Journal of Applied Science and Advanced Technology
Journal Homepage : https://jurnal.umj.ac.id/index.php/JASAT

Comparison of Genetic Algorithm Optimization with Support Vector Machine (SVM) for Weather Forecast

Siska Andriani[1*], **Fajar Delli Wihartiko**[2]

[1,2]*Universitas Pakuan, Jalan Raya Ciheuleut, Bogor Po Box 425, Indonesia*

ARTICLE INFO

JASAT use only:
Received date : 24 January 2024
Revised date : 04 February 2024
Accepted date : 25 May 2024

Keywords:
Optimation,
Weather Forecast,
LVQ,
GA,
SVM.

ABSTRACT

Weather forecasts are one of the important factors for daily activities. It can be used for daily work activities such as farming, aviation, production and distribution. The Meteorology, Climatology and Geophysics Agency makes weather forecasts based on weather parameters, namely temperature, air pressure, solar radiation, humidity and rainfall. The weather forecast class is divided into 5 classes, namely Cloudy, Rainy, Sunny, Cloudy Rainy and Cloudy Sunny. In this research, a comparison of weather forecast models using Learning Vector Quantization optimization and Genetic Algorithms will be made with weather forecast models using the Support Vector Machine method. The data used in this research is weather data at the Citeko Class III Climatology and Geophysics station, the data used is data from the last 3 years. Then the data is divided into training and test data using percentage split with a division of 65% used for training data and 35% used for test data. After making the model using the LVQ-GA and SVM methods, a comparison of the model test results was carried out, from the test results the accuracy value was calculated using a confusion matrix for each model. The accuracy result of the LVQ-GA optimization weather forecast model was 73%, while the weather forecast model using the SVM method obtained an accuracy value of 81.5%, thus the results from SVM were better.

Introduction

The rapid growth of data accumulation causes a lot of data to be collected but cannot be processed properly so that good information is not obtained. By applying artificial technology, one of the artificial methods will be used to make weather forecasts, namely the Artificial Neural Network method. Artificial Neural Networks are a suitable method for weather forecasting based on existing weather parameters. Weather predictions require data for at least the last year, this data is taken every day at the times determined by the BMKG. Optimization is the process of finding the best value based on predetermined criteria. Forecasting is an activity that predicts an event in the future. According to the World Meteorological Organization (WMO), weather is the condition of the atmosphere at a certain place and time that can be measured and observed, and includes parameters such as temperature, humidity, air pressure, wind and precipitation [1].

Genetic Algorithm (GA) optimization is carried out when weather forecast modeling has been carried out using the ANN method. ANN in its modeling relies on previous data which is used as learning data which can then be used for future predictions due to learning from past data. The ANN method used is Learning Vector Quantization (LVQ). The weather forecast application using LVQ was created as a tool to predict weather which is in fact difficult to predict, as well as assisting BMKG in data management, especially for determining the classification of weather predictions in the Citeko Bogor class III meteorological and geophysical station area.

The weather forecast application using the LVQ method was created as a tool for predicting the weather. To obtain optimal prediction results in the research, prediction results will be optimized using GA so that this research is named forecast optimization using the LVQ-GA method. Apart

* Corresponding author.

E-mail address: siska.andriani@unpak.ac.id

DOI: https://dx.doi.org/10.24853/JASAT.6.3.83-90

from that, this research will also create weather forecast modeling using SVM. So apart from getting weather forecast modeling using the LVQ method which is optimized using GA, this research will also make a comparison of algorithm methods for weather forecasting, namely using SVM. This is done so that we can find out which algorithm method is better to use in terms of weather forecasting, so that the weather forecast results obtained are better. Winarti has already used the ANN method for forecasting. In 2018, ANN was used to forecast rain events in the city of Pontianak. The ANN used is Backpropagation ANN [2].

Weather forecast modeling using the LVQ method optimized using GA has been carried out in research entitled Optimizing Weather Forecasts using LVQ with Genetic Algorithms [3] and the novelty of this research is that weather forecast modeling was created using the SVM method as a comparison for weather forecast results. Classification problems can be solved using the SVM classification method.

