

**Bambang Purnomosidi Dwi
Putranto**¹

²Master in Information Technology,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: bpdp@utdi.ac.id

Moh. Abdul Kholik

Department of Computer Engineering,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: riki@utdi.ac.id

Muhammad Agung Nugroho

Informatics,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: m.agung.n@utdi.ac.id

Danny Kristanto

Informatics,
Universitas Teknologi Digital Indonesia,
Yogyakarta, Indonesia
email: danny@utdi.ac.id

Polynomial Regression Method and Support Vector Machine Method for Predicting Disease Covid-19 in Indonesia

The COVID-19 pandemic has become a major threat to the entire country. According to the WHO report, COVID-19 is a severe acute respiratory syndrome transmitted through respiratory droplets resulting from direct contact with patients. This study of data history is then processed using data mining prediction methods, namely the Polynomial Regression method compared to the Support Vector Machine method. Of the two methods will be sought the most accurate method by testing accuracy with MAE, MSE, and also MAPE to get the results of covid-19 predictions in Indonesia. Based on the comparison of test results through various scenarios against both methods, the Polynomial Regression method obtained the smallest test value, resulting in an accuracy value of MAE = 4146.025749867596, MSE = 19031800.02642069, MAPE = 0.006174164877416524. Polynomial regression is the best-recommended method.

KeyWords: COVID-19, predictions, Polynomial Regression, Support Vector Machine, accuracy

This Article was:

submitted: 19-06-23
accepted: 11-07-23
publish on: 14-07-23

How to Cite:

Putranto B.P.D., et al, "Polynomial Regression Method and Support Vector Machine Method for Predicting Disease Covid-19 in Indonesia", Journal of Intelligent Software Systems, Vol.2, No.1, 2023, pp.18-23, [10.26798/jiss.v2i1.931](https://doi.org/10.26798/jiss.v2i1.931)

1 Introduction

At the end of 2019, the world was shocked by the discovery of several patients with unusual pneumonia originating from in southern China's Hubei province[1], this was confirmed by the World Health Organization (WHO) representative office located in Beijing in December 31, 2019. The results of the investigation found that a group of patients exposed to pneumonia were declared infected with a new type of coronavirus, better known as Coronavirus Disease (COVID-19)[2].

Coronaviruses Disease 19 (Covid-19) is spreading very quickly in almost all countries currently affected by the outbreak of transmission of this virus. Covid-19 is a burden on public health and safety services[3]. This is based on the lack of ways to control this virus so that transmission and spread are more widespread, while treatment is still little known for this disease as well as the development of vaccines that are still being worked on[4]. Researchers in various countries continue to explore the characteristics of Covid-19, including the discovery of positive patient cases without showing symptoms but still transmitting, an increasingly long incubation period, and a virus that continues to mutate[5].

The increase in cases that continues to increase every day resulted in the health system in Indonesia is currently very depressed with

the condition of the surge of Covid-19 positive patients. Not only that, but the soaring positive number of Covid-19 also has an impact on the level of bed occupancy rate (BOR) is getting higher, this condition can cause the health system in Indonesia to become paralyzed if the number of hospital bed availability has reached above 80% for several days[5].

The relatively high increase in new cases shows that Indonesia's Covid-19 pandemic has not passed its peak[6]. Efforts can be made by modelling the cumulative number of positive cases of COVID-19. This predictive model can be used as projection and optimization for the government to take policies and evaluate existing policies[7]. Although in reality, there is indeed no prediction that has a true level of 100%. But error rates or errors in prediction or forecasting can be minimized, by looking for methods that can produce forecasting with high accuracy values. In Indonesia, research on covid-19 prediction methods is still not much including research conducted by Nuraini, using time series models based on growth methods, Rustan and Handayani research, using modified SEIR (Susceptible Exposed Infectious Recovered) methods, research conducted by Parhusip using SVM, Bayesian and Gaussian methods, research conducted by Arianto using Reverse and Fuzzy Propagation methods[8][9][10][11]. Therefore, researchers are interested in comparing polynomial regression prediction methods and support vector machine methods to get the most accurate prediction results in predicting the spread of covid-19 in Indonesia.

