Comparison of Linear Regression and Polynomial Regression for Predicting Rice Prices in Lhokseumawe City

Muhammad Iqbal[1]*, Rozzi Kesuma Dinata[2], Rizki Suwanda[3]
Department of Informatics Engineering, Universitas Malikussaleh [1], [2], [3]
Lhokseumawe, Indonesia
muhammad.200170138@mhs.unimal.ac.id[1], rozzi@unimal.ac.id[2], rizkisuwanda@unimal.ac.id[3]

Abstract*— Rice is a strategic food commodity in Indonesia, and its price fluctuations significantly impact inflation, economic stability, and poverty levels. Accurate price prediction is, therefore, essential for effective policymaking. The objective of this research is to develop a system for predicting the price of rice in Lhokseumawe City, employing a comparison of the accuracy of linear and polynomial regression models. To this end, daily price data from the Strategic Food Price Information Center (PIHPS) from 2020 to 2024 were utilized, with both models being implemented in Python. The findings indicate that 4th-order polynomial regression exhibited optimal performance, attaining a mean absolute percentage error (MAPE) of 1.85%, a mean absolute error (MAE) of 205.23, and a root mean squared error (RMSE) of 284.88. Conversely, the implementation of linear regression resulted in substantially elevated error metrics, with a mean absolute percentage error (MAPE) of 5.16%, a mean absolute error (MAE) of 553.91, and a root mean square error (RMSE) of 614.14. The findings indicate that 4th-order polynomial regression is a substantially more effective model for predicting rice prices in Lhokseumawe. The latter's superiority suggests that local rice price dynamics are characterized by significant non-linear patterns, rendering it a more robust tool for capturing data volatility and supporting data-driven policy.

***Keywords*— Rice Price, Prediction, Linear Regression, Polynomial Regression**

I. Introduction

Agriculture is pivotal for human life, primarily by ensuring food security. In Indonesia, rice stands out as the most critical commodity, serving as the staple food for the majority of the population [1]. As a primary component of household consumption, the availability and price stability of rice significantly impact public welfare, particularly for low-income communities [2] [3].

Daily and monthly fluctuations in rice prices are influenced by various factors, including weather conditions, production levels, distribution logistics, and government policies. Uncontrolled price increases can significantly drive inflation, affect economic stability, and exacerbate poverty levels. Consequently, developing accurate prediction models is imperative for informed decision-making and effective price control strategies [4] [5].

In this context, a relevant approach is data mining. This semi-automated process utilizes statistical, mathematical, artificial intelligence, and machine learning techniques to discover hidden patterns or knowledge in data [6]. Two popular methods in data mining for prediction are linear regression and polynomial regression. Linear regression is a statistical method that models the linear relationship between independent and dependent variables [7]. In contrast, polynomial regression is an extension of linear regression that can capture non-linear relationship patterns with curve shapes [8].

The choice between these methods involves a fundamental trade-off. Linear regression offers simplicity and clear interpretability but is limited to linear relationships and risks underfitting when data patterns are complex [9] [10]. Conversely, polynomial regression provides greater flexibility in modeling non-linear trends but at the cost of higher model complexity, which can lead to a significant risk of overfitting and interpretation difficulties if not carefully calibrated [10] [11].

The debate over the superiority of linear versus non-linear models for price prediction is ongoing. A substantial body of research highlights the effectiveness of non-linear models. For instance, a 4th-order polynomial model successfully tracked non-linear trends in the Nasdaq100 stock market [12]. and demonstrated significantly higher accuracy (R^2 = 0.99) compared to a linear model (R^2 = 0.69) in the real estate sector [13]. Similarly, a polynomial model showed a marginal advantage in predicting banking stock prices in Indonesia [11].

However, the superiority of non-linear models is not absolute. In certain contexts, the simplicity of linear models yields more reliable results. A 2024 study on national rice prices in Indonesia found that linear regression achieved a lower mean absolute percentage error (MAPE) of 6.29% compared to 6.88% for polynomial regression [14]. This finding reinforces the hypothesis that the optimal model choice is highly dependent on the specific characteristics of the data rather than assumptions about model complexity.

Despite this existing literature, most prior research has focused on other commodities or analyzed rice prices at a national scale. This national-level focus limits the

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

generalizability of findings to regional contexts. Price dynamics at the local level, such as in Lhokseumawe City, often exhibit unique characteristics driven by local supply chains, consumer behavior, and specific micro-economic factors. These local data patterns are critical determinants of a model's predictive success, yet they remain underexplored.

