

Climate Prediction Using RNN LSTM to Estimate Agricultural Products Based on Koppen Classification

Novia Andini^{1*}, Wiranto Herry Utomo²

Information Technology, Faculty of Computing, President University

Email: ¹novia.andini@student.president.ac.id, ²wiranto.herry@president.ac.id

Abstract – The yield of an agricultural process is very important and influential, where the harvest is used as a support for human life both as food and a source of income. Many factors can influence the success of agriculture, such as human resources, seed quality and climate that is going on around in the surrounding area. One of the important factors is which climate, the accuracy of determining the climate for agriculture will affect the results obtained. The wrong prediction in determining the future climate will cause crop failure due to incompatibility with the type of plant. In this era, many technologies have been able to predict climate, one of which is technology machine learning that has many types and techniques, which machine learning technology has been widely used in predicting many things. This study aims to predict the climate in an area which is intended to determine crop yields based on the Koppen classification, and also the prediction based on several parameters such as temperature, humidity, duration of sun exposure and rainfall. And the results of this study is have a loss of 0.006 and with the MAPE value as an indicator of the percentage error and as an indicator for determining the accuracy of the prediction results, which is 3.29%, which means that it is included in the very accurate category in predicting climate to estimate agricultural yields.

Keywords –Climate Prediction, Long Short Term Memory (LSTM), Recurrent Neural Network (RNN), Koppen Classification

I. INTRODUCTION

In Foodstuff is one of the important things for human life that can be produced from the agricultural process [1]. Based on it, the results of agriculture are one of the fields that are very close and can support the advancement of economic progress, the better the results obtained, the better the level of the economy [2]. To produce good quality crops, there are many aspects that need to be considered, such as from the climate aspect [3]. The climate itself is the weather conditions in an area that exist at a certain time and is influenced by several factors such as temperature and humidity [4].

Judging from these aspects, climate can be categorized into several types such as rainy season and dry season, which every season will be affected from every aspect such as temperature, sunlight or humidity [5]. Other than that, changing climates can be influenced by changes in rainfall and also ambient temperature that occurs over a certain period of time [6].

The process of climate forecasting is a difficult thing to do in view of the uncertain climate change. This makes the yields that will be obtained at a certain time period difficult to know. Incorrect climate forecasts in an area can cause considerable losses in agriculture, such as decreasing crop quality, crop failure, decreasing the number of crops and damaged crops [7].

In this era, there have been many technologies capable of predicting the climate in agriculture, one of which is machine learning technology [8]. Machine learning technology is one of the machine lessons that is intended so that a machine can have the ability to predict and analyze and recognize a pattern [9]. Machine learning itself has many methods such as K-means, Recurrent Neural

Networks, Decision Trees, Artificial Neural Networks, etc [10].

Several previous studies have examined the machine learning technology used to predict climate. Machine learning was used to predict maize and soybean yields using the Convolutional Neural Network and Recurrent Neural Network methods. The parameters used are weather, performance, agricultural and soil management [11].

Previous research has also studied machine learning used to predict climate to be able to see crop yields with parameters temperature, rainfall, cloud cover, humidity and the method used, namely the Decision Tree [12]. Several other studies have also made predictions using the Recurrent Neural Network method which is used to predict bad weather [13], the temperature for the surrounding area [14], wind velocity [15], temperature for daily [16] and rain [17].

There have been many previous studies that have examined the use of technology learning to predict climate in determining agricultural yields. However, there has not been any research that examines the use of machine learning using the RNN method with Long Short Term Memory and the application of several parameters such as humidity, temperature, rainfall and solar radiation combined with the use of the Koppen classification in classifying climate types. RNN itself is a learning machine that conducts learning by reviewing previous information. Whereas LSTM is a type of Recurrent Neural Network method that is able to overcome the weaknesses of RNN so that it can improve performance and increase accuracy [18]. On this basis, this method is used, in which climatic conditions are closely related to climatic conditions in the past [19].

The purpose of this research is to predict the climate which is intended to see the yield from agriculture

according to Koppen's classification using the RNN LSTM method using temperature, humidity, rainfall and solar radiation as parameters.

II. RESEARCH METHODOLOGY

The object of this research is the climate in the city of Bandung, West Java, the results of rice farming and the Recurrent Neural Network method for the prediction process. In this study, there are several steps that will be carried out in predicting the climate to determine agricultural yields. The steps taken in predicting climate in this study can be seen in Figure 1.

