

# Comparative Analysis of ARIMAX, Support Vector Regression, and Linear Regression for Rice Price Prediction Using Weather-Based Time Series Features

Fatimah Tuzahra<sup>1\*</sup>, Marcelinus Yosep Teguh Sulistyono<sup>2\*</sup>

\* Sistem Informasi, Universitas Dian Nuswantoro, Semarang  
[112202206899@mhs.dinus.ac.id](mailto:112202206899@mhs.dinus.ac.id)<sup>1</sup>, [teguh.sulistyono@dsn.dinus.ac.id](mailto:teguh.sulistyono@dsn.dinus.ac.id)<sup>2</sup>

## Article Info

### Article history:

Received 2026-01-27

Revised 2026-04-24

Accepted 2026-04-29

### Keyword:

Rice Price Prediction,  
Support Vector Regression,  
Linear Regression,  
ARIMAX,  
Time Series,  
Machine Learning.

## ABSTRACT

Rice prices are an important indicator of economic stability and food security in Indonesia. Rice price fluctuations are influenced by various factors, including weather conditions, making accurate predictive models necessary to capture data patterns effectively. This study aims to compare the performance of ARIMAX, Support Vector Regression (SVR), and Linear Regression (LR) methods in predicting medium and premium rice prices using a time series approach with rainfall as an external variable. The data used consist of daily rice price data from 2023–2025 obtained from the Department of Industry and Trade of Central Java Province and rainfall data from NASA POWER. The study applies preprocessing, feature lag, and time-based validation (time-based split), while also using Naïve Forecasting as a baseline model. Model evaluation was conducted using MAE, RMSE, and R<sup>2</sup> Score. The results show that Naïve Forecasting achieved the best performance for both commodities. Among the proposed models, SVR demonstrated the best performance for medium rice with an MAE of 26.03 Rupiah/kg, RMSE of 33.57 Rupiah/kg, and R<sup>2</sup> of 0.85, while Linear Regression showed relatively good performance for premium rice with an MAE of 42.54 Rupiah/kg, RMSE of 62.74 Rupiah/kg, and R<sup>2</sup> of 0.92. Meanwhile, ARIMAX did not achieve optimal performance. The Wilcoxon Signed-Rank Test results indicate that the performance differences among models are statistically significant (p-value < 0.001). This study contributes through the implementation of a comparative time series evaluation framework that integrates feature lag, external variables, time-based validation, and baseline forecasting for food commodity price prediction.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Harga beras merupakan indikator penting dalam stabilitas ekonomi nasional dan kesejahteraan masyarakat, terutama di negara agraris seperti Indonesia. Selain sebagai pangan pokok, beras juga berperan dalam menjaga ketahanan pangan dan memengaruhi tingkat inflasi nasional. [1] Fluktuasi harga beras yang tidak terkendali dapat berdampak langsung pada daya beli masyarakat, khususnya kelompok berpendapatan rendah, serta berpotensi menimbulkan ketidakseimbangan dalam distribusi pangan secara nasional. [2] Oleh karena itu, pemahaman terhadap dinamika perubahan harga beras menjadi sangat penting sebagai dasar

dalam merumuskan kebijakan perdagangan dan pengelolaan produksi pangan yang lebih efektif. Perubahan harga beras dipengaruhi oleh berbagai faktor, seperti kondisi cuaca, produksi pertanian, distribusi, serta dinamika permintaan dan penawaran. Kompleksitas faktor tersebut menyebabkan pola harga menjadi dinamis dan sulit diprediksi secara konvensional. Oleh karena itu, pendekatan machine learning seperti Support Vector Regression (SVR) dan Linear Regression (LR) digunakan untuk meningkatkan akurasi prediksi. [3] Beberapa penelitian terdahulu telah dilakukan oleh [4] untuk memprediksi data berbasis *time series*. Penelitian berjudul Comparative Study of Support Vector

Regression and Long Short-Term Memory for Stock Price Prediction membandingkan metode SVR dan LSTM dalam memprediksi harga saham. Hasil penelitian menunjukkan bahwa SVR lebih unggul dalam prediksi jangka pendek dengan tingkat error yang lebih rendah, sedangkan LSTM lebih baik dalam menangkap pola tren jangka menengah sehingga menghasilkan prediksi yang lebih stabil pada periode yang lebih panjang.