To address this gap, the objective of this study is to comparatively evaluate the accuracy and effectiveness of linear and polynomial regression methods in predicting daily rice prices in Lhokseumawe City. The primary objective is to determine which model offers optimal precision for local data patterns. The findings are expected to yield two main contributions: theoretically, to enrich the literature on food price prediction with empirical evidence at the micro-level, and practically, to provide a scientific basis for local governments to formulate more accurate, data-driven price control policies.

II. METHODOLOGY

This research uses a quantitative methodological approach, focusing on data in numerical form. The steps of the research process are as follows:

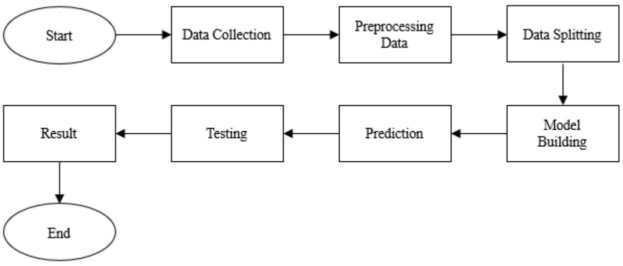


Fig 1. Research Process

A. Data Collection

This research uses Strategic Food Price Information Center (PIHPS) data. The analyzed data is on the development of rice prices in Lhokseumawe City from 2020 to 2024. The data was collected by downloading the data provided by PIHPS through the "Traditional Market Price Table by Commodity." Secondary data collection techniques were carried out through documentation by recording the obtained data for further analysis.

B. Preprocessing Data

The preprocessing stage is a step in the data preparation process before data mining is carried out. This stage is critical because approximately 50% of the data mining process occurs during this stage [15]. The first step in data preprocessing is data cleaning, which involves removing missing or duplicate data [16]. Preprocessing ensures that data quality is maintained before proceeding to the next stage.

C. Data Splitting

Data splitting is dividing a dataset into two or more parts. In this stage, the dataset was divided using an 80:20 ratio, with 80% designated as the training set and 20% as the testing set. This process first separated the independent (x) and dependent variables (y). The model was then trained on the larger training set to learn underlying data patterns. At the same time, the testing set was reserved for an unbiased evaluation of its performance on new, unseen data, thereby measuring its predictive accuracy [17].

D. Models Building

The modeling process is carried out using training data. In this case, the Linear Regression and Polynomial Regression models are implemented using the Python programming language and manual calculations. Linear regression is suitable for analyzing time series rice price data [18]. Meanwhile, polynomial regression was chosen for its ability to identify and analyze nonlinear patterns in the data.

1) *Linear Regression*

Manual calculations are performed using Python to determine the intercept and slope values at this stage. The intercept is the distance between the origin and the point where the regression line intersects the vertical axis. Concurrently, the term "slope" is employed to denote the gradient of the regression line. This slope is calculated by measuring the angle between the horizontal and regression lines [16]. The linear regression equation with the least squares method is as follows [19]:

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X).(\sum XY)}{n(\sum X^2) - (\sum X)^2} \quad (1)$$

$$b = \frac{n(\sum XY) - (\sum X).(\sum Y)}{n(\sum X^2) - (\sum X)^2} \quad (2)$$

$$Y = a + bX \quad (3)$$

In this model, *y* denotes the dependent variable, *x* represents the independent variable, *a* indicates the intercept, and *b* denotes the slope.

2) *Polynomial Regression*

At this stage, calculations are performed by accumulating the impact of each predictor variable (x) that is increased to a certain degree [17]. At this stage, manual calculations are performed using the Python programming language. In order to maximize the accuracy of the prediction, the coefficient value is calculated manually using the Gaussian elimination method. The formula for polynomial regression is as follows [19]:

$$Y = b_0 + b_1 X + b_2 X^2 + \cdots + b_n X^n + \epsilon \quad (4)$$

Assuming y is the predicted value:

$$Y = b_0 + b_1 X + b_2 X^2 + \cdots + b_n X^n \quad (5)$$

In this model, the dependent variable is indicated by y. The intercept, b_0, is represented by b_0. The slopes, b_1, b_2, …, b_n, indicate the degree of the polynomial, n. The independent variable, x, is designated by x. The error factor, ε, is

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

indicated by ε.

E. Prediction

Subsequent to the building of the linear and polynomial regression models, their application to the testing data was undertaken for the purpose of prediction. Each model was utilized to predict the price of rice for the designated time period corresponding to the test data.

F. Testing

During the testing phase, model predictions are compared with the actual rice price values in the testing data set. Three evaluation metrics are used: mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE).

