Analysis of rainfall classification over Tanah Laut district based on global climate indicators using support vector machine method

N Hayah¹, O Soesanto¹, M A Rahman²

¹Universitas Lambung Mangkurat, Indonesia
²BMKG Stasiun Klimatologi Kelas I Banjarbaru, Indonesia

aya.nrlhyh@gmail.com

Abstract. The Support Vector Machine (SVM) classification method can be applied in various fields, one of which is meteorology and climatology in rainfall forecasting. Thus, a study was conducted by classifying rainfall to recognize the relationship between global phenomena and rainfall and the results of applying the classification using the SVM method to rainfall in the Tanah Laut Regency. The analysis is carried out using the SVM Multiclass concept with 4 categories of rainfall classification: low, medium, high, and Extreme. The kernel used in SVM is the RBF kernel with optimization parameters used, namely Cost (C) 1,5,10,15 and Gamma (γ) 1,5,10,15. The dataset formed is based on the annual period, climatic conditions, and seasonality. The Spearman Rank correlation test describes the relationship between global phenomena and rainfall with a correlation range of (−0.1456) – (0.43144) for the entire dataset. The implementation of the SVM classification method shows that the Cost (C) 10 and Gamma (γ) ≥ 5 parameters obtained the highest accuracy of 100% on the training data. In contrast, in testing the data testing, the accuracy was good, namely the accuracy of 78.00% in La Nina and 81.38% in seasonal periods.

1. Introduction

Indonesia is a region that has a tropical climate, with this tropical climate resulting in very large evaporation of water into the air, so that the intensity of rainfall becomes unstable. High rainfall intensity or what is often called extreme rain can cause flooding, while low intensity rainfall can cause drought which will potentially lead to fire disasters. Such as the flood and fire disasters that occurred in Tanah Laut Regency which was quoted through the website bpbd.tanahlautkab.go.id which stated that in the last 5 years there have been approximately 112 floods and 1267 fires. This should be a concern for researchers to be able to contribute to seeing the effects that occur globally in the Tanah Laut Regency area. The climate in Indonesia is influenced by several factors including global factors such as El Nino/La Nina, Southern Oscillations, and Dipole Mode Event (DME) or Indian Ocean Dipole (IOD).

Classification is one of the sciences found in machine learning, such as the Support Vector Machine method, where the method can be applied in meteorology and climatology, one of which is rainfall. In a previous study conducted by Siregar in 2017 modeling using the SVM method for classification of rainfall. Then in 2019 and 2020, research was carried out using a comparison between the SVM and Naive Bayes methods and the determinant in determining the classification of rainfall. From the
conclusion of previous studies that the SVM method produces better accuracy than other methods. [2], [3], [4]

Based on the description above, a study was carried out using the SVM method in determining the classification of rainfall status in Tanah Laut Regency based on 4 classifications of rainfall status according to the intervals of rainfall categories in the BMKG, namely low, medium, high, and extreme. While the variables used are factors that influence the climate in Indonesia globally in the form of index values such as ONI (Ocean Nino Index), DMI (Dipole Mole Index), and OSI (Southern Oscillation Index).

2. Literature Review

2.1. Rainfall

Rainfall can be interpreted as the height of rainwater that collects in a flat place, does not seep, does not flow, does not evaporate, and has units of millimeters (mm) [5]. Rainfall is the amount of water that falls on the ground during a certain period if no removal occurs by the process of evaporation, flow and infiltration, which measured in units of height [6]

2.2. El Nino Southern Oscillation

ENSO is a phenomenon of deviation from sea surface temperature in the Pacific Ocean near the central and eastern equator. ENSO is a non-periodic Global Climate System. El Nino is the warm ENSO phase and La Nina is the cold ENSO phase. El Nino is identified through the increase in sea surface temperature in the Equatorial Pacific waters, while La Nina is the opposite condition in the same region. El Nino can cause a decrease in sea surface temperature in Indonesian waters and La Nina tends to increase sea surface temperature in Indonesian waters [7]. To know the phenomenon of ENSO used several indices, namely ONI (Oceanic Nino Index) and SOI (Southern Oscillation Index). Oceanic Nino Index (ONI) is based on Temperature Sea Level (SST) of the region's average Nino 3.4, and is the primary measure for monitor, assess and predict ENSO. While the determination of the SOI index is based on the difference in air pressure at sea level between Tahiti and Darwin. [8]

