# Weather Forcast using Learning Vector Quantization Methods

Siska Andriani[*] , Kotim Subandi

siska.andriani@unpak.ac.id, kotim.subandi@unpak.ac.id

Universitas Pakuan, Jalan Raya Ciheuleut Pobox 425

**Abstract.** Weather forecasting is one of the important factors in daily life, as it can affect the activities carried out by the community. The study was conducted to optimize weather forecasts using artificial neural network methods. The artificial neural network used is a learning vector quantization (LVQ) method, in which artificial neural networks based on previous research are suitable for prediction. The research is modeling weather forecast optimization using the LVQ method. Models with the best accuracy can be used in terms of weather forecasts. Based on the results of the training that has been done in this study produces the best accuracy on the LVQ method which is to produce 72%.

**Keywords:** weather forcast, learning vector quantization.

## 1 Introduction

Forecast is an activity that predicts an event in the future. Weather is a state of the atmosphere which is expressed by the value of various parameters, including temperature, pressure, wind, humidity and various phenomena of rain, in a place or region during a short period of time, minutes, hours, days, months, seasons, years. Weather forecasting is an activity that predicts the weather in the future, weather forecasts can predict the potential for rain that will occur in the future [1].

By applying artificial technology, one of the artificial methods will be used to perform weather forecasts, namely the Artificial Neural Network (ANN). Artificial Neural Network is a suitable method in terms of weather forecasting based on existing weather parameters. The ANN method used is Learning Vector Quantization (LVQ). The prediction classes to be made are cloudy, rainy, cloudy rain, sunny and cloudy.

Basically the above method has been widely used in terms of prediction one example is prediction optimization used for rainfall forecasting optimization [2], in the study used JST reverse propagation and genetic algorithms. Other research that has previously been done is optimization of the determination of the quality of cow's milk [3], in the study used lvq method that vector weight of lvq is optimized using genetic algorithms. Results obtained in previous studies show good results for Artificial Neural Network (ANN) use.

To make predictions, it requires a classification of the elements required in terms of prediction of the field. In predicting the weather, at least one year of data is needed, the data is taken every day with a set hour, the data is obtained from the Meteorology and Climatology Agency (BMKG) where this research was conducted. The data used to generate training data in this study were BMKG daily weather data for 3 years. ANN in modeling relies on previous data which is used as learning data which can then be used for future predictions because of learning from past data.

Procedia of Social Sciences and Humanities

Proceding of the International Conference on Intellectuals' Global Responsibility 2020 (ICIGR):
Science for Handling the Effects of Covid-19, Facing the New Normal, and Improving Public Welfare

## 2  Methods

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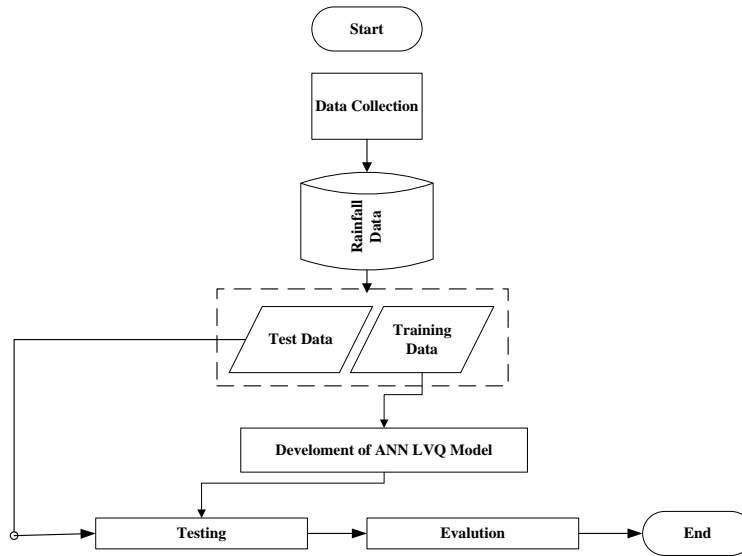


**Fig.1.**  Flowchart Of Research Stage

### 2.1 Data Collection

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

### 2.2 Data Sharing

Data sharing is done to divide weather data into training data and test data. The training data will be used to generate the model and the test data will be used for evaluating the performance of the predictive optimization model made. Data sharing uses a percentage split with the provisions of 70% for training data and 30% for test data.

