Journal of Information Technology and Cyber Security 3(2) July 2025: 95-113 EISSN 2987-386X | PISSN 2987-3878
DOI: 10.30996/jitcs.12353

A Comparison of Polynomial Regression and Support Vector Regression for Predicting the Consumer Price Index Based on Food Commodity Prices in East Java, Indonesia

Ayu Adelina Suyono [1,*]

[1] Department of Information Systems, Universitas KH. A. Wahab Hasbullah, Indonesia
[*] Corresponding author: adelinasuyono23@gmail.com

Received: 14 December 2024
Accepted: 3 February 2025
Revised: 24 January 2025
Available online: 21 September 2025

To cite this article: Suyono, A. A. (2025). A Comparison of Polynomial Regression and Support Vector Regression for Predicting the Consumer Price Index Based on Food Commodity Prices in East Java, Indonesia. *Journal of Information Technology and Cyber Security, 3*(2), 95-113. https://doi.org/10.30996/jitcs.12353

Abstract

Food price fluctuations occur almost daily and directly affect purchasing power as well as the stability of regional and national economies. As one of the largest provinces in Indonesia, East Java, which significantly contributes to national GDP, has diverse economic structures and highly sensitive to price changes. Given this situation, government needs more accurate prediction methods to monitor Consumer Price Index (CPI) movement as a basis for establishing more appropriate economic strategy and policy. This study aims to compare the performance of Polynomial Regression (PR) and Support Vector Regression (SVR) in predicting CPI using food price data from SISKAPERBAPO for the 2014 - 2020 period, covering regencies and cities in East Java. To ensure the quality of the analysis, missing values were removed. A Pearson's r correlation analysis was then conducted to assess the relationships between food prices and CPI. The model obtained was then evaluated using mean squared error (MSE), root mean square error (RMSE), Mean absolute percentage error (MAPE), and computation time. The results shows that third order PR achieved higher accuracy with MAPE of 0.3% (training) and 3.4% (testing), while SVR performed lower with MAPE of 5.9% (training) and 6.0% (testing). In addition, PR was more computationally efficient than SVR. These findings underscore PR as a more reliable method for predicting CPI using complex regional food data.

Keywords: Consumer Price Index, food commodities, Pearson Correlation Coefficient, Polynomial Regression, Radial Basis Function, Support Vector Regression.

1. Introduction

Daily fluctuations in food commodity prices are difficult to avoid. For example, in early 2022, the prices of cooking oil, chilies, onions, and eggs rose significantly (Aida & Nugroho, 2022). This situation suppressed consumer purchasing power and national income (Sari, Muslihah, Mutohari, & Sari, 2025). As one of Indonesia's largest provinces, contributing approximately 14.86% to the national Gross Domestic Product (GDP), East Java is also among the regions most significantly affected by the decline in national income (Rahmadi, 2018).

The diverse economical structure, range from agriculture, fisheries, industry, trade, and services, giving the province a strategic role in Indonesia's economy as well as different consumption patterns across regions that sensitive to price changes. In April 2024, East Java recorded year on year (y o y) inflation of 3.25% (BPS Provinsi Jawa Timur, 2024). To address these challenges, the government uses the Consumer Price Index (CPI) as a primary instrument to monitor inflation and deflation. Computed from the prices of various commodities, CPI serves as a key indicator for policy formulation. However, a sole monitoring activity is not sufficient. CPI forecasts are needed to anticipate potential inflation early so that appropriate policy responses can be prepared.

Numerous studies have attempted to forecast CPI using different methods. Double Exponential Smoothing (DES) has demonstrated high performance, for example, achieving a Mean Absolute Percentage

Error (MAPE) of 0.32% for Pontianak's CPI (January 2020–September 2023) (Pramesti, Sadikin, Imro'ah, & Maulida, 2023) and 0.76% for Yogyakarta's CPI (2014–2022) (Asmaradana & Widodo, 2023). These results indicate that DES is capable of providing accurate predictions for regional CPI data, consistent with the findings of Chen, Do, Nguyen, & Doan (2018), who argued that the smaller the MAPE, the better the model's performance.

On the other hand, Support Vector Regression (SVR) is also a promising alternative. Prakoso (2019) reported MAPE of 8.70% for national CPI (2003–2017) using an SVR model, which is lower in accuracy than DES for local CPI with good linearity. Nevertheless, SVR has strengths in handling nonlinearity, mitigating overfitting, and working effectively with high dimensional data (Haoyuan, Yizhong, Chenglong, Jian, & Lijun, 2023; Zhang & O'Donnell, 2020), typical of economics and finance. Using prices of 34 food commodities as predictors and CPI as the target, and analyzing correlations via Pearson's method, SVR can handle complex time series data (R, Sudarmin, & Rais, 2022; Rohmah, Ardiantoro, Putra, & Hartati, 2019), although its performance depends on feature weighting, training time (Xie, Xie, & Zhu, 2021), and appropriate hyperparameter selection (Hsia & Lin, 2020).

Polynomial Regression (PR) has also been employed for prediction. A prior study by Zhou, Yan, & Zhang (2024) reported that PR can fit training data exceptionally well, achieving nearly perfect R^2 values with very low Mean Absolute Error (MAE) and (MAPE, which indicates a tendency toward overfitting. In contrast, SVR often demonstrates more stable performance on test data, suggesting a better ability to handle non-linear relationships and outliers. These findings were derived from a soybean branching study conducted using Smart-breeding systems and specialized breeding methods, which analyzed 1,918 samples and 42,291 Single Nucleotide Polymorphisms (SNPs) that revealed genetic variations.

Based on this background, there is a research gap in CPI modeling that directly compares PR and SVR on complex regional economic data. Therefore, this study focuses on predicting CPI using SVR and PR and comparing their performance to identify the superior model and to support more accurate economic forecasting strategies.

2. Literature Review

2.1. Consumer Price Index (CPI)

The Consumer Price Index (CPI) is an economic indicator that reflects changes in the prices of goods and services commonly purchased by households (Badan Pusat Statistik, 2020). An increase in the CPI indicates inflation, a condition in which prices rise and purchasing power declines. To capture household consumption patterns, Badan Pusat Statistik (BPS, Statistics Indonesia) conducts the Household Budget Survey (SBH).

According to BPS (2024), SBH is designed to capture society's general consumption patterns. Two main outputs are produced: (i) a commodity basket listing the goods and services most consumed in a district/city and (ii) the corresponding consumption values (expenditures). These inform CPI updates for consumer level inflation measurement.

In 2018, BPS conducted SBH in 90 cities consist of 34 provincial capitals and 56 other relatively large economies (BPS Provinsi Jawa Timur, 2021). n East Java, eight regions were covered: Jember, Banyuwangi, Sumenep, Kediri, Malang, Probolinggo, Madiun, and Surabaya (BPS Provinsi Jawa Timur, 2021). The result shows that the number of basket items ranged from 248 to 478 goods/services. CPI uses 2018 = 100 as the base year and is classified into 11 groups and 43 subgroups.

This study focuses on the Food, Beverages, and Tobacco group, especially the Food subgroup. From among the recorded commodities, 34 food items most commonly consumed by East Java residents were selected, following prior research (Rohmah, Ardiantoro, Putra, & Hartati, 2019). These commodities form the basis for analyzing consumption patterns and price trends in East Java.

The data used in the study of Rohmah, Ardiantoro, Putra, & Hartati (2019) was obtained from SISKAPERBAPO administered by the government of East Java. The data consists of price and availability of the commodities, including rice (e.g., Bengawan, Mentik, IR64), local sugar, palm-based oil (e.g., 2 litres package oil, bulk oil), meat (e.g., beef, broiler chicken meat, free-range chicken meat), chicken egg, sweetened condensed milk (e.g., Bendera, Indomilk), instant milk powder (e.g., Bendera, Indomilk), dried corn kernels, Iodized salt (e.g., brick, fine), Medium Grade Wheat Flour (i.e., segitiga biru), Imported Soybeans, Instant Noodles (i.e., Indomie with Chicken Curry Flavor), chily (e.g., regular chily, bird's eye chily), shallots, garlic, salted anchovies, green beans, peanuts, cassava, cabbage, potatoes, tomatoes, carrots, and green beans.

