Published online on the page: https://journal.orange-institute.org/index.php/cister/index

International Journal CISTER

| ISSN (Online) 2963-2994 |

Orange Institute

Rainfall prediction system with predictors of temperature humidity, wind speed and air temperature using r programming language

Anita Sari [a], *Susanto* [b,*], *Harma Oktafia Lingga Wijaya* [c],

[a,b,c] *Bina Insan University, Lubuklinggau, Indonesia*

ARTICLE INFORMATION

Article History:
Received: 14 May 2023
Final Revision: 18 May 2023
Published *Online*: 18 May 2023

KEYWORDS

Predictions
Rainfall
Multiple Linear Regression

CORRESPONDENCE

E-mail: susanto@univbinainsan.ac.id*

A B S T R A C T

Rainfall is a climatic factor that can affect human survival, especially in the agricultural sector. Lubuklinggau City is one of the agricultural areas in the province of South Sumatra. Information about rainfall is needed by farmers in Lubuklinggau City. In determining planting time, generally farmers in the city of Lubuklinggau are only based on observing conditions, without looking at information about climate or rainfall. This is caused by the lack of information about rainfall provided by the government. A prediction system is a system that can process or estimate systematically about something that is most likely to happen in the future based on past and present information that is owned, so that the error (the difference between something that happened and the estimated result) can be minimized. Generally in a prediction system there is a calculation method used. One method that is often used in prediction systems is multiple linear regression. Multiple linear regression analysis was carried out to determine the direction and how much influence the independent variables have on the dependent variable. From the research results, the test values obtained were MSE (Mean Squere Error) = 0.8206, RMSE (Root Mean Squere Error) = 0.9059 and MAPE (Mean Absolute Percentage Error) = 39.886. Based on the calculation results mape value is equal to 39.886.

I. Introduction

Rainfall is a climatic factor that can affect human survival, especially in the agricultural sector. Lubuklinggau City is one of the agricultural areas in the province of South Sumatra. Information about rainfall is needed by farmers in Lubuklinggau City. In determining planting time, generally farmers in the city of Lubuklinggau are only based on observing conditions, without looking at information about climate or rainfall. This is caused by the lack of information about rainfall provided by the government.

A prediction system is a system that can process or estimate systematically about something that is most likely to happen in the future based on past and present information that is owned, so that the error (the difference between something that happened and the forecast result) can be minimized.

Generally in a prediction system there is a calculation method used. One method that is often used in prediction systems is multiple linear regression. Multiple linear regression analysis was carried out to determine the direction and how much influence the independent variables have on the dependent variable.

The purpose of this study is to help farmers and the government of Lubuklinggau City accurately predict rainfall which will be used to increase agricultural yields.

II. Method

Multiple Linear Regression is a linear regression model involving more than one independent variable or predictor. In this study, the data used were data on rainfall, temperature, temperature and humidity. The temperature used was data from January 1, 2022 to September 30, 2022. The workflow of the multiple linear regression method used is as follows:

https://doi.org/10.56481/cister.v2i01.157

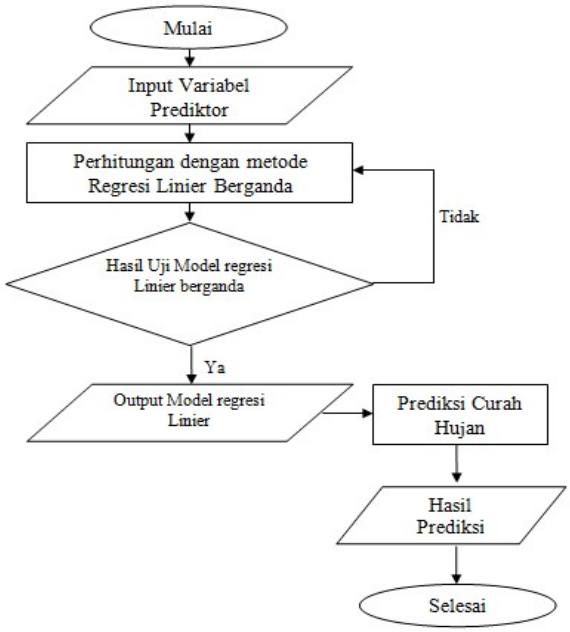

Figure 1. The workflow of the multiple linear regression method.

The test carried out in predicting the amount of rainfall using the multiple linear regression method is testing the accuracy of the prediction results. The accuracy of a prediction is determined by how big this deviation or error occurs between the predicted data and the actual data or actual data. Errors in formulating a prediction are not only caused by an element of error but also by the inability of a forecasting model to recognize other elements in the data series that affect the magnitude of the deviation in the prediction.