2.3. Data Mining

Data mining is an application process in extracting data. Data mining is also an analysis of a set of data in finding unexpected relationships and to summarize data in a new way so that it can be easily understood. [9]

2.4. Spearman Rank Correlation

Spearman correlation is included in the nonparametric statistical category correlation analysis. Spearman introduced this correlation in 1904. The assumption is that the data is a random sample consisting of n pairs of numerical or non-numeric observations and which is used at least has an ordinal scale to be analyzed by ranking data. The data analyzed on the Rank Spearman correlation does not have to meet the normal distribution [10]. Here is the formula for the Rank Spearman correlation:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$  \hspace{2cm} (1)

2.5. Support Vector Machine

In 1992 for the first time SVM was introduced by Vapnik as a series of superior concepts in the field of pattern recognition [11]. SVM has a linear principle, but now SVM has been developed to work on non-linear problems. The way SVM works on non-linear problems is to incorporate the kernel concept in a high-dimensional space. In this dimensional space, we will look for a separator or what is often called a hyperplane. Hyperplanes can maximize the distance or margin between data classes. The best hyperplane between the two classes can be found by measuring the margin and finding the maximum point. Efforts to find the best hyperplane as a class separator are the core of the process in the SVM method. In SVM there is a function called the kernel. This kernel function is used to solve non-linear
problems. The function of the kernel allows to implement a model in a higher dimensional space (feature space) [12]. Several Kernel functions are often used in the SVM literature, including the following [13]:

a. The polynomial kernel is a kernel that is often used for image classification.
\[ K(x_i, x_j) = (\gamma x_i^T x_j + d)^p, \gamma > 0 \]  
(2)

b. The RBF kernel, commonly known as the Radial Basis Function, is commonly used for valid (available) data and is the default in SVM tools.
\[ K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2), \gamma > 0 \]  
(3)

c. The Tangent Hyperbolic kernel is the most commonly used kernel for neural networks.
\[ K(x_i, x_j) = \tanh(x_i^T x_j + d) \]  
(4)

2.6. One Against All

One-Against-All (OAA) is one approach to the problems in multiclass in the SVM method. In addition, an approach that can be used for multiclass SVM problems is an approach using OAO or One Against One. In this study, the approach used is the OAA approach. OAA works by making the class of the i-th data positive and negative for data that does not come from the i-class. As in Table 1, an example of a classification problem with the OAA approach is described with 4 classes. [14]

| Table 1. Example of 4 Class Classification Problems |
|----------------------------------|
| \( y_i = 1 \) | \( y_j = -1 \) | Kernel Hypothesis |
| Class 1 | not Class 1 | \( f^1(x) = (w^1)x + b^1 \) |
| Class 2 | not Class 2 | \( f^2(x) = (w^2)x + b^2 \) |
| Class 3 | not Class 3 | \( f^3(x) = (w^3)x + b^3 \) |
| Class 4 | not Class 4 | \( f^4(x) = (w^4)x + b^4 \) |

2.7. Confusion Matrix

A confusion Matrix or error matrix is a matrix that displays a visualization of the performance of the classification algorithm using the data in the matrix. It compares the prediction classification to the actual classification in the form of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) of information. The confusion matrix for the two-class classifier system is as follows. [15]

| Table 2. Confusion Matrix |
|-------------------------|
| Predicted Value | Positives | Negatives |
| Positives | TP | FP |
| True Positive | False Positive |
| Negatives | FN | TN |
| False Negative | True Negative |

3. Method

3.1 Source Of Data

This study uses quantitative data sourced from secondary data obtained from the Class 1 Banjarbaru Meteorology, Climatology and Geophysics Agency (BMKG), the NOAA website, namely the National Weather Service, which can be accessed via http://www.cpc.ncep.noaa.gov/ and https://psl.noaa.gov/, as well as the Bureau of Meteorology website, which can be accessed at http://www.bom.gov.au/.