1) *Mean Absolute Error (MAE)*

In statistics and predictive modeling, mean absolute error (MAE) is one of the evaluation methods used to measure the average absolute value of the difference between the actual and predicted values, MAE value is always positive or zero; the smaller the value, the more accurate the model. Conversely, a large MAE indicates a high level of prediction error, general form of the mean square error equation is as follows [18]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|y - \hat{Y}_i| \quad (6)$$

Explanation:
n : number of data points
y : actual data
$\hat{Y}_i$: predicted value

2) *Root Mean Squared Error (RMSE)*

The root mean square error (RMSE) method calculates the error level in the estimation results. This error indicates the difference between the estimated and actual values. RMSE measures the error rate of analytical calculation results using methods such as training and testing data. A high RMSE value indicates a large discrepancy between the predicted and actual values, resulting in low accuracy. Conversely, an RMSE close to zero indicates high prediction accuracy because the model's predictions are closer to the actual values [19]. The general form of the RMSE equation is as follows:

$$RMSE\ atau\ RMDE = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{Y}i)^2}{n}} \quad (7)$$

Explanation:
n : number of data points
y : actual data
$\hat{Y}_i$: predicted value

3) *Mean Absolut Percentage Error (MAPE)*

The mean absolute percentage error (MAPE) is a method used to measure the accuracy of a model. It calculates the percentage difference between the actual and forecast data [20]. This method first calculates the absolute error for each period, then divides the actual value by the forecast value to find the average absolute percentage error. A prediction is considered excellent if its MAPE value is less than 10%. The MAPE criteria are shown in the following table:

TABLE I. MAPE CRITERIA

MAPE	Description
<10%	Very Good
10% - 20%	Good
20% - 50%	Fair
50%	Poor

The formula for this testing method is shown below:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y-y_i|}{y}\ x\ 100\% \quad (8)$$

Explanation:
n : number of data points
y : actual data
$\hat{Y}_i$: predicted value

G. Result

The results stage summarizes and interprets the entire research process. The goal of this analysis is to develop a more accurate model based on Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) values.

III. DISCUSSION AND RESULT

A. Data Collection

The data were collected in September 2024 from the Strategic Food Price Information Center (PIHPS) website. The data set under consideration encompasses the price of rice from January 1, 2020, to August 30, 2024. The dataset under consideration contains a total of 1,217 rows and 4 columns, with attributes including "Date," "Semua Provinsi," "Aceh," and "Kota Lhokseumawe." The following is a detailed list of the contents of the dataset:

	Date	Semua Provinsi	Aceh	Kota Lhokseumawe
0	2020-01-01	11800.0	10300.0	10000.0
1	2020-01-02	11800.0	10300.0	10000.0
2	2020-01-03	11850.0	10300.0	10000.0
3	2020-01-06	11800.0	10350.0	10000.0
4	2020-01-07	11800.0	10350.0	10000.0
...	...	...	...	...
1213	2024-08-26	15350.0	13900.0	13450.0
1214	2024-08-27	15350.0	13900.0	13450.0
1215	2024-08-28	15250.0	13900.0	13450.0
1216	2024-08-29	15300.0	13900.0	13450.0
1217	2024-08-30	15350.0	14100.0	13450.0

1218 rows × 4 columns

Fig 2. Dataset

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

B. Preprocessing Data

The data preprocessing procedure encompasses two primary phases: data selection and data cleaning. In the data selection stage, the process of selecting attributes from the entire dataset is carried out. The following attributes were utilized in the prediction process: the "Date" column and "Kota Lhokseumawe.". The results of attribute selection at the data selection stage are shown in the following figure:

	Date	Kota Lhokseumawe
0	2020-01-01	10000.0
1	2020-01-02	10000.0
2	2020-01-03	10000.0
3	2020-01-06	10000.0
4	2020-01-07	10000.0
...	...	...
1213	2024-08-26	13450.0
1214	2024-08-27	13450.0
1215	2024-08-28	13450.0
1216	2024-08-29	13450.0
1217	2024-08-30	13450.0

1218 rows × 2 columns

Fig 3. Result of Data Selection

As illustrated in Figure 3 the attributes employed in the prediction process include the "Date" column and "Kota Lhokseumawe". These two attributes exhibit a close relationship, wherein alterations in the value of the Lhokseumawe City attribute are influenced by the time period represented by the Date attribute. Subsequent to the selection of data, the subsequent stage of the process is data cleaning.