There are various ways to calculate the magnitude of the error, some of which are the mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE).

Linear Regression is a statistical method that makes predictions using the development of mathematical relationships between variables, namely the dependent variable (Y) and the independent variable (X). The steps to be taken can be seen below:

1. Preparation / processing of research data sets
2. Formation of a Linear Regression model (model created based on training data)
3. The next step is to find the determinant value of each matrix
4. The next step is to find the values b1,b2,b3 and b4
5. Determine the equation or multiple linear regression model. After getting the values b1.b2,b3 and b4, we can determine the multiple linear regression model, namely Y = -8042.984685 + 503.3675399 X1 + 0.104996463 X2 + 127.3281779 X3
6. MAE, MSE and MAPE testing
   From the existing calculations obtained:
   a. MSE (Mean Squere Error) = 0,8206
   b. RMSE (Root Mean Squere Error) = 0,9059
   c. MAPE (Mean Absolute Percentase Error) = 39,886

III. Results and Discussion

The library declaration in the R programming language functions so that we can use the functions of the library. The display of the R library declaration can be seen in the following figure:

```
library(readxl)

## Warning: package 'readxl' was built under R version 4.0.5

library(GGally)

## Warning: package 'GGally' was built under R version 4.0.5

## Loading required package: ggplot2

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

library(dplyr)
```

Figure 2. Display of the r library declaration

The read_excel command in the R programming language is used to access data sets in Microsoft Excel format. The read_excel command can be seen in the following image:

curah_hujan <dbl>	temp <dbl>	suhu <dbl>	kec_angin <dbl>
1.300000	26.7	84	3
1.200000	26.0	86	3
1.600000	27.2	82	3
1.800000	26.3	84	3
1.100000	27.6	83	3
1.500000	27.6	84	2
1.600000	27.0	81	3
2.330000	27.0	85	3
2.285714	26.2	88	3
1.200000	26.1	86	4

1-10 of 273 rows Previous 1 2 3 4 5 6 .. 28 Next

Figure 3. Display of the read excel command

The rmarkdown data command in the R programming language is used to turn a data analysis into high-quality documents, reports, presentations and dashboards. In this case the data to be displayed is population data, the rmarkdown command for population data can be seen in the following figure:

```
rmarkdown::paged_table(Data_curah_hujan)
```

Figure 4. Display of the rmarkdown command

https://doi.org/10.56481/cister.v2i01.157

E-ISSN : 2963-2994

The dependent(x) and independent(y) variable declaration commands in the R programming language can be seen in the following figure:

```
x1=(Data_curah_hujan$temp)
x2=(Data_curah_hujan$suhu)
x3=(Data_curah_hujan$kec_angin)
y=(Data_curah_hujan$curah_hujan)
```

Figure 5. Display of the x variable declaration command

The command to display the data length of dependent and independent variables in the R programming language can be seen in the following figure:

```
length(x1)
## [1] 273
length(x2)
## [1] 273
length(x3)
## [1] 273
length(y)
## [1] 273
```

Figure 6. Display command displays the length of the data variable x

The simple linear regression equation declaration command in the R programming language can be seen in the following figure:

```
Persamaan_Regresi <-(lm(curah_hujan~temp+suhu+kec_angin,data=Data_curah_hujan))
Persamaan_Regresi
```

```
##
## Call:
## lm(formula = curah_hujan ~ temp + suhu + kec_angin, data = Data_curah_hujan)
##
## Coefficients:
## (Intercept)         temp         suhu    kec_angin
##    4.588695    -0.091179    -0.004794    -0.039215
```

Figure 7. Orders and results of multiple linear regression equations

The command to predict the results of multiple linear regression equations for rainfall data can be seen in the following figure:

```
data <- data.frame(actual=c(Data_curah_hujan),
                   forecast=c(Data_prediksi))
data
```

Figure 8. Display of the rainfall prediction command

```
##     Prediksi_curah_hujan
## 1               1.633869
## 2               1.688106
## 3               1.597868
## 4               1.670340
## 5               1.556602
## 6               1.591023
## 7               1.620897
## 8               1.601721
## 9               1.660282
## 10              1.639773
## 11              1.642987
## 12              1.740932
## 13              1.768756
## 14              1.754204
## 15              1.663966
## 16              1.604464
## 17              1.725439
## 18              1.769696
## 19              1.711606
## 20              1.710026
## 259             1.683142
## 260             1.503306
## 261             1.636612
## 262             1.583316
## 263             1.588580
## 264             1.544571
## 265             1.530189
## 266             1.560756
## 267             1.619317
## 268             1.581905
## 269             1.630238
## 270             1.693593
## 271             1.623171
## 272             1.694703
## 273             1.690379
```