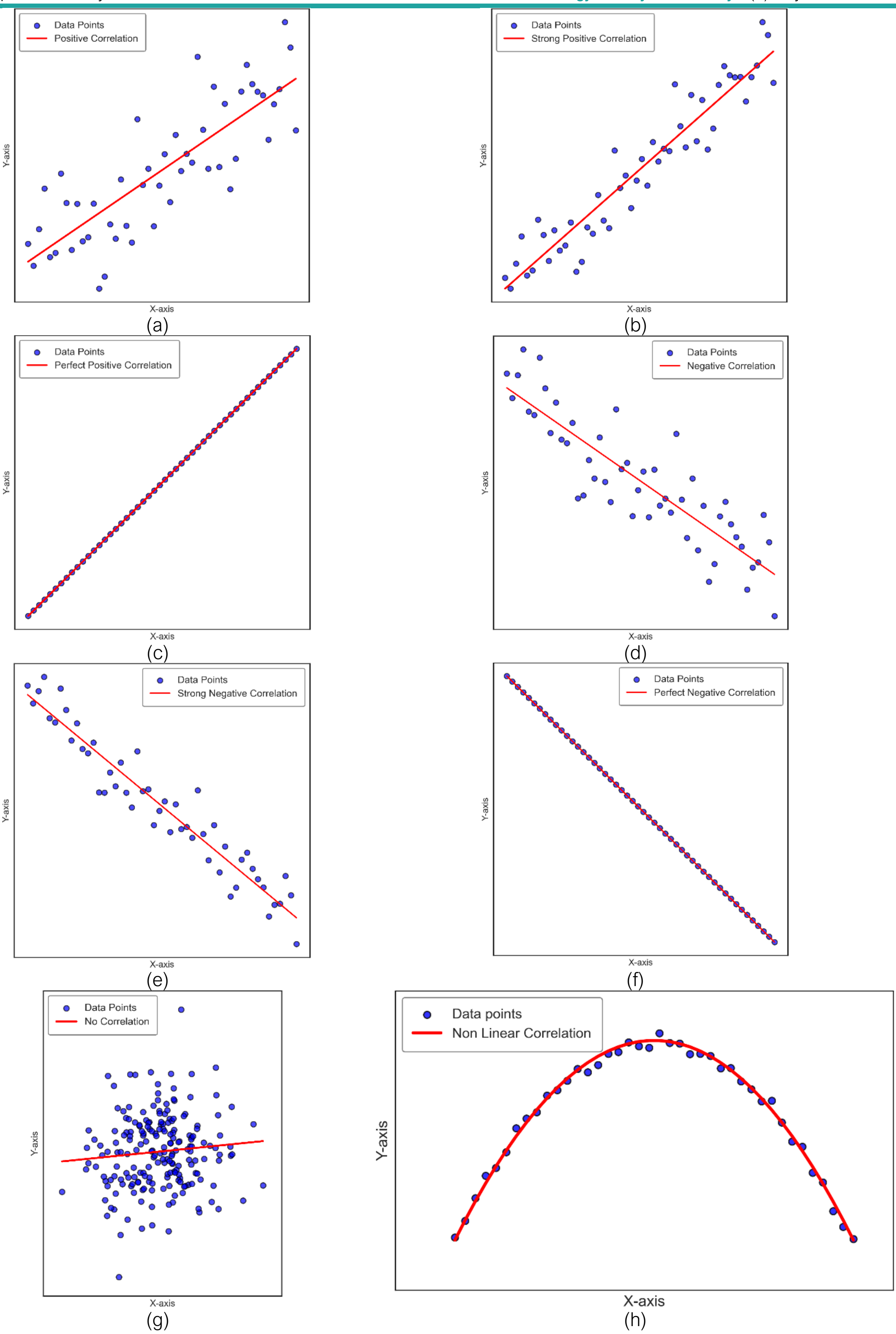


Fig. 1. **(a)** Positive correlation (Gogtay & Thatte, 2017). **(b)** Strong positive correlation (Delmo, Villarica, & Vinluan, 2022). **(c)** Perfect positive correlation (Delmo, Villarica, & Vinluan, 2022). **(d)** Negative correlation (Gogtay & Thatte, 2017). **(e)** Strong negative correlation (Delmo, Villarica, & Vinluan, 2022). **(f)** Perfect negative correlation (Delmo, Villarica, & Vinluan, 2022). **(g)** No correlation (Gogtay & Thatte, 2017). **(h)** Nonlinear correlation (Kumar & Chong, 2018).

Table 1
General guidelines for interpreting correlation coefficients according to Hinkle, Wiersma & Jurs.

Correlation Value	Interpretation
0.90 to 1.00 (−0.90 to −1.00)	Very strong positive or negative correlation
0.70 to 0.90 (−0.70 to −0.90)	Strong positive or negative correlation
0.50 to 0.70 (−0.50 to −0.70)	Moderate positive or negative correlation
0.30 to 0.50 (−0.30 to −0.50)	Weak positive or negative correlation
0.00 to 0.30 (0.00 to −0.30)	Negligible or very weak correlation

Table 2
General guidelines for interpreting correlation coefficients according to Sugiyono.

Correlation Value	Interpretation
0.80-1.000	Very strong correlation
0.60-0.799	Strong correlation
0.40-0.599	Moderate correlation
0.20-0.399	Weak correlation
0.00-0.199	Negligible or very weak correlation

Table 3
General guidelines for interpreting correlation coefficients according to Salkind.

Correlation Value	Interpretation
0.8-1.0	Very strong relationship
0.6-0.8	Strong relationship
0.4-0.6	Moderate relationship
0.2-0.4	Weak relationship
0.0-0.2	Weak or no relationship

2.2. Pearson Correlation Coefficient (r)

In statistics, correlation is used to assess the likelihood and strength of association between two variables (Wisniewski & Brannan, 2024). According to Bewick, Cheek, & Ball and Mata & Milner in Wisniewski & Brannan (2024), the choice of correlation analysis depends on the type of data, whether interval/continuous or ordinal/ranked.

Pearson Product-Moment Correlation (PPMC), also commonly known as the Pearson correlation coefficient (r) or Pearson's r (Park, et al., 2020; Wisniewski & Brannan, 2024), is one of the most widely used methods. It is most suitable for continuous variables, particularly interval or ratio data (Bewick, Cheek, & Ball, 2003; Mata & Milner, 2021).

The coefficient ranges from –1 to +1, where 0 indicates no linear association. Values approaching ±1 indicate strong linear relationships (positive when moving in the same direction; negative when moving in opposite directions) (Mata & Milner, 2021). Correlation quantifies the strength of association but does not imply causation (Soyer, 2017).

Visual inspection can be performed using scatterplots, where each data pair (x, y) is represented as a single point, as shown in Fig. 1. Several widely cited interpretation guides are presented in Table 1 (Hinkle, Wiersma, & Jurs, cited in Mukaka, 2012), Table 2 (Sugiyono, cited in Larasati, Susilorini, Surjowardojo, & Wahyuni, 2024), and Table 3 (Salkind, 2010).

2.3. Support Vector Regression (SVR)

SVR is the regression counterpart of Support Vector Machines, designed to handle continuous outputs—including time series forecasting. According to Vanitha & Kasthuri (2024), SVR aims to discover the most accurate function to predict continue output values based on their inputs.

SVR can use linear or nonlinear kernels (Vanitha & Kasthuri, 2024), with kernel choice adapted to data characteristics and problem complexity. This way, SVR can effectively work in a condition prone to overfitting (Ashoka, Rekha, & Sudha, 2024).

SVR has been applied to daily or monthly CPI using real time inputs such as daily commodity prices and foreign exchange rates, as well as staple food prices (Budiastuti, Nugroho, & Hariadi, 2017). Its ability to process continuous data makes SVR a reliable method for predictive analysis, both in economics and data science.

SVR has been applied in various fields, including daily and monthly CPI predictions. These predictions utilize real-time data, such as daily commodity prices, foreign exchange rates (Budiastuti, Nugroho, & Hariadi, 2017), and staple food prices (Editya, Kurniati, Septianto, Lisdiyanto, & Al Haromainy, 2021).

Table 4
Example of daily food commodity price data.

City/District	Date	Commodity	Price (IDR/kg)
Jember	01/01/2014	Onion	33,000
Jember	01/01/2014	Garlic (White Onion)	11,000
Jember	01/01/2014	Rice (IR64 variety)	7,500
Jember	01/01/2014	Broiler Chicken Meat	24,000
Jember	01/02/2014	Onion	33,000
Jember	01/02/2014	Garlic (White Onion)	10,000
Jember	01/02/2014	Rice (IR64 variety)	7,500
Jember	01/02/2014	Broiler Chicken Meat	25,000
...	...	...	...
Jember	31/12/2014	Onion	33,333
Jember	31/12/2014	Garlic (White Onion)	20,000
Jember	31/12/2014	Rice (IR64 variety)	10,167
Jember	31/12/2014	Broiler Chicken Meat	33,000

Note: Prices are reported in Indonesian Rupiah (IDR) per kilogram (kg).

Table 5
Example of monthly Consumer Price Index (CPI) data for Jember (2014–2020).

Month	2014	2015	2016	2017	2018	2019	2020
January	164.80	177.70	124.94	131.88	133.04	136.18	104.11
February	168.53	122.28	127.45	129.79	133.51	134.77	105.48
March	172.92	120.60	127.83	128.24	131.76	133.91	104.50
April	171.31	119.92	126.96	127.56	132.23	135.82	103.43
May	167.62	117.81	126.27	128.75	132.74	137.45	102.67
June	168.50	118.63	125.81	128.59	135.68	135.76	103.10
July	177.12	118.67	126.77	128.37	134.10	136.89	102.71
August	179.85	122.26	128.13	127.57	133.15	136.70	101.33
September	177.82	123.70	126.38	127.46	131.44	134.47	101.07
October	175.12	123.67	126.15	125.93	131.59	134.72	101.17
November	173.15	122.85	124.40	127.46	132.81	135.87	102.96
December	177.70	123.84	125.48	130.63	135.51	138.26	104.56

Note: CPI values are based on data from Statistics Jawa Timur (BPS Provinsi Jawa Timur) and reflect the monthly Consumer Price Index for Jember. The base year and scaling factor follow the official BPS methodology.

2.4. Polynomial Regression (PR)

PR is a variant of linear regression in which the relationship between independent variables x and dependent variable y is represented by an n order polynomial (Altair RapidMiner, 2025). This approach is relevant when relationships are nonlinear (Myśliwiec, et al., 2025). In Altair RapidMiner (2025), The variable y is treated as a label attribute, while the variable x is a set of regular attributes used to predict the value of y. Thus, PR can be viewed as a special case of multiple linear regression (Ostertagová, 2012).

While PR is more flexible than simple linear regression, higher polynomial orders increase the risk of overfitting. In this condition, the model no longer captures the underlying patterns in the data, but instead adapts to the noise. As a result, the model's performance may be excellent on the training data, but it may decline drastically when tested on new data (Cybellium, 2023).

3. Methodology

This section explains the research methodology employed in the study. The methodology comprises data collection and processing, algorithm design, analysis, and evaluation. All processes were carried out using Python in a Jupyter Notebook environment.

3.1. Data Collection

The data collection was conducted through indirect observation using secondary data provided by the government of East Java. This dataset is publicly available on the official website of SISKAPERBAPO (https://siskaperbapo.jatimprov.go.id) and consists of daily prices of 34 food commodities, classified according to Rohmah, Ardiantoro, Putra, & Hartati (2019). The data cover eight regions/cities in East Java, selected based on the 2018 SBH coverage area, and span from January 1, 2014, to December 31, 2020, resulting in a total of 20,453 records. In addition, the Consumer Price Index (CPI) data were obtained from BPS Jawa Timur (https://jatim.bps.go.id), published annually, and used as the output of the machine learning system.