Subsequently, the data undergoes a thorough cleansing process, entailing the identification and resolution of missing values within the dataset. The results of this check are shown in the following figure:

```
[6] # Check missing values
    df.isnull().sum()
```

	0
Date	0
Semua Provinsi	0
Aceh	0
Kota Lhokseumawe	0

dtype: int64

Fig 4. Result of Data Cleaning

As illustrated in Figure 4, the data set under consideration is devoid of any empty or missing values, thereby rendering it suitable for subsequent analysis and prediction processes.

C. Data Splitting

At this stage, data separation is carried out to support the modeling process. The data separation encompasses the independent variable (*x*), designated as the Date attribute, and the dependent variable (*y*), denoted as the Lhokseumawe City attribute. The objective of this process is to ensure that the data is well structured, thereby facilitating its utilization in the analysis and development of prediction models. The following source code segment illustrates the process of data separation into *x* and *y* variables:

```
[7] # Prepare data for regression
    X = df['Date'].values.reshape(-1, 1)  # Independent variable
    Y = df['Kota Lhokseumawe'].values     # Dependent variable
```

Fig 5. Separating Dependent and Independent Variables

Subsequent to the separation of the data into independent variables (*x*) and dependent variables (*y*), the dataset was segmented into 80% training and 20% test data. The proportion of data points obtained for training and testing was 974 and 244, respectively. The fundamental objective of this division is to ensure that the evaluation of the model is conducted objectively. The model will be constructed using the training data, while its performance and generalization ability will be accurately measured using the test data, which is entirely new to the model. The figure below illustrates the visualization of the training and test data distributions:

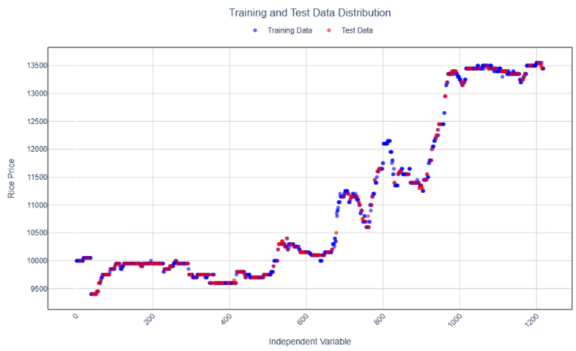


Fig 6. Training and Testing Data Distribution

As illustrated in Figure 6, the allocation of rice price data into two sets: a training set, represented by the blue area, and a testing set, represented by the red area. The intersection of these two data sets across the range of independent variables signifies a valid representation of price fluctuation patterns and trends. This balanced distribution of data is fundamental to ensure that the developed model has robust generalization capabilities and can be objectively evaluated against unseen data, guaranteeing the reliability of model performance estimates.

D. Model Building

At this stage, the Linear Regression and Polynomial Regression models were created using the Python programming language by utilizing the Google Colab platform.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

1) *Linear Regression*
A linear regression model is developed using the prepared training data at this stage. This training process aims to ascertain the most suitable model parameters, specifically the intercept value (a) as the initial cut-off point and the slope value (b) denoting the slope of the regression line. Preliminary analysis of the modeling results indicates that the intercept value (a) is 8805.79649552436, and the slope value (b) is 3.62721857. It is possible to formulate the final linear regression equation for predicting the price (Y) based on these two parameters as follows:

$$Y = 8805.79649552436 + 3.62721857X \quad (9)$$

2) *Polynomial Regression*
Polynomial regression is a form of nonlinear regression in which the relationship between the independent and dependent variables is expressed as a polynomial function. This model facilitates the use of higher polynomial orders, enabling the capture of intricate relationship patterns. In the present study, the focus is on polynomials of orders ranging from second to fifth. The following steps are outlined in the development of the Polynomial Regression model:

a) *2nd-order Polynomial Regression*
For a second-order polynomial model, the relationship between variables is represented by a quadratic curve (a second-degree polynomial). After training the model, three primary coefficients were determined: the intercept (b_0), the linear coefficient (b_1), and the quadratic coefficient (b_2).

The developed model yielded the following parameter values: an intercept: (b_0) of 9947.783130437068, a linear coefficient (b_1) of -2.0219701, and a quadratic coefficient (b_2) of 2.0219701. Consequently, the second-order polynomial regression equation for predicting the value of (Y) can be expressed as follows:

$$Y = 9947.783130437068 - 2.0219701X + 2.0219701X^2 \quad (10)$$

b) *3rd-order Polynomial Regression*
At the third order, the relationship between the independent and dependent variables is expressed as a third-degree or cubic polynomial. Consequently, the model encompasses not only the variables x and x^2, but also x^3. Training this model determines four distinct coefficients: the intercept (b_0), linear (b_1), quadratic (b_2), and cubic (b_3).