Figure 9. Display of the command prediction results

Data plots are used to describe the distribution of the research data used. The Plot Data command in the R programming language can be seen in the following figure:

```
plot(Data_curah_hujan)
```

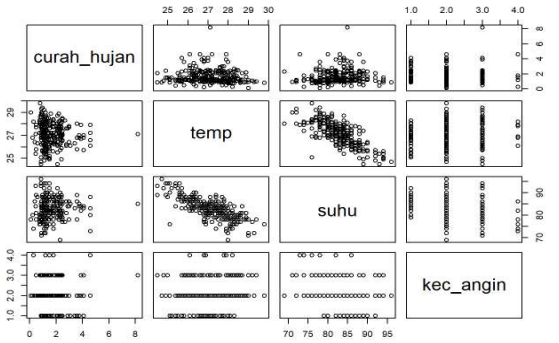

Figure 10. View of the rainfall data plot command

The ggcorr data command is used to describe correlations between data sets. The ggcorr Data command in the R programming language can be seen in the following figure:

```
ggcorr(Data_curah_hujan,label = T, size=3, label_size = 3, hjust=0.95)+
  labs(
    title="Matriks Korelasi Data set"
  )+
  theme_minimal()+
  theme(
    plot.title = element_text(hjust = 0.5),
    axis.title=element_text(size=8,face="bold"),
    axis.text.y=element_blank()
  )
```

Figure 11. The ggcorr command

E-ISSN : 2963-2994

https://doi.org/10.56481/cister.v2i01.157

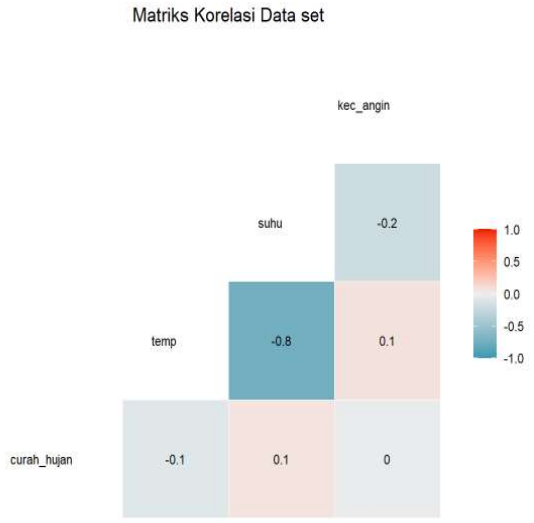

Figure 12. The results of the correlation matrix.

IV. Conclusion

Based on the existing discussion, it can be concluded that the research is:

a. From predictions using the multiple linear regression method, the equation model is Y = -8042.984685 + 503.3675399 X1 + 0.104996463 X2 + 127.3281779 X3, this equation will be used to predict future rainfall.
b. From the measurement of the linear regression equation model, the value of MSE (Mean Squere Error) = 0.8206, RMSE (Root Mean Squere Error) = 0.9059 and MAPE (Mean Absolute Percentage Error) = 39.886 %

V. Suggestion

The development of this prediction system should be added to other predictor variables and also use other methods related to rainfall so as to get more accurate results.

