision	Test	Precision	Test	Precision
1						
2	T 1	61.40%	T 1	66.67%	T 1	66.67%
3	T 2	59.72%	T 2	55.68%	T 2	55.68%
4	T 3	55.30%	T 3	64.76%	T 3	64.76%
5	T 4	53.66%	T 4	58.89%	T 4	58.89%
6	T 5	54.95%	T 5	66.33%	T 5	66.33%
7	T 6	52.63%	T 6	59.09%	T 6	59.09%
8	T 7	56.00%	T 7	49.41%	T 7	49.41%
9	T 8	68.52%	T 8	58.62%	T 8	58.62%
10	T 9	62.50%	T 9	65.96%	T 9	65.96%
11	T 10	62.50%	T 10	65.96%	T 10	65.96%

From the testing result with 10 testings randomly has generates highest precision level, it is K-NN algorithm.

Table 8 Recall Value Result

No	SVM		NB		KNN	
	Recall		Recall		Recall	
	Test	Value	Test	Value	Test	Value
1	T 1	50.00%	T 1	51.92%	T 1	64.42%
2	T 2	87.50%	T 2	74.17%	T 2	76.67%
3	T 3	46.24%	T 3	51.61%	T 3	52.69%
4	T 4	68.22%	T 4	53.27%	T 4	63.55%
5	T 5	67.35%	T 5	62.24%	T 5	54.08%
6	T 6	50.00%	T 6	30.00%	T 6	65.00%
7	T 7	21.74%	T 7	79.35%	T 7	56.52%

No	SVM		NB		KNN	
	Recall		Recall		Recall	
	Test	Value	Test	Value	Test	Value
8	T 8	25.00%	T 8	23.91%	T 8	45.65%
9	T 9	40.66%	T 9	29.67%	T 9	56.04%
10	T 10	61.11%	T 10	61.11%	T 10	68.89%

From the testing result with 10 testings randomly has generates highest recall level, it is K-NN algorithm. According to the data that has been tested, the result of the data explains the level of accuracy, recall, and precision. Here below is the graphic of overall data that has been tested by using rapid miner.

1. Accuracy

The graphic result of accuracy value with result of data training and testing.

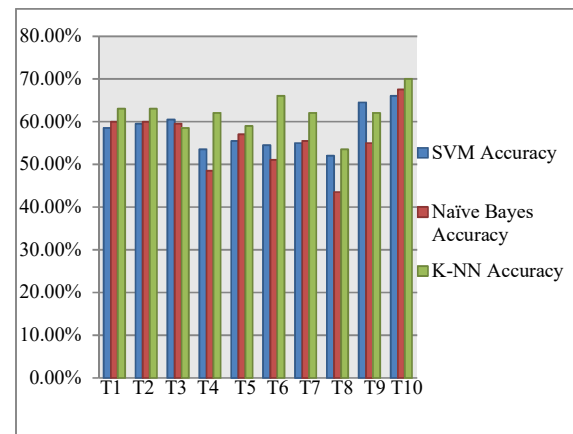


Figure 4 Grafik Accuracy

2. Precision

The graphic result of precision value with result of data training and testing.

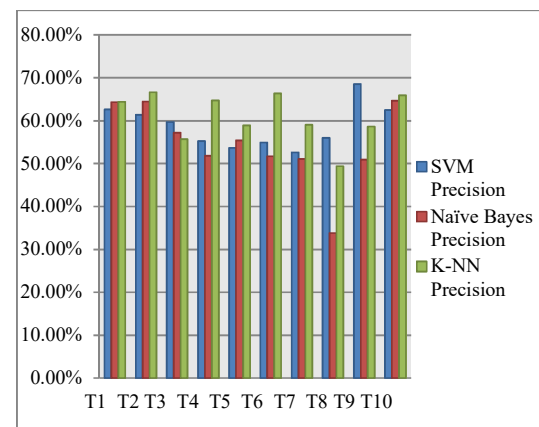


Figure 5 Grafik Precision

The graphic result of recall value with result of data training and testing.



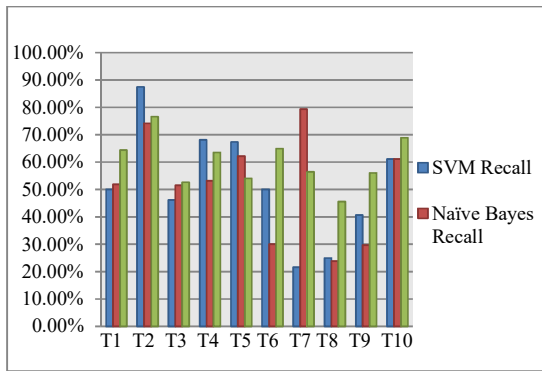


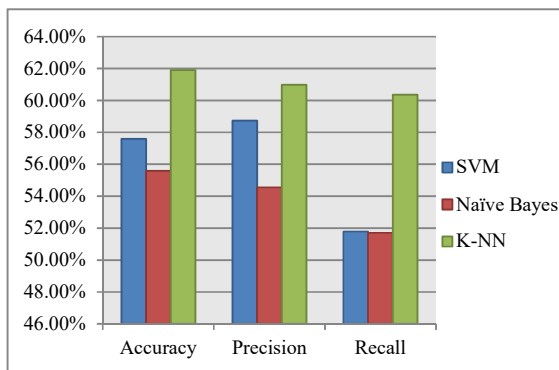
Figure 6 Grafik Recall

The test results from the gold data produce the level of accuracy, precision and recall of each algorithm, which can be explained as follows. The SVM algorithm has an accuracy rate of 57.59%, precision 58.73% and recall 51.78% while the Naive Bayes algorithm has an accuracy rate of 55.59% precision 54.55% and recall 51.70% and which has a level accuracy. The best precision and recall in comparing the 3 algorithms in testing this gold data is the K-NN algorithm which has an accuracy value of 61.90%, precision 60.98% and recall 60.35% and can be seen from table 4.7 the results of 3 the algorithm.

Table 8 Accuracy Value Result

No	Accuracy			
	Algorithm	Accuracy	Precissin	Recall
1	Naive Bayes	55.59%	54.55%	51.70%
2	SVM	57.59%	58.73%	51.78%
3	KNN	61.90%	60.98%	60.35%

From the test results by conducting 10 random tests with the highest level of accuracy, precision and recall being K-NN, the following is an overview of the graph of the results of testing 3 algorithms.



Picture 6 Graphic of Testing Result Value

#### Analysis

Based on the results of the gold data analysis and the results of the gold data testing that was carried out, the data testing resulted in good results with the composition of the 6 data attributes tested, then the data was tested with the SVM algorithm, K-NN and Naive Bayes produce good results above 50% and based on the results obtained

in this study, the algorithm that produces good accuracy, precision and recall levels of the 3 algorithms is K-NN. In gold investment so that we can analyze and the results obtained from testing by doing random or random tests by generating the average value of the algorithm SVM has 57.59% accuracy, 58.73% precision and 51.78% recall, while the Naive Bayes algorithm has 55.59% accuracy, 54.55% precision and 51.70% recall and which has the best level of accuracy, precision and recall in comparing the 3 algorithms in testing this gold data is the K-NN algorithm which has 61.90% accuracy, 60.98% precision and 60.35% recall. The results of the K-NN algorithm have quite good results from the 3 tests of the algorithm.

From the testing of the naive Bayes algorithm, it was carried out randomly or randomly in the test and the results of the naive Bayes algorithm test had an accuracy value of 55.59%, precision 54.55% and recall 51.70%. From the prediction results, the data tested or read from the Naive Bayes algorithm has a fairly low gold data classification value or in the process of predicting data up will be read down and data down will be read up, so from the results of this study, random testing was carried out with the same data using 200 testing data and had 6 attributes used because gold data in general can be categorized as good results because the data has 6 attributes that have information about gold investment.

#### IV. CONCLUSION

Based on the test results using the SVM, K-NN and Naive Bayes algorithms, the following conclusions can be drawn:

In testing the gold data, the data tested resulted in good results with the composition of the 6 data attributes being tested, then the data was tested with the SVM, K-NN and Naive Bayes algorithms produce good results above 50%, testing by conducting random tests or random by producing an average value of the SVM algorithm has an accuracy rate of 57.59%, precision 58.73% and recall 51.78% while the Naive Bayes algorithm has an accuracy level of 55.59%, precision is 54.55% and recall is 51.70% and which has a level of accuracy. The best precision and recall in comparing the 3 algorithms in testing this gold data is the K-NN algorithm which has 61.90% accuracy, 60.98% precision and 60.35% recall. the results of the K-NN algorithm have fairly good results from the 3 tests of the algorithm and the naive Bayes algorithm test has an accuracy value of 55.59%, precision 54.55% and recall 51.70%. From the prediction results, the data tested or read from the naive Bayes algorithm has a fairly low gold data classification value or in the process of predicting the data up will be read down and the data down will be read up. Based on the research conducted, this research can provide some suggestions as follows:

Maximize or add more specific and more attributes in Naive Bayes, SVM and K-NN classifications, further research is needed by testing with other algorithms such as C.45, C.50 and so on in order to obtain comparisons with the highest level of accuracy in making classifications on Naive Bayes, SVM and K-NN and further research to improve accuracy, precision and recall in classifying by conducting experiments on each parameter.



## REFERENCES

- [1] Guntur, M., Santony, J., & Yuhandri, Y. (2018). Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(1), 354-360
- [2] Mahena, Y., Rusli, M., & Winarso, E. (2015). Prediksi Harga Emas Dunia Sebagai Pendukung Keputusan Investasi Saham Emas Menggunakan Teknik Data Mining. *Kalbiscentia J. Sains dan Teknol*, 2(1), 36-51.
- [3] Saputro, N. D. (2015). Penerapan Algoritma Support Vector Machine untuk Prediksi Harga Emas. *Jurnal Informatika Upgris*, 1(1 Juni).
- [4] Witjaksono, A. A. (2010). *Analisis Pengaruh Tingkat Suku Bunga SBI, Harga Minyak Dunia, Harga Emas Dunia, Kurs Rupiah, Indeks Nikkei 225, dan Indeks Dow Jones terhadap IHSG (studi kasus pada IHSG di BEI selama periode 2000-2009)* (Doctoral dissertation, UNIVERSITAS DIPONEGORO)
- [5] Budiyo, E. P., Nerfita Nikentari, S. T., Sallu, S., & Kom, S. ANALISA KLASIFIKASI KADAR KARAT EMAS MENGGUNAKAN METODE K-NEAREST NEIGHBOURS (KNN).
- [6] Hidayat, R. N. (2013). *Implementasi Jaringan Syaraf Tiruan Perambatan Balik untuk Memprediksi Harga Logam Mulia Emas Menggunakan Algoritma Lavenberg Marquardt* (Doctoral dissertation, Diponegoro University).
- [7] Fajrul Falah (2015). Rancang bangun aplikasi prediksi pergerakan harga emas logam mulia dengan menggunakan metode Backpropagation
- [8] Sari, Y. (2017). Prediksi Harga Emas Menggunakan Metode Neural Network backpropagation Algoritma Conjugate Gradient. *Jurnal Eltikom*, 1, 2
- [9] Faustina, R. S., Agoestanto, A., & Hendikawati, P. (2017). Model Hybrid ARIMA-GARCH Untuk Estimasi Volatilitas Harga Emas Menggunakan Software R. *UNNES Journal of Mathematics*, 6(1), 11-24
- [10] Zhu, Y., & Zhang, C. (2018). Gold price prediction based on pca-ga-bp neural network. *Journal of Computer and Communications*, 6(7), 22-33.
- [11] 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 e-ISSN, 2548, 964X*
- [12] Hadavandi, E., Shavandi, H., & Ghanbari, A. (2010, August). A genetic fuzzy expert system for stock price forecasting. In *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery* (Vol. 1, pp. 41-44). IEEE.
- [13] Rahmawati, N. E. (2012). Prediksi Data Time Series Menggunakan Fuzzy Inference System dan Particle Swarm Optimization (Studi Kasus: Prediksi Harga Emas)
- [14] Rusbariand, S. P., Masodah, R., & Herawati, S. (2012). Analisis pengaruh tingkat inflasi, harga minyak dunia, harga emas dunia, dan kurs rupiah terhadap pergerakan jakarta islamic index di bursa efek indonesia. In *Prosiding Seminar Nasional* (Vol. 1, pp. 724-740).
- [15] SYARAT-SYARAT, U. M. S., SATU, M. G. S. S., & ADIB, A. M. (2009). Pengaruh Inflasi, Suku Bunga Domestik, Suku Bunga Luar Negeri dan Kurs Terhadap Indeks Harga Saham (Studi pada JII dan IHSG tahun 2005-2007).





- [16] Sodik, A. (2016). Ka jian Historis Tentang Dinar dan MaMata UaUang Berstandar Emas. *IQTISHADIA*, 8(2).
- [17] Iman, N. (2009). Investasi Emas. *Jakarta: Daras Books*.
- [18] Indriasari, T. (2011). Pengaruh harga minyak dunia, nilai tukar rupiah dan tingkat suku bunga SBI terhadap Jakarta Islamic Index (JII). *Pengaruh harga minyak dunia, nilai tukar rupiah dan tingkat suku bunga SBI terhadap Jakarta Islamic Index (JII)/Titik Indriasari*
- [19] Sidarta, W. (2010). Pengaruh gejolak harga minyak mentah terhadap IHSG
- [20] Muhammad Wildan, M. W. (2016). *PRODUK MURABA> HA> H LOGAM INVESTASI ABADI DI PEGADAIAN SYARIAH PERSPEKTIF HUKUM ISLAM (Studi Kasus di PT. Pegadaian Syariah Cabang Purwokerto)* (Doctoral dissertation, IAIN PURWOKERTO)
- [21] Witjaksono, A. A. (2010). *Analisis Pengaruh Tingkat Suku Bunga SBI, Harga Minyak Dunia, Harga Emas Dunia, Kurs Rupiah, Indeks Nikkei 225, dan Indeks Dow Jones terhadap IHSG (studi kasus pada IHSG di BEI selama periode 2000-2009)* (Doctoral dissertation, UNIVERSITAS DIPONEGORO)
- [22] Christina, C. (2012). Prediksi Harga Emas Menggunakan Metode Neuro-Fuzzy Tipe 2. *Telkom University, Bandung*