Table 6
Sample of integrated dataset: Daily food commodity prices and corresponding CPI values.

City/District	Date	Onion (IDR/kg)	Garlic (IDR/kg)	Bengawan Rice (IDR/kg)	IR64 Rice (IDR/kg)	…	CPI
Jember	01/01/2014	33,000	11,000	8,500	7,500	…	164.80
Jember	02/01/2014	33,000	11,000	8,500	7,500	…	164.80
Jember	01/03/2014	33,000	10,000	8,500	7,500	…	164.80
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Surabaya	31/12/2020	27,600	23,400	12,620	9,450	…	104.61

Table 7
Proportion of training and testing data splits.

Training Proportion (%)	Testing Proportion (%)	Number of Training Samples	Number of Testing Samples
90	10	17,750	1,973
80	20	15,778	3,945
70	30	13,806	5,917

Daily commodity price data were summarized into monthly averages to align with the CPI reporting period. These data were then integrated with the consumption quantity data used by BPS to calculate the CPI. The CPI values were calculated using the Laspeyres formula, as described in BPS's annual publications, where commodity prices (P) and quantities (Q) were combined to derive the monthly index. This method transformed daily commodity price data into monthly CPI data.

The processing resulted in an integrated dataset of commodity prices and CPI values, which was then used for correlation analysis and predictive modeling with regression methods. All data were stored in Microsoft Excel format (.xlsx) to support data preprocessing and further exploration. The input data structure is shown in Table 4, while the output data structure is shown in Table 5.

3.2. Data Preprocessing

Data preprocessing is conducted to ensure the sufficiency of the data (i.e., clean, structured, and ready to analyze). The process consists of three steps:

1) Missing value handling: In the commodity price and CPI data, several entries with zero or empty values were considered missing values in the analysis context. These missing values could potentially interfere with the modeling results, so they were removed from the dataset. After the cleaning process, the data was reduced from 20,453 to 19,723 complete entries ready to be used as model input.
2) Data combination and integration: At this stage, daily commodity price data, which serves as the input variable (x), is combined with CPI data, which serves as the target variable (y). Furthermore, data collected from eight districts/cities in East Java Province is combined into a single .xlsx file. The results of this integration form a unified dataset that is used in the model training and testing process. An example of the integrated data structure can be seen in Table 6.
3) Dataset grouping – training and testing set: To build and evaluate the performance of the SVR model, the data was divided into two subsets: a training set for training the model and a testing set for testing prediction accuracy. In this study, experiments were conducted using varying data split ratios, with 70%:30% being a commonly used ratio (Joseph, 2022), 80%:20% based on Pareto principle (Joseph, 2022), and 90%:10% which has been proven obtaining best results in various algorithms and datasets (Bukaita, Celis, & Gurram, 2024).

The purpose of using these ratio variations is to evaluate changes in model performance based on different proportions of training and test data, and to ensure that model results are independent of a single data split scenario. A summary of the data split variations used can be seen in Table 7.

3.3. Variable Correlation Analysis and Regression Methods

3.3.1. Correlation analysis using Pearson Product Moments

At this stage, a correlation analysis was conducted to evaluate the relationship between food commodity prices as the independent variable (x) and the CPI value as the dependent variable (y) using Pearson's r. The x variable covers 34 types of food commodities, while the y variable represents the CPI value. The results of the correlation calculations were then visualized in the form of a heatmap and horizontal bar chart, implemented using the Seaborn and Matplotlib libraries in Python, as shown in Fig. 2 and Fig. 3.

Based on Fig. 2, the visualization interpretation is based on color intensity. A darker bluish-green color (approaching +1) indicates a strong positive correlation, meaning that an increase in the price of one commodity tends to be followed by an increase in the price of another commodity and an increase in the CPI. Conversely, a darker brown color (approaching -1) indicates a strong negative correlation, meaning

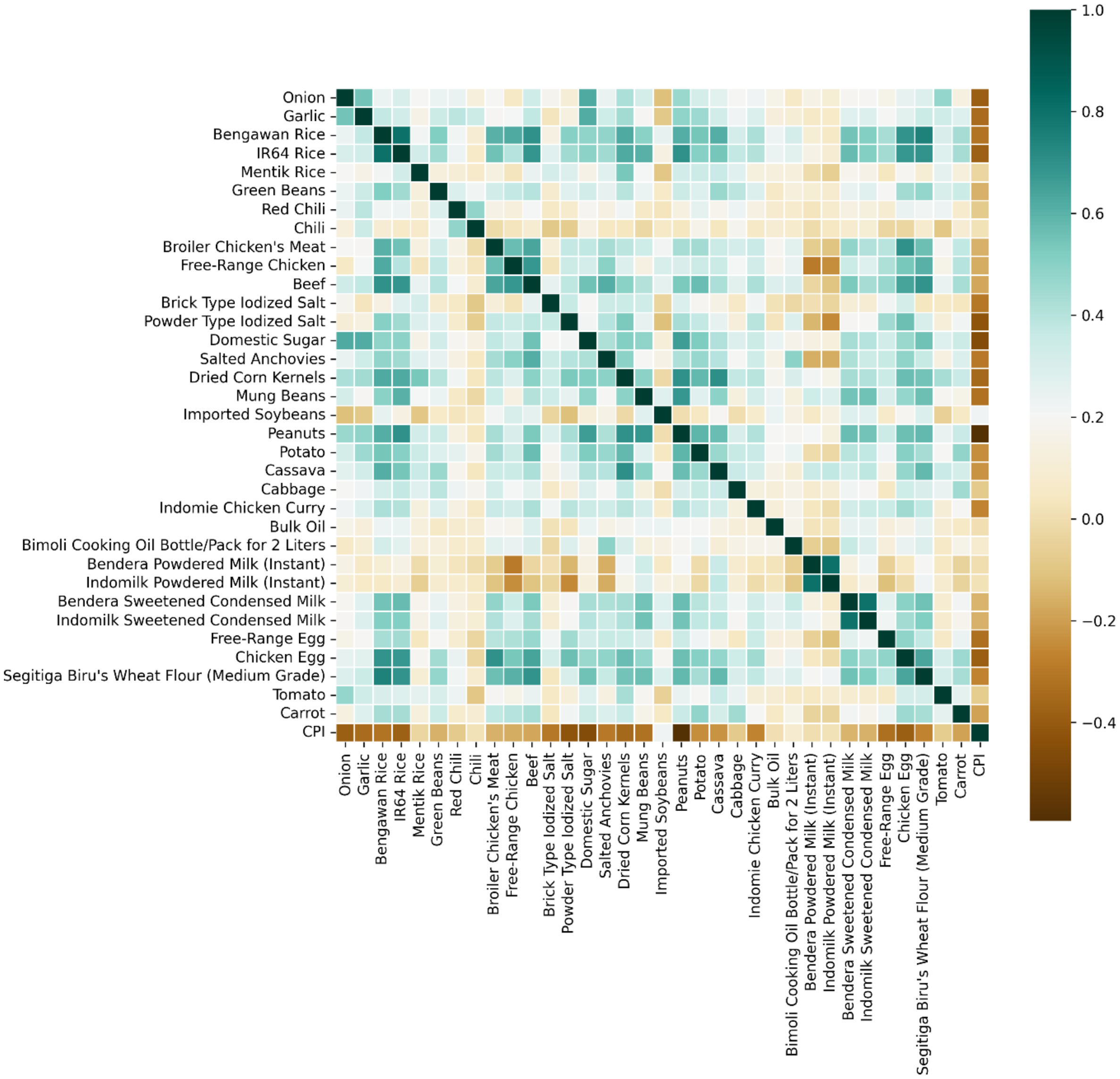


Fig. 2. Heatmap of the Pearson correlation coefficient matrix.

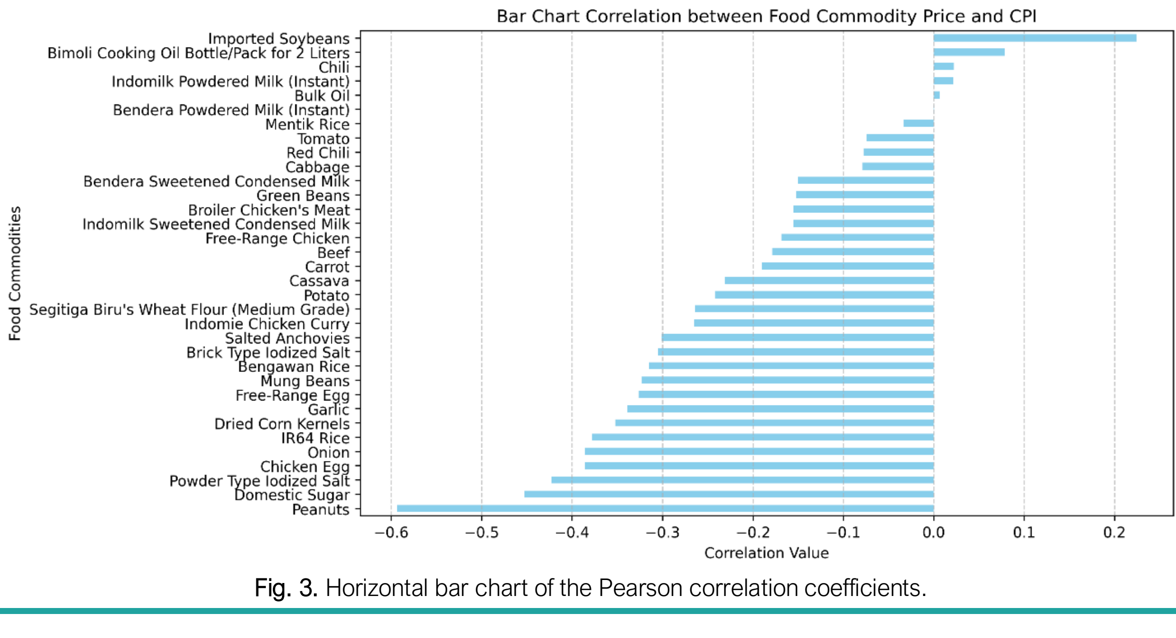


Fig. 3. Horizontal bar chart of the Pearson correlation coefficients.