Methods

This research was conducted using daily weather data from BMKG Class III Citeko Cisarua Bogor. The data used as a training data generator is BMKG daily weather data for 3 years. The research stages carried out are: (1) Data collection, (2) Data sharing, (3) weather prediction optimization modeling using the LVQ method with genetic algorithms, (4) Modeling using SVM, (5) testing, analysis and evaluation. The flow diagram of these stages is shown in **Figure 1**.

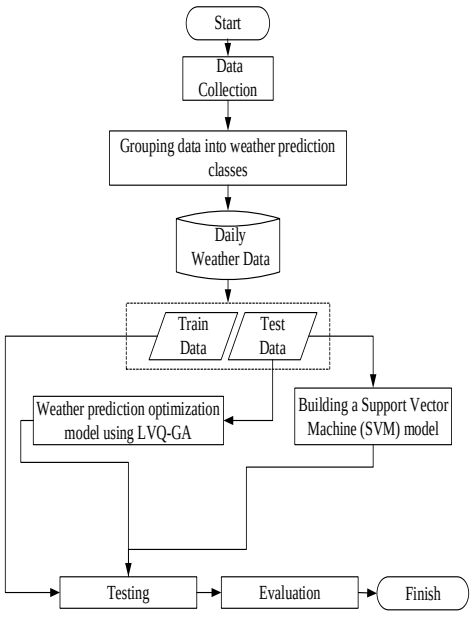

Figure 1. Research methods

1. Data Colection

The data used in this research is BMKG Class III Citeko Bogor daily weather data for 3 years. BMKG Class III Citeko is located in Citeko Village, Cisarua District, Bogor.

2. Data Division

Data division is carried out to divide weather data into training data and test data. Training data will be used to generate the model and test data will be used to evaluate the performance of the prediction optimization model created. Data distribution uses a percentage split with the provisions of 65% for training data and 35% for test data.

3. Weather Forecast Optimization Modeling using LVQ-GA

Artificial Neural Network (Artificial Neural Network) is a computer system that imitates the way the human brain thinks and is able to complete a number of calculation processes. There are several types of neural networks, but almost all of them have the same components. Neural networks consist of several neurons just like the human brain, and these neurons are interconnected with each other or are known as weights.

(**Figure 2**) shows the LVQ network in the research carried out, namely with 5 units in the input layer, and 2 units (neurons) in the output layer [4]. LVQ is an ANN with higher accuracy and faster computing time [5].

Genetic Algorithms are inspired by Darwin's theory of evolution. GA is a search algorithm based on natural system mechanisms, namely genetics and natural selection. In the GA application, the solution variable is encoded into a string structure that represents a sequence of genes, which is a characteristic of the problem solution [6]. In the genetic algorithm development cycle looking for the 'best' solution (chromosome), there are several processes as follows:

1. Initialization
2. Reproduction
3. Evaluation
4. Selection

The solution to a problem must be mapped (encoded) into a chromosome string [7]. GA is different from conventional search techniques, GA comes from a set of randomly generated solutions. This set is called a population. Individuals represent problems that have been encoded in chromosome

form. Individuals will compete to reproduce, which consists of crossover and mutation, to produce offspring that have better traits or genes than their parents. Next, individuals and their offspring will be selected based on fitness that has been previously evaluated. Individuals who have good fitness will pass selection to the next generation [8].

The selection used is elitism. Sorting fitness from largest to smallest then taking individuals with high fitness, this is based on what was done by [9] and [10]. At this stage, we carry out weather forecast modeling by optimizing LVQ weights using GA. Increasing LVQ performance can be done by optimizing weights, one way is using the genetic algorithm method [11].

4. Building a Support Vector Machine (SVM) model

The SVM method is rooted in statistical learning theory, the results of which are very promising to provide better results than other methods [12]. SVM works by mapping data into a high-dimensional feature space so that data points can be categorized, even though the data cannot be separated linearly. The separator between categories is found, then the data is transformed in such a way that the separator can be described as a hyperplane. After that, the characteristics of the new data can be used to predict which groups should have new records. SVM uses a linear model as a decision boundary with the following general form:

$$y(x) = w^T f(x) + b \quad (1)$$

where x is the input vector, w is the weight parameter, f(x) is the basis function, and b is a bias.