3.2 Research Variables

In the following, the dependent and independent variables used in this study are presented in Table 3.
Table 3. Research Variables

| Variable Type          | Variable Name          | Scale    |
|------------------------|------------------------|----------|
| Independent Variables (X) | Southern Oscillation Index (SOI) | Interval |
|                        | Dipole Mole Indeks (DMI)  | Interval |
|                        | Ocean Nino Index (ONI)    | Interval |
| Dependent Variable (Y)  | Rainfall               | Ratio    |

3.3 Research Procedures

This research procedure uses several stages, which will be described as follows:

1) Pre-processing data.
   a. Identify missing values in each observation post of rainfall in Tanah Laut Regency. Only the rainfall post which produces the lowest percentage of missing value will be analyzed.
   b. Imputing missing values on the selected rainfall post data by replacing the missing data.
   c. Changing the type of rainfall data from ratio to ordinal is grouped based on the rainfall data group determined by the BMKG.

2) Identify the relationship between rainfall and global scale climate indicators, namely DMI, SOI, and ONI, by analyzing the correlation between variables X and Y using Spearman correlation analysis.

3) Doing Support Vector Machine (SVM) Analysis
   a. Classifying the types of data is divided into two divisions, namely, with training data of 80% and testing data of 20%.
   b. After the data sharing is done, the training data will be carried out with the SVM process with the multiclass concept using the RBF kernel function.
   c. The next step is to obtain an accuracy where the highest accuracy value is the best classification method for this research data so that the model will be used for prediction on testing data using R software.
   d. Then form a confusion matrix on the testing data.
   e. Conclude with the results of the classification formed through the confusion matrix table.

4. Result and Discussion

4.1 Preprocessing Data

The first step taken in data analysis is data preparation because in data recording, especially in recording rainfall data, sometimes problems often occur such as damaged measuring devices, displacement of rainfall gauges, unavailability of rainfall recorders, so that there will be causes a data to be missing. This also happened at the rainfall observation posts in Tanah Laut Regency. Thus, it is necessary to preprocess the data before conducting the analysis, including identifying missing values and imputing data. The results of the identification of missing values in rainfall data at 8 observation posts in Tanah Laut Regency in table 1:

Table 4. Identification of Missing Value in Rainfall Data in Tanah Laut District

| No | Post Name      | n (Length Data) | Missing Data | (%) Missing Data |
|----|----------------|-----------------|--------------|-----------------|
| 1  | Kintap         | 468             | 89           | 19%             |
| 2  | Kurau          | 468             | 104          | 22.22%          |
| 3  | Panyipatan     | 468             | 157          | 33.55%          |
| 4  | Pabahanan      | 468             | 386          | 82.45%          |
| 5  | SMPK Pelaihari | **468**         | **19**       | **4%**          |
| 6  | Telaga         | 468             | 381          | 81.41%          |
| 7  | Takisung       | 468             | 160          | 34.18%          |
| 8  | Tambang Ulang  | 468             | 123          | 26.28%          |
Based on the identification of missing values in the rainfall posts in Tanah Laut Regency above, this study uses the rainfall observation posts at SMPK Pelaihari because they have a low percentage of missing values, which is 4%. The next stage of data pre-processing is imputing the missing data found. Various ways can be done to fill in this missing value, one of which is to fill in the average value of the rainfall, but it can also be done with statistical analysis techniques. However, in this study, the missing data were imputed with the value of the analysis conducted by BMKG Class 1 Banjarbaru experts at the Pelaihari Junior High School rainfall observation post on the missing value. So that the data becomes intact and there are no missing values. For global-scale climate indicator variables such as ONI, DMI, and SOI, no checking for missing values is carried out because the data is constantly updated, so that this stage is not required for these variables.