Penelitian lain oleh Ayu Adelina Suyono dkk [5] membahas prediksi Indeks Harga Konsumen (IHK) komoditas makanan menggunakan metode SVR dengan data 34 komoditas pada periode 2014–2020. Tujuan penelitian tersebut adalah membangun model prediksi IHK yang akurat. Hasil penelitian menunjukkan bahwa model SVR dengan kernel polynomial memberikan performa terbaik dengan nilai Mean Absolute Percentage Error (MAPE) sebesar 4,31% yang termasuk kategori sangat baik.

Selain itu, penelitian lain membandingkan beberapa algoritma seperti LSTM, Random Forest, dan SVR dalam memprediksi harga beras dan menunjukkan bahwa pendekatan machine learning mampu meningkatkan akurasi prediksi harga komoditas pangan. [6] Penelitian lain juga menerapkan metode regresi linear dalam memprediksi harga beras premium dan menghasilkan performa prediksi yang cukup baik berdasarkan nilai MAE dan MSE. [7]

Namun demikian, sebagian besar penelitian tersebut masih memiliki beberapa keterbatasan. Pertama, banyak studi hanya berfokus pada satu model tanpa melakukan evaluasi komparatif antara pendekatan linear dan non-linear dalam satu kerangka analisis yang terpadu. Kedua, sebagian besar penelitian masih hanya menggunakan data historis tanpa mempertimbangkan faktor eksternal yang dapat memengaruhi harga. Ketiga, penerapan konsep time series seperti fitur lag, validasi berbasis waktu (time-series split), serta penggunaan baseline model masih jarang diterapkan secara menyeluruh. Padahal, komponen tersebut penting untuk menangkap karakteristik temporal dan meningkatkan objektivitas evaluasi model.[3] [8]

Berdasarkan keterbatasan tersebut, penelitian ini mengusulkan pendekatan yang mengintegrasikan metode machine learning dan time series dalam satu kerangka analisis. Penelitian ini menggunakan metode Support Vector Regression (SVR), Linear Regression (LR), serta ARIMAX yang memungkinkan penggunaan variabel eksternal. Variabel eksternal yang digunakan adalah curah hujan yang berpotensi memengaruhi produksi dan distribusi beras. Selain itu, penelitian ini menerapkan fitur lag untuk menangkap pola historis, validasi berbasis waktu untuk menjaga urutan data, serta model baseline berupa naive forecast sebagai pembanding dalam evaluasi kinerja model.

Kontribusi utama penelitian ini adalah pengembangan kerangka evaluasi komparatif berbasis time series yang mengintegrasikan model machine learning dan model statistik dalam satu sistem analisis terpadu. Kerangka ini menggabungkan fitur autoregressive (lag), variabel eksternal berupa curah hujan, serta baseline forecasting untuk

meningkatkan objektivitas evaluasi model. Penelitian sebelumnya umumnya hanya membandingkan model machine learning tanpa menerapkan validasi temporal, baseline forecasting, dan integrasi variabel eksternal secara bersamaan. Penelitian ini menggabungkan ketiga aspek tersebut dalam satu kerangka evaluasi time series sehingga menghasilkan evaluasi yang lebih realistis terhadap prediksi harga pangan khususnya beras. Selain itu, penelitian ini juga mengevaluasi dua karakteristik komoditas beras yang berbeda, yaitu beras medium dan beras premium, sehingga memberikan pemahaman yang lebih komprehensif mengenai pengaruh struktur data terhadap performa model prediksi.

Sejalan dengan urgensi permasalahan, fluktuasi harga beras yang berdampak pada inflasi pangan menuntut adanya model prediksi yang akurat sebagai dasar pengambilan keputusan. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja metode SVR, Linear Regression, dan ARIMAX dalam memprediksi harga beras medium dan premium menggunakan data time series. Evaluasi dilakukan menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan  $R^2$  Score untuk menentukan model terbaik dalam memodelkan pergerakan harga beras.