The simplest form of linear model for decision boundaries is:

$$y(x) = w^T x + w \quad (2)$$

Where x is the input vector, w is the weight vector and w0 is the bias. So, the decision boundary is y(x)=0, which is a hyperplane of dimension (D-1)

1. An input vector x will be classified into class 1 (R1) if y(x)³0, and class 2 (R2) if y(x)<0
2. If x and x lie on the decision boundary (DS), then y(xA)=y(xB)=0 or wT (x -x) = 0, so w is perpendicular to all vectors in DS. In other words w determines the orientation of DS
3. The distance from the starting point to DS is -w0 /||w||. In other words w0 determines the location of DS. The distance of any vector x to DS and in the direction w is y(x)/||w||. [13]

In this study, the Support Vector Machine (SVM) method was used to forecast weather into five classes, namely cloudy, rainy, sunny, cloudy rain and partly cloudy. The process sequence starts with the data being transformed into a matrix and given weights, then the training and testing process is carried out on the data using the SVM method to obtain results that will be evaluated to calculate the accuracy value of the results. The parameters in SVM are C or kernel and Gamma.

Classification problems can be solved using the SVM classification method. This is done by finding a line that can separate two groups of data. The best hyperplane separator can be obtained by measuring the hyperplane margin and its maximum point. Margin is the distance between the hyperplane and the closest data from each class. The data closest to the best hyperplane is called the Support Vector [14].

5. Testing and Evaluation

Testing is carried out by comparing the actual value with the predicted value. Next, evaluation of the prediction results is carried out using a confusion matrix. The confusion matrix is a method for evaluating the level of success of data when called against its data class. After the two forecast results have been tested, the next step is to compare the two forecast results so that it can be seen which method is suitable for weather forecasting so that the weather forecast results obtained are better.

The confusion matrix is a method for evaluating the level of success of data when called against its data class [15]. Comparison of observation class data with prediction class data is shown in **Table 1**, while the formula for accuracy is shown in Equation 3.

Table 1. Comparison of observation class data with forecast class data

	Positive Forecast Class Data	**Negative Forecast Class Data**
Positive Observation Class Data	TP	TN

Negative Observation Class Data	FP	FN

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} X100\% \quad (3)$$

Where:
TP (True Positive) = observation data and prediction data produce positive values.
TN (True Negative) = observed data and predicted data produce negative values.
FP (False Positive) = observation data and prediction data produce positive values.
FN (False Negative) = observation data and prediction data produce negative values.

Results and Discussions

1. Study Area Data

This research was conducted at the BMKG Class III Citeko Station. An example of weather forecast characteristics with actual data is shown in **Figure 2.**

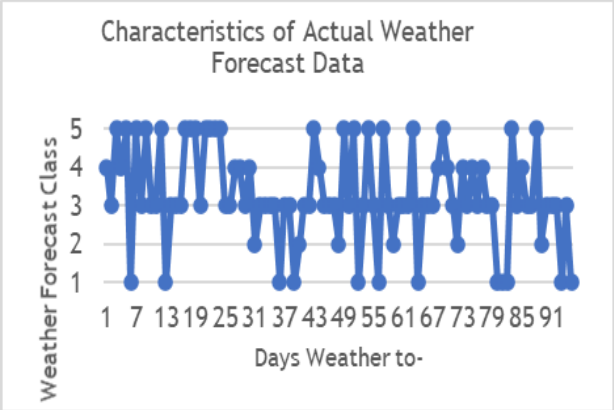

Figure 2. Graphics of actual weather forecast data

2. Data Division

Weather forecast data is divided into 2 data parts, namely Training Data and Test Data. In this study, data division was carried out using percentage split. Where the best division is 65% of the data used as training data and 35% used as test data. The data used was 1096 data divided into 712 data as training data and 384 data as test data. **Figure 3** shows the distribution of the weather forecast class data used.