2 Research Methodology

2.1 Type of Research. This research according to theory is included in the type of action research (action research)[12]. This research aims to develop the most effective research methods. In this study, the authors used a quantitative approach that emphasized their analysis of numerical data (numbers) and then processed it with statistical methods. This research links the theory of Polynomial Regression and Support Vector Machine algorithms with the application to predict the spread of Covid-19 in Indonesia. This study intends to find the most efficient method of predicting the spread of Covid-19 in Indonesia

¹Corresponding Author.

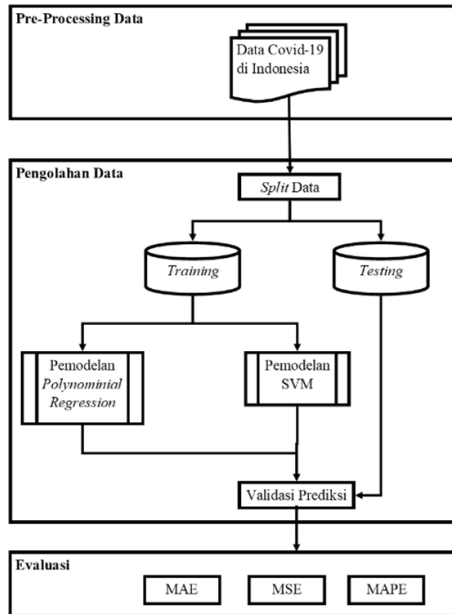


Fig. 1 Research Flow Diagram

2.2 Research Flow Diagram. In achieving the research objectives, it is necessary to plan the stages of activities to be carried out. This stage will be a guideline in the work of research.

2.3 Dataset Collections. This stage will explain the data set that will be used in this study. The data used in this study was positive patient data, cured and also died. Data obtained from the official website of Covid-19 Indonesia <https://kawalcovid19.id/> is data from 34 provinces taken in the form of patient history data from April 01, 2020, to January 01, 2021. Table 1 shows data on some cases of Covid-19 in Indonesia.

2.4 Data Pre-processing. Data preparation is the first step to the data mining process. This process is known as data pre-processing. In the preparation of data or data pre-processing there are four stages, namely: Data cleaning, data reduction, data transformation, and data integration. If it is associated with the CRISP-DM methodology, then the preparation of data is in the third stage, namely data preparation. Data preparation or data pre-processing is done to find out the quality of the data we get. Good data has specifications such as accurate accuracy, completeness of data, consistent data, reliable data and also easy-to-understand data.

2.5 Data Cleaning. Data circulating in cyberspace is actually still in a dirty state, meaning that a lot of data is potentially inaccurate or incorrect, such as instrument errors, human errors, or transmission errors that require data cleaning. Some of them are incomplete data, noise data, inconsistent and duplicate data. Likewise, with covid-19 data, there is also incomplete data so it needs to be done pre-processing data.

Kasus terdapat	Aceh	Bali	Banten	Bengkulu	DIY	Jakarta	Jambi	Jawa Barat	Jawa Tengah	Jawa Timur	Kalimantan Barat	Kalimantan Timur	Kalimantan Tengah	Kalimantan Selatan	Kalimantan Utara	Kepulauan Riau	Nusa Tenggara Barat	Sumatera Selatan	Sumatera Barat	Sulawesi Utara	Sumatera Utara	Sulawesi Tenggara	Sulawesi Selatan	Sulawesi Tengah	Lampung	Riau	Maluku Utara	Maluku	Papua Barat	Papua	Sulawesi Barat	Nusa Tenggara Timur	Gorontalo							
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 2 Covid-19 Case

