Desimal: Jurnal Matematika Vol 3 No 3 (2020) 227-234

Contents lists available at DJM
DESIMAL: JURNAL MATEMATIKA
p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm
http://ejournal.radenintan.ac.id/index.php/desimal/index

The Application of Non-linear Cubic Regression in Rice Yield Predictions

Retno Tri Vulandari*, Hendro Wijayanto, Afan Lathofy

STMIK Sinar Nusantara Surakarta, Indonesia

ARTICLE INFO

Article History
Received : 11-06-2020
Revised : 07-07-2020
Accepted : 06-09-2020
Published : 20-09-2020

Keywords:
Polynomial Non-Linear Regression, Third-Degree, Rice Yields.

*Correspondence: E-mail: retnotv@sinus.ac.id

Doi: 10.24042/djm.v3i3.6580

ABSTRACT

The rice yields have fluctuated in Wonogiri Regency. This occasion happened in 2016-2018. Therefore, a prediction is needed to know whether rice yields will increase or decrease in the following year. The purpose of this study was to apply the polynomial non-linear regression method of third-degree in predicting rice yields. This study utilized the Unified Modeling Language (UML) as the system design, black-box testing as the functional testing, and MSE testing as the validity testing. The computed data was data of 2016-2018. The results showed that the prediction of 2017-2019 using the harvested area model produced more accurate calculations. The harvested area model produced the same MSE value in manual and application calculations, which were 405433,1349 in 2017, 312677,7798 in 2018, and 171183.6347 in 2019. The polynomial non-linear cubic regression is a solution to predict rice yields. The output of the application is the prediction information for rice yields

http://ejournal.radenintan.ac.id/index.php/desimal/index

INTRODUCTION

Wonogiri Regency is one of the food buffer districts in Central Java. Wonogiri Regency produces various kinds of food plants every year. One of the food crops that is produced is rice. Indonesian statistical data for 2018 shows that Wonogiri is the eighth largest rice-producing district of 35 districts/cities in Central Java. Rice productivity in 2018 reached 58.56 kw/ha with a production of 351,576 tons and a harvest area of 60,033.4 ha (BPS, 2019).

Rice yields in three years from 2016-2018 in Wonogiri Regency experienced ups and downs and tended to be unstable. In February 2018, farmers who had been producing rice in Wonogiri, and several regions experienced crop failures which resulted in reduced harvests and increased rice prices (Purnomo, 2019). Therefore, it is necessary to have a prediction to find out the yield of rice in the future whether an increase or decrease as well as a reference to determine policies in controlling rice prices and stock.

This study uses data from the Wonogiri Regency Agriculture Office and the Central Statistics Agency website. The data used include rice yield data (tons), harvested area data (ha), and irrigation data (ha). While the method used is a non-linear cubic regression. The method was chosen because the relationship between the dependent variable and the independent variable is not linear, meaning that the rate of change of the dependent variable due to the rate of change of the dependent variable is not constant for the value of certain independent variables (Prasetyo, 2018).

Studies with the theme of rice yield prediction are not the first thing to do. Many studies use the same topic, but the used method or independent variables are different. As one of the studies conducted by Suryanto & Muqtadir. The study uses a casual linear regression method to predict rice yields and uses MEA as its forecasting error. The study also uses population registration as the independent variable (Suryanto, 2019).

The rice yield prediction study was also conducted by Asih, Setyaningsih, & Midyanti. The authors use harvested area, planted area, productivity, and rainfall as the independent variables. The result of this study is an application that predicts rice yields in each of the three periods using an interval regression method with neural fuzzy (Asih, 2017).

There is also a study by Hermawan and Vulandari. The author compares the prediction of rice yields in the Sukoharjo Regency, which is affected by irrigation and rainfall. Forecasting models used in this study by applying the cubic regression method and applying MSE & MAPE as forecast errors. Based on the calculations, the rice yield prediction will be more accurate if it calculated using the irrigation model (Vulandari & Hermawan, 2018).

Also, the study from Nurudin uses the least square time series as a forecasting method and uses harvest month, rainfall, and pests as the independent variables (Nurudin, 2017).

Based on these thoughts, the authors are interested in researching to predict rice yields that are affected by harvested area or irrigation using a non-linear polynomial regression method with a degree of three in Wonogiri Regency. Calculation of forecasting error is done using Mean Squared Error (MSE). The MSE calculation results of the harvest and irrigation area models are compared, where the smallest MSE value is an accurate forecast model (Suryanto, 2019).

METHOD

The research method is the stages that the researcher goes through in researching so that the results achieved do not deviate from the goal.