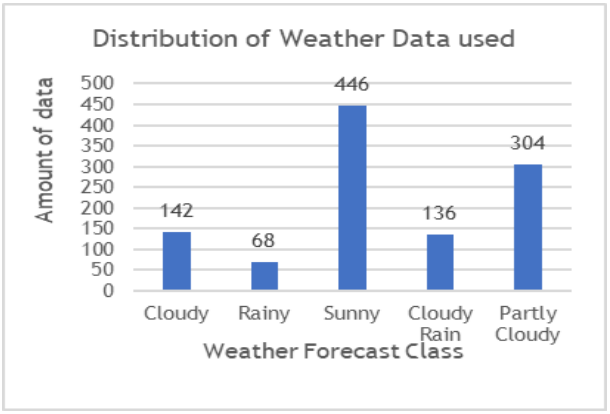

Figure 3. Distribution of weather forecast data for each class

3. Weather Forecast Optimization Modeling using LVQ-GA

A Weather prediction optimization modeling uses the LVQ-GA method, namely in this case GA plays a role in optimizing the best weights produced by LVQ so as to obtain new weights which will later be used to make weather forecasts. The GA parameters used in this research are as follows:

1. Solution representation: Binary String
2. Population: 60 individuals
3. Number of generations: 100 generations
4. Crossover: Simple crossover (0.6)
5. Mutation: Binary Mutation (0.0333)

This GA model is based on a model carried out [16]. A comparison of weather forecast results using the combined LVQ-GA method with actual data can be seen in **Figure 4**.

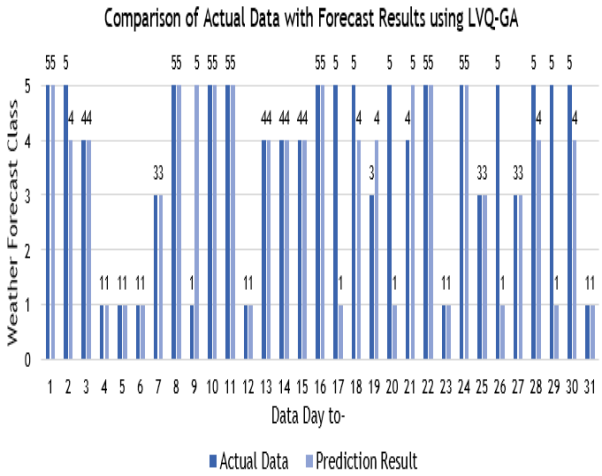

Figure 4. Comparison of LVQ-GA prediction results with actual data

Figure 4 shows a comparison of actual data with LVQ-GA predicted data. From the 31 test data samples taken, it can be seen that there are several results that do not match the actual data, namely in test data 2, 9, 17, 18, 19, 20, 21, 26, 28, 29 and 30.

4. Weather Forecast Modeling using SVM

Weather forecast modeling using SVM in the research was carried out to be a comparison of the results obtained previously with LVQ-GA.

The SVM parameters used in this research are as follows:

1. C : 1.0
2. Gamma : 1/(5*data variance)
3. Kernel : Polykernel
4. Parameter tolerance: 0.001
5. Mutation : Binary Mutation (0.0333)

A comparison of weather forecast results using the SVM method with actual data can be seen in **Figure 5.**

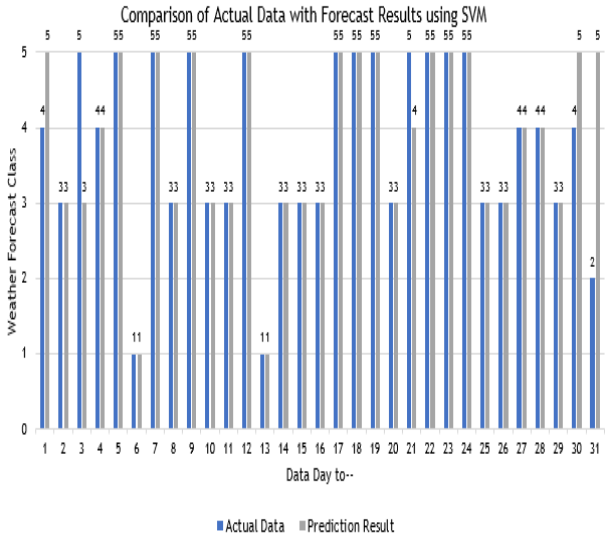

Figure 5. Comparison of SVM prediction results with actual data

Figure 5 shows a comparison of actual data with SVM prediction data. From the 31 test data samples taken, it can be seen that there are several results that do not match the actual data, namely test data 1, 3, 21, 30 and 31.