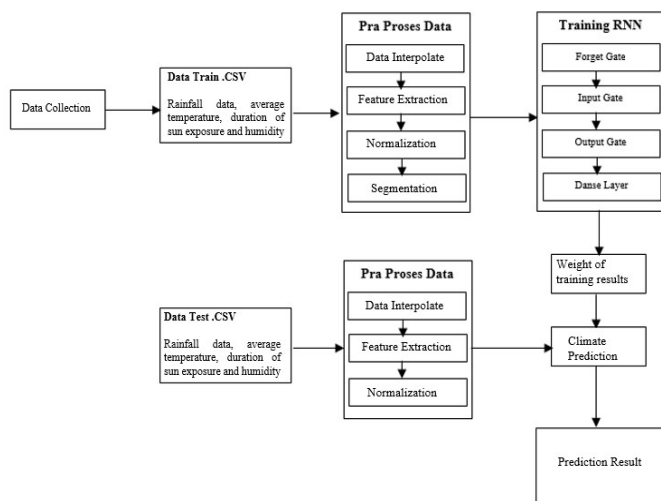


Figure 1. Predicting Climate Step

A. Data Collection

The selection of data in this study is based on the use of Koppen theory which will be used to categorize climate types. The data used in this study are rainfall data, average temperature, duration of sun exposure and humidity for the last 10 years from 2010 to 2020. The data is obtained from the daily climate data of BMKG Bandung City, West Java Province, Bandung Geophysical Station through the official website. BMKG in file format (.csv). This data is daily data, so the amount of data obtained is 3650 x 4 climate parameters = 14600 data

B. Data Training

The training data process will be carried out using the RNN LSTM method which was previously carried out data processing or called data preprocessing, which aims to make the data used for training better and appropriate.

1. Preprocessing Training Data

In this study, the preprocessing was carried out in four stages, namely interpolation, feature extraction, segmentation, and normalization.

- Data Interpolation

Interpolation is the process of finding the value between several known data points. The purpose of interpolation is to correct data that is not measured or not recorded by BMKG by finding the middle value between two values.

- Feature Extraction

Feature extraction is a process to find the largest value in each variable, so that the data will become monthly data with the maximum value of each variable.

- Normalization

Normalization is the process of converting data into normal form. Normalization is needed when the data is very large, very small, or has different units. In this process, the Min-Max normalization is carried out by scaling in the range of zero to one [20].

$$Z = (x - \min()) / (\max() - \min()) \quad (1)$$

X - Data to be normalized,

max - The highest data in the column,

min - The lowest data in the column

- Segmentation

Segmentation is the process of separating and grouping data from raw data into data needed by the system. In this study, data that has become monthly data will be grouped into one year or 12 months with an overlap process, where the one and the next training data has a difference of one month. For example, there are 120 months of climate parameter data, then the distribution of data starts from the 1st month to the 12th month which is used as the first training data, the second month to the 13th month as the second training data, the third month to the 14th month as the third training data, and so on until the 109th training data, namely the 112th month to the 120th month

2. RNN LSTM Training

In the data normalization stage, 109 sets of data were produced to be used as input for the training process using Recurrent Neural Networks. Each data set has 48 total data obtained from 12 months x 4 variables, so that the neurons for input are 36 units. These neurons are connected to neurons contained in the hidden layer, where in the hidden layer there are cells, this process is what distinguishes between ordinary RNN and LSTM. In this cell, there are several steps to be able to produce output.

C. Climate Prediction

At this prediction stage, the MLP architecture will be used in the Recurrent Neural Networks method to predict climate, with the input data used in the input layer, namely rainfall, average temperature, duration of sun exposure and humidity. The MLP architecture designed in the Recurrent Neural Networks method can be seen in Figure 2.



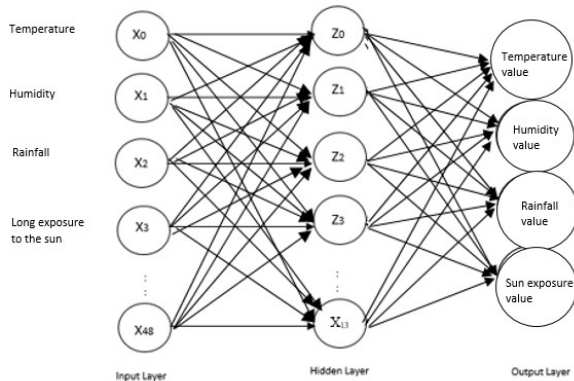


Figure 2. MLP Architecture for Climate Prediction

As can be seen in Figure 2 regarding the MLP architecture used in this study to predict climate, where there are 48 neurons in the input layer obtained from 12 months x 4 climate parameters. There are 13 neurons in the hidden layer which are calculated using equation 2.9 and in the output layer there is one neuron that represents the output of the predictions. The testing process uses the weight of the training results that have been stored in the form of a file, and the testing process will produce an output value in the form of a climate prediction value for each variable. Then the climate value will be entered into the range of values that have been determined using the Koppen Classification