Table 1 Covid-19 Data in Indonesia

Province	Case	Recovery	Deaths
Aceh	9712	7996	388
Bali	37458	34890	1045
Banten	33832	26768	699
Bangka Belitung	8308	7495	131
Bengkulu	5212	4828	152
DI Yogyakarta	30954	25249	745
DKI Jakarta	365419	352787	6141
Jambi	5835	4777	88
Jawa Barat	237251	205366	2855
Jawa Tengah	163460	117591	7072
Jawa Timur	135803	124044	9578
Kalimantan Barat	5457	4665	33
Kalimantan Timur	60868	54458	1426
Kalimantan Tengah	15627	13309	374
Kalimantan Selatan	24976	21982	789
Kalimantan Utara	10613	8870	163
Kepulauan Riau	8924	8537	225
Nusa Tenggara Barat	9142	7359	348
Sumatera Selatan	16782	15032	817
Sumatera Barat	30520	28813	672
Sulawesi Utara	15235	12292	499
Sumatera Utara	26321	22942	889
Sulawesi Tenggara	10171	9521	203
Sulawesi Selatan	58771	55260	901
Sulawesi Tengah	10716	9154	273
Lampung	13425	12091	712
Riau	33099	31123	806
Maluku Utara	4238	3543	118
Maluku	7214	6622	110
Papua Barat	7991	7281	130
Papua	18788	10112	197
Sulawesi Barat	5338	4699	110
Nusa Tenggara Timur	11376	8807	312
Gorontalo	5017	4695	141
Total	1443853	1272958	39142

From the data on Figure 2 can be known there are some areas that have missing data or blank data, therefore this can be a problem in the future in the data processing. In the pandas library there are several functions that can replace empty values, namely fillna(). Using fillna(), the missing value can be replaced with a special value or an aggregate value such as mean, median. To fill in the empty data in the dataset is done by replacing the empty data with the number '0'.

2.6 Covid-19. Coronavirus has been known since the 1930s as a group of viruses that not only infect humans but also infect various types of mammals[14]. Coronavirus is a member of the subfamily Orthocoronavirinae in the family Coronaviridae and its order Nidovirales[14]. Coronaviruses are classified into four types: Alphacoronavirus (α -CoV), Betacoronavirus (β -CoV), Gammacoronavirus (γ -CoV), and Deltacoronavirus (δ -CoV) according to phylogenetic relationships and genomic structure[15]. Coronaviruses are a group of positive-stranded RNA viruses that encase their genomes the largest among other RNA viruses, and the number of open-read skeletons (ORFs) varies between six and eleven[16][17]. To date, there are seven coronaviruses (HCoV) that have been indicated consisting of two α -CoV (HCoV-229E and HCoV-NL63) and five β -CoV (HCoVOC43 [lineage A], HCoV-HKU1 [lineage A], severe acute respiratory syndrome CoV

(SARS-CoV) [lineage B], Middle East Respiratory Syndrome CoV (MERS-CoV) [lineage C], and SARS-CoV-2 [lineage B][18]. SARS-CoV-2 or Covid-19 was first discovered in the China city of Wuhan. These viral infections generally cause symptoms that resemble the common cold, but symptoms can be more severe in infants, the elderly, and people with immune system disorders[19].

2.7 Polynomial Regression. Polynomial regression is a statistical method in data mining that aims to observe and facilitate relationships between variables[20]. The relationship between these variables is primarily the relationship in the functional form between those variables. Polynomial regression is the result of the modification of multiple linear regression models. In general, polynomial regression is expressed as[21]:

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (1)$$

For the polynomial equation of order 2, the matrix is obtained as follows:

$$\begin{bmatrix} na_0(\sum_{i=1}^n x_i) + (\sum_{i=1}^n x_i^2)a_1 + (\sum_{i=1}^n x_i^3)a_2 = \sum_{i=1}^n x_i \\ (\sum_{i=1}^n x_i)a_0 + (\sum_{i=1}^n x_i^2)a_1 + (\sum_{i=1}^n x_i^3)a_2 = \sum_{i=1}^n (x_i y_i) \\ (\sum_{i=1}^n x_i^2)a_0 + (\sum_{i=1}^n x_i^3)a_1 + (\sum_{i=1}^n x_i^4)a_2 = \sum_{i=1}^n (x_i^2 y_i) \end{bmatrix} \quad (2)$$

2.8 Support Vector Machine. Theoretically, the Support Vector Machine was developed for classification problems with two classes in search of the best hyperplane. The hyperplane is a separating function between two classes in space input. Figure 3 is shown some data that is a member of two classes. Positive classes are notated with +1 and negative classes are denoted with -1[22].