4.2 Identify the Relationship of Global Climate Indicators with Rainfall

The results of the correlation analysis between global climate indicators such as ONI, DMI, and SOI with rainfall based on annual datasets can be seen in table 5.

| Factor | Correlation | P-Value | 5% |
|--------|-------------|---------|----|
| ONI    | -0.1791284  | 0.0097300 | Significant |
| DMI    | -0.1479422  | 0.0013280 | Significant |
| SOI    | 0.2179593   | 0.0019550 | Significant |

Based on the table of correlation assessment criteria, it can be seen that the results of the correlation analysis between global climate indicators and rainfall are very weak correlations, namely \( r < 0.25 \). The correlation value produced by ONI and DMI with damaging rainfall is \(-0.1791284\) and \(-0.1479422\), indicating an inverse relationship. If the ONI and DMI values increase, the intensity of rainfall will decrease, whereas the intensity of rainfall will also increase if the ONI and DMI values decrease. The correlation generated by SOI is \(0.2179593\), which has a positive value that indicates that the relationship is unidirectional. This means that the SOI value will increase along with the La Nina phase, which will generally increase rainfall in Tanah Laut Regency, while if the SOI value decreases, it is common for the El Nino phase to occur which causes rainfall in Tanah Laut Regency to decrease. The DMI produces the lowest correlation. For Indonesia in the middle part of the DMI, the value is not too influential because the DMI value will only affect rainfall with extreme values. Then analysis of global climate indicators is carried out with rainfall based on quarterly this is an analysis carried out to see the correlations that occur in phases such as El Nino, Normal, and La Nina. The results of the resulting analysis in table 6.

| Factor | Event | Correlation | P-Value | 5% |
|--------|-------|-------------|---------|----|
| ONI    | El Nino | 0.1012855 | 0.2650000 | Not Significant |
|        | Normal | -0.0663895 | 0.3281000 | Not Significant |
|        | La Nina | -0.2850701 | 0.0012150 | Significant |
|        | El Nino | -0.2822488 | 0.0015620 | Significant |
| DMI    | Normal | 0.0362262 | 0.5939000 | Not Significant |
|        | La Nina | -0.1646021 | 0.0655000 | Not Significant |
|        | El Nino | 0.0876340 | 0.3351000 | Not Significant |
| SOI    | Normal | 0.1459876 | 0.0308000 | Significant |
|        | La Nina | 0.1988385 | 0.0256100 | Significant |

In the El Nino event, ONI and SOI have an insignificant correlation, which means that ONI and SOI do not affect El Nino and rainfall in the Tanah Laut Regency. While in the La Nina phase, ONI and
SOI have a significant relationship, ONI with a negative correlation value of -0.2850701, which indicates an inverse relationship, while in SOI, a positive correlation value of 0.1988385 indicates a unidirectional relationship. Only the SOI variable has a significant effect in ordinary events, with a correlation of 0.1459876 with a fragile correlation category. The analysis carried out for the seasonal period in table 7.

| Factor | Season | Correlation | P-Value | 5%  |
|--------|--------|-------------|---------|-----|
| ONI    | DJF   | -0.1408422  | 0.1299000 | Not Significant |
|        | MAM   | 0.0399942   | 0.6686666 | Not Significant |
|        | JJA   | -0.3196274  | 0.0004434 | Significant |
| DMI    | MAM   | -0.0978592  | 0.2939000 | Not Significant |
|        | JJA   | -0.0378750  | 0.6852000 | Not Significant |
|        | SON   | -0.4049266  | 0.0059420 | Significant |
|        | DJF   | 0.2235000   | 0.1134014 | Not Significant |
| SOI    | MAM   | 0.3779000   | 0.0822608 | Not Significant |
|        | JJA   | 0.0001036   | 0.3511644 | Not Significant |
|        | SON   | 0.0000386   | 0.4816013 | Not Significant |