Table 8
Hyperparameter tuning results using Grid Search Optimization for all data comparisons.

Hyperparameter	70%:30%	80%:20%	90%:10%
C	2.0	2.0	2.0
ε	1.0	1.0	1.0
γ	'scale'	'scale'	'scale'
Kernel	' rbf '	' rbf '	' rbf '

that an increase in the price of one commodity tends to be followed by a decrease in the price of another commodity and a decrease in the CPI. A light brown color (around 0) indicates no correlation or only a very weak correlation. Thus, the relationship pattern between commodity prices can be mapped more clearly, thus helping to understand the relationship between food price dynamics and CPI fluctuations.

When examined as a whole in Figs. 2 and 3, most commodities exhibit a negative correlation with the CPI. However, not all commodities follow this pattern. Several commodities exhibit a positive correlation, including soybean export/import, Bimoli Cooking Oil (2L), bird's eye chili, Indomilk Instant Milk Powder, and Bulk Cooking Oil. This positive correlation indicates that increases in the prices of these commodities tend to be followed by increases in the CPI.

In general, the result indicates a relatively weak correlation between food commodity prices and the CPI. The correlation coefficients obtained ranged from 0.000103 to 0.593384, falling into the very low, low, and moderate categories.

3.3.2. Prediction using Support Vector Regression

This study uses the kernel of Radial Basis Function (RBF). This kernel is selected based on empirical evidence that suggested the superiority of RBF which is consistently seen in both daily and monthly CPI data and applies to both training and testing data as reported in the study of Budiastuti, Nugroho, & Hariadi (2017).

The kernel is controlled by three main hyperparameters: Cost (C), Gamma (γ), and Epsilon (ε) (Budiastuti, Nugroho, & Hariadi, 2017). These parameters is important in shaping the characteristics of the models. The complexity of the kernel is controlled by Gamma, generalization capability to balance bias and variance is controlled by Cost, and the level of tolerance to prediction error is controlled by Epsilon. In this study, Cost is determined in the range of $C \in \{0.1, 0.5, 1.0, 1.5, 2.0\}$, Gamma is defined by $\gamma \in \{scale = \frac{1}{p\sigma^2}, auto = \frac{1}{p}\}$, where p refers to the number of variables and σ^2 represents varians of independent variable.

The hyperparameter tuning process was performed using Grid Search Optimization (GSO) combined with K-Fold Cross Validation ($k = 5$). To increase reliability, this procedure was repeated three times for each data ratio. After the correlation analysis stage, the study continued with the application of the SVR method to build a predictive model. The implementation was carried out in the Python programming language, utilizing several libraries as follows:

- SVR: Built using the sklearn.svm module from the scikit-learn library, which is used to model nonlinear relationships between input and output variables.
- GridSearchCV: Part of the sklearn.model_selection module, used to systematically search for the best hyperparameter combination based on a specified grid.
- Time: A standard Python library used to record model execution time (t0 to t1), so that computational duration can be measured and evaluated.

The stages of predicting the CPI value using SVR with an RBF kernel are as follows:

1) Identify optimal hyperparameters using GSO with K-Fold Cross Validation.
2) Experiment with various C values to tune the RBF-based SVR.
3) The best hyperparameters were obtained at $C = 2$, Gamma = scale, and Epsilon = 1.0, as shown in Table 8. The SVR implementation was performed with an RBF kernel using $C = 2$, Gamma = scale, and Epsilon = 1.0 based on the training and testing data.
4) Experiments were conducted on several data ratio scenarios: 90%:10%, 80%:20%, and 70%:30% (training:testing).

After implementation, the prediction results were evaluated to assess model performance.

3.3.3. Prediction using Polynomial Regression

One of the main challenges in Polynomial Regression (PR) is determining the appropriate order. PR can have up to six orders, where the order indicates the degree of the polynomial used to represent the relationship between the independent and dependent variables. In this study, the PR model was implemen-

Table 9
Interpretation of MAPE values.

MAPE	Interpretation
<10%	Highly accurate forecasting ability
10-20%	Moderately accurate forecasting ability
20-50%	Acceptably accurate forecasting ability
>50%	Inaccurate forecasting ability

ted using three different orders: Order 1, Order 2, and Order 3. Several Python libraries were used during the implementation process:

1) LinearRegression: The scikit-learn library (sklearn.linear_model) was used to create a simple linear regression model.
2) Time: The standard Python library, with the time.time() function, was used to record the start time (t0) and end time (t1) of the model execution.

The PR model was implemented using three different levels or orders: Order 1, Order 2, and Order 3. The order of the model was selected by gradually increasing the degree of the polynomial. The polynomial order used was chosen based on experimental results showing improved model performance on the training data. This process does not involve finding the optimal order through further model evaluation methods, but instead relies on analyzing model performance for each degree of the tested polynomial. Each order is trained separately using the training data. After training, the models are evaluated using the test data to determine the best performing model. This is followed by a visualization process comparing the actual data and the predicted data.

3.4. Prediction Evaluation

To evaluate the quality of the prediction models, three different metrics were used: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and MAPE (Singh, Singh, Gupta, Alotaibi, & Malik, 2025), as follows.

1) MSE calculates the average of the squared differences between actual and predicted values (Singh, Singh, Gupta, Alotaibi, & Malik, 2025). This metric indicates the overall magnitude of error.
2) RMSE is the square root of MSE (Ladjal, et al., 2025). It measures the difference between predicted and actual values in regression analysis (Geetha, Mala, Priyanka, & Uma, 2024). A smaller RMSE value indicates lower error rates and predictions closer to the actual values (Singh, Singh, Gupta, Alotaibi, & Malik, 2025). However, a good RMSE is not only reflected by small values but also by consistent results across training and testing datasets. A large gap between training and testing RMSE may indicate overfitting.
3) MAPE evaluates the model's accuracy by expressing the difference between predicted and actual values in percentage terms (Singh, Singh, Gupta, Alotaibi, & Malik, 2025). The interpretation of MAPE is presented in Table 9 (Chang, Wang, & Liu, 2007).

Among those three metrics, RMSE is used as a key indicator in model evaluation due to sensitiveness of the metric to high level errors (Singh, Singh, Gupta, Alotaibi, & Malik, 2025). RMSE gives a more clear picture of the model quality, especially in detecting outlier or high value of deviation. In this case, the smaller RMSE value is selected as a main indicator to determine the quality of the models.

3.5. Analysis of variance (ANOVA)

This section will present the results of statistical tests on all evaluation data using Analysis of Variance (ANOVA). In this test, two hypotheses are established:

- Null Hypotesis (H0): There is no significant difference in the evaluation results.
- Alernative Hypotesis (H1): There is significant difference in the evaluation results.

Decisions are made based on the p-value. If the p-value is less than the significance level ($\alpha = 0.05$). If H0 is rejected, it can be concluded that there is a significant difference between the values produced by the evaluation method. Conversely, if the p-value is greater than α, hen H0 is rejected and it can be concluded that there is no significant difference.

The steps for the ANOVA test to compare model performance based on MSE, RMSE, and MAPE are as follows:

- Organizing Data: Organizing data from the evaluation results table into a format ready for ANOVA analysis.
- Performing the ANOVA Test: A one-way ANOVA test is applied to each metric (MSE, RMSE, MAPE) to assess whether there are significant differences between the models.
- Results and Interpretation: Presenting the results of the statistical test and providing an interpretation

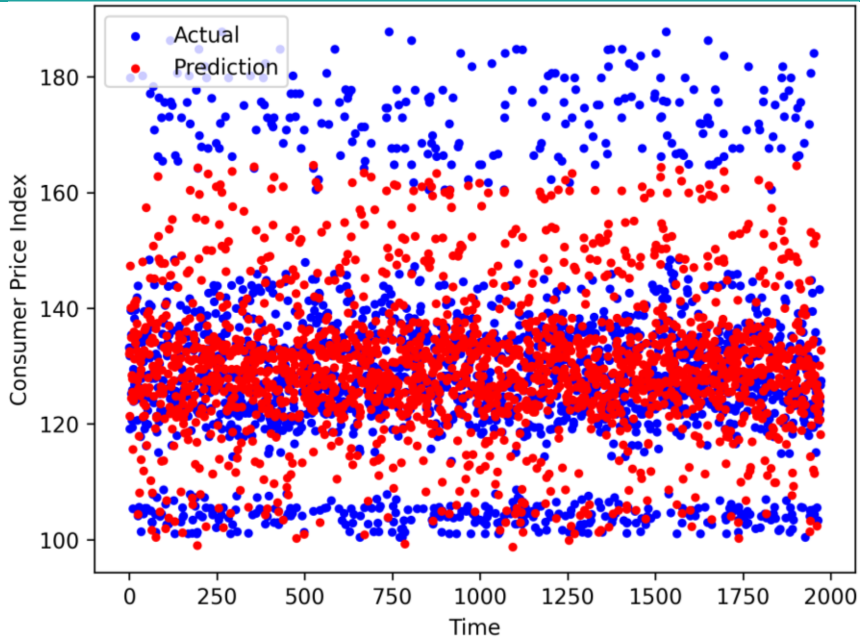


Fig. 4. Comparison between actual and predicted values using a 90%:10% train–test split.