## II. METODE

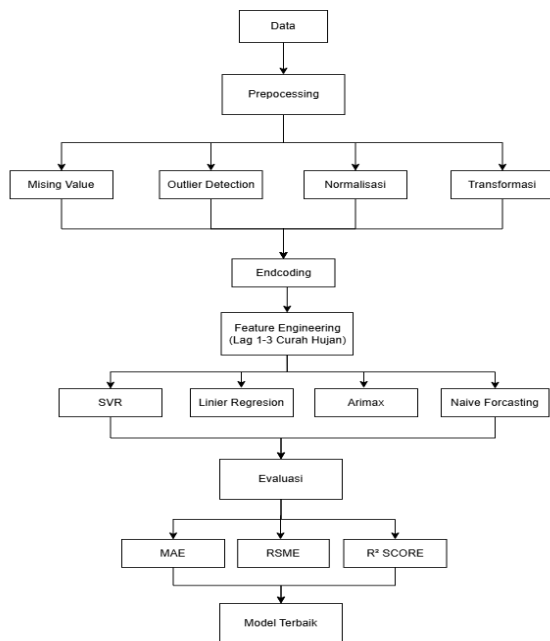
Penelitian ini menggunakan pendekatan supervised learning berbasis regresi dan time series untuk memodelkan serta memprediksi harga beras sebagai variabel dependen. Metode yang digunakan meliputi Support Vector Regression (SVR), Linear Regression (LR), serta ARIMAX sebagai pengembangan model time series dengan variabel eksternal.

Tahapan penelitian meliputi pengumpulan data, preprocessing, feature engineering, pemodelan, evaluasi, dan penentuan model terbaik.

### A. Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data Kebutuhan Pokok Masyarakat (KEPOKMAS) yang diperoleh dari Dinas Perindustrian dan Perdagangan Provinsi Jawa Tengah. Data yang digunakan mencakup periode tahun 2023 hingga 2025 dengan total 1.476 observasi data harian dan berasal dari total 33 komoditas, namun penelitian ini hanya memfokuskan analisis pada dua komoditas, yaitu beras medium dan beras premium. Data disusun dalam bentuk data harian (time series) dengan atribut tanggal bulan, tahun, komoditi, harga. Di mana harga (Rp/Kg) merupakan data kuantitatif berskala rasio yang menjadi variabel target (Y), sedangkan komoditi merupakan data kualitatif berskala nominal yang membedakan jenis beras. Selain variabel utama, penelitian ini juga menambahkan variabel eksternal berupa curah hujan sebagai faktor yang dapat memengaruhi produksi dan distribusi beras. Variabel ini digunakan untuk semua model yaitu SVR, LR, dan ARIMAX agar dapat melihat pengaruh faktor eksternal terhadap fluktuasi harga. Data disusun dalam bentuk time series harian karena memiliki resolusi yang lebih tinggi

dibandingkan data agregasi bulanan, sehingga mampu merepresentasikan perubahan harga secara lebih detail.



Gambar 1 Tahapan Penelitian

Pembagian data dilakukan menggunakan metode time series split dengan rasio 80:20 tanpa pengacakan (*shuffle = False*), di mana 80% data awal digunakan sebagai data pelatihan dan 20% data akhir sebagai data pengujian. Pendekatan ini digunakan untuk menjaga urutan waktu sehingga model dapat dievaluasi secara realistis dalam memprediksi data masa depan.

TABEL I  
DATA KOMODITAS

Bulan	Tahun	Komoditas	Tanggal	Harga
1	2023	Beras Medium	02/01/23	10.818
1	2023	Beras Premium	02/01/23	12.296
1	2023	Beras Medium	03/01/23	10.855
.....	.....	.....	.....	.....
.....	.....	.....	.....	.....
12	2025	Beras Premium	30/12/25	14.682
12	2025	Beras Medium	31/12/25	13.143
12	2025	Beras Premium	31/12/25	14.680

**B. Preprocessing**

Preprocessing data merupakan tahap awal yang bertujuan untuk meningkatkan kualitas data sebelum digunakan dalam proses pemodelan. Data time series umumnya memiliki berbagai permasalahan seperti missing value, outlier, dan perbedaan skala, sehingga perlu dilakukan penanganan yang

tepat. Data dibagi menggunakan time-based split dengan rasio 80:20 tanpa pengacakan (*shuffle = False*) untuk menjaga urutan temporal antara data latih dan data uji. Dengan demikian, proses evaluasi menjadi lebih realistis dibandingkan pembagian data secara acak. Seluruh proses preprocessing seperti normalisasi dan transformasi dilakukan setelah pembagian data untuk menghindari terjadinya data leakage.