Several independent variables affect rice yields, including soil fertility, seed selection, land area, rainfall, harvested area, water availability, and so on. The selected independent variables were the irrigation variable and the harvested variable. The reason why those variables were chosen was that they directly or indirectly affect the rice yields. There are several methods to predict the causal relationship model, including regression, correlation, econometrics, and input-output methods. The selected prediction method is the non-linear cubic regression. Besides, it has a non-linear relation between the independent variable and the dependent variable.

The data source used in this study was the secondary data. The data obtained were the data of rice yields, harvested area, and irrigation in the form of reports from the Agriculture Service Office of Wonogiri Regency and BPS. The data collection technique was the literature study. This technique was used to find the information and knowledge from the literature related to the object being studied, such as the basic concept of forecasting, forecasting error, the application of a non-linear cubic regression, etc.

The polynomial non-linear cubic regression is a form of the polynomial with k=3. The equation for the regression is :

$$\hat{Y} = b_0 + b_1X + b_2X^2 + b_3X^3 \tag{1}$$

Where:

$\hat{Y}$ = the predicted score (forecast)
X = the independent variable
b_0, b_1, b_2, b_3 = the regression coefficient

The equation is converted into a matrix using the formula below:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 & {X_1}^2 & {X_1}^3 \\ 1 & X_2 & {X_2}^2 & {X_2}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & {X_n}^2 & {X_n}^3 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{2}$$

To estimate the coefficient b_0, b_1, b_2, and b_3, use the equation below:

$$\begin{aligned} &Y = X\,b \\ &X^T Y = X^T X\,b \\ &(X^T.X)^{-1}\,X^T Y = (X^T.X)^{-1}\,(X^T.X)\,b = I\,b \\ &b = (X^T.\,X)^{-1}\,X^T\,Y \end{aligned} \tag{3}$$

From equation (3), the value of X^T. X can be given as the following equation:

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \\ {X_1}^2 & {X_2}^2 & \cdots & {X_n}^2 \\ {X_1}^3 & {X_2}^3 & \cdots & {X_n}^3 \end{bmatrix} \begin{bmatrix} 1 & X_1 & {X_1}^2 & {X_1}^3 \\ 1 & X_2 & {X_2}^2 & {X_2}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_n & {X_n}^2 & {X_n}^3 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} n & \sum X & \sum X^2 & \sum X^3 \\ \sum X & \sum X^2 & \sum X^3 & \sum X^4 \\ \sum X^2 & \sum X^3 & \sum X^4 & \sum X^5 \\ \sum X^3 & \sum X^4 & \sum X^5 & \sum X^6 \end{bmatrix} \tag{4}$$

From equation (3), the value of X^T. Y can be given as the following equation:

$$X^T Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \\ {X_1}^2 & {X_2}^2 & \cdots & {X_n}^2 \\ {X_1}^3 & {X_2}^3 & \cdots & {X_n}^3 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} \sum Y \\ \sum XY \\ \sum X^2 Y \\ \sum X^3 Y \end{bmatrix} \tag{5}$$

As indicated in equation (4) and (5), the coefficient b_0, b_1, b_2, and b_3 can express the following equation:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} n & \sum X & \sum X^2 & \sum X^3 \\ \sum X & \sum X^2 & \sum X^3 & \sum X^4 \\ \sum X^2 & \sum X^3 & \sum X^4 & \sum X^5 \\ \sum X^3 & \sum X^4 & \sum X^5 & \sum X^6 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y \\ \sum XY \\ \sum X^2 Y \\ \sum X^3 Y \end{bmatrix} \tag{6}$$

The analysis stage was divided into two, namely the dataset and analysis of application methods. The dataset used in this study was the data on rice yields (tons), harvested area (ha), and irrigation (ha) of each sub-district in Wonogiri District within three years from 2016-2018. The 2016-2017 data are used as the training data while the 2018 data was used as the test data. The data was then calculated manually. The analysis of the application of the method was done by explaining the steps in applying the non-linear polynomial regression method with a degree.

The Design stage was done by designing the system design and user interface design. The system design in this study was carried out using modeling based on scenarios/UML including use case diagrams, sequence diagrams, class diagrams, and activity diagrams. The interface design was used as a reference in making the application. The interface consisted of: login interface, main page interface, initial data interface, coefficient calculation interface, forecasting calculation interface, error interface per district, results in comparison interface, application user interface, profile interface, and instruction interface.

This stage explained the functional and additional requirements of the application to be built, the software and hardware requirements needed to build the application, and the steps for making the application.

The tests were in the form of testing the system's functionality using black-box testing and testing its validity using Mean Squared Error (MSE). The MSE values from the prediction of harvest and irrigation area were compared to the smallest MSE value

RESULTS AND DISCUSSION

The results and discussion chapter consists of five sub-chapters, namely the selection of variables and methods, manual calculation, system implementation, functional testing, and validity testing.

a) Calculation of prediction for 2019 based on 2018 irrigation data

The first step to take is to find the values of X^2, X^3, X^4, X^5, and X^6 each sub-district. Then, calculate the $\sum X$, $\sum X^2$, $\sum X^3$, $\sum X^4$, $\sum X^5$, and $\sum X^6$. Put those values into equation (4).

$$X^T X = \begin{bmatrix} 25 & 22212 & 28316230 & 41036149404 \\ 22212 & 28316230 & 41036149404 & 6{,}349 \text{ x } 10^{13} \\ 28316230 & 41036149404 & 6{,}349 \text{ x } 10^{13} & 1{,}023 \text{ x } 10^{17} \\ 41036149404 & 6{,}349 \text{ x } 10^{13} & 1{,}023 \text{ x } 10^{17} & 1{,}699 \text{ x } 10^{20} \end{bmatrix}$$

Change the matrix of X^TX into an inverse matrix.

$$(X^T X)^{-1} = \begin{bmatrix} 0{,}402831959 & -0{,}001562218 & 1{,}573 \text{ x } 10^{-6} & -4{,}609 \text{ x } 10^{-10} \\ -0{,}001562218 & 8{,}731 \text{ x } 10^{-6} & -1{,}012 \text{ x } 10^{-8} & 3{,}213 \text{ x } 10^{-12} \\ 1{,}573 \text{ x } 10^{-6} & -1{,}012 \text{ x } 10^{-8} & 1{,}274 \text{ x } 10^{-11} & -4{,}270 \text{ x } 10^{-15} \\ -4{,}609 \text{ x } 10^{-10} & 3{,}213 \text{ x } 10^{-12} & -4{,}270 \text{ x } 10^{-15} & 1{,}488 \text{ x } 10^{-18} \end{bmatrix}$$

Find the values of XY, X^2Y, and X^3Y each sub-district, and then calculate the $\sum Y$, $\sum XY$, $\sum X^2Y$, dan $\sum X^3Y$. Put those values into equation (5).

$$X^T Y = \begin{bmatrix} 351576 \\ 400748793 \\ 5{,}602 \text{ x } 10^{11} \\ 8{,}513 \text{ x } 10^{14} \end{bmatrix}$$

Calculate the regression coefficients by multiplying the inverse matrix of X^TX with X^TY as in equation (6).

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 4498{,}238759 \\ 11{,}95097081 \\ -0{,}001143492 \\ 1{,}472 \text{ x } 10^{-7} \end{bmatrix}$$

After calculating the regression coefficients, then calculate the predicted score ($\hat{Y}$) or the 2019 prediction in each sub-district with equation (1) :

$\hat{Y}$ =4498,238759 + 11,95097081*X_i + -0,001143492*X_i^2 + 1,47296E-07*X_i^3

Prediction calculation results are shown in Table 1.

Table 1. The 2019 Prediction Results Based on 2018 Irrigation Data

No	Sub-district	$\hat{Y}$ (tons)
1.	Pracimantoro	12701,92457
2.	Paranggupito	4498,238759
3.	Giritontro	7130,995697
4.	Giriwoyo	13402,59592
5.	Batuwarno	6207,536656
6.	Karangtengah	5305,663582
7.	Tirtomoyo	20120,32841
8.	Nguntoronadi	6670,964965
9.	Baturetno	19363,68247
10.	Eromoko	19325,20026
11.	Wuryantoro	14884,85824
12.	Manyaran	8153,935422
13.	Selogiri	21380,5479
14.	Wonogiri	14034,34551
15.	Ngadirojo	24849,04521
16.	Sidoharjo	21859,61808
17.	Jatiroto	11019,98222
18.	Kismantoro	12974,68597
19.	Purwantoro	17486,6641
20.	Bulukerto	8221,545802
21.	Slogohimo	20823,51491
22.	Jatisrono	18658,67467
23.	Jatipurno	11621,01393
24.	Girimarto	21953,29965
25.	Puh Pelem	8927,137128

b) Calculation of prediction for 2019 based on 2018 harvested area data

The calculation of rice yield prediction is calculated by finding the regression coefficient beforehand using equation (6). The regression coefficient values based on the harvested area are obtained as follows:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} -40{,}56108492 \\ 5{,}761955537 \\ 2{,}542 \text{ x } 10^{-5} \\ 3{,}792 \text{ x } 10^{-9} \end{bmatrix}$$

Prediction calculation results are shown in Table 2.

Use Case Diagram from the application of rice yield prediction illustrates the interaction of the actor(s) with the system (Sukamto, 2015). The diagram is depicted in Figure 1.

Table 2. The 2019 Prediction Results Based on 2018 Harvested Area Data

No	Sub-district	Ŷ (tons)
1.	Pracimantoro	8007,376138
2.	Paranggupito	-40,56108492
3.	Giritontro	2116,430272
4.	Giriwoyo	16572,15705
5.	Batuwarno	5357,064413
6.	Karangtengah	7198,499918
7.	Tirtomoyo	22560,5051
8.	Nguntoronadi	11247,43749
9.	Baturetno	15263,8599
10.	Eromoko	22383,0127
11.	Wuryantoro	12933,79629
12.	Manyaran	15135,47341
13.	Selogiri	24910,15477
14.	Wonogiri	11694,58421
15.	Ngadirojo	23948,3513
16.	Sidoharjo	21134,01892
17.	Jatiroto	8390,426797
18.	Kismantoro	12009,74978
19.	Purwantoro	16796,53744
20.	Bulukerto	12313,9439
21.	Slogohimo	20986,05031
22.	Jatisrono	18145,06807
23.	Jatipurno	13966,18975
24.	Girimarto	22023,485
25.	Puh Pelem	6522,388161

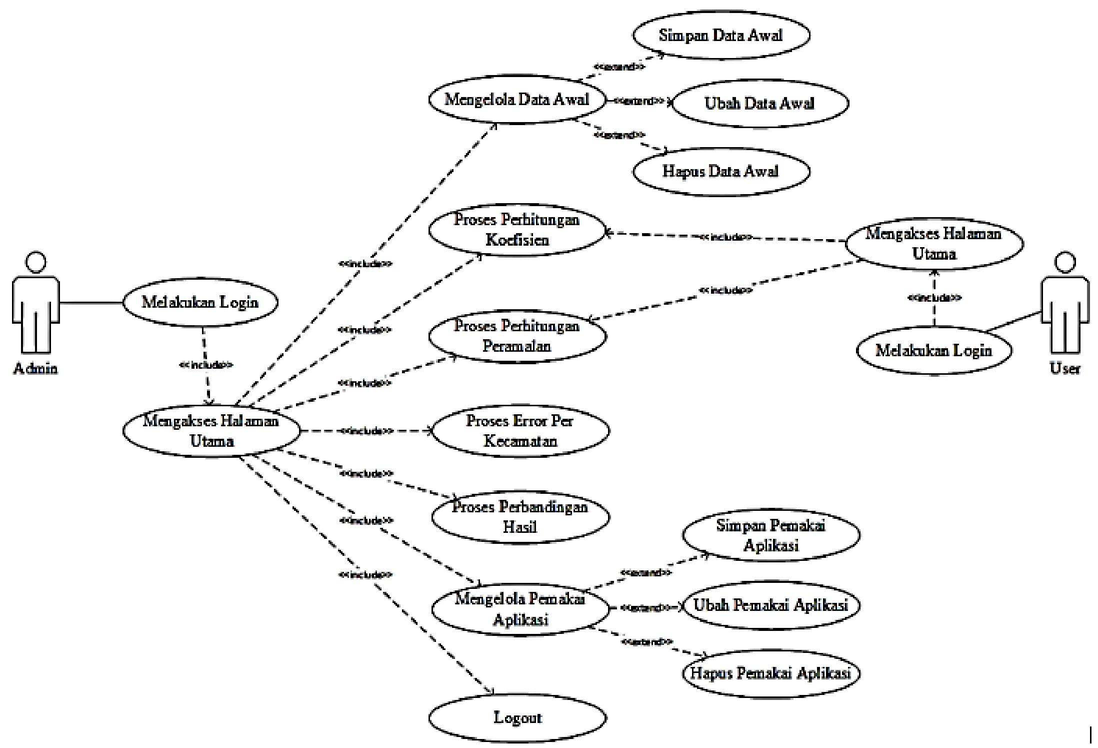

Figure 1. Use Case Diagram of the Application of Rice Yield Prediction

Meanwhile, the description of the actor(s) is explained in Table 3.

Table 3. Actor(s) Description

No	Actors	Description
1.	Admin	The actor who has admin privileges is part of the land and irrigation section. Actors can access all of the menus; which are the initial data menu, calculation of menu coefficients, calculation of menu forecasting, error per menu district, comparison of menu results, user menu applications, profile menu, and instruction menu.
2.	User	Actors who have user privileges are part of the food crops section and the field of food security. Actors can only access two menus; which are the coefficient calculation menu and forecasting calculation menu.

The first thing before calculating MSE is to calculate the error value per data (Y-Ŷ). The MSE value for irrigation model:

$$MSE = \frac{\sum(Y - \hat{Y})^2}{n} = \frac{256440115{,}6}{25} = 10257604{,}62$$

The MSE value for harvested model:

$$MSE = \frac{\sum(Y - \hat{Y})^2}{n} = \frac{4279590{,}867}{25} = 171183{,}6347$$

Based on the MSE calculation above, it can be concluded that the use of the harvested area as a calculation of the predicted rice yields in 2019 will be more accurate with a smaller value of 171183,6347.

The MSE values of predicted rice yields in 2017-2019 based on irrigation and harvested area are shown in Table 4.

Table 4. MSE Calculations

No	Year	MSE Irrigation	MSE Harvested Area
1.	2017	14806093,58	405433,1349
2.	2018	13196779,89	312677,7798
3.	2019	10257604,62	171183,6347

Based on Table 4, it can be concluded that the prediction of rice yields in 2017-2019 will be more accurate by using the harvested area model because the results of the MSE calculation of the harvested area model have a smaller value than the irrigation model. The study by Hermawan and Vulandari on the prediction of rice yields in the Sukoharjo Regency discovered that the rice yields was affected by the irrigation and rainfall. The forecasting model used in this study was the cubic regression method and applying MSE & MAPE as forecast errors. Based on the calculations, the rice yield prediction will be more accurate if it is calculated using the irrigation model.

CONCLUSIONS AND SUGGESTIONS

Based on the discussion from the previous chapter, it can be concluded that the use of the harvest area model as a prediction calculation in 2017-2019 was more accurate because it had a smaller MSE value than the MSE value of the irrigation model. The harvest area model produced the same MSE value in manual and application calculations, specifically 405433.1349 in 2017; 312677,7798 in 2018; and 171183,6347 in 2019. The developed product of this study was the predictive application of rice yield using a non-linear polynomial regression method with a degree of three by comparing the prediction results between the irrigation model and the harvested area.

It is suggested for further studies to use other variables so that the results of forecasting with other calculated variables can produce a smaller error value. It is also expected for further studies to use other methods, compare prediction results and predictive error values between the two methods, and improve the application. The application can be improved by providing additional data upload features so that the users do not have to input the initial data manually, but can import data directly into the application. This research is suggested to be a guide and reference for further research.

REFERENCES

Asih, K. (2017). Aplikasi prediksi produksi padi menggunakan regresi interval dengan neural fuzzy di Kabupaten Kubu Raya. *Jurnal Coding, Sisten Komputer UNTAN*, 108-118.

BPS. (2019). *Rekapitulasi luas panen dan produksi tanaman pangan.* Wonogiri: Badan Pusat Statistik.

Hutabarat, M. (2018). Penerapan algoritma backpropagation dalam memprediksi produksi tanaman padi sawah menurut Kabupaten/Kota di Sumatera Utara. *SemanTIK Vol 4 No 1* , 77-86.

Lestari, S. (2019). *Peramalan stok sparepart menggunakan metod eleast square.* Aceh: CV. Sefa Bumi Persada.

Nurudin, A. (2017). Aplikasi prediksi hasil panen padi dengan metode least square. *Simki-Techsain Vol 1 No 1*, 1-11.

Prasetyo, D. (2018). Analisis perencanaan kapasitas produksi menggunakan rough cut capacity planning untuk produk . *Jurnal Sistem Indonesia*, 77-90.

Purnomo, A. (2019, Oct 15). *Harga beras kabupaten wonogiri.* Retrieved from rr.co.id: https://rri.co.id/surakarta/post/berita/496113

Sukamto, R. A. (2015). *Rekayasa perangkat lunak terstruktur dan berorientasi objek.* Bandung: Informatika.

Suryanto, A. (2019). Penerapan metode mean absolute error (MEA) dalam algoritma regresi linear untuk prediksi produksi padi. *SAINTEKBU Vol 11 No 1*, 78-83.

Vulandari, R. T., & Hermawan, D. D. (2018). Perbandingan hasil panen padi dipengaruhi rata-rata curah hujan atau irigasi dengan model regresi nonlinier kubik di Kabupaten Sukoharjo. *TIKomSiN Vol 6 No 1*, 6-10.