### 2.3 Weather Forecast Modeling using LVQ

An Artificial Neural Network is a computer system that mimics the thinking of the human brain that is capable of completing a number of calculation processes. There are several types of neural networks but almost all of them have the same components. Neural network consists of several neurons like the human brain, and these neurons are related to one another or known as weights. LVQ is a method for training supervised competitive layers. The competitive layer will learn automatically to classify a given input vector. **Figure 2** shows the LVQ network in the research carried out, namely with 6 units in the input layer, and 2 units (neurons) in the output layer [4]. Learning Vector Quantization (LVQ) is ANN with higher accuracy and faster computation time [5].
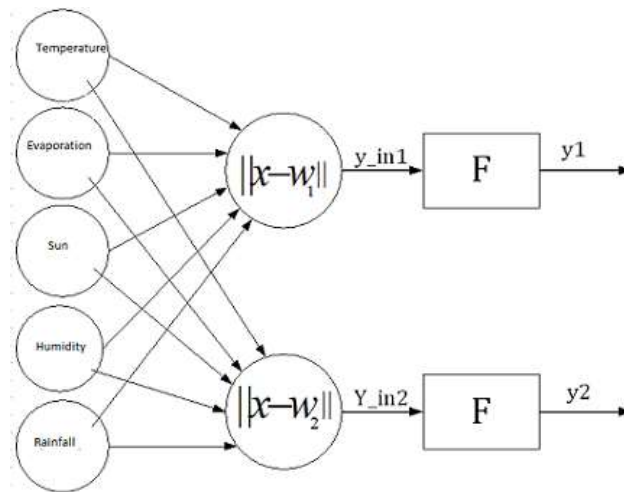
Procedia of Social Sciences and Humanities

Proceding of the International Conference on Intellectuals' Global Responsibility 2020 (ICIGR):
Science for Handling the Effects of Covid-19, Facing the New Normal, and Improving Public Welfare

**Fig.2.** Network Architecture LVQ

The Learning Vector Quantization (LVQ) method is divided into two stages, namely training and testing. The following are the steps for the algorithm:

**a. Training Algorithm**
- Set the initial weight of the j-input variable towards the i-class (cluster): Wij, with i = 1,2,…, K; and j = 1,2,…, m.
- Set Maximum epoh: MaxEpoh
- Set the learning rate Parameter: α
- Set learning rate reduction: Decα
- Set the allowed learning rate: Minα
- Enter input data:; with i = 1,2,…, n; and j = 1,2,…, m
- Enter targets in the form of classes:; with k = 1,2,…, n
- Set initial condition: epoh = 0
- Work if: (epoh ≤ MaxEpoh) and (α ≥ Minα)

  a. epoh = epoh + 1

  b. Work for i = 1 through n

  1. Define J such that | - | minimum; where j = 1,2,…, K

    2. Fix it with the following conditions:

  If T = then, = + α (-)

  If T ≠ then, = - α (-)

    3. Reduce the α value (α reduction can be done by: α = α - Decα; or by: α = α - α * Decα)
- After training, the final weight (W) will be obtained, this weight will be used for the simulation or testing process [5].

**b. Simulation (testing)Algorithm**
- Enter the data to be tested, for example:; with i = 1,2, .., np; and j = 1,2,…, m.
- Work for i = 1 to np

    1.  Determine such J to Q% - Q% minimum; with j = 1,2,…, K

    2.  J is class for $x_i$

The structure of the LVQ ANN used in this study is shown in table 1.

# Procedia of Social Sciences and Humanities

Proceding of the International Conference on Intellectuals' Global Responsibility 2020 (ICIGR):
Science for Handling the Effects of Covid-19, Facing the New Normal, and Improving Public Welfare

**Table 1.** LVQ structured used

| Characteristics | Specification |
|---|---|
| Architecture | *2 Layer* |
| *Neuron Input* | *5* |
| *Neuron Output* | 1 (wheather prediction) |
| *Epoch* | *1000* |
| *Learning rate* | *0.01 0.1 0.2* |
| *Lr_declay* | *0.2* |
| Error Tolerance | *0.001* |
| Artificial Neural Network Algirithm | *Learning Vector Quantization* |

**2.4 Testing, Analysis and Evaluation**

The test is done by comparing the actual value with the predicted value. Furthermore, the evaluation of the prediction results is carried out with a confusion matrix. Confusion matrix is a method for evaluating the success rate of a data when it is called against its data class [6].

## 3 Result and Discusion

The data attributes consist of numeric and categoric. The numerical attributes as input parameters are: temperature, evaporation, sunlight, humidity and rainvol. While the categorical attributes, namely as the output of the weather forecast, include: Cloudly (C), Partly Cloudly (PC), Sunny (S), Rain (R) and Cloudly rain (CR). The sample data used is 1096 data, the data for each class is shown in **Figure 3.**
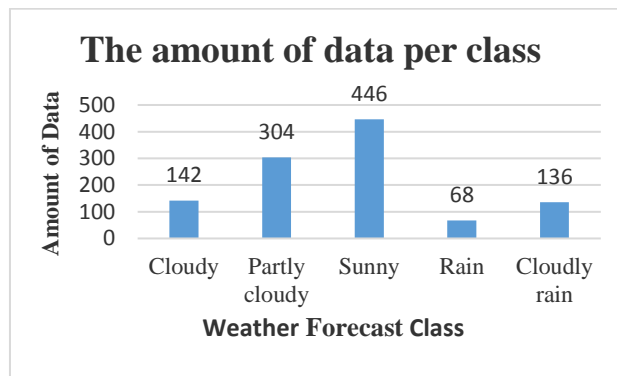