The resulting model yielded the following parameters: an: intercept (b_0) of 10293.254164805698, a linear coefficient (b_1) of -5.47311619, a quadratic coefficient (b_2) of 1.17225001, and a cubic coefficient (b_3) of -3.87524452. Therefore, the third-order polynomial regression equation for predicting the value of (Y) can be written as follows:

$$Y = 10293.254164805698 - 5.47311619X + 1.17225001X^2 - 3.87524452X^3 \quad (11)$$

c) *4th-order Polynomial Regression*
In the fourth order, the relationship between the independent and dependent variables is expressed in the form of a fourth-degree or quartic polynomial, similar to the previous orders. which includes terms up to x^4. Training this model determines five coefficients: the intercept (b_0), linear (b_1), quadratic (b_2), cubic (b_3), and quartic (b_4).

The developed model yielded the following parameters: an intercept (b_0) of 9731.313125044031, a linear coefficient (b_1) of 3.98026709, a quadratic coefficient (b_2) of -2.34633746, a cubic coefficient (b_3) of 4.12228435, and a quartic coefficient (b_4) of -1.8550564. Therefore, the fourth-order polynomial regression equation for predicting the value of (Y) can be written as follows:

$$Y = 9731.313125044031 + 3.98026709X - 2.34633746X^2 + 4.12228435X^3 - 1.8550564X^4 \quad (12)$$

d) *5th-order Polynomial Regression*
At the fifth order, the relationship between the independent and dependent variables is expressed in the form of a higher-degree polynomial, namely a fifth-degree or quintic polynomial, which incorporates terms up to x^5. The training process for this model determines six coefficients: the intercept (b_0), linear (b_1), quadratic (b_2), cubic (b_3) quartic (b_4), and quintic (b_5).

The developed model yielded the following parameters: an Intercept (b_0): 9.902910107716594, linear coefficient (b_1) of -7.29617638, quadratic coefficient (b_2) of -1.47187878, cubic coefficient (b_3) of -5.64689117, quartic coefficient (b_4) of 2.39139853, quintic coefficient (b_5) of -1.37466803, Thus, the 5th-order polynomial regression equation to predict the value of Y is written as follows:

$$Y = 9902.910107716594 - 7.29617638 - 1.47187878X^2 - 5.64689117X^3 + 2.39139853X^4 - 1.37466803 \quad (13)$$

IV. PREDICTION

At this stage, predictions are made using linear regression and polynomial regression methods by utilizing the previously developed models.

1) *Linear Regression Prediction*
In the course of the modeling process, the intercept (a) and

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

the slope (b) must first be determined. Subsequently, the prediction is calculated by evaluating the y value based on the formula. The calculation utilizes test data comprising 244 entries. The prediction results obtained using the linear regression method are enumerated in the subsequent table:

TABLE II. RICE PRICE PREDICTION USING LINEAR REGRESSION

	Actual	**Predicted**
542	10300	10771.75
260	10000	9748.87
44	9400	8965.39
1009	13200	12465.66
585	10150	10927.72
…	…	…
…	…	…
588	9800	10332.86
341	9900	9690.84
83	9600	9023.43
1503	13450	12701.43
89	9650	9037.94

The following graph illustrates the outcomes of predicting the price of rice utilizing a linear regression model:

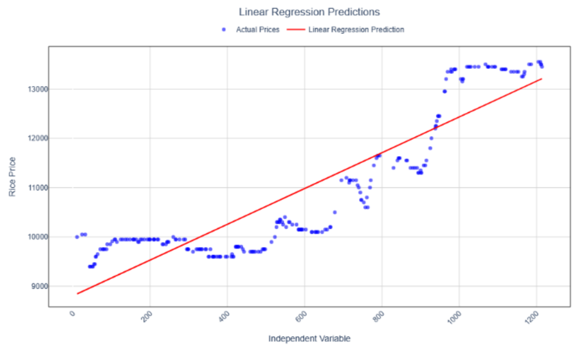


Fig 7 Visualization of Linear Regression Models

Figure 7 above illustrates a visualization of rice price prediction utilizing a linear regression model. The blue dots represent the actual price of rice, while the red line signifies the predicted price based on the regression model. This model produces a line that demonstrates the tendency of price increases over time.