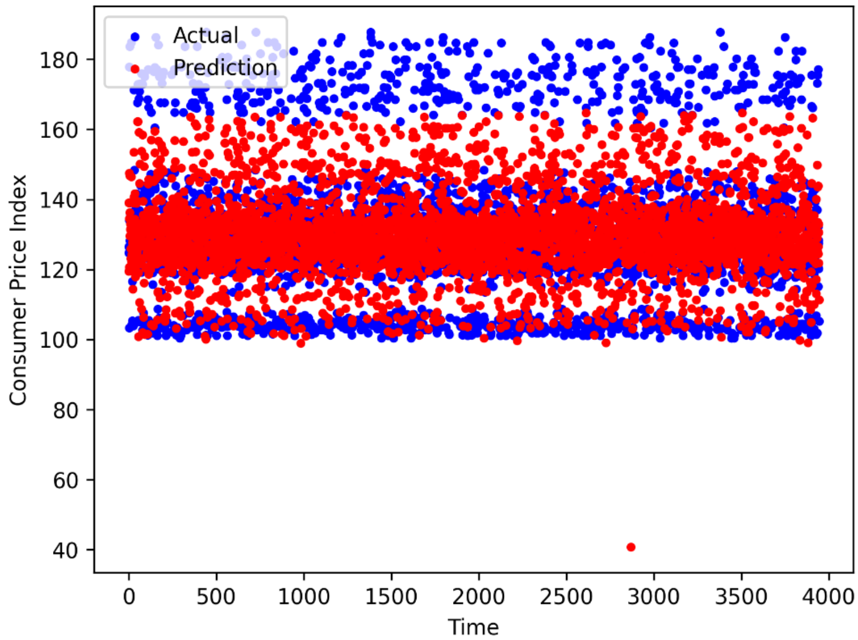


Fig. 5. Comparison between actual and predicted values using an 80%:20% train–test split.

of any significant differences, if found.

4. Results and Discussion

4.1. Prediction of the CPI value using SVR

Figs. 4, 5, and 6 show the comparison between actual and predicted values. The blue dots represent the actual values, while the red dots represent the predicted values. The performance evaluation of the SVR algorithm is presented in Table 10. The CPI predicted by the SVR model is then compared with the actual data published by BPS to assess the model's accuracy. A detailed comparison between the actual and predicted data is provided in Table 11.

Based on the evaluation results in Table 10, model performance tends to decline as the training data proportion decreases. Specifically, the Mean Squared Error (MSE) increases with decreasing training data. For example, the MSE values for training data proportions of 90%, 80%, and 70% are 137.025, 141.589, and 147.679, respectively. Higher MSE values indicate greater prediction errors.

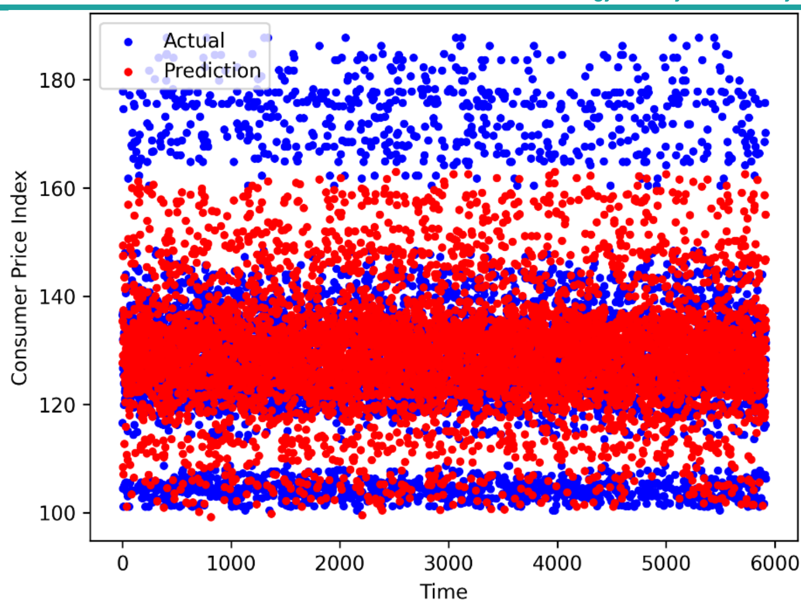


Fig. 6. Comparison between actual and predicted values using a 70%:30% train–test split.

Table 10
SVR evaluation results with different train–test splits.

Performance Indicators	Train 90%	Test 10%	Train 80%	Test 20%	Train 70%	Test 30%
MSE	137.025	139.775	141.589	146.540	147.679	148.002
RMSE	11.705	11.822	11.899	12.105	12.152	12.165
MAPE	0.059	0.060	0.060	0.061	0.061	0.061
Compile Time (s)	740.403	5.229	582.651	4.657	443.349	9.167

Table 11
Comparison of actual and predicted values using SVR.

City/District	Time	Actual Value	Predicted		
			90%:10%	80%:20%	70%:30%
Jember	January 2014	164.8	**159.0448**	157.5187	156.5632
Jember	February 2014	168.53	**160.9732**	154.3498	158.6616
Jember	March 2014	172.92	**157.3313**	155.8586	155.7209
Jember	April 2014	171.31	**154.6495**	153.2031	152.9014
Jember	May 2014	167.62	**153.1426**	151.6543	150.7115
Jember	June 2014	168.5	**150.015**	148.5763	147.5757
Jember	July 2014	177.12	**146.0624**	144.3325	143.1529
⋮	⋮	⋮	⋮	⋮	⋮
Surabaya	December 2020	104.61	126.8231	**122.6174**	122.9898

Further evaluation using RMSE shows a similar pattern. The RMSE values for training data proportions of 90%, 80%, and 70% are 11.705, 11.899, and 12.152, respectively. MAPE follows a similar pattern, with values of 0.059, 0.060, and 0.061 for each training data proportion. Based on these three metrics, it can be concluded that the SVR model performs better with a larger training data proportion.

In terms of computational efficiency, SVR exhibits relatively long processing times. The processing time increases with the increase in the training data proportion. The training process with 90% training data took 740.403 seconds, while the testing process took 5.229 seconds. Experiments with 80% and 70% training data proportions resulted in training times of 582.651 seconds and 443.349 seconds, respectively, with testing times of 4.657 seconds and 9.167 seconds. This indicates that a smaller training data proportion results in faster processing times, but with lower prediction accuracy.

4.2. Prediction of the CPI value using Polynomial Regression

The Polynomial Regression (PR) evaluation process begins with the Order 1 model. The comparison results between actual and predicted values are shown in Table 12 and Fig. 7. Next, experiments were conducted with the Order 2 PR model, which was evaluated using MSE, RMSE, and MAPE. The evaluation results are then visualized in Table 13 and Fig. 8. The final experiment used Order 3 PR, with a visualization of the comparison between actual and predicted data shown in Table 14 and Fig. 9.

Table 12
Comparison of actual and predicted values using PR (order 1).

City/District	Time	Actual Value	Predicted Value
Jember	January 2014	164.8	157.0354
Jember	February 2014	168.53	162.6269
Jember	March 2014	172.92	162.8232
Jember	April 2014	171.31	164.7891
Jember	May 2014	167.62	165.3217
Jember	June 2014	168.5	162.1291
Jember	July 2014	177.12	160.6939
⋮	⋮	⋮	⋮
Surabaya	December 2020	104.61	120.844

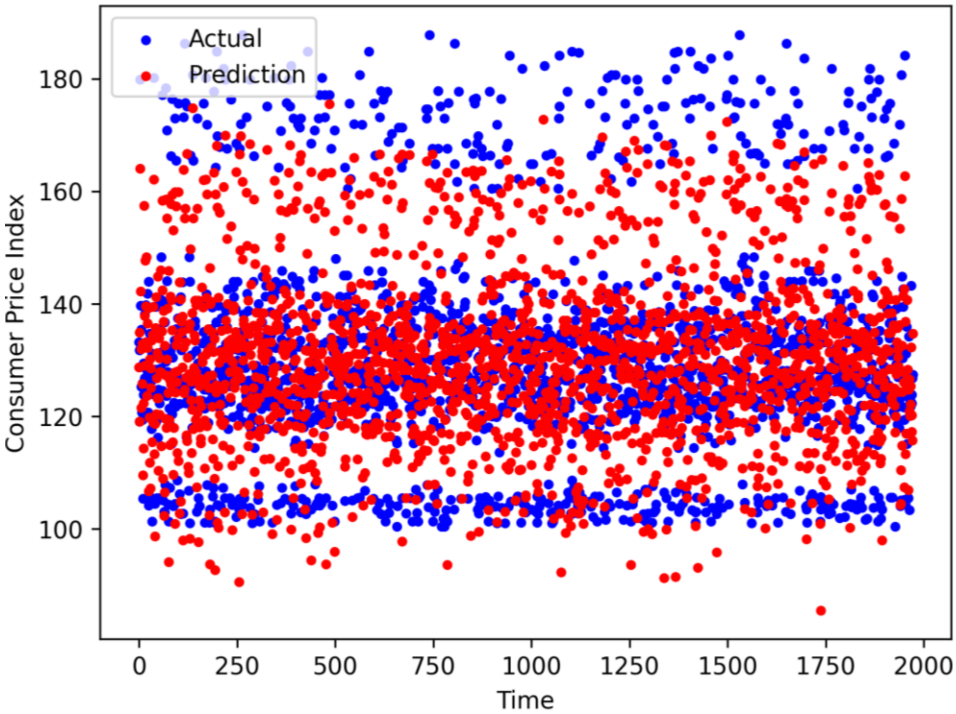


Fig. 7. Comparison of actual and predicted values using PR (order 1).

Table 13
Comparison of actual and predicted values using PR (order 2).