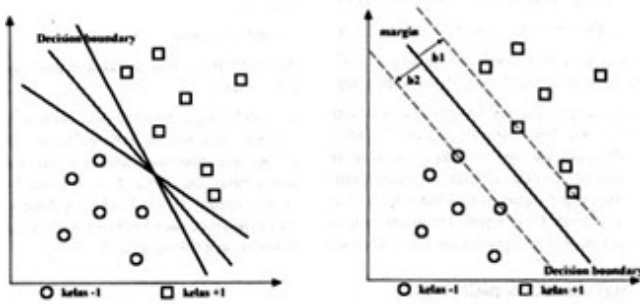


Fig. 3 Hyperplane

Support Vector Machine (SVM) classifiers simply as non-linear mappings to convert original training data into higher dimensions. In this new dimension, it will look for hyperplanes to be linearly separated, and classify them into higher dimensions through corresponding non-linear mapping. In this method, try to find the best classification function that can separate two different classes[22]. The best hyperplanes are those located in the middle between two sets of objects in two categories[23].

2.9 Support Vector Machine Linear. In a linear support vector machine separator is a linear function $\vec{x}_i \in R^d$. Data is notated with, each level is denoted as $y_i \in \{-1, +1\}$ to $i = 1, 2, \dots, l$, which l is the amount of data. It is estimated that classes -1 and +2 can part perfectly with the d -dimensional hyperplane d , which is interpreted as follows[24]:

$$\vec{w} \cdot \vec{x} + b = 0 \quad (3)$$

Patterns \vec{x}_i including class -1 (negative sample) can be formulated as patterns that meet the following inequalities:

$$\vec{w} \cdot \vec{x} + b \leq -1 \quad (4)$$

While the pattern including the +1 class (positive sample) meets inequality

Data x_i entered into class -1 is the following inequality data:

$$\vec{w} \cdot \vec{x} + b \leq +1 \quad (5)$$

To get the maximum margin value, it can be done by optimizing the distance value between the hyperplane and the closest point, that is $\frac{1}{\|\vec{w}\|}$. It can be formulated as a Quadratic Programming (QP) problem, which is to find the minimum point of equation (5) by paying attention to constraint equations (10).

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad (6)$$

$$y_i(\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0, \forall_i \quad (7)$$

This problem can be solved by sharing computational techniques, including using Lagrange Multiplier.

$$L(w, b, a) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^l a_i (y_i ((x_i \cdot w_i + 1) - 1)) \quad (8)$$

a_i Lagrange Multipliers, positive value ($y_i \geq 0$). The optimal value of the equation (8) can be calculated by minimizing L against \vec{w} and b , and maximizing L against a_i . With properties at the optimal point of gradient $L = 0$, equation (8) can be modified as maximization of problems that contain only a_i , such as equation (9) as follows:

$$\sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j \vec{x}_i \vec{x}_j \quad (9)$$

With constraint:

$$a_i \geq 0 (i = 1, 2, 3, \dots, l) \sum_{k=0}^n a_k y_k = 0 \quad (10)$$

From the results of these calculations, it can be concluded that most are positive. Data that correlate with positive ones are called support vectors[25].

The above explanation is based on the assumption that the two types can be perfectly separated by hyperplanes (linear separable). However, usually, the two categories in the input space cannot be completely separated (non-linear separable). This results in the non-fulfilment of the limitations in equation (10), thus cannot be optimized. To solve this problem, the support vector machine was reconstructed by introducing soft margin technology.

In soft margin, equation (5) is notified by entering the slack variable ξ_i ($\xi_i > 0$) as follows[26]:

$$y_i(x_i \cdot w + b) \geq 1 - \xi_i, \forall_i \quad (11)$$

Thus the equation (6) is changed into:

$$\min_{\vec{w}} \tau(w, \xi) = \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (12)$$

Parameter C was chosen to control the trade-off between margin and classification error ξ . A large C value would give a greater penalty to the classification error[27].

2.10 Support Vector Machine Non Linear. Generally, problems in the real world are rarely linearly separable, but non-linear. To resolve nonlinear issues, the support vector machine is modified to include kernel functions. In non-linear support vector machines, data is first \vec{x} mapped in $\phi(\vec{x})$ a high-dimensional vector room function[26]. The hyperplane separating the two classes can be constructed. Figure 4 shows that the function maps each data in the input space to a new vector space with a higher dimension (third dimension) so that the two classes can be linearly separated by a hyperplane. The mathematical notation of this mapping is as follows:

$$[H]\phi : \mathbb{R} \rightarrow \mathbb{R}^d \text{ d} < \text{q} \quad (13)$$

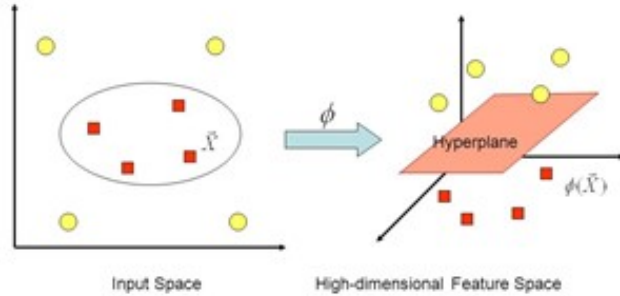


Fig. 4 Function Φ

Furthermore ϕ , the learning process of finding a point on a support vector machine depends only on the results of the multiplication of points from data that have been converted into new high-dimensional spaces, namely $\phi(\vec{x}_i) \cdot \phi(\vec{x}_j)$ [27]. Since transformations are usually unknown and difficult to understand, the $K(\vec{x}_i, \vec{x}_j)$ function that implicitly defines transformation ϕ can be used to replace point multiplication calculations[27]. This is called the kernel technique, and is formulated as follows:

$$K(\vec{x}_i, \vec{x}_j) = \phi(\vec{x}_i) \cdot \phi(\vec{x}_j) \quad (14)$$

$$\int (\phi(\vec{x}_i) \cdot \vec{w} \cdot \phi(\vec{x}) + b) \quad (15)$$

$$= \sum_{i=1, \vec{x}_i \in SV}^n a_i y_i \phi(\vec{x}) \cdot \phi(\vec{x}_i) + b \quad (16)$$

$$= \sum_{i=1, \vec{x}_i \in SV}^n a_i y_i K(\vec{x}, \vec{x}_i) + b \quad (17)$$

SV in the equation above is intended by a subset of the training set selected as a support vector, in other words data \vec{x}_i , the corresponding data on $a_i \geq 0$ [27].

2.11 Evaluation of the prediction performance. In the output resulting from the prediction results, it is necessary to evaluate each algorithm to find the best accuracy value. Some methods can be used as follows:

- (1) Mean Absolute Error (MAE) is one of the methods used to measure the accuracy of forecasting models. Mae values show the average of absolute errors between forecasting/predicting results and real values[28]. Mae is described as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - f_i| \quad (18)$$

- (2) Mean Squared Error (MSE) is the average square error between the actual value and the forecasting value. The Mean Squared Error method is generally used to check the estimated value of errors in forecasting[29]. A low Mean Squared Error value or a near-zero mean squared error value indicates that the forecasting results are in accordance with actual data and can be used for forecasting calculations in the coming period.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad (19)$$

- (3) Mean Absolute Percentage Error (MAPE) is calculated using absolute errors in each period divided by the real observation value for that period. Then, evenly the absolute percentage error. This approach is useful when the size or size of the prediction variable is important in evaluating the accuracy of the prediction[30]. MAPE indicates how much error in predicting is compared to real value. The equation is as follows:

$$MAPE = \frac{1}{2} \sum |PE| \quad (20)$$

$$PE = \frac{Y_t - F_t}{Y_t} \cdot 100$$

3 Result and Discussion

3.1 Comparative Analysis Polynomial Regression dan SVM. Polynomial Regression and Support Vector Machine (SVM) methods are methods used in solving prediction problems with different algorithms. Polynomial regression is a method of predicting objects based on learning data with poly (curved) line patterns. Support Vector Machine (SVM) uses SVR in the case of regression.

In the comparison of the two methods, the parameters used in the training process are different. This is seen in the Polynomial Regression method and Support Vector Machine used different degree parameters.

After obtaining the parameters that provide the highest accuracy value in this study. The next trial was conducted on the same amount of training data, which is 256 data with parameters degree = 6, for Polynomial Regression and degree = 2 for Support Vector Machine. The results obtained from the training and testing process using the amount of data as much as 256 data in both methods are shown the results of predictions in the next 10 days using polynomial regression algorithms and support vector machines as follows:

Table 2 SVM Predictions

Date	Confirmed Cases in Indonesia
1/2/2021	685186
1/3/2021	690222
1/4/2021	695276
1/5/2021	700349
1/6/2021	705439
1/7/2021	710548
1/8/2021	715675
1/9/2021	720820
1/10/2021	725983
1/11/2021	731165

Table 3 Polynomial Regression Predictions

Date	Confirmed Cases in Indonesia
1/2/2021	766454
1/3/2021	776220
1/4/2021	786241
1/5/2021	796527
1/6/2021	807086
1/7/2021	817928
1/8/2021	829062
1/9/2021	840498
1/10/2021	852247
1/11/2021	864317

To better understand the accuracy comparison can be seen in the image below:



Fig. 5 Polynomial Regression Prediction

Figure 5 shows that the blue line is real data and the line is purple as a result of adjacent predictions that indicate that the prediction results with Polynomial Regression are accurate because they are close to the original data.



Fig. 6 SVM Prediction

Figure 6 shows that the blue line as real data and the purple line as the result of the prediction show not close together, so the SVM method is not accurate for the prediction of covid-19 in Indonesia. From some of the above comparisons from the results of the predictions of the authors concluded that the Polynomial Regression method is more suitable to be used to predict covid-19 based on covid-19 patient data in Indonesia using degree 6.

3.1.1 Comparative Analysis Prediction Accuracy. After making predictions on covid-19 patient data with Polynomial Regression and Support Vector Machine methods, the next step is to compare the prediction accuracy results of the two methods used to find out which method is best to use to predict the number of covid-19 positive patients in Indonesia.

To test or evaluate the accuracy and performance of polynomial regression and support vector machine data mining methods, mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE) is used. But before that, the dataset that has been done pre-processing is divided into training data (training data) and data testing (test data) first by using split data on the sklearn library module. As for the accuracy results of each method as follows:

Table 4 Accuracy of Polynomials Regression Predictions

Test Method	Polynomial Regression	
	Result	Degree
Mean Absolute Error	4146.025749867596	6
Mean Suared Error	19031800.02642069	6
Mean Absolute Persetage Error	0.67%	6

Table 5 Accuracy of Support Vector Machine Predictions

Test Method	Support Vector Machine	
	Result	Degree
Mean Absolute Error	48683.01806620052	2
Mean Suared Error	2528778193.6754	2
Mean Absolute Persetage Error	7%	2

Table 4 and Table 5 show a comparison of the accuracy of prediction results using methods in Polynomial Regression and Support Vector Machine. The trial was conducted using 276 data as training and testing data. The results of the trial with the Polynomial Regression method showed that from the overall data testing, the results of the Calculation of Average Absolute Error accuracy of 4146.025749867596, Mean Squared Error amounted to 19031800.02642069, and by 0.67%. While the trial with the Support Vector Machine method resulted in the calculation of the accuracy of Mean Absolute Error of 48683.01806620052, Mean Squared Error of 2528778193.6754, and Mean Absolute Percentage Error by 7%.

When viewed from the performance in the running program process, in Table 4 it appears that the Polynomial Regression method is much more accurate than using the Support Vector Machine method. This is based on the striking number difference from the Mean Absolute Error accuracy measurement, Mean Squared Error from the method has an assessment that the method of calculation results that are close to zero then the calculation is more accurate. As for Mean Absolute Percentage Error for Polynomial Regression by 0.67% and Support Vector Machine by 7% which is the smallest value can be concluded more accurately.

Therefore, the author concluded that the Polynomial Regression method is well used to predict the spread of covid-19 in Indonesia, based on the accuracy calculations that have been done.

4 Conclusion

Based on the results of the analysis that has been done, covid-19 cases in Indonesia have not shown when it will end, even early in 2021 covid-19 cases in Indonesia became a record number of people exposed to covid-19. The Prediction Model in this study used Polynomial Regression and Support Vector Machine-sized algorithms. The amount of data in this is 276, data training 256, and

data testing 20. The SVM method in this study used the Poly kernel with a gamma parameter = 1, epsilon = 1, degree = 2, and optimum C = 10, so it has an accuracy of MEA = 45119.32514, MSE = 2186704154.880, and MAPE = 0.07008. Of the several prediction accuracy tests using the SVM method with parameters as mentioned is the most optimal result. The Polynomial Regression method in this study used degree=6 and produced an accuracy value of MAE= 4146.02574, MSE= 19031800.0264, MAPE = 0.006174. Thus, Polynomial Regression have better result than Support Vector Machine.

References

- [1] Brown, E. E., Kumar, S., Rajji, T. K., Pollock, B. G., and Mulsant, B. H., 2020, "Anticipating and Mitigating the Impact of the COVID-19 Pandemic on Alzheimer's Disease and Related Dementias," doi: [10.1016/j.jagp.2020.04.010](https://doi.org/10.1016/j.jagp.2020.04.010).
- [2] Ezzikouri, S., Nourlil, J., Benjelloun, S., Kohara, M., and Tsukiyama-Kohara, K., 2020, "Coronavirus disease 2019—Historical context, virology, pathogenesis, immunotherapy, and vaccine development," *Human Vaccines and Immunotherapeutics*, **16**(12), pp. 2992–3000.
- [3] Villavicencio, C. N., Macrohon, J. J. E., Inbaraj, X. A., Jeng, J. H., and Hsieh, J. G., 2021, "Covid-19 prediction applying supervised machine learning algorithms with comparative analysis using weka," *Algorithms*, **14**(7), p. 7.
- [4] Telaumbanua, D., 2020, "Urgensi Pembentukan Aturan Terkait Pencegahan Covid-19 di Indonesia," *QALAMUNA: Jurnal Pendidikan, Sosial, dan Agama*, **12**(01), pp. 59–70.
- [5] Bappenas, 2021, "Proyeksi COVID-19 di Indonesia," .
- [6] Baihaqi, W. M., Dianingrum, M., and Ramadhan, K. A. N., 2019, "Regresi Linier Sederhana untuk Memprediksi Kunjungan Pasien di Rumah Sakit Berdasarkan Jenis Layanan dan Umur Pasien," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, **10**(2), pp. 671–680.
- [7] Pratikto, F. R., 2020, "Prediksi Akhir Pandemi COVID-19 di Indonesia dengan Simulasi Berbasis Model Pertumbuhan Parametrik," *Jurnal Rekayasa Sistem Industri*, **9**(2), pp. 63–68.
- [8] Nuraini, N., Khairudin, K., and Apri, M., 2020, "Modeling Simulation of COVID-19 in Indonesia based on Early Endemic Data," *Communication in Biomathematical Sciences*, **3**(1), pp. 1–8.
- [9] Rustan, R. and Handayani, L., 2020, "the Outbreak'S Modeling of Coronavirus (Covid-19) Using the Modified Seir Model in Indonesia," *Spektra: Jurnal Fisika dan Aplikasinya*, **5**(1), pp. 61–68.
- [10] Parhusip, H. A., 2020, "Study on COVID-19 in the World and Indonesia Using Regression Model of SVM, Bayesian Ridge and Gaussian," *Jurnal Ilmiah Sains*, **20**(2), p. 49.
- [11] Arianto, F. S. D. and P. N., 2020, "Prediksi Kasus Covid-19 Di Indonesia Menggunakan Metode Backpropagation Dan Fuzzy Tsukamoto," *Jurnal Teknologi Informasi*, **4**(1), pp. 120–127.
- [12] Gunawan, R., 2004, *Makalah untuk Pertemuan Dosen UKDW yang akan melaksanakan penelitian pada tahun 2005*, Metodologi Tindakan Action Research.
- [13] Kholis, M. N., Fraternali, and Wahidin, L. O., 2020, "Prediksi Dampak Covid-19 Terhadap Pendapatan Nelayan Jaring Insang Di Kota Bengkulu," *ALBACORE Jurnal Penelitian Perikanan Laut*, **4**(1), pp. 001–011.
- [14] Velavan, T. P. and Meyer, C. G., 2020, "The COVID-19 epidemic," *Tropical Medicine and International Health*, Vol. 25, Med. Int. Heal, vol. 25, no. 3, pp. 278–280.
- [15] Cui, J., Li, F., and Shi, Z. L., 2019, "Origin and evolution of pathogenic coronaviruses," *Nature Reviews Microbiology*, **17**(3), pp. 181–192.
- [16] Jonassen, C. M., Kofstad, T., Larsen, I. L., Løvland, A., Handeland, K., Follestad, A., and Lillehaug, A., 2005, "Molecular identification and characterization of novel coronaviruses infecting graylag geese (*Anser anser*), feral pigeons (*Columbia livia*) and mallards (*Anas platyrhynchos*)," *Journal of General Virology*, **86**(6), pp. 1597–1607.
- [17] Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., Zhu, H., Zhao, W., Han, Y., and Qin, C., 2019, "From SARS to MERS, thrusting coronaviruses into the spotlight," *Viruses*, **11**(1), p. 59.
- [18] Hao, Z., Singh, V. P., and Xia, Y., 2018, "Seasonal Drought Prediction: Advances, Challenges, and Future Prospects," *Reviews of Geophysics*, **56**(1), pp. 108–141.
- [19] Tyrrell, D. A., Bynoe, M. L., and Hoorn, B., 1968, "Cultivation of "Difficult" Viruses from Patients with Common Colds," *British Medical Journal*, Vol. 1, Med. J, vol. 1, no. 5592, pp. 606–610.
- [20] Sun, B., Liu, H., Zhou, S., and Li, W., 2014, "Evaluating the performance of polynomial regression method with different parameters during color characterization," *Mathematical Problems in Engineering*, **2014**, p. 3.
- [21] Santosa, G., 2014, *Data Mining Pemodelan Regresi Polinomial Terhadap IHSG dan Uji Keterikatan dengan Kenaikannya*, October, Universitas Kristen Duta Wacana.
- [22] Ritonga, A. S. and Purwaningsih, E. S., 2018, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengeelasan Smaw (Shield Metal Arc Welding);," *Ilmiah Edutic*, **5**(1), pp. 17–25.
- [23] Cristianini, N. and Shawe-Taylor, J., 2000, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Vol. 30, [Online]. Available.
- [24] Campbell, C. and Yiming, Y., 2011, *Learning with Support Vector Machines (Synthesis Lectures on Artificial Intelligence and Machine Learning)*, Learning with Support Vector Machines (Synthesis Lectures on Artificial Intelligence and Machine Learning.
- [25] Sain, S. R. and Vapnik, V. N., 1996, *The Nature of Statistical Learning Theory*, Vol. 38, Springer-Verlag, USA.
- [26] Wang, W., Men, C., and Lu, W., 2008, "Online prediction model based on support vector machine," *Neurocomputing*, **71**(4-6), pp. 550–558.
- [27] Brereton, R. G. and Lloyd, G. R., 2010, "Support Vector Machines for classification and regression," *Analyst*, **135**(2), pp. 230–267.
- [28] Subagyo, P., 2002, *Forecasting, Konsep dan Aplikasi*, Yogyakarta, BPPE UGM, Yogyakarta.
- [29] Maricar, A. M., 2019, "Analisa Perbandingan Nilai Akurasi Moving Average dan Exponential Smoothing untuk Sistem Peramalan Pendapatan pada Perusahaan XYZ," *Jurnal Sistem dan Informatika (JSI)*, **13**(2), pp. 36–45.
- [30] Nabillah, I. and Ranggadara, I., 2020, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," *JOINS (Journal of Information System)*, **5**(2), pp. 250–255.