2) *Polynomial Regression Prediction*

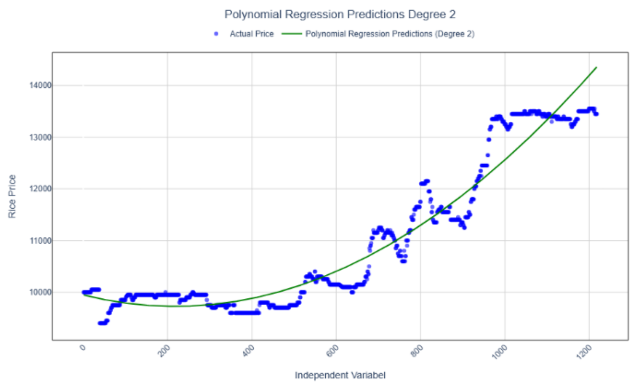
Polynomial Regression Model that was constructed in the preceding stage encompasses polynomial orders ranging from second to fifth. Subsequent to the modeling process, which generates intercept values and regression coefficients, predictions are derived utilizing the formula equation to calculate the y value. The process of prediction is contingent upon the utilization of testing data for calculation. The testing data utilized is the splitting result data, which amounts to 277 data points. The ensuing table presents the outcomes of employing the polynomial regression method for prediction purposes.

TABLE III. RICE PRICE PREDICTION USING POLYNOMIAL REGRESSION

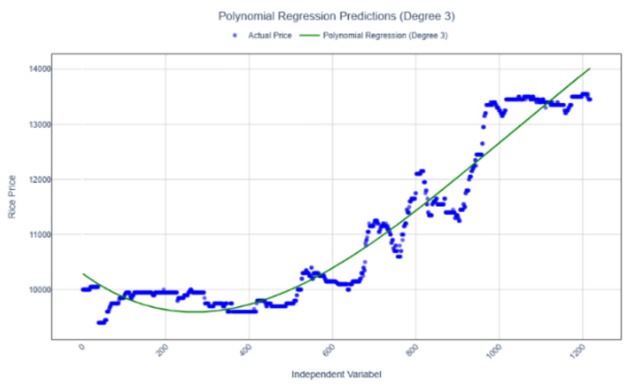
	Actual	**2nd-order**	**3rd-order**	**4th-order**	**5th-order**
542	10300	10213,19	10153.46	9958,56	9992,16
260	10000	9735,33	9594.57	9819,82	9797,11
44	9400	9867,79	10074.80	9864,46	9899,67
1009	13200	12625,43	12724.51	12978,24	13013,74
585	10150	10350,81	10327.38	10110,29	10127,59
…	…	…	…	…	…
…	…	…	…	…	…
421	9800	9917,87	9777.62	9741,56	9790,1
244	9900	9730,31	9599.43	9838,67	9806,12
60	9600	9843,15	10006.23	9894,32	9896,69
1074	13450	13121,43	13135.97	13328,25	13383,77
64	9650	9837,36	9989.97	9900,44	9895,79

The following visualization shows the predicted rice prices in Lhokseumawe City using the Polynomial Regression model.

(a)

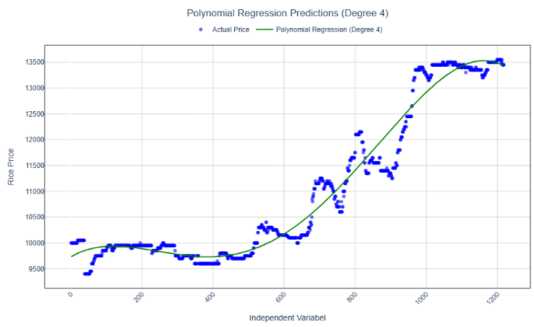


(b)



p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

(c)



(d)

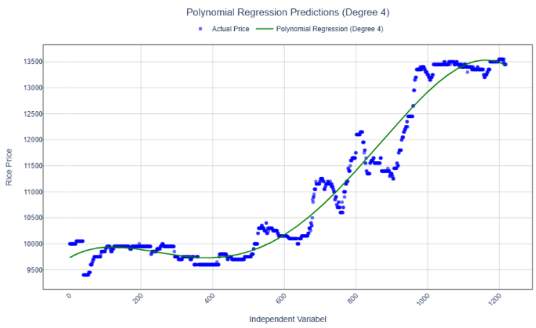


Fig 8 Visualization of Polynomial Regression

As illustrated in Figure 8, the visualization of rice price prediction using Polynomial Regression of each order employed is demonstrated. In this figure, the blue dots represent the actual price of rice, while the green line illustrates the trend of rice prices based on predictions from the Polynomial Regression model.