City/District	Time	Actual Value	Predicted Value
Jember	January 2014	164.8	157.0345
Jember	February 2014	168.53	162.6269
Jember	March 2014	172.92	162.8232
Jember	April 2014	171.31	164.7891
Jember	May 2014	167.62	165.3217
Jember	June 2014	168.5	162.1291
Jember	July 2014	177.12	160.6939
⋮	⋮	⋮	⋮
Surabaya	December 2020	104.61	103.310

Based on the results of the PR model implementation for CPI with Orders 1, 2, and 3, the performance evaluation is summarized in Table 15. In the first experiment using Order 1, the MSE values for the training and test data were 124,454 and 118,690, respectively. The RMSE values were 11,155 and 10,894, indicating relatively high error. Evaluation using MAPE yielded 0.064 (6.4%) for the training data and 0.063 (6.3%) for the test data, also indicating limited accuracy.

Experiments with Order 2 showed a significant performance improvement compared to Order 1. The MSE values for the training and test data were 17,548 and 50,801, respectively, while the RMSE values were 4,189 and 7,127. Despite the improvements, the prediction error was still far from zero. The MAPE value also increased, namely 0.021 (2.1%) for training data and 0.024 (2.4%) for testing data.

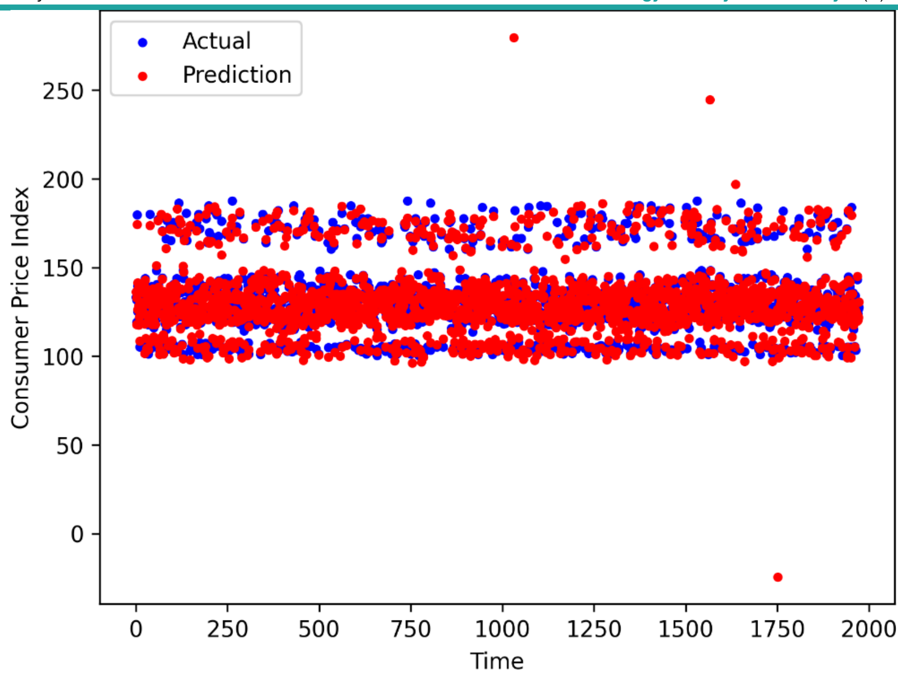


Fig. 8. Comparison of actual and predicted values using PR (order 2).

Table 14
Comparison of actual and predicted values using PR (order 3).

City/District	Time	Actual Value	Predicted Value
Jember	January 2014	164.8	157.0345
Jember	February 2014	168.53	162.6269
Jember	March 2014	172.92	162.8232
Jember	April 2014	171.31	164.7891
Jember	May 2014	167.62	165.3217
Jember	June 2014	168.5	162.1291
Jember	July 2014	177.12	160.6939
⋮	⋮	⋮	⋮
Surabaya	December 2020	104.61	104.67

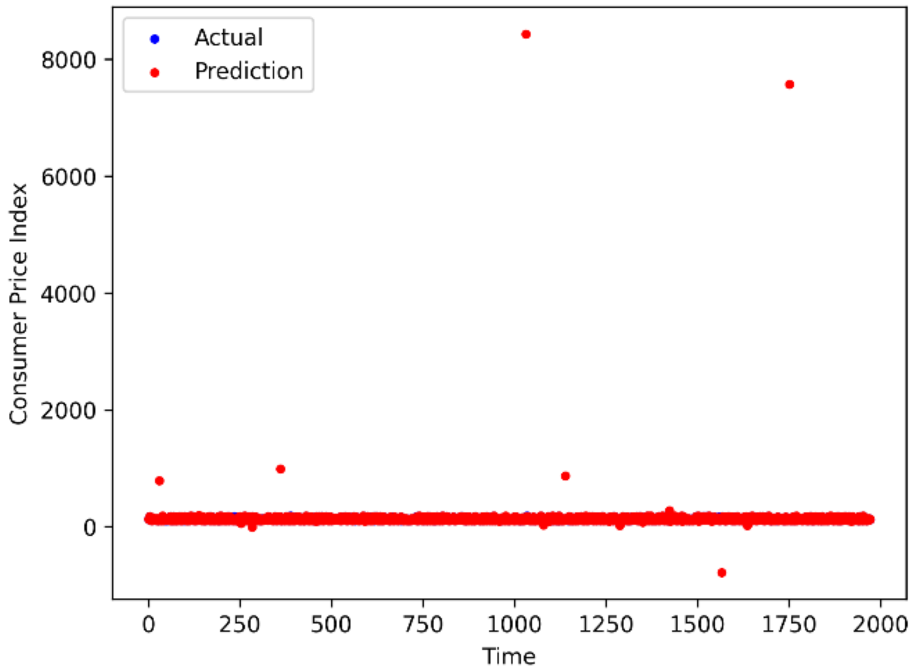


Fig. 9. Comparison of actual and predicted values using PR (order 3).

Table 15
PR model evaluation results using MSE, RMSE, and MAPE.

Performance Indicators	Order 1		Order 2		Order 3	
	Train	Test	Train	Test	Train	Test
MSE	124.454	118.690	17.548	50.801	2.136	3,530.197
RMSE	11.155	10.894	4.189	7.127	1.461	59.415
MAPE	0.064	0.063	0.021	0.024	0.003	0.034
Compile time	0.044	0.004	1.817	0.041	144.591	0.220

Table 16
Comparison of SVR and PR order 3 results.

Performance Indicators	SVR		PR Order 2	
	Train	Test	Train	Test
MSE	137.025	139.775	2.136	3530.197
RMSE	11.705	11.822	1.461	59.415
MAPE	0.059	0.060	0.003	0.034
Compile time	740.403	5.229	144.591	0.220

Table 17
Statistical test results on evaluation data using the ANOVA test.

Performance Indicators	F-Statistic	P-Value
MSE	0.877	0.548
RMSE	0.514	0.759
MAPE	10.938	0.006

The best performance was achieved with Order 3. The MAPE values for the training and test data were 0.003 (0.3%) and 0.034 (3.4%), respectively. A comparison between the predicted results and the actual data, as seen in Fig. 9, shows nearly perfectly parallel lines. This confirms that the Polynomial Regression model achieved optimal performance at Order 3.

Further, the model was also evaluated based on processing time (Compile Time). For Order 1, the processing time was 0.044 seconds for the training data and 0.004 seconds for the test data. For Order 2, the processing time was 1.817 seconds for the training data and 0.041 seconds for the test data, respectively. Meanwhile, for Order 3, the processing time increased significantly to 144.591 seconds for the training data and 0.220 seconds for the test data, reflecting the higher model complexity.

4.3. Comparison of the CPI value prediction using SVR and PR

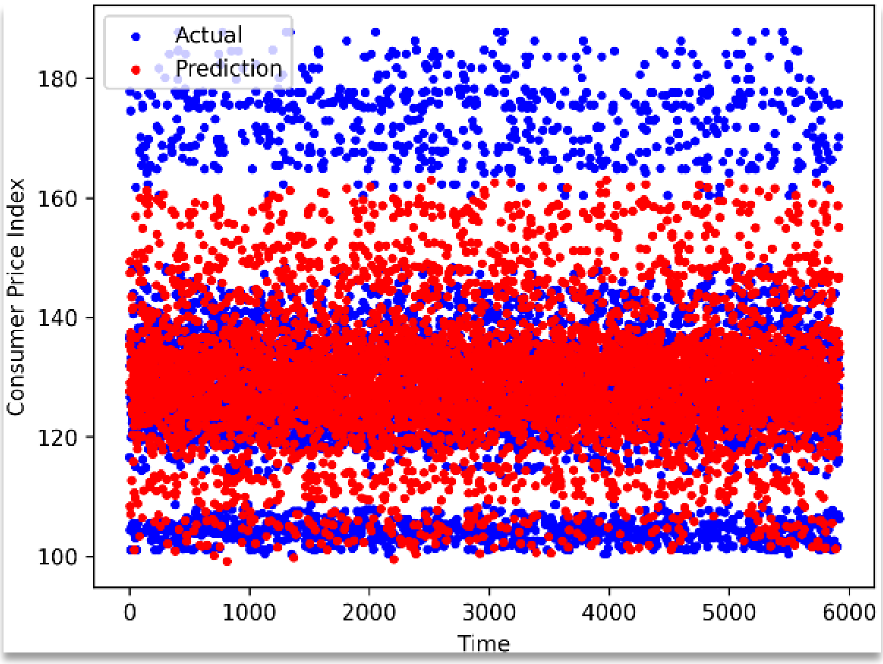
Based on the test results using various performance indicators, both models demonstrated MAPE values below 10% which indicates high model reliability. This indicates that both SVR and Polynomial Regression (PR) were successfully applied to CPI prediction.

The SVR model was implemented using an RBF kernel shows the best performance at a train-test split ratio of 90%:10%. The MAPE values obtained were 5.9% for the training data and 6% for the test data. In a condition where the PR model was tested with three orders, it is shown that the best performance was achieved at Order 3, resulting in a MAPE of 0.3% for the training data and 3.4% for the test data. A comparison of the performance of the two models presented in Table 16 shows that PR outperforms SVR, with a lower MAPE value approaching zero.