D. Prediction Result

This study will produce a prediction of the climate based on four parameters such as rainfall data, average temperature, duration of sun exposure and humidity for the next one month and then the prediction results will be categorized into climate types based on Koppen's theory. The prediction result can be seen in Figure 3.

```

temperature      : [236.68836449]
kelembapan      : [4169.87289564]
curah hujan     : [4976.57872939]
terik matahari  : [39.21663761]
prediksi koppen iklim =tropis
    
```

Figure 3. MLP Architecture for Climate Prediction

III. RESULTS AND DISCUSSION

The test consisted of the method effect test, namely between the optimization model effect test, and the test for the effect of the number of epochs. At this stage an analysis of each tested parameter is carried out. Before test it,, the training process has been carried out using data that has passed the data preprocess.

Table 2. Data training for first batch

Training Data								
Data - 1				...	Data - 12			
SU	KL	CH	LP	...	SU	KL	CH	LP
24,6	91	86	8	...	24	92	78	6,7
24,9	93	82,4	6,6	...	24,4	85	17,5	7,4

24,6	94	94	7,5	...	25,2	88	20,5	8
26	85	27	8	...	25,2	90	73,5	7,4
25,4	94	92	7,7	...	25,3	86	56	7,9
24,9	94	27,4	7,9	...	24,3	87	44,4	8
24,9	94	61	7,7	...	24,3	84	36,9	8
...

- SU - Temperature
- KL - Humidity
- CH - Rainfall
- LP - Duration of sun exposure

The training process will be carried out using the data in the Table 2 by using the RNN LSTM architecture. In addition, training is also carried out using two optimizers, namely Adam and Sigmod with several epoch including 50, 100, 200 and also used 0,001 for learning rate.

A. Optimization Model Influence Test

The optimization model effect test is conducted to determine which optimization model is more suitable for predicting climate. This optimization model is used to update the weights during training using two optimization models, namely Adam optimizer and SGD optimizer. The accuracy results from testing the two optimization models can be seen in Table 3.

Table 3. Test the Influence of the Optimization Model

No	Optimizer	Loss	MAPE(%)
1	Adam	0,006	32,9
2	SGD	0,011	13,1

After testing, it was found that the Adam optimization method has a lower loss than SGD. Graph of SGD and Adam test results can be seen in Figure 3 and Figure 4.

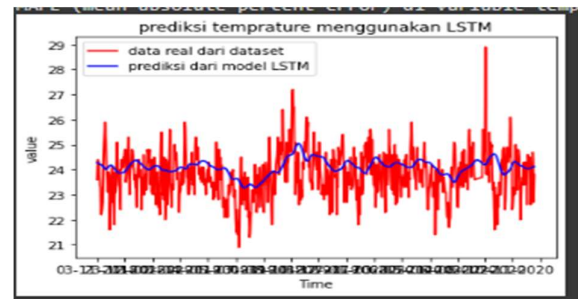


Figure 4. Results of SGD Loss and MAPE

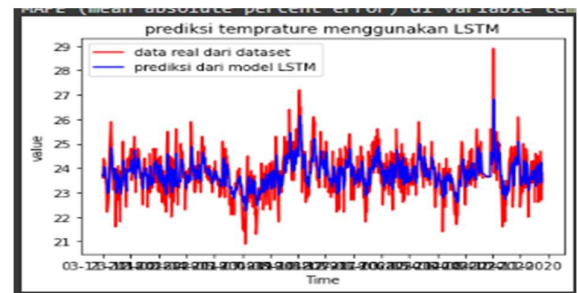


Figure 5. Adam's Loss and MAPE Results

It is found that Adam's optimization has a lower loss than SGD optimization, which is 0.006 and with MAPE



measurement as an indicator of the percentage error of the prediction results, which is 3.29%, which means that it is included in the very accurate category. Meanwhile, SGD itself has a loss of 0.011 and MAPE of 13.1%, which means it is in the good category

B. Test of the Influence of the Number of Epochs

This study used 200 epochs with low loss results. The results of the epoch test can be seen in Table 4.

Table 4. Test for the Effect of the Number of Epochs

No	Epoch	Loss	MAPE(%)
1	50	0,009	7,83
2	100	0,007	3,29
3	200	0,006	2,36

It was found that the use of epoch 50 has the highest loss value with a value of 0.009 and a MAPE of 7.83%. Meanwhile, the epoch which had the lowest loss value was 200 and MAPE 2.36%.