**1) Missing value**

Missing value adalah kondisi ketika suatu atribut dalam dataset tidak memiliki nilai atau bernilai kosong (*null*). Pada data time series harga beras, missing value dapat terjadi akibat keterlambatan pencatatan, kesalahan sistem, atau tidak tersedianya data pada tanggal tertentu. Keberadaan missing value dapat mempengaruhi akurasi model prediksi apabila tidak ditangani dengan baik. [9]

Dalam penelitian ini, penanganan *missing value* dilakukan menggunakan metode mean imputation, yaitu menggantikan nilai yang hilang dengan nilai rata-rata dari data yang tersedia pada atribut tersebut.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \tag{1}$$

Dimana Persamaan 1 adalah  $\bar{x}$  = nilai rata-rata,  $x_i$  = nilai data ke-I, dan n = jumlah data valid.

Nilai missing kemudian digantikan dengan  $\bar{x}$  agar kontinuitas data time series tetap terjaga.

**2) Outlier Detection**

Outlier adalah nilai data yang menyimpang jauh dari pola umum data lainnya. Dalam data harga beras, outlier dapat disebabkan oleh lonjakan harga ekstrem akibat gangguan distribusi, bencana, atau kesalahan pencatatan. Outlier perlu diidentifikasi karena dapat menyebabkan bias pada hasil prediksi. [10] Pada penelitian ini, deteksi *outlier* dilakukan menggunakan metode Interquartile Range (IQR).

$$IQR = Q3 - Q1 \tag{2}$$

Persamaan 2 adalah batas outlier ditentukan dengan:

$$\text{Lower Bound} = Q_1 - 1.5 \times IQR \tag{3}$$

$$\text{Upper Bound} = Q_3 + 1.5 \times IQR \tag{4}$$

Nilai yang berada di luar batas tersebut dikategorikan sebagai outlier dan dapat dihapus atau ditangani sesuai kebutuhan analisis. [11]

**3) Normalisasi**

Normalisasi adalah proses penyamaan skala data numerik agar berada dalam rentang tertentu. Normalisasi sangat penting dalam metode SVR, karena algoritma ini sensitif terhadap perbedaan skala data. [12]

Dalam penelitian ini digunakan metode Min-Max Normalization, yang mentransformasikan nilai data ke dalam rentang [0,1].

$$x^1 = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

Persamaan 5 menunjukkan bahwa  $x_1$  adalah nilai hasil normalisasi,  $x$  adalah nilai asli,  $x_{\min}$  adalah nilai minimum, dan  $x_{\max}$  nilai maksimum. Hasil normalisasi akan berada pada rentang nilai yang sama sehingga dapat meningkatkan stabilitas proses pelatihan model [4]

#### 4) Transformasi

Transformasi data bertujuan untuk mengubah bentuk distribusi data agar lebih stabil dan mendekati distribusi normal. Pada data harga beras, transformasi digunakan untuk mengurangi efek fluktuasi yang tajam. [13]

Salah satu metode transformasi yang umum digunakan adalah logaritmik:

$$x^1 = \log(x) \quad (6)$$

Persamaan 6 menunjukkan bahwa,  $x$  nilai data asli (harga beras dalam rupiah per kilogram),  $x^1$  adalah nilai data setelah transformasi, dan  $\log$  adalah fungsi logaritma (basis 10 atau natural log, tergantung implementasi).

#### 5) Encoding

Encoding adalah proses mengubah data kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning. Dalam penelitian ini, atribut komoditi (beras medium dan beras premium) merupakan data kategorikal. [14]

Metode encoding yang digunakan adalah Label Encoding, yaitu:

Beras Medium = 0  
Beras Premium = 1

$$f(x) = \begin{cases} 0, & \text{jika } x = \text{Beras Medium} \\ 1, & \text{jika } x = \text{Beras Premium} \end{cases} \quad (7)$$

Persamaan 7 menunjukkan bahwa  $x$  nilai kategorikal asli pada variabel komoditi,  $f(x)$  adalah nilai numerik hasil encoding dan Angka 0 dan 1 digunakan sebagai label numerik, bukan sebagai nilai matematis yang menunjukkan tingkatan harga atau kualitas.