**Fig. 3.** The Number of sample for each class

Weather forecasting using the LVQ method is carried out by generating weather training data with the ANN structure shown in Table 1. After that, tests are carried out on the model that has been built so that the weather forecast results are obtained. Comparison of weather prediction results using the LVQ method with actual data can be seen in **Figure 4**.
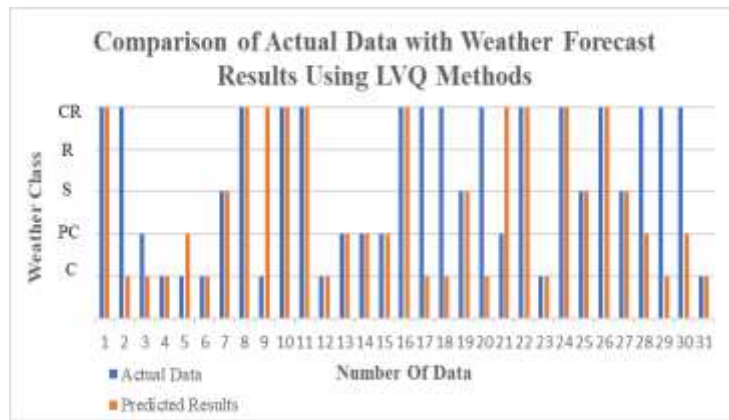
Procedia of Social Sciences and Humanities

Proceding of the International Conference on Intellectuals' Global Responsibility 2020 (ICIGR):
Science for Handling the Effects of Covid-19, Facing the New Normal, and Improving Public Welfare

**Fig. 4.** Comparison of the predicted LVQ results with actual data

Figure 4 shows the comparison of the actual data with the predicted data. From the sample test data taken as many as 31 data, it can be seen that there are some results that are not in accordance with the actual data, namely the test data 2, 3, 5, 9, 17, 18, 20, 21, 28, 29 and 30.

The results of training conducted using LVQ get an accuracy score of 72%. The process of finding accuracy, pecision and recall values is carried out using the confucius matrix method. The results of the confucion matrix calculation are shown in **Figure 5**.

|  | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.59 | 0.91 | 0.72 | 43 |
| 1 | 0.82 | 0.84 | 0.83 | 134 |
| 2 | 0.67 | 0.64 | 0.66 | 91 |
| 3 | 0.75 | 0.15 | 0.25 | 20 |
| 4 | 0.65 | 0.59 | 0.62 | 41 |
| accuracy |  |  | 0.72 | 329 |
| macro avg | 0.70 | 0.62 | 0.61 | 329 |
| weighted avg | 0.73 | 0.72 | 0.70 | 329 |

**Fig. 5.** The results of the confucion matrix calculation

The results showed that the daily weather data of BMKG citeko cisarua bogor region has a mismatch if modeled using lvq methods. This is because BMKG Citeko weather data has unbalanced data for each class. so that to get better results must be done balancing the data of each predictive class.

## 4 Conclusion

This research has succeeded in making a weather forecast model using the LVQ methods. The weather parameters used are temperature, evaporation, sunlight, humidity and rainvol. While the weather forecast class consists of 5 classes, namely Cloudly (C), Partly Cloudly (PC), Sunny (S), Rain (R) and Cloudly rain (CR). The weather data obtained previously was modeled

using the LVQ methods. After testing with 31 samples of test data, the accuracy value obtained is 72% for the LVQ methods.

## Acknowledgements

## References

[1] Gibbs. 1987. Pengertian Cuaca.
[2] A. Kustiyo, A. Buono, N. Apriyanti. 2006. Optimasi Jaringan Syaraf Tiruan dengan Algoritma Genetika untuk Peramalan Curah Hujan. Jurnaillmiah IImu Komputer. Edisi 6 Nol.4 No.1; 1-9.
[3] K. Widyawati, B.D. Setiawan, P.P Adikara. T. 2018. Optimasi Vektor Bobot Learning Vector Quantization Menggunakan Algoritme Genetika untuk Penentuan Kualitas Susu Sapi. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol. 2, No. 1, Januari 2018, hlm. 217-225.
[4] S. Kusumadewi. 2004. Membangun Jaringan Syaraf Tiruan menggunakan *Matlab* dan *Excell Link.* Penerbit Graha Ilmu.
[5] D. M. Wuryandari, I. Afrianto. 2012. Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation Dan Learning Vector Quantization Pada Pengenalan Wajah.
[6] J. Han, M. Kamber, J. Pei. 2012. *Data Mining: Concept and Techniques.* Ed ke-3. Amsterdam (NL): Elsevier, Morgan Kaufmann.