References

[1] T. P. Purboningtyas, A. H. Dharmawan, and E. I. K. Putri, "The Impact of Climate Variability on The Livelihood Structure of Farmers Households and Patterns of Adaptation," Sodality: Jurnal Sosiologi Pedesaan, vol. 6, no. 3. 2019, doi: 10.22500/sodality.v6i3.21514.
[2] A. Izzah and R. Widyastuti, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression untuk Pencegahan Data Outlier," Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control. pp. 141–150, 2017, doi: 10.22219/kinetik.v2i3.268.
[3] F. J. Kaunang, R. Rotikan, and G. S. Tulung, "Pemodelan Sistem Prediksi Tanaman Pangan Menggunakan Algoritma Decision Tree," CogITo Smart Journal, vol. 4, no. 1. p. 213, 2018, doi: 10.31154/cogito.v4i1.115.213-218.
[4] S. C. Wohon, D. Hatidja, and N. Nainggolan, "Penentuan Model Regresi Terbaik Dengan Menggunakan Metode Stepwise (Studi Kasus : Impor Beras Di Sulawesi Utara) Determining the Best Regression Model Using Stepwise Method (Case Study : Rice Imports in North Sulawesi)," Jurnal Ilmiah Sains, vol. 17, no. 2. p. 81, 2017.
[5] S. Sulistyono and W. Sulistiyowati, "Peramalan Produksi dengan Metode Regresi Linier Berganda," PROZIMA (Productivity, Optimization and Manufacturing System Engineering), vol. 1, no. 2. pp. 82–89, 2017, doi: 10.21070/prozima.v1i2.1350.
[6] A. Setiawan and Z. Zarenuddin, "Sistem Peminjaman Barang Inventaris Media Perkuliahan Berbasis Qrcode Sebagai Pendamping Sistem Konvensional Untuk Pengelolaan Barang," Technologia: Jurnal Ilmiah, vol. 10, no. 3. p. 134, 2019, doi: 10.31602/tji.v10i3.2119.
[7] Nurhendi, "Sistem Informasi Broker Asuransi Pada PT. Visi Bersama Serantau," Jurnal Inovasi Penelitian, vol. 1, no. 10. pp. 1–208, 2021.
[8] Zakky, "Pengertian Sistem | Definisi, Unsur-Unsur, Jenis-Jenis, dan Elemennya," Zona Referensi, 2021. https://www.zonareferensi.com/pengertian-sistem/.
[9] T. Sutabari, Konsep Sistem Informasi. Andi, 2012.
[10] Adhi, "Analisis Perancang Sistem," Bab I, no. 2010. pp. 1–16, 2020.
[11] M. Kafil, "Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 3, no. 2. pp. 59–66, 2019, doi: 10.36040/jati.v3i2.860.
[12] A. Ardiana and A. EP Yunus, "SISTEM PREDIKSI PENENTUAN JENIS TANAMAN SAYURAN BERDASARKAN KONDISI MUSIM DENGAN PENDEKATAN METODE TREND MOMENT." 2017.
[13] R. A. R. JULEHA, "Analisa Intensitas Hujan," no. 1. 2008.
[14] A. Kurniawan, "Evaluasi Pengukuran Curah Hujan Antara Hasil Pengukuran Permukaan (AWS, HELLMAN, OBS) dan Hasil Estimasi (Citra Satelit =GSMaP) Di Stasiun Klimatologi Mlati Tahun 2018," J. Geogr. Edukasi dan Lingkung., vol. 4, no. 1, pp. 1–7, 2020.
[15] F. A. Mushi, "Curah Hujan: Pengertian, Klasifikasi, Pengukuran, dan Alat Ukur," Forester Act, 2021. https://foresteract.com/curah-hujan/.
[16] Tukidi, "Karakter Curah Hujan Di Indonesia," Jurnal Geografi, vol. 7, no. 2. 2010, [Online]. Available: http://journal.unnes.ac.id/nju/index.php/JG/article/view/84.

https://doi.org/10.56481/cister.v2i01.157

E-ISSN : 2963-2994

[17] T. N. Padilah and R. I. Adam, "Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang," FIBONACCI: Jurnal Pendidikan Matematika dan Matematika, vol. 5, no. 2. p. 117, 2019, doi: 10.24853/fbc.5.2.117-128.

[18] Syilfi, D. Ispriyanti, and D. Safitri, "Analisis Regresi Linier Piecewise Dua Segmen," Jurnal Gaussian, vol. 1, no. 1. pp. 219–228, 2012.

[19] I. M. Yuliara, "Regresi linier berganda 1.," Journal Article. pp. 1–6, 2016, [Online]. Available: http://www.mendeley.com/research/regresi-linier-berganda-1/.

[20] F. SUSANTI, "Pengertian, Sejarah, Fitur, Kelebihan dan Kekurangan Bahasa Pemrograman R." 2021, [Online]. Available: https://rstudio-pubs-static.s3.amazonaws.com/821986_d3a5240c1c9a4f459e47bf1043e80707.html.

[21] Andry brew, "Data Analytics Menggunakan Orange," Menulis Yang Terpikir. 2015, [Online]. Available: https://andrybrew.blog/2015/06/14/data-analytics-menggunakan-orange/.

[22] E. Dewi et al., "Prediksi Curah Hujan Di Kabupaten Majalengka Dengan Menggunakan Algoritma Regresi Rainfall Prediction in Majalengka District Using Regression Algorithm," Informasi □, vol. 67, no. 1. 2019.

[23] G. I. Merdekawati and Ismail, "Prediksi Curah Hujan Di Jakarta Berbasis Algoritma Levenberg Marquardt," Jurnal Ilmiah Informatika Komputer, vol. 24, no. 2. pp. 116–128, 2019, doi: 10.35760/ik.2019.v24i2.2366.

[24] A. Rachmawati, "Prediksi Curah Hujan di Kota Pontianak Menggunakan Parameter Cuaca sebagai Prediktor Pada Skala Bulanan, Dasarian dan Harian," Positron, vol. 5, no. 2. 2015, doi: 10.26418/positron.v5i2.11755.

https://doi.org/10.56481/cister.v2i01.157
E-ISSN : 2963-2994