E. Testing

The testing process entails the calculation of the discrepancy between the observed and anticipated values, employing three distinct evaluation methodologies: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The Linear Regression Model yielded an MAE of 554.07, an RMSE of 614.30, and a MAPE of 5.16%. Conversely, the 4th-order Polynomial Regression Model yielded optimal results, exhibiting an MAE of 205.23 and a MAPE of 1.85%. This model demonstrated superior prediction accuracy in comparison to other polynomial models of varying orders. While the 5th-order model demonstrated the optimal Root Mean Square Error (RMSE) value of 281.02, there was a marginal decline in accuracy. Consequently, the 4th-order polynomial regression model is regarded as the most optimal model for predicting rice price data in Lhokseumawe City.

F. Result

The results of the prediction made by the linear regression and polynomial regression methods can be analyzed using three main evaluation metrics: The following error metrics will be examined: mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The following table presents a comparison of the two methods based on the three metrics outlined:

TABLE IV. COMPARISON OF TEST VALUES

Model	MAE	RMSE	MAPE
Linear Regression	553.91	614.14	5.16%
2nd-order Polynomial Regression	286.50	359.55	2.54%
3rd-order Polynomial Regression	280.43	345.92	2.53%
4th-order Polynomial Regression	205.23	284.88	1.85%
5th-order Polynomial Regression	210.25	281.02	1.90%

A comparative analysis of linear and polynomial regression models revealed a substantial superiority of the polynomial approach in predicting rice prices. The linear regression model was deficient, as evidenced by the elevated error rates across all evaluation metrics: MAE 553.91, RMSE 614.14, and MAPE 5.16%.

In contrast, the polynomial models that were examined were capable of significantly reducing the error, suggesting the presence of a robust nonlinear relationship in the data. The 4th-order polynomial model demonstrated the optimal performance, with an MAE of 205.23, an RMSE of 284.88, and a MAPE of 1.85%. This model showed a substantially reduced prediction error, with a decrease of over 60% compared to the linear model. While the 5th-order model demonstrated a marginally superior Root Mean Square Error (RMSE) value of 281.02, the selection of the 4th-order model as the optimal model was predominantly influenced by its lower Mean Absolute Percentage Error (MAPE) and Mean Absolute Deviation (MAD) values. This finding suggests that the 4th-order model is accurate, more stable, and less sensitive to outliers than the fifth-order model. Consequently, 4th-order polynomial regression is recommended as the most reliable and effective predictive model for daily rice price dynamics in Lhokseumawe.

V. CONCLUSION

The findings of this study indicate that the 4th-order polynomial regression model is the most effective and stable method for predicting daily rice prices in Lhokseumawe City, with a low prediction error rate (MAPE 1.85%). The model's capacity to capture nonlinear price fluctuations significantly surpasses the accuracy of linear regression models, thereby establishing it as a superior instrument for price analysis. These findings contribute to the scientific basis for local governments in designing proactive and data-driven price stabilization policies. Making precise predictions enables stakeholders to foresee price surges, strategize market interventions, ensure food security, and regulate regional inflation. However, this study is not without its limitations. Firstly, it relies exclusively on historical price data, excluding external variables such as supply data, distribution costs, and weather factors. Consequently, further research is recommended to develop a multivariate model that integrates these factors. Furthermore,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025

exploring more sophisticated forecasting methodologies, such as ARIMA or other machine learning models, could be a prospective research trajectory to enhance prediction accuracy.

References

[1] V. Arinal and M. Azhari, “Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia,” *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 341–346, Sep. 2023.

[2] Y. P. P. Pasaribu and I. M. K. Karo, “ANALISIS PREDIKSI HARGA BERAS DI INDONESIA DENGAN METODE MONTE CARLO,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, Art. no. 4, Jun. 2024, doi: 10.36040/jati.v8i4.9962.

[3] N. Nafi’iyah and P. A. Wulandari, “Prediksi Harga Beras Berdasarkan Kualitas Beras dengan Metode Long Short Term Memory,” vol. 7, no. 2, 2022.

[4] A. R. Anandyani, D. K. A. Astutik, N. Bariroh, and A. Indrasetianigsih, “PREDIKSI RATA-RATA HARGA BERAS YANG DIJUAL OLEH PEDAGANG BESAR (GROSIR) MENGGUNAKAN METODE ARIMA BOX JENKINS,” *TEKNOSAINS*, vol. 15, no. 2, p. 151, Aug. 2021, doi: 10.24252/teknosains.v15i2.17721.

[5] M. Andriyani, S. Nurwilda, D. Z. Haq, and D. C. R. Novitasari, “PREDIKSI HARGA BERAS PREMIUM TAHUN 2024 MENGGUNAKAN METODE GRADIENT BOOSTED TREES REGRESSION,” *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, vol. 18, no. 2, Art. no. 2, Aug. 2024, doi: 10.47111/jti.v18i2.14859.

[6] N. Nurdin, “ANALISA DATA MINING DALAM MEMPREDIKSI MASYARAKAT KURANG MAMPU MENGGUNAKAN METODE K-NEAREST NEIGHBOR,” *JITET*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4131.

[7] M. Sholeh, E. K. Nurnawati, and U. Lestari, “Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 8, no. 1, Art. no. 1, Jan. 2023, doi: 10.14421/jiska.2023.8.1.10-21.

[8] D. A. Ferryan, P. K. Intan, and M. Hafiyusholeh, “Peramalan Harga Minyak Mentah di Indonesia dengan Metode Regresi Polinomial,” *JIMT*, vol. 19, no. 1, pp. 13–18, Jun. 2022, doi: 10.22487/2540766X.2022.v19.i1.15779.

[9] S. M. Putri, R. Novita, M. Mustakim, and M. Afdal, “Perbandingan Algoritma Linear Regression, Support Vector Regression, dan Artificial Neural Network untuk Prediksi Data Obat,” *bits*, vol. 6, no. 1, Jun. 2024, doi: 10.47065/bits.v6i1.5184.

[10] V. K. Sahu, “Linear vs. Polynomial Regression: Understanding the Differences,” GeeksforGeeks. Accessed: Sep. 04, 2024. [Online]. Available: https://www.geeksforgeeks.org/linear-vs-polynomial-regression-understanding-the-differences/

[11] S. A. L. Satriyo, A. R. Pratama, and Rahmat, “Perbandingan metode linear regresi dan polynomial regresi untuk memprediksi harga saham studi kasus Bank BCA,” *INFOTECH : Jurnal Informatika & Teknologi*, vol. 4, no. 1, Art. no. 1, Jun. 2023, doi: 10.37373/infotech.v4i1.602.

[12] G. Cohen, “Polynomial Moving Regression Band Stocks Trading System,” *Risks*, vol. 12, no. 10, p. 166, Oct. 2024, doi: 10.3390/risks12100166.

[13] I. P. Putra and I. K. G. Suhartana, “Perbandingan Akurasi Algoritma Regresi Linier, Regresi Polinomial, dan Support Vector Regression Pada Model Sistem Prediksi Harga Rumah,” vol. 1, 2022.

[14] F. A. Bilawa, H. Hikmayanti, and R. Rahmat, “Prediksi Harga Beras Medium Di Indonesia Dengan Membandingkan Metode Regresi Linear Dan Regresi Polinomial,” *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, vol. 9, no. 2, Art. no. 2, Aug. 2024, doi: 10.30645/jurasik.v9i2.810.

[15] Z. Setiawan *et al.*, *BUKU AJAR DATA MINING*. PT. Sonpedia Publishing Indonesia, 2023.

[16] S. Karbala and I. Ali, “MEMPREDIKSI HARGA BERAS ECERAN MENGGUNAKAN ALGORITMA REGRESI LINIER,” *jati*, vol. 7, no. 3, pp. 1554–1559, Oct. 2023, doi: 10.36040/jati.v7i3.6901.

[17] R. Rahman, A. Sudiarjo, and Y. Sumaryana, “PREDIKSI UPAH MINIMUM PROVINSI 10 TAHUN KEDEPAN DENGAN MENGGUNAKAN MODEL POLYNOMIAL REGRESSION,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, Art. no. 4, Aug. 2024, doi: 10.36040/jati.v8i4.9889.

[18] R. Putra and A. Sindar, “Perkiraan Harga Beras Premium DKI Jakarta Menggunakan Regresi Linier,” *Journal of Information Engineering and Educational Technology*, vol. 6, pp. 80–85, Dec. 2022, doi: 10.26740/jieet.v6n2.p80-85.

[19] A. E. Putra and A. Juarna, “Prediksi Pro duksi Daging Sapi Nasional dengan Meto de Regresi Linier dan Regresi Polinomial,” *Jurnal Ilmiah Komputasi*, vol. 20, no. 2, Art. no. 2, Jun. 2021.

[20] R. Heni, Solihin, J. Supratman, and R. Muhendra, “Pengembangan model peramalan penjualan menggunakan metode regresi linier dan polinomial pada industri makanan ringan (Studi Kasus: CV. Stanley Mandiri Snack),” *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, Art. no. 2, Jul. 2023, doi: 10.37373/tekno.v10i2.456.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i3.2396, Copyright ©2025
Submitted : Mey 23, 2025, Revised : June 4, 2025, Accepted : June 11, 2025, Published : July 28, 2025