IV. CONCLUSION

This research produces climate prediction using RNN LSTM. This prediction can provide results in the form of climate forecasts for the next day. The results of this prediction are in the form of climate types belonging to tropical climates, dry climates, temperate climates, continental climates and arctic climates. This prediction can be used by the user in predicting the future climate as a benchmark to determine which farm will produce good yields for that climate. Climate prediction in this study produces low loss using Adam's optimization and learning rate of 0.001 and the number of epochs is 200, which results in a loss of 0.006 and with the MAPE value as an indicator of the amount of error percentage and as an indicator of determining the accuracy of the prediction results, which is 3.29%. means it falls into the very accurate category.

REFERENCES

- [1] Abbas, Z., Al-Shishtawy, A., Girdzijauskas, S., & Vlassov, V. (2018). Short-Term Traffic Prediction Using Long Short-Term Memory Neural Networks. *2018 IEEE International Congress on Big Data (BigData Congress)*, 57-65.
- [2] Adedeji, O., Reuben, O., & Olatoye, O. (2014). Global Climate Change. *Journal of Geoscience and Environment Protection*(2), 114-122.
- [3] Beck, H., Zimmermann, N. E., McVicar, T., & Vergopolan, N. (2018). Present and Future Köppen-Geiger climate Classification Maps at 1-Km Resolution.
- [4] Chenn, D., & Chen, H. W. (2013). Using the Koppen Classification to Quantify Climate Variation and Change: An Example for 1901–2010. *Environmental Development*, 63-79.
- [5] Rippke, U., Ramirez-Villegas, J., Jarvis, A., & Vermeulen, S. J. (2016). Timescales of transformational climate change adaptation in Sub-Saharan African agriculture. *Nature Climate Change*.
- [6] Feng, Q. Y., Vasile, R., Segond, M., Gozolchiani, A., Wang, Y., Abel, M., . . . Dijkstra, H. A. (2016). ClimateLearn: A Machine-Learning Approach for Climate Prediction Using Network Measures. *Geosci Model Dev Discuss*.
- [7] Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Weather Forecasting Using Merged Long Short-term Memory Model. *Bulletin of Electrical Engineering and Informatics*, 7(3), 377-385
- [8] Fente, D. N., & Singh, D. K. (2018). Weather Forecasting Using Artificial Neural Network. *2nd International Conference on Inventive Communication and Computational Technologies (ICICCT)*.
- [9] Huang, Y. (2019). A Prediction Scheme for Daily Maximum and Minimum Temperature Forecasts Using Recurrent Neural Network and Rough set. *IOP Conference Series: Earth and Environmental Science ICAESEE*.
- [10] Jahan, I., Sajal, S. Z., & Nygard, K. E. (2019). Prediction Model Using Recurrent Neural Networks. *IEEE International Conference on Electro Information Technology (EIT)*.
- [11] Jin, J., Li, M., & Jin, L. (2015). Data Normalization to Accelerate Training for Linear Neural Net to Predict Tropical Cyclone Tracks. *Mathematical Problems in Engineering*, 8.
- [12] Kaunang, F. J., Rotikan, R., & Tulung, G. S. (2018). Pemodelan Sistem Prediksi Tanaman Pangan Menggunakan Algoritma Decision Tree. *Cogito Smart Journal*, vi(1).
- [13] Khaki, S., Wang, L., & Archontoulis, S. V. (2019). A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, x.
- [14] Kumar, N., Kaur, G., & Aditi. (2017). Wind Speed Prediction using Neural Network. *International Journal of Advanced Production and Industrial Engineering IJAPIE-SI-IDCM*, 608, 36-41.
- [15] Liu, Y., Wang, Y., Yang, X., & Zhang, L. (2017). Short-Term Travel Time Prediction by Deep Learning: A Comparison of Different LSTM-DNN Models. *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*.
- [16] Mahalingam, U., Elangovan, K., Dobhal, H., Valliappa, C., Shrestha, S., & Kedam, G. (2019). A Machine Learning Model for Air Quality Prediction for Smart Cities. *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*.
- [17] Mubyarto. (1989). *Ekonomi pertanian*. Jakarta: LP3ES.
- [18] Naveen Kumar Arora. (2019). Impact of Climate Change on Agriculture Production and Its Sustainable Solutions. *Environmental Sustainability*(2), 95-96.
- [19] Nilsson, & J, N. (2015). *Introduction to Machine Learning*. Stanford University.
- [20] Nilsson, N. J. (2015). *Introduction to Machine Learning*.