#### 6) Feature Lag

Dalam penelitian ini digunakan fitur lag untuk menangkap pengaruh nilai historis terhadap harga beras saat ini. [15] Fitur lag yang digunakan meliputi lag 1, lag 2, dan lag 3, yang merepresentasikan harga pada satu, dua, dan tiga hari sebelumnya.

$$(x_{t-1}, x_{t-2}, x_{t-3})$$

Penggunaan lag dilakukan melalui eksperimen dengan beberapa skenario, yaitu lag 1, lag 1–2, dan lag 1–3. Setiap

skenario diuji menggunakan model SVR, Linear Regression, dan ARIMAX untuk melihat pengaruhnya terhadap hasil prediksi. Evaluasi dilakukan menggunakan MAE, RMSE, dan  $R^2$  Score untuk menentukan kombinasi lag terbaik yang menghasilkan performa prediksi paling optimal. Feature lag digunakan untuk menangkap ketergantungan nilai masa lalu terhadap nilai saat ini. Selain itu, penelitian ini menggunakan time-based split tanpa pengacakan data (shuffle = False) untuk menjaga urutan kronologis data, sehingga model dilatih menggunakan data masa lalu dan diuji pada data masa depan. Hal ini dilakukan untuk menghindari data leakage dan memastikan evaluasi model lebih realistis sesuai karakteristik data time series.

#### C. Pemodelan

Pemodelan prediksi adalah proses membangun dan menggunakan model matematis atau algoritma pembelajaran mesin untuk memperkirakan nilai atau kejadian di masa depan berdasarkan pola yang dipelajari dari data historis. Proses ini meliputi pengumpulan dan pengolahan data, pemilihan variabel yang relevan, pelatihan model, serta evaluasi kinerja model menggunakan metrik tertentu agar hasil prediksi dapat digunakan sebagai dasar pengambilan keputusan yang lebih akurat dan berbasis data. [16]

##### 1) SVR

Support Vector Regression (SVR) merupakan pengembangan dari Support Vector Machine (SVM) yang digunakan untuk permasalahan regresi. SVR bertujuan untuk mencari fungsi regresi terbaik dengan toleransi kesalahan tertentu ( $\epsilon$ ). [17]

Fungsi regresi SVR dirumuskan sebagai :

$$f(x) = w^T x + b \quad (8)$$

Dengan fungsi objektif :

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (9)$$

Persamaan 8 dan 9 menunjukkan  $w$  adalah vektor bobot,  $b$  adalah bias,  $C$  merupakan parameter regularisasi yang mengontrol keseimbangan antara kompleksitas model dan kesalahan prediksi, sedangkan  $\xi_i + \xi_i^*$  merupakan slack variable yang merepresentasikan deviasi kesalahan terhadap batas toleransi  $\epsilon$ .

Pada penelitian ini digunakan kernel Radial Basis Function (RBF) karena mampu menangkap pola non-linear pada data time series harga beras yang bersifat fluktuatif. Parameter Support Vector Regression (SVR) ditentukan menggunakan metode Grid Search dengan pembagian data berbasis time series split, di mana data dibagi menjadi data latih dan data uji secara kronologis untuk menjaga urutan waktu. kombinasi parameter yang diuji meliputi nilai  $C = \{10, 50, 100\}$ ,  $\gamma = \{0.001, 0.01, 0.1\}$ , dan  $\epsilon = \{0.01, 0.1, 0.5\}$ . Setiap kombinasi parameter dievaluasi menggunakan data validasi dengan metrik Mean Absolute Error (MAE), Root Mean

Square Error (RMSE), dan R<sup>2</sup> Score. Berdasarkan hasil pengujian, diperoleh bahwa kombinasi C = 10, gamma = 0.001, dan epsilon = 0.01 memberikan performa paling stabil pada kedua komoditas. Meskipun terdapat kombinasi lain yang menghasilkan error lebih rendah pada salah satu komoditas, konfigurasi tersebut tidak menunjukkan konsistensi performa secara keseluruhan. Oleh karena itu, parameter tersebut dipilih karena memberikan keseimbangan terbaik antara akurasi prediksi dan kemampuan generalisasi model.

## 2) Linear Regression

Linear Regression merupakan metode statistik yang digunakan untuk memodelkan hubungan linear antara variabel independen dan variabel dependen. [18] Penggunaan LR bertujuan untuk mengetahui sejauh mana model sederhana mampu menangkap pola data dibandingkan model non-linear seperti SVR.