The table 7 shows that the results of the correlation analysis between global climate indicators and rainfall in the DJF and MAM seasons, there is no significant correlation. In this DJF to MAM season, the ONI and DMI values can be said to be uncorrelated. When viewed from the correlation value produced between SOI and rainfall, the correlation is very weak, although not significant at alpha 5%, which means that it can be said that seasonally SOI has little effect on rainfall that occurs in Tanah Laut Regency.

During the JJA and SON seasons, ONI has a significant correlation with a reasonably strong correlation. The DMI only has a significant correlation in the SON season with a strong correlation because DMI affects extreme rainfall in the southern part of South Kalimantan. While the SOI does not have a significant correlation, SOI does not have an influence on the rainfall that occurs in Tanah Laut Regency. The negative correlation value on ONI and DMI indicates that there is an inverse relationship. If the ONI and DMI values are more significant, the rainfall that occurs will decrease and vice versa.

### 4.3 Analysis of Support Vector Machine (SVM) Methods

In this study, using the Support Vector Machine (SVM) method, the data will be separated into training data and testing data. The data is separated by 80% for training data and 20% for testing data from the total data. The kernel used in this SVM method is the RBF kernel. The RBF kernel function is a kernel that can be used for continuous and non-linear data. In the RBF kernel function that will be used in this study, the Cost (C) and Gamma (γ) parameters will be optimized to find the best accuracy where Cost (C) 0 and Gamma (γ) 0. The parameters that will be used later is the parameter Cost (C) 1,5,10,15 and Gamma (γ) 1,5,10,15.

This study uses the SVM One Against All concept because the rainfall categories formed are divided into 4 classes. Where to distinguish each class, this method requires 3 different levels. At level 1, it functions to separate class 1, namely the "Low" category, which is worth 1, with other classes, which are worth -1. Then the level 2 class separates the "Medium" class 2, worth 1, with other classes worth -1. After that, level 3 functions to separate class 3 "High" worth 1, with level 4 serving to separate class level 4 "Extreme" worth -1.

Furthermore, to determine which data belongs to which class, the data will be processed. If the data value is 1 at level 1, then the data is class 1, and if it is not worth 1 and is worth -1, it will continue
the calculation process back to the next level. This process will continue to be repeated until the data is
worth 1 at a certain level or processed to the last level, namely level 3. For a value of -1 at level 3, the
data is included in the class 4 category.

This SVM analysis is also carried out as with the previous analysis, namely testing the datasets
that have been formed, namely annual data, seasonal data, event data, and combined seasonal event data.
The steps for calculating SVM will be described as follows:

The dataset used in each dataset has been compiled based on annual, quarterly, and seasonal
periods.

a. Divide the data into 80% for training data and 20% for testing.
b. The first equation to be formed is so that the training data in class 1 is labeled +1, and data in classes
2, 3, and 4 are labeled -1 based on what has been described by Sembiring [14]
c. Initialize the parameter value with Cost (C) 1, 5, 10, 15 as alpha value and Gamma (γ) 1, 5, 10, 15.
d. Calculate the RBF kernel function for training data with the following formula:

\[ K(x_i, x_j) = \exp(-\gamma * \|x_i - x_j\|^2) \]
e. After the training data is mapped using the kernel function, the values of \(a\) and \(b\) are then used in the
SVM equation to predict the classification of the testing data.

The calculation using the equation \(f_1(x)\) obtained a row vector with a size of 1 x n (as much as testing
data), which is a prediction of a temporary classification. Next, form the second equation \(f_2(x)\),
\(f_3(x)\), and \(f_4(x)\), with the same steps as the equation.