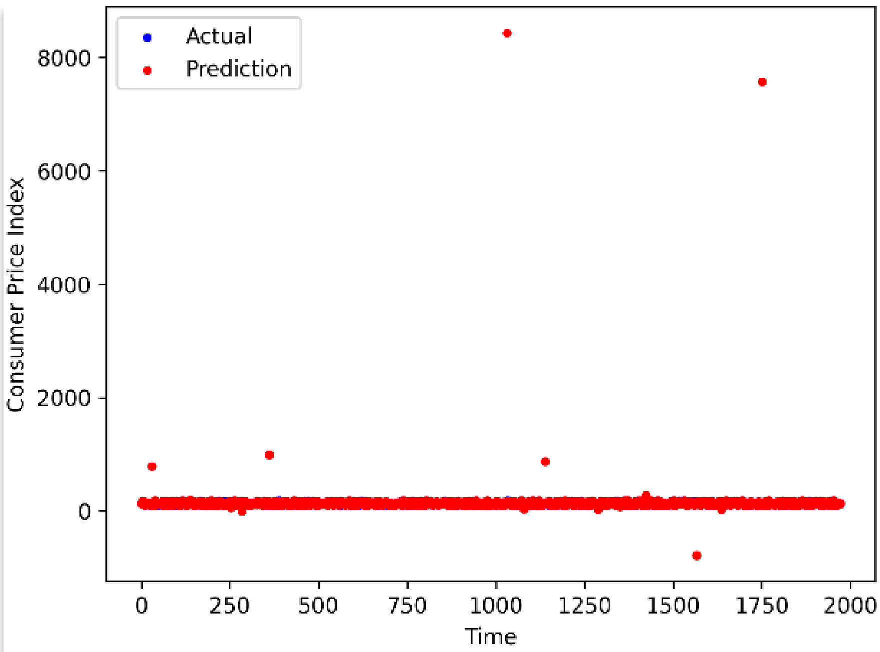
The assessment of the implementation speed was conducted by comparing both SVR and PR models. SVR required 740.403 seconds for the training data and 5.229 seconds for the test data, while PR required 144.591 seconds to process the training data and 0.220 seconds for the testing data. Overall, PR Order 3 outperformed SVR in terms of accuracy and processing efficiency.

The evaluation results were also conducted using ANOVA test (see Table 17). The statistical results showed no significant difference in MSE and RMSE between the models (p-value >0.05) even though there was a significant difference in MAPE with PR especially with higher orders, indicates a better performance in PR.

A comparison of the prediction results between SVR and 3rd-order PR shows a significant difference with a better performance of SVR (Fig. 10), indicated by a low error rates (Putranto, Kholik, Nugroho, & Kriestanto, 2023). However, in CPI prediction, SVR produced a higher error rate than 3rd-order PR. This is because SVR requires variables with strong correlations, while this study used 34 commodities with low to moderate correlations. The problem of perfect multicollinearity is one reason PR does not always produce satisfactory results in practice (Marsh & Cormier, 2001). However, even though the correlation between va-



(a)



(b)

Fig. 10. Comparison of results: (a) SVR and (b) PR order 3 method.

riables was far from ±1, the 3rd-order PR model still achieved a low error rate, suggesting that, while PR has limitations and does not always guarantee reliable predictions, this model is capable of producing reliable predictions when the polynomial degree is increased.

Increasing the polynomial degree does increase model complexity, but a complex model does not always guarantee generalizability, as error minimization may only be applicable to the current dataset. Various previous studies have explored time series, machine learning, and deep learning methods for predicting the CPI. The results of this study indicate that both the 3rd-order PR and SVR perform better than previous methods. In particular, the 3rd-order PR model is capable of producing more accurate and reliable predictions, even on datasets with high complexity and low correlation. With a lower MAPE approaching zero, this study makes a significant contribution to CPI prediction, overcoming the limitations of methods used in previous studies.

5. Conclusions

This study aims to compare the performance of PR and SVR in predicting the CPI based on food commodity prices obtained from SISKAPERBAPO managed by the government of East Java. This topic is important because CPI prediction plays a strategic role in understanding regional economic dynamics and supporting data-driven decision-making in the food sector.

Using data cleaning to address missing values, Pearson's r correlation analysis, and model performance evaluation using MSE, RMSE, MAPE, and computational time metrics, this study presents a measurable comparison between PR and SVR. The results indicate that PR (order 3) is more accurate than SVR, with a MAPE of 0.3% on the training data and 3.4% on the testing data, as well as better computational efficiency. In contrast, SVR performs less well, with a MAPE of 5.9% on the training data and 6% on the testing data.

The implication of this study is that the PR method can be a more reliable and practical alternative for predictive analysis of the CPI based on food data. However, this study still has limitations, primarily because it failed to consider other control variables that also influence the CPI, such as income levels, fiscal and monetary policies, exchange rate fluctuations, production costs, and seasonal factors. Therefore, future research is recommended to incorporate these variables to provide more comprehensive predictions and provide a stronger contribution to economic policymaking at the regional and national levels.

6. Declaration of AI and AI assisted technologies in the writing process

During the preparation of this work, the author used ChatGPT to assist in improving readability through proofreading. After using this tool, the author reviewed, verified, and edited the content as necessary, and takes full responsibility for the accuracy and integrity of the final publication.

7. CRediT Authorship Contribution Statement

Ayu Adelina Suyono: Conceptualization; Methodology; Data curation; Formal Analysis; Writing - Original Draft; Writing - Review & Editing; Visualization; Supervision.

8. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

9. Acknowledgments

The authors would like to thank the anonymous referees for their helpful comments and suggestions.

10. Data Availability

The daily food commodity price data are available on the official website of Siskaperbapo (Food Commodity Price Information System, East Java Province) at https://www.siskaperbapo.jatimprov.go.id, while the Consumer Price Index (CPI) data are available on the official website of Statistics Indonesia, East Java Province, at https://jatim.bps.go.id.

11. References

Aida, N. R., & Nugroho, R. S. (2022, 3 1). *Update Harga Pangan 1 Maret 2022: Telur, Cabai, Minyak Goreng hingga Gula Pasir Naik*. Retrieved 8 21, 2025, from Kompas.com: https://www.kompas.com/tren/read/2022/03/01/143000165/update-harga-pangan-1-maret-2022--telur-cabai-minyak-goreng-hingga-gula

Altair RapidMiner. (2025, 9 5). *Polynomial Regression (AI Studio Core).* Retrieved from Altair RapidMiner: https://docs.rapidminer.com/2024.0/studio/operators/modeling/predictive/functions/polynomial_regression.html

Ashoka, D. V., Rekha, P. M., & Sudha, P. R. (2024). Smart Power Management in Data Centers Using Machine-Learning Techniques. In R. Nagarajan, S. K. Narayanasamy, R. Thirunavukarasu, & P. Raj, *Intelligent Systems and Sustainable Computational Models: Concepts, Architecture, and Practical Applications* (pp. 1-13). New York, US: CRC Press. doi:https://doi.org/10.1201/9781003407959

Asmaradana, A. A., & Widodo, E. (2023). Penerapan Metode Peramalan Double Exponential Smoothing pada Indeks Harga Konsumen Kota Yogyakarta. *Emerging Statistics and Data Science Journal (ESDS)., 1*(1), 30-36. Retrieved 8 24, 2025, from https://journal.uii.ac.id/esds/article/download/27021/

14715/84183

Badan Pusat Statistik. (2020). *Indeks Harga Konsumen di 82 Kota di Indonesia (2012=100) 2019.* Jakarta, Indonesia: Badan Pusat Statistik. Retrieved 9 5, 2025, from https://www.bps.go.id/id/publication/2020/04/09/91a62fdd238c2b440752e161/indeks-harga-konsumen-di-82-kota-di-indonesia--2012-100--2019.html

Badan Pusat Statistik. (2024). *Diagram Timbang Indeks Harga Konsumen Hasil Survei Biaya Hidup 2022* (Vol. 8). Jakarta, Indonesia: Badan Pusat Statistik. Retrieved 9 5, 2025, from https://www.bps.go.id/id/publication/2024/02/05/f164320774fe896914f6fe1e/diagram-timbang-indeks-harga-konsumen-hasil-survei-biaya-hidup-2022.html

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical Care: Sepsis and Severe Infection, 7*(6), 451–459. doi:https://doi.org/10.1186/cc2401

BPS Provinsi Jawa Timur. (2021). *Indeks Harga Konsumen 8 Kota di Provinsi Jawa Timur 2020.* Surabaya: BPS Provinsi Jawa Timur. Retrieved 8 24, 2025, from https://jatim.bps.go.id/id/publication/2021/04/23/4932c8cfa622e3c2618e04d8/indeks-harga-konsumen-8-kota-di-provinsi-jawa-timur-2020.html

BPS Provinsi Jawa Timur. (2024, 5 2). *April 2024 inflasi Year on Year (y-on-y) Provinsi Jawa Timur sebesar 3,25 persen. Inflasi tertinggi terjadi di Sumenep sebesar 4,87 persen*. Retrieved from Badan Pusat Statistik Provinsi Jawa Timur: https://jatim.bps.go.id/id/pressrelease/2024/05/02/1405/april-2024-inflasi-year-on-year--y-on-y--provinsi-jawa-timur-sebesar-3-25-persen--inflasi-tertinggi-terjadi-di-sumenep-sebesar-4-87-persen-.html

Budiastuti, I. A., Nugroho, S. M., & Hariadi, M. (2017). Predicting daily consumer price index using support vector regression method. *2017 15th International Conference on Quality in Research (QiR) : International Symposium on Electrical and Computer Engineering.* Nusa Dua, Bali, Indonesia: IEEE. doi:https://doi.org/10.1109/QIR.2017.8168445

Bukaita, W., Celis, G. G., & Gurram, M. (2024). Training-Testing Data Ratio Selection for Accurate Time Series Forecasting: A COVID-19 Case Study. *Proceedings of the Future Technologies Conference (FTC) 2024. 3*, pp. 227–246. Springer. doi:https://doi.org/10.1007/978-3-031-73125-9_14

Chang, P.-C., Wang, Y.-W., & Liu, C.-H. (2007). The development of a weighted evolving fuzzy neural network for PCB sales forecasting. *Expert Systems with Applications, 32*(1), 86-96. doi:https://doi.org/10.1016/j.eswa.2005.11.021

Chen, J.-F., Do, Q. H., Nguyen, T. V., & Doan, T. T. (2018). Forecasting Monthly Electricity Demands by Wavelet Neuro-Fuzzy System Optimized by Heuristic Algorithms. *Information, 9*(3). doi:https://doi.org/10.3390/info9030051

Cybellium. (2023). *Mastering Probability and Statistics: A Comprehensive Guide to Learn Probability and Statistics.* Cybellium.