Conclusions

Judging from the forecast results using the LVQ-GA and SVM models shown in (Figure 4) and (Figure 5), it shows that the weather forecast results produced by the SVM model are better than those produced by the LVQ-GA forecast model. After carrying out the test, the next step is to evaluate the two models that have been created so that the comparison of the results obtained is known. Evaluation is carried out using the confusion matrix method with the formula listed in equation 3, from the predetermined distribution of training and test data, namely 65% for training data and 35% for test data. So, overall testing was carried out from the test data, namely 35% with a total of 384 data, so the evaluation results were obtained, namely for the forecast model using LVQ-GA, it got an accuracy result of 73%, while the forecast model using SVM produced an accuracy value of 81.5%.

This research has succeeded in comparing the optimization of weather forecast models using the LVQ-GA method with SVM. The actual weather data used is data from the BMKG Class III Climatology and geophysics station Citeko Bogor for 3 years with weather attributes or parameters consisting of temperature, evaporation, sun, humadity and rainfall. The weather forecast class consists of 5 classes, namely cloudy, rainy, sunny, sunny, cloudy and cloudy rain. Weather data is divided into training data and test data using a percentage split of 65% training data and 35% test data. So, of the total 1096 data, 712 were used as training data and 384 as test data. The accuracy value obtained by the model using LVQ-GA was 73%, while the accuracy value of the weather forecast model using SVM obtained an accuracy value of 81.5%. There is a difference in the accuracy results of the two models that have been created. Thus, modeling using SVM can be said to be better for use as a model for weather forecasting in this research because the accuracy results obtained from testing and evaluation are better.

Acknowledgment

Thank you to LPPM Pakuan University for providing support and support in this research process, so that this research can be completed well.

Funding

This research was funded by LPPM Pakuan University, an internal research grant for novices lecturers.

Author Contributions

The author's contribution to this research is as follows: Siska Andriani and Fajar Delli.;

methodology, Siska Andriani.; software, Siska Andriani.; validation, Siska Andriani and Fajar Delli.; formal analysis, Siska Andriani.; investigation, Siska Andriani and Fajar Delli.; resource persons, Siska Andriani and Fajar Delli.; data curation, Siska Andriani.; writing—preparation of original draft, Siska Andriani.; writing—review and editing, Siska Andriani.; visualization, Siska Andriani; supervision, Siska Andriani.; project administration, Siska Andriani.; obtaining funding, Siska Andriani and Fajar Delli. All authors have read and approved the published version of the manuscript.

Conflicts of Interest

The authors declare there are no conflicts of interest and the funders had no role in the study design; in the collection, analysis, or interpretation of data; in script writing; or in the decision to publish the results.

References

[1] WMO, "World Meteorological Organization." https://public.wmo.int/en/our-mandate/weather (accessed Dec. 22, 2023).

[2] V. Winarti, M. I. Jumarang, and A. Apriansyah, "Prakiraan Kejadian Hujan di Kota Pontianak dengan Metode JST-Logika Fuzzy," *Prism. Fis.*, vol. 6, no. 2, pp. 117–123.

DOI: https://dx.doi.org/10.26418/pf.v6i2.26522

[3] S. Andriani and K. Subandi, "Weather Forcast Optimization Using Learning Vector Quantization Methods with Genetic Algorithms," *J. Appl. Sci. Adv. Technol (JASAT).*, vol. 3, no. 2, pp. 45–50, 2020.
DOI: https://doi.org/10.24853/jasat.3.2.45-50

[4] S. Kusumadewi, "Membangun Jaringan Syaraf Tiruan Menggunakan MATLAB & EXCEL LINK," *Yogyakarta Graha Ilmu*, 2004.