In SVM analysis based on annual data, the amount of data used is 468 data divided into 80% as training
data, which is 351 data, and 20% as testing data, which is 117 data from the total data. The parameters
used are based on the optimization parameters that have been determined previously. The results
obtained from trial and error in Table 8.

| Parameter | Cost (C) | Training Accuracy | Gamma (γ) |
|-----------|----------|-------------------|-----------|
| \(\gamma = 1\) | \(\gamma = 5\) | \(\gamma = 10\) | \(\gamma = 15\) |
| C = 1     | 0.5641   | 0.7151            | 0.7863    | 0.8376    |
| C = 5     | 0.6382   | 0.8917            | 0.963     | 0.9972    |
| C = 10    | 0.6581   | 0.9345            | 0.9858    | \textbf{0.9972} |
| C = 15    | 0.6809   | 0.9544            | 0.9915    | 0.9972    |

Based on Table 8 above, it is obtained for the value of Cost and. If seen in Table 7, it can be interpreted
that the higher the value of Cost and, the higher the accuracy value that will be obtained. Cost = 15 and
\(\gamma = 1\) are the parameters that produce the most excellent accuracy value, while Cost = 1 and \(\gamma = 1\) are the
parameters that produce the smallest accuracy value.

Furthermore, the training data that has been processed, the best parameters generated by the
training data are used to test the data for testing to test its accuracy. The classification results obtained
for testing data in Table 9.
Table 9. Classification Results of Annual Dataset Testing with the Best Parameters

| Observation | Prediction | Low | Medium | High | Extrem |
|-------------|------------|-----|--------|------|--------|
| Low         | 5          | 23  | 3      | 0    |
| Medium      | 10         | 31  | 3      | 0    |
| High        | 5          | 17  | 5      | 1    |
| Extrem      | 4          | 8   | 1      | 1    |

Table 9 is the classification result using the best parameters obtained in the training process with $\text{Cost} = 15$ and $\gamma = 15$ for data testing. In this analysis, there are 42 precise data classifications where for the "Low" rain status, the observation gets the proper classification, which is 5 data, the suitable classification for the "Medium" rain status is 31 data, the suitable classification for the "High" rain status is 5 data, and proper classification on the rain status “Extreme” of 1 data. While 75 data are not correctly classified.

Table 10. Comparison of Annual Dataset Accuracy

| Best Parameter | Accuracy  | Error    |
|----------------|-----------|----------|
| Cost           | Gamma     | Testing  | Training | Testing |
| 15             | 15        | 99.72 %  | 35.89 %  | 0.28%    | 64 %    |

In the above equation, the accuracy results for testing data are 0.358974 or 35.90%. Furthermore, the error value obtained is 0.641026 or 64.10%. In the SVM analysis with this annual dataset, the accuracy generated by the training accuracy is excellent, while the testing accuracy gets poor accuracy. This is due to the varying amount of data between classes. Besides that, the correlation between variables and rainfall has a very weak correlation, causing data inconsistency in placing classes. The accuracy results generated in the seasonal and quarterly periods in table 11 and 12.

Table 11. Comparison of Accuracy of Climate Condition Dataset

| Event       | Best Parameter | Accuracy | Error    |
|-------------|----------------|----------|----------|
|              | Cost | Gamma | Testing | Training | Testing |
| El Nino     | 10   | 10    | 100%    | 45.83%   | 54.17%   |
| Normal      | 5    | 15    | 100%    | 48.84%   | 51.16%   |
| La Nina     | 10   | 5     | 100%    | 52%      | 48%      |

Table 12. Comparison of Seasonal Dataset Accuracy

| Event | Best Parameter | Accuracy | Error    |
|-------|----------------|----------|----------|
|       | Cost | Gamma | Testing | Training | Testing |
| DJF   | 5    | 10    | 100%    | 52.17%   | 47.83%   |
| MAM   | 5    | 15    | 100%    | 60.87%   | 39.13%   |
| JJA   | 10   | 5     | 100%    | 56.52%   | 63.48%   |
| SON   | 15   | 5     | 100%    | 60.43%   | 81.38%   |
5. Conclusion
The conclusions obtained from this research are:

a. In general, the relationship between global climate indicators and rainfall in Tanah Laut District is very weakly correlated when analyzed throughout the year due to different seasonal effects in each data variable. When examined more at the seasonal level, the event found a relationship in the seasonal period with a correlation range of 0.2235 to 0.43144 with a reasonably strong correlation category, and an event period ranging from 0.1456 to 0.2850 with a weak correlation category.

b. In applying the Support Vector Machine (SVM) method, a combination of parameter optimization experiments was carried out on the classification of rainfall in Tanah Laut Regency so that the resulting changes inaccuracy could be seen. The optimization of the parameters used is by using Cost (C) 1,5,10,15, Gamma (γ) 1,5,10,15. The SVM method used is using the SVM multiclass concept with the RBF kernel. With the increase in accuracy in the large values of Cost (C) and Gamma (γ), it produces the best model, so that the application of the SVM method with global climate indicator variables has the potential to be very good to be applied in BMKG operational activities.

c. The highest level of accuracy obtained in this study is 100% on the training data in the seasonal and event period datasets with variable values of Cost (C) 10 and Gamma (γ) 5. The accuracy results in this study are excellent. Meanwhile, when it was carried out on testing data, it obtained a sufficient accuracy, namely an accuracy of 78.00% in the La Niña event period and 81.37% in the seasonal period. The thing that causes the accuracy produced at the testing stage to be lower than the training in this study is because the data used is not evenly distributed in each class and is also influenced by the level of closeness or the relationship between variables to rainfall.

d. SVM is one method that can assist in determining the classification of the status of rainfall in Tanah Laut district in order to be an alternative method for the BMKG in disaster preparedness, both in the form of fires and floods, which are measured in the intensity of the falling rainfall.

References
[1] Hamada J I, Yamanaka D, M. M, J. F, S. W, A. P and Sribimawati T 2002 J. Meteorol. Soc. Japan. Ser 80 285–310
[2] Siregar A B 2017 Pemodelan support vector machine untuk klasifikasi curah hujan bulanan di kabupaten indramayu
[3] Ilmawan L B and Mude M A 2020 Ilk. J. Ilm. 12 154–161
[4] Rahmadina R P and Paramita N L P S P 2019 Perbandingan metode klasifikasi Support Vector Machine (SVM), Naive Bayes Classifier (NBC), dan diskriminan pada data forest fires. May 2019. [Online]. Available: https://www.researchgate.net/publication/341446064
[5] Laia M L and Setyawan Y 2020 P J. Stat. Ind. dan Komputasi 5 51–61
[6] Prawaka F, Zakaria A and Tugiono S 2016 J. Rekayasa Sipil dan Desain 4 397–406
[7] Arifin M 2010 Modul klimatologi (Jawa Timur: Fakultas Pertanian Universitas Brawijaya)
[8] Philander S G 1989 El Niño, La Niña, and the Southern Oscillation (San Diego: Academic Press)
[9] Zakir A, Sulistya W and Khotimah M K 2010 Perspektif operasional cuaca tropis (Jakarta: Pusat Penelitian dan Pengembangan BMKG)
[10] Daniel W W 1989 Statistika nonparametrik terapan (Jakarta: PT Grafindo)
[11] Santosa B 2007 Data mining teknik pemanfaatan data untuk keperluan bisnis (Yogyakarta: Graha Ilmu)
[12] Nugroho A S 2003 Bioinformatika dan pattern recognition
[13] Karatzoglou A, Smola A, Hornik K and Zeileis A 2004 J. Stat. Softw. 11 1–20
[14] Sembiring K 2007 Tutorial Support Vector Machine Bahasa Indonesia. Sekolah Teknik Elektro dan Informatika (Institut Teknologi Bandung)
[15] Awad M and Khanna R 2015 Efficient learning machines: theories, concepts, and applications for engineers and system designers (Springer nature)