Delmo, J. A., Villarica, M. V., & Vinluan, A. A. (2022). Classification of Coffee Variety using Electronic Nose. *2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA). 18*, pp. 248-253. Selangor, Malaysia: IEEE. doi:https://doi.org/10.1109/CSPA55076.2022.9782056

Editya, A. S., Kurniati, N., Septianto, T., Lisdiyanto, A., & Al Haromainy, M. M. (2021). Effect of Kernel in Support Vector Regression Method to Predict Surabaya Consumer Price Index Trend. *2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE).* Purwokerto, Indonesia: IEEE. doi:https://doi.org/10.1109/ICITISEE53823.2021.9655891

Geetha, A. V., Mala, T., Priyanka, D., & Uma, E. (2024). Multimodal Emotion Recognition with Deep Learning: Advancements, challenges, and future directions. *Information Fusion, 105*. doi:https://doi.org/10.1016/j.inffus.2023.102218

Gogtay, N. J., & Thatte, U. (2017). Principles of Correlation Analysis. *Journal of The Association of Physicians of India, 65*(3), 78-81. Retrieved from https://www.kem.edu/wp-content/uploads/2012/06/9-Principles_of_correlation-1.pdf

Haoyuan, S., Yizhong, M., Chenglong, L., Jian, Z., & Lijun, L. (2023). Hierarchical Bayesian support vector regression with model parameter calibration for reliability modeling and prediction. *Reliability Engineering & System Safety, 229*. doi:https://doi.org/10.1016/j.ress.2022.108842

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences.* Houghton Mifflin. Retrieved 9 6, 2025, from https://library.wur.nl/WebQuery/titel/1944963

Hsia, J.-Y., & Lin, C.-J. (2020). Parameter Selection for Linear Support Vector Regression. *IEEE Transactions on Neural Networks and Learning Systems, 31*(12), 5639-5644. doi:https://doi.org/10.1109/TNNLS.2020.2967637

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: An ASA Data Science Journal, 15*(4), 531-538. doi:https://doi.org/10.1002/sam.11583

Kumar, S., & Chong, I. (2018). Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States. *International Journal of Environmental Research and Public Health, 15*(12), 2907. doi:https://doi.org/10.3390/ijerph15122907

Ladjal, B., Nadour, M., Bechouat, M., Hadroug, N., Sedraoui, M., Rabehi, A., . . . Agajie, T. F. (2025). Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: a study on solar potential in Ghardaia, Algeria. *Scientific Reports, 15*. doi:https://doi.org/10.1038/s41598-025-94239-z

Larasati, R., Susilorini, T. E., Surjowardojo, P., & Wahyuni, R. D. (2024). Correlation of Linear Body Size With Body Condition Score and Body Weight of Participated Cow Used in the Progeny Test in East Java. *Animal Production : Indonesian Journal of Animal Production, 26*(2), 70-81. doi:https://doi.org/10.20884/1.jap.2024.26.2.298

Marsh, L. C., & Cormier, D. R. (2001). *Spline Regression Models.* Sage.

Mata, D. A., & Milner, D. A. (2021). Statistical Methods in Experimental Pathology: A Review and Primer. *The American Journal of Pathology, 191*(5), 784-794. doi:https://doi.org/10.1016/j.ajpath.2021.02.009

Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal, 24*(3), 69–71. Retrieved 9 6, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC3576830/

Myśliwiec, P., Szawara, P., Kubit, A., Zwolak, M., Ostrowski, R., Derazkola, H. A., & Jurczak, W. (2025). FSW Optimization: Prediction Using Polynomial Regression and Optimization with Hill-Climbing Method. *Materials, 18*(2), 448. doi:https://doi.org/10.3390/ma18020448

Ostertagová, E. (2012). Modelling using Polynomial Regression. *Procedia Engineering, 48*, 500-506. doi:https://doi.org/10.1016/j.proeng.2012.09.545

Park, K., Rothfeder, R., Petheram, S., Buaku, F., Ewing, R., & Greene, W. H. (2020). Linear Regression. In *Basic Quantitative Research Methods for Urban Planners.* Taylor & Francis.

Prakoso, B. H. (2019). Implementasi Support Vector Regression pada Prediksi Inflasi Indeks Harga Konsumen. *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer, 19*(1), 155-162. doi:https://doi.org/10.30812/matrik.v19i1.511

Pramesti, S. A., Sadikin, U. A., Imro'ah, N., & Maulida, F. (2023). Prediksi Indeks Harga Konsumen Kota Pontianak Menggunakan Metode Double Exponential Smoothing dan Analysis Trend. *Equator: Journal of Mathematical and Statistical Sciences, 2*(2), 37-47. Retrieved 8 24, 2025, from https://jurnal.untan.ac.id/index.php/EMSS/article/view/73302

Putranto, B. P., Kholik, M. A., Nugroho, M. A., & Kriestanto, D. (2023). Polynomial Regression Method and Support Vector Machine Method for Predicting Disease Covid-19 in Indonesia. *Journal of Intelligent Software System, 2*(1), 18-23. doi:http://dx.doi.org/10.26798/jiss.v2i1.931

R, I., Sudarmin, S., & Rais, Z. (2022). Analisis Support Vector Regression (SVR) dengan Kernel Radial Basis Function (Rbf) untuk Memprediksi Laju Inflasi di Indonesia. *Variansi: Journal of Statistics and Its Application on Teaching and Research, 4*(1), 30-38. Retrieved 8 24, 2025, from https://jurnalvariansi.unm.ac.id/index.php/variansi/article/view/13

Rahmadi, D. R. (2018, 5 7). *Lessons dari Jawa Timur*. Retrieved 8 21, 2025, from Fakultas Ekonomi dan Bisnis Universitas Brawijaya: https://feb.ub.ac.id/lessons-dari-jawa-timur/

Rahmawati, R. (2020). *Implementasi Algoritme Support Vector Regression dan Particle Swarm Optimization dalam Peramalan Penjualan.* Bandung, Indonesia: Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI. Retrieved 8 24, 2025, from https://library.likmi.ac.id/book/109

Rohmah, M. F., Ardiantoro, L., Putra, I. K., & Hartati, R. S. (2019). Meramal Indeks Harga Konsumen Kabupaten di Jawa Timur dengan Metode Support Vector Regression Data Mining. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)* (pp. C-30 – C-36). Yogyakarta, Indonesia: Universitas Islam Indonesia. Retrieved 8 24, 2025, from https://journal.uii.ac.id/Snati/article/view/13434

Salkind, N. J. (2010). *Statistics for People who (think They) Hate Statistics: Excel 2007 Edition.* SAGE Publications.

Sari, D. N., Sasmito, B., & Hadi, F. (2023). Estimasi Produktivitas Kopi Menggunakan Citra SPOT-7 Dengan Transformasi Indeks Vegetasi. *Jurnal Geodesi Undip (JGU), 12*(1), 20-29. Retrieved 8 24, 2025, from https://ejournal3.undip.ac.id/index.php/geodesi/article/view/36910

Sari, L. P., Muslihah, M., Mutohari, R., & Sari, R. N. (2025). Dampak Krisis Keuangan Global Terhadap Pertumbuhan Ekonomi di Indonesia. *Moneter: Jurnal Ekonomi dan Keuangan, 3*(1), 90-103. doi:https://doi.org/10.61132/moneter.v3i1.1094

Singh, U., Singh, S., Gupta, S., Alotaibi, M. A., & Malik, H. (2025). Forecasting rooftop photovoltaic solar power using machine learning techniques. *Energy Reports, 13*, 3616-3630. doi:https://doi.org/10.1016/j.egyr.2025.03.005

Soyer, P. (2017). Misuse of semantics and basic statistical terms in original articles. *Diagnostic and Interventional Imaging, 98*(12), 825-826. doi:https://doi.org/10.1016/j.diii.2017.11.006

Vanitha, G., & Kasthuri, M. (2024). *Basic Guide for Machine Learning Algorithms and Models.* Madurai, India: SK Research Group of Companies.

Wisniewski, S. J., & Brannan, G. D. (2024). *Correlation (Coefficient, Partial, and Spearman Rank) and Regression Analysis.* StatPearls Publishing. Retrieved 9 6, 2025, from https://www.ncbi.nlm.nih.gov/sites/books/NBK606101/

Xie, M., Xie, L., & Zhu, P. (2021). An Efficient Feature Weighting Method for Support Vector Regression. *Mathematical Problems in Engineering*. doi:https://doi.org/10.1155/2021/6675218

Zhang, F., & O'Donnell, L. J. (2020). Chapter 7 - Support vector regression. In A. Mechelli, & S. Vieira, *Machine Learning: Methods and Applications to Brain Disorders* (pp. 123-140). London, United Kingdom: Elsevier & Academic Press. doi:https://doi.org/10.1016/B978-0-12-815739-8.00007-9

Zhou, W., Yan, Z., & Zhang, L. (2024). A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction. *Scientific Reports, 14*. doi:https://doi.org/10.1038/s41598-024-55243-x