[5] I. Afrianto, "Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation Dan Learning Vector Quantization Pada Pengenalan Wajah," *KOMPUTA J. Komput. dan Inform.*, vol. 1, no. 1, 2012.

[6] A. Kustiyo, A. Buono, and N. Apriyanti, "Optimasi jaringan syaraf tiruan dengan algoritma genetika untuk peramalan curah hujan," 2006.

http://repository.ipb.ac.id/handle/123456789/12819

[7] F. D. Wihartiko, A. Buono, and B. P. Silalahi, "Integer programming model for optimizing bus timetable using genetic algorithm," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 166, no. 1.

DOI: 10.1088/1757-899X/166/1/012016.

https://iopscience.iop.org/article/10.1088/1757-899X/166/1/012016

[8] R. Sulistiyorini and W. F. Mahmudy, "Penerapan algoritma genetika untuk permasalahan optimasi distribusi barang dua tahap," *DORO Repos. J. Mhs. PTIIK Univ. Brawijaya*, vol. 5, no. 12, 2015. http://repository.ub.ac.id/id/eprint/146137

[9] Arniantya, Raissa, Budi Darma Setiawan, dan Putra Pandu Adikara. 2017. "Optimasi Vektor Bobot Pada Learning Vector Quantization Menggunakan Algoritme Genetika Untuk Identifikasi Jenis Attention Deficit Hyperactivity Disorder Pada Anak". *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2 (2):679-87. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/909.

[10] A. A. Dwi K, "Implementasi Algoritma Genetika untuk Optimasi LVQ pada Penentuan Kelayakan Kredit (Studi Kasus: Bank X)," Universitas Brawijaya, 2018.

[11] M. Hendry, R. M. Dayawati, and A. T. Wibowo, "Analisis dan Implementasi Optimasi Jaringan Saraf Tiruan dengan Menggunakan Algoritma Genetika untuk Pendiagnosaan Penyakit Stroke (Studi Kasus: RS. Dr. M. Djamil Padang Sumbar)." Telkom University.

[12] Tumini. 2019. "Prediksi Kelulusan Mahasiswa Dengan Metode Support Vector Machine (SVM) Studi Kasus STMIK Cikarang". *Jurnal Informatika SIMANTIK* 4 (2):50-61.

[13] V. Jakkula, "Tutorial on support vector machine (svm)," *Sch. EECS, Washingt. State Univ.*, vol. 37, no. 2.5, p. 3, 2006.

[14] M. B. Nuraedah and A. A. Kasim, "Quadratic support vector machine for the bomba traditional textile motif classification,"

Indones. J. Electr. Eng. Comput. Sci, vol. 11, no. 3, pp. 1004–1014, 2018. DOI: http://doi.org/10.11591/ijeecs.v11.i3.pp1004-1014

[15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, third edition*. 2016.

[16] B. Kusumoputro, "Pengembangan Sistem Pengenalan Wajah Secara 3 Dimensi Menggunakan Hemisphere Structure of Neural Networks dan Optimasi Struktur Menggunakan Algoritma Genetika," 2004.

[17] S. Puri and S. P. Singh, An Efficient Hindi Text Classification Model Using SVM, no. March 2021. Springer Singapore, 2019.

[18] L. G. Irham, A. Adiwijaya, and U. N. Wisesty, "Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine," J. Media Inform. Budidarma, vol. 3, no. 4, p. 284, 2019, doi: 10.30865/mib.v3i4.1410.DOI: http://dx.doi.org/10.30865/mib.v3i4.1410

[19] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," Neurocomputing, vol. 313, pp. 47–64, 2018. https://doi.org/10.1016/j.neucom.2018.05.002

[20] Sudianto, Sudianto, Asti Dwi Sripamuji, Imada Riski Ramadhanti, Risa Riski Amalia, Julian Saputra, and Bagas Prihatnowo. 2022. "Penerapan Algoritma Support Vector Machine Dan Multi-Layer Perceptron Pada Klasifikasi Topik Berita". *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI* 11 (2):84-91.