Model regresi linear dinyatakan sebagai :

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (10)$$

Persamaan 10 menunjukkan y adalah variabel dependen (harga beras), x adalah variabel independent (waktu),  $\beta_0$  adalah konstanta  $\beta_1$  koefisien regresi dan  $\varepsilon$  adalah error. [19]

## 3) Arimax

ARIMAX merupakan pengembangan dari metode ARIMA (AutoRegressive Integrated Moving Average) yang digunakan untuk pemodelan data deret waktu dengan mempertimbangkan variabel eksternal (*exogenous variables*). [20] Metode ini sangat sesuai digunakan pada prediksi harga komoditas karena mampu menggabungkan pola historis data dengan faktor-faktor lain yang memengaruhi, seperti inflasi, cuaca, atau distribusi. Dalam penelitian ini, ARIMAX diimplementasikan menggunakan pendekatan SARIMAX (Seasonal ARIMAX) dengan memasukkan variabel eksogen berupa curah hujan serta fitur lag dari curah hujan. Penambahan variabel lag dilakukan untuk menangkap efek keterlambatan (time-lag effect), yaitu kondisi di mana pengaruh variabel cuaca terhadap harga tidak terjadi secara langsung pada waktu yang sama, tetapi berdampak pada periode berikutnya. Dengan demikian, model tidak hanya menangkap hubungan kontemporer, tetapi juga hubungan temporal antara variabel eksternal dan harga beras.

Model ARIMAX dinyatakan sebagai:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=1}^m B_k X_{t,k} \varepsilon_t \quad (11)$$

Persamaan 11 menunjukkan  $y_t$  adalah nilai variabel dependen pada waktu ke-t, c yaitu konstanta,  $\phi_i$  parameter autoregressive (AR),  $\theta_j$  adalah parameter moving average (MA),  $X_{t,k}$  variabel eksogen ke-k pada waktu ke-t,  $B_k$  koefisien variabel eksogen,  $\varepsilon_t$  error pada waktu ke-t, p orde autoregressive, d orde differencing dan q adalah orde moving average.

Parameter model SARIMAX pada penelitian ini ditetapkan menggunakan konfigurasi awal (p,d,q) = (2,1,2). Penentuan nilai parameter dilakukan melalui percobaan beberapa kombinasi nilai parameter dengan mempertimbangkan performa model pada data pelatihan. Selain itu, pemilihan parameter juga mengacu pada karakteristik data time series yang memiliki pola tren, sehingga digunakan differencing pada orde d = 1 untuk mengatasi non-stationary. Pendekatan ini dipilih untuk memperoleh model yang relatif stabil tanpa meningkatkan kompleksitas model secara berlebihan. Data dibagi menggunakan pendekatan time-based split untuk menjaga urutan temporal data sehingga tidak terjadi data leakage.

## 4) Naïve Forecasting

Naïve Forecasting merupakan model baseline paling sederhana yang digunakan sebagai pembanding dalam penelitian ini. Prinsip dasarnya adalah bahwa nilai pada periode berikutnya diprediksi sama dengan nilai pada periode sebelumnya.

$$y_{t+1} = y_t \quad (12)$$

Artinya, harga beras pada waktu berikutnya diasumsikan sama dengan harga saat ini. Metode ini digunakan sebagai pembanding untuk menilai apakah model seperti SVR, Linear Regression, dan ARIMAX mampu menghasilkan prediksi yang lebih baik.

## D. Evaluasi

### 1) RMSE

RMSE digunakan untuk mengukur besar kesalahan prediksi antara nilai aktual dan nilai prediksi.

$$RMSE = \frac{1}{n} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

Persamaan 11 menunjukkan n adalah jumlah data observasi,  $y_i$  nilai aktual harga beras pada data ke-i,  $\hat{y}_i$  nilai hasil prediksi model pada data ke-I,  $(y_i - \hat{y}_i)^2$  adalah selisih antara nilai actual dan nilai prediksi (error),  $(y_i - \hat{y}_i)^2$  kuadrat error untuk memastikan nilai selalu positif dan memberi penalty lebih besar pada error yang besar,  $\sum_{i=1}^n$  adalah seluruh error kuadrat, dan  $\frac{1}{n}$  adalah rata-rata error kuadrat. Nilai RMSE menunjukkan seberapa jauh rata-rata prediksi model menyimpang dari nilai aktual. Semakin kecil nilai RMSE, maka semakin baik kinerja model dalam memprediksi harga beras. [21]

### 2) MAE

Mean Absolute Error (MAE) digunakan untuk mengukur rata-rata besar kesalahan absolut antara nilai prediksi dan nilai aktual, sehingga dapat menggambarkan tingkat ketepatan model dalam melakukan prediksi. Semakin kecil nilai MAE yang dihasilkