ICEBIT

Experiment Time Series Forcasting Using Machine Learning (Case study : Stock Value Prediction)

Agung Muliawan
Institute of Technology and Sains Mandala
Jember, Indonesia
089610345572, +628
agung.muliawan@itsm.ac.id

Difari Afreyna Fauziah
Institute of Technology and Sains Mandala
Jember, Indonesia
085777583880, +628
difa@itsm.ac.id

Ferry Wiranto
Institute of Technology and Sains Mandala
Jember, Indonesia
08991087383, +628
ferry@itsm.ac.id

ABSTRACT

The rapid development of technology has resulted in a high need for information so it is necessary to present information, one of which is the field of stocks. The use of data mining is necessary for predicting dynamic stock values. Predicting stock values by looking at certain variables using machine learning methods provides benefits for stock players. This study uses 5 attributes, namely open, low, high, close and volume on stock value. The dataset used is the shares of MNC, Hotel Sahid and XL on the Indonesia Stock Exchange during the previous 4 years. The algorithm used in this study uses machine learning including *Neural Network* (NN) and *Support Vector Machine* (SVM). The results obtained using the SVM algorithm with the smallest error value of 2,993 +/- 3,070 on MNC shares that are not that far from SVM with an error value of 3.7208 +/- 4.042. When compared with the 2 datasets of Hotel Sahid and XL shares, it is found that the NN algorithm has a smaller error value than the SVM with a long distance. So it can be concluded that the use of predictions using the Neural Network (NN) algorithm is generally better than SVM

Keywords : Time series, forecasting, machine learning, prediction, computing.

1.   INTRODUCTION

The capital market is an organized system of financial transactions consisting of financial sector institutions and all valuable documents of ownership. A capital market is a place for buying and selling company ownership certificates or shares between sellers and buyers. One of the benefits of the capital market is that it provides opportunities for the general public to participate in economic activities, especially investing. The term that is often heard from the capital market is investment. Investment is a form of delaying consumption from the present to the future where there is an uncertain value risk. One of the most popular investments today among the public is stocks. Shares are securities issued by a company whose income depends on the company issuing the shares. Stocks can also be said to be instruments that are widely used or chosen by investors because stocks can provide attractive economic benefits (Himawan et al., 2022). In its implementation, shares can also be said to be shareholders as part owners of the company (Maulana & Kumalasari, 2019). The more shares owned by a person, the greater the amount of money given to the company.

Dynamic stock prices make investors have different views from one another, causing investors to lose when they buy, sell, or retain the shares they own. Apart from this dynamic nature, the risks that occur in stocks such as capital loss, liquidation risk,, and delisting shares need to be considered.

The capital market in Indonesia is a developing capital market and is vulnerable to global economic conditions and world capital markets. The development of technology and the flow of information is very fast and continues to grow by providing various kinds of needs and information, one of which is the value of shares (Muliawan et al., 2022). In recent yearswith , the development of the information economy, the value of Indonesian stocks, has slowed down, causing stock prices to experience insignificant conditions so quickly. Factors that influence changes in stock prices such as inflation rates, deposit rates, company profits, corporate finance an,d marketing strategies (R. Kusumodestoni & Sarwido, 2017). With several possible changes to these factors, stock prices in the stock market run randomly. This resulted in the value of the stock index that was recorded on the IDX increasing and decreasing (Himawan et al., 2022). Therefore, an analysis of the movement of stock values is needed to assist investors in making stock investment decisions. The movement of stock values tends to be non-linear, therefore investors need a method that can carry out an accurate analysis of stock movements. This is what causes the prediction of changes in stock prices to become the main topic in discussing stock investment among beginners to investors (Nusa Mandiri University et al., 2022)

Stock prediction is one of the important issues in stock trading depending on the information provided at any time. This is needed inanalytermingdetermine the future value of a company's shares. The use of the right method affects the accuracy of stock predictions so it is needed by investors in protecting and benefiting from trading indexes. One method commonly used in determining stock price predictions is using the forecasting method or forecasting in machine learning. Machine Learning is a computer or program that has intelligence in managing data that is useful in providing information as material for consideration indecision-makingg osupportn support. One of the machine learning methods is support vector matching (SVM), Neural Networks (NN).

In this study, researchers used a neural network model (NN) and a support vector machine (SVM). Neural Network is an algorithm formulated by Werbos and popularized by Rumelhart & McClelland. A neural network is a type of artificial neural network that uses supervised learning methods. The NN algorithm uses ANN or artificial neural

networks to process input and output data to produce the required weight values. This algorithm processes the incorrect output to be corrected so that the aggregate is reduced and processed by ANN so that it is expected to produce values that are close to correct (R. Kusumodestoni & Sarwido, 2017). This method also uses non-linear computation elements which are also called neurons which are organized as connected networks to recognize certain patterns or classifications.

Meanwhile, the Support Vector Machine (SVM) algorithm is a hypothetical learning method in the form of a linear function in a high-dimensional feature space. The SVM algorithm is based on optimization theory with the application of statistical learning. This model is one of the less expensive techniques and has better performance in various text classifications and handwriting recognition. Previous research was conducted by (R. Kusumodestoni & Sarwido, 2017) by comparing the SVM and NN methods by predicting stock prices at PT. Interpan Pacific Futures obtained better results using the NN algorithm with SVM with a value of 0.503 +/- 0.009 (micro 503) while using the SVM model obtained results of 0.477 +/- 0.008 (micro: 0.477).

Another study was conducted by (Sari, 2016) predicting social media stock prices with the Support Vector Machine (SVM) algorithm which was optimized with Particle Swarm Optimization (PSO). The DOT kernel accuracy results yield an accuracy value of 94.8% for training and 94.6% for testing.

Other research conducted by (R. H. Kusumodestoni & Suyatno, 2015) made forex predictions using the Neural Network (NN) model. The results of his study show that forex time series data can be predicted using the NN model with a prediction accuracy of 0.431+/- 0.0096 using data per 1 minute.

The purpose of this study is to predict stock values using 2 machine learning methods, namely the support vector machine (SVM) and neural network (NN) algorithms in which the result that has the smallest error is the algorithm with the highest accuracy. The data used is MNC, Hotel Sahid and XL stock data for the previous 4 years (2017-2021) to see the accuracy of the method on different data. Stock price prediction using this algorithm trial is expected to assist investors in making decisions in the capital market. In addition, researchers will retest whether the NN model is better than SVM.

2. Methodology

The research method used in this study consisted of several stages starting from data collection, making NN calculations, SVM modeling and conducting evaluation and validation to see the accuracy value. The following is a detailed explanation of each stage
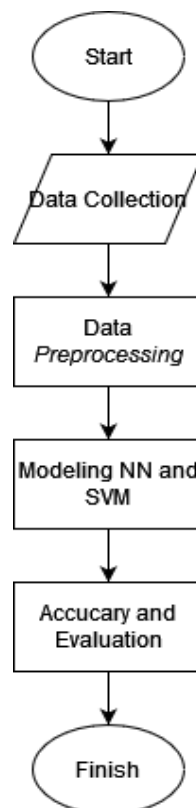

Fig 1. Metodelogi Penelitian

2.1 Data Colleciton

n this study, the researcher used data on the value of stock prices at 3 different PTs, namely MNC, Hotel Sahid and XL with an interval of 4 years from 28-03-2017 to 28-03-2021. The attributes used in this study consist of six attributes, namely date, open, high, low, close and volume. The following is an explanation of each attribute used in this study :

Tabel 1. Detail Attribute

No	Nama Atribut	Format	Keterangan
1	Date	Date	The time of the stock transaction
2	Open	Integer	The opening price at that time
3	High	Integer	The highest price at that time
4	Low	Integer	The lowest price at that time
5	Close	Integer	The closing price at that time
6	Volume	Integer	The value of the sale or purchase transaction at that time

The amount of data taken is 1012 data per share with 3 shares for a total of 3036. The data taken is an open dataset taken on the site: https://finance.yahoo.com/quote. The following is the raw data to be retrieved

1	Date,Open,High,Low,Close,Adj Close,Volume
2	2017-03-29,1935.000000,1935.000000,1925.000000,1925.000000,1486.526611,6100
3	2017-03-30,1930.000000,1940.000000,1930.000000,1930.000000,1490.387695,18000
4	2017-03-31,1930.000000,1930.000000,1930.000000,1930.000000,1490.387695,68800
5	2017-04-03,1930.000000,1940.000000,1920.000000,1940.000000,1498.109863,7800
6	2017-04-04,1940.000000,1940.000000,1915.000000,1935.000000,1494.248779,10000
7	2017-04-05,1935.000000,1935.000000,1935.000000,1935.000000,1494.248779,0
8	2017-04-06,1930.000000,1930.000000,1930.000000,1930.000000,1490.387695,8000
9	2017-04-07,1930.000000,1930.000000,1920.000000,1920.000000,1482.665527,2800
10	2017-04-10,1925.000000,1925.000000,1875.000000,1875.000000,1447.915649,5600
11	2017-04-11,1875.000000,1920.000000,1875.000000,1900.000000,1467.221069,11300
12	2017-04-12,1900.000000,1900.000000,1900.000000,1900.000000,1467.221069,0
13	2017-04-13,1900.000000,1920.000000,1900.000000,1920.000000,1482.665527,2000
14	2017-04-17,1920.000000,1930.000000,1915.000000,1920.000000,1482.665527,86700
15	2017-04-18,1930.000000,1930.000000,1920.000000,1920.000000,1482.665527,44900
16	2017-04-19,1920.000000,1920.000000,1920.000000,1920.000000,1482.665527,0
17	2017-04-20,1920.000000,1945.000000,1900.000000,1905.000000,1509.608765,127600
18	2017-04-21,1875.000000,1900.000000,1870.000000,1875.000000,1525.062744,9000

Fig 2. Raw data used

The data taken is still in the form of raw data so it needs to be preprocessed. The purpose of the preprocessing process is also to improve the results and accuracy of data mining for the better. Before being processed, the preprocessing process is carried out first. The steps of the preprocessing process start with data cleaning, data integration and data reduction. Data cleaning is done by cleaning empty and inconsistent values (missing values and noisy), the second is the data integration process which combines data into one storage then data reduction is carried out where the number of attributes used is too large so it needs to be removed or deleted according to the attributes needed. used (R. Kusumodestoni & Sarwido, 2017). The following is a sample data used after the preprocessing process above:

Tabel 2. Data that has gone through the preprocessing process (Hotel Sahid)

Data i	Date	Open	High	Low	Close	Volume
1	29/03/2017	920	920	920	920	920
106	29/08/2017	1120	1125	1100	1125	1125
215	29/01/2018	1250	1250	1250	1250	1250
354	10/08/2018	1565	1700	1565	1700	1700
489	15/02/2019	4770	4800	4760	4760	4760
573	13/06/2019	3740	3740	3740	3740	3740
712	27/12/2019	2680	2790	2660	2790	2790
861	07/08/2020	3800	3800	3700	3700	3700
958	06/01/2021	2330	2330	2330	2330	2330
1013	26/03/2021	2430	2430	2260	2260	2260

After the data preprocessing process, data classification will be carried out using the Split Validation technique with a comparison of data training and testing of 70:30. Then from these two data will be tested using the Neural Network (NN) and Support Vector Machine (SVM) methods to get the value that has the highest value / smallest error.

2.2   Algorithm *Neural Network* (NN)

In this study, the data that has been prepared from the preprocessing process will be processed using the Neural Network (NN) model. The use of this method is because this method is a method that produces a fairly high prediction accuracy. Adding large amounts of data increases the accuracy of the resulting predictions because this method uses learning by calculating the previous time value. In this model training data and neural network (backpropagation) are tested using data testing with a ratio of 70 : 30 with the aim of measuring the level of prediction accuracy.

2.3   Algorithm *Support Vector Machine* (SVM)

Support Vector Machine (SVM) is a classification method that is divided into two categories, namely the classification of support vector machines and regression of support vector machines (Setiawan et al., 2018). In this study, the Support Vector Machine algorithm method is used with the regression category which can produce accurate predictions using C (cost) and kernel parameters. The advantages of this algorithm can select models automatically and have no problem with overfitting. Overfitting itself is an unwanted machine behavior when accurate predictions are produced but the results will not be accurate if done with new data. This method is good for prediction because it can minimize classification errors and deviations in the training data . In the next process, a search for the best value of the parameters is then carried out to compare the variables to get the best predictive results.

2.4   Evaluation and Validation

At the evaluation and validation stage of the dataset test results carried out using the NN and SVM methods using the rapid miner, the results showed that the accuracy of the predictions corresponded to the variables used. This evaluation uses the RMSE or Root Mean Square Error method. RMSE is a forecasting evaluation method by adding the squared error or the difference between the real value and the forecast value (Budiman, 2016). The following is RMSE formula which is defined as follows (Fadilah et al., 2020):

$$MSE = \sum \frac{(Y' - Y)^2}{n}$$

Diketahui :

Y' : Forecast data

Y : Real data

n : rows ddata

3.   Result and Discussion

3.1   Modeling

This study uses data that has been downloaded on the page: https://finance.yahoo.com/quote. The raw data that has been processed is then modeled using Rapid Miner to test the prediction of MNC, Hotel Sahid and XL stock prices to obtain a level of accuracy. The following is the design of the Rapid Miner model used in the Neural Network and Support Vector Machine algorithm testing models
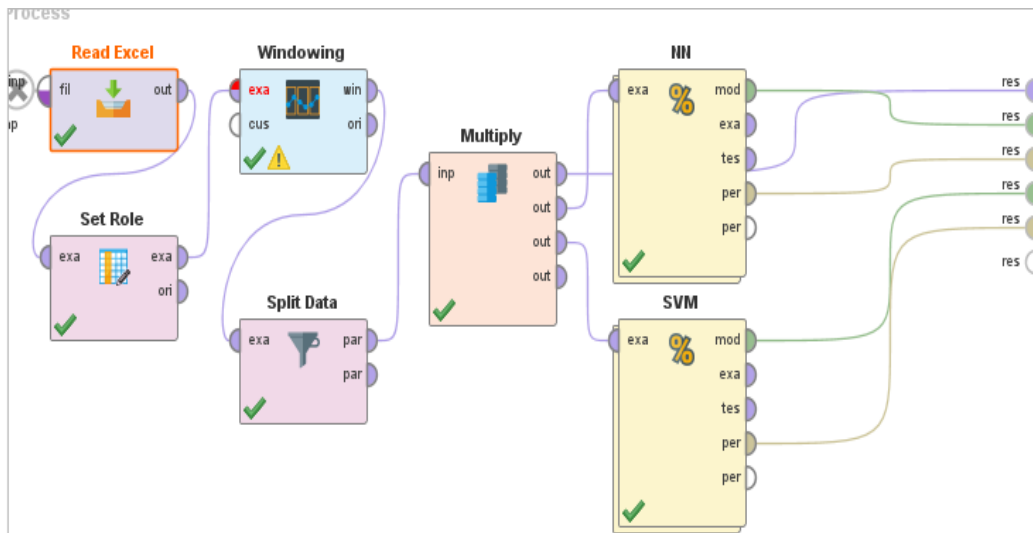

Fig 3. Modeling with Rapid Miner

Figure 3 describes the stock price prediction process using Rapid Miner. Data that is processed using the Read Excel operator is then configured for Set Role as the borrowing of the attribute used (date) and Prediction as the target role. Furthermore, windowing is carried out by adjusting the number of processes to calculate the time series used, in this study using the H-5 measurement, which means that the past 5 days are processed to display forecasting results and display the target to be predicted (close attribute). Furthermore, a data split was carried out with the aim of separating the amount of training and testing data, in this study the data was separated by a percentage of 70: 30 training and testing data. Then the Multiply model is used to process 2 methods with the same data, solving is done using 2 Neural Network (NN) methods and Support Vector Machine (SVM). The following is an example of an image resulting from a graph using the Neural Network method on Hotel Sahid stock:

Fig 4. Graph Prediction Hotel Sahid

3.2   Accuration and Evalution

Furthermore, from the experiment of 3 stocks (MNC, Hotel Sahid and XL) the accuracy performance value was obtained using Root Mean Squared or RMSE. The following is a sample image of the accuracy performance test on Hotel Sahid stock using the Neural Network (Performance 2) and Support Vector Machine (Performance 3) methods
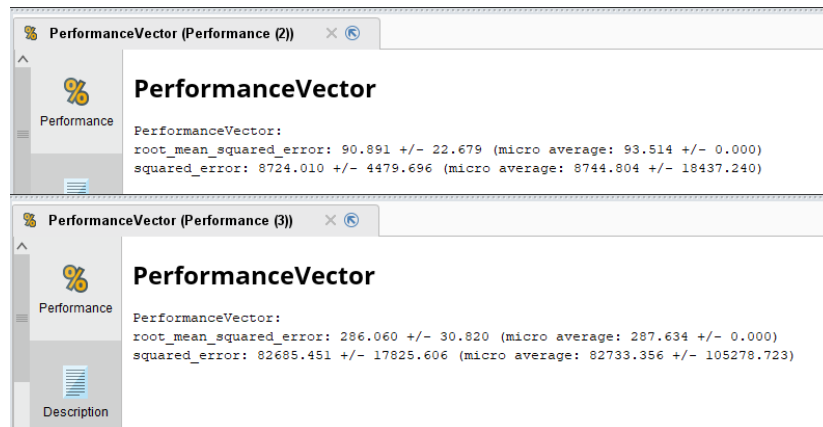

Fig 5. Testing NN dan SVM to Hotel Sahid

The results of this study compare neural network algorithms and support vector machines using 3 datasets (MNC, Hotel Sahid and XL). The following are the results of the accuracy and error rate of the method proposed in this study using 3 datasets

Table 3. Result Accuration *Neural Network* and *Support Vector Machine*

No		Algorithm	
	Name Stock	*Neural Network (NN)*	*Support Vector Machine* (SVM)
1	MNC	3.7208 +/- 4.042	2.993 +/- 3.070
2	Hotel Sahid	118.881 +/- 35.029	988.700 +/- 126.198
3	XL	90.891 +/- 22.679	286.060 +/- 30.820

4.   Kesimpulan

After conducting research on stock price prediction using three different stock datasets (MNC, Hotel Sahid and XL) using stock value data starting from 29-03-2017 to 26-03-2021 as many as 1012 records per share using two Neural Network (NN) methods and Support Vector Machine (SVM). Data comparison training and 70.30 tests continued with the accuracy of the RMSE test. The results obtained using the SVM algorithm with the smallest error value of 2,993 +/- 3,070 on MNC shares that are not that far from SVM with an error value of 3.7208 +/- 4.042. When compared with the 2 datasets of Hotel Sahid and XL shares, it is found that the NN algorithm has a smaller error value than the SVM with a long distance. So it can be concluded that the use of predictions using Neural Network (NN) algorithms is generally better than SVM. From the test results, the accuracy of the Neural Network algorithm is higher than the

Support Vector Machine algorithm. But it is impossible if there are other methods that have better accuracy than the NN method

5. REFERENCES

Budiman, H. (2016). Analisis Dan Perbandingan Akurasi Model Prediksi Rentet Waktu Support Vector Machines Dengan Support Vector Machines Particle Swarm Optimization Untuk Arus Lalu Lintas Jangka Pendek. *Systemic: Information System and Informatics Journal*, *2*, 19–24. https://doi.org/10.29080/systemic.v2i1.103

Fadilah, W. R. U., Agfiannisa, D., & Azhar, Y. (2020). Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine. *Fountain of Informatics Journal*, *5*(2), Article 2. https://doi.org/10.21111/fij.v5i2.4449

Himawan, I., Nurdiawan, O., & Dwilestari, G. (2022). Irvan Himawan PREDIKSI HARGA SAHAM DENGAN ALGORITMA REGRESI LINIER DENGAN RAPIDMINER. *JURSIMA (Jurnal Sistem Informasi dan Manajemen)*, *10*(3), Article 3. https://doi.org/10.47024/js.v10i3.475

Kusumodestoni, R. H., & Suyatno, S. (2015). PREDIKSI FOREX MENGGUNAKAN MODEL NEURAL NETWORK. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, *6*(2), Article 2. https://doi.org/10.24176/simet.v6i2.453

Kusumodestoni, R., & Sarwido, S. (2017). KOMPARASI MODEL SUPPORT VECTOR MACHINES (SVM) DAN NEURAL NETWORK UNTUK MENGETAHUI TINGKAT AKURASI PREDIKSI TERTINGGI HARGA SAHAM. *Jurnal Informatika Upgris*, *3*. https://doi.org/10.26877/jiu.v3i1.1536

Maulana, R., & Kumalasari, D. (2019). ANALISIS DAN PERBANDINGAN ALGORITMA DATA MINING DALAM PREDIKSI HARGA SAHAM GGRM. *Jurnal Informatika Kaputama (JIK)*, *3*(1), Article 1. https://doi.org/10.1234/jik.v3i1.134

Muliawan, A., Badriyah, T., & Syarif, I. (2022). Membangun Sistem Rekomendasi Hotel dengan Content Based Filtering Menggunakan K-Nearest Neighbor dan Haversine Formula. *Technomedia Journal*, *7*, 231–247. https://doi.org/10.33050/tmj.v7i2.1893

Sari, E. P. (2016). Model Prediksi Harga Saham Media Sosial Berdasarkan Algoritma Svm Yang Dioptimasikan Dengan Pso. *Pilar Nusa Mandiri*, *12*(2), 161–170.

Setiawan, F. R., Umbara, R. F., & Rohmawati, A. A. (2018). Prediksi Pergerakan Harga Saham Dengan Metode Support Vector Machine (svm) Menggunakan Trend Deterministic Data Preparation (Studi Kasus Saham Perusahaan Pt Astra International Tbk, Pt Garuda Indonesia Tbk, Dan Pt Indosat Tbk). *eProceedings of Engineering*, *5*(3), Article 3. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/7540

Universitas Nusa Mandiri, Fitri, E., & Riana, D. (2022). ANALISA PERBANDINGAN MODEL PREDICTION DALAM PREDIKSI HARGA SAHAM MENGGUNAKAN METODE LINEAR REGRESSION, RANDOM FOREST REGRESSION DAN MULTILAYER PERCEPTRON. *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, *6*(1), 69–78. https://doi.org/10.46880/jmika.Vol6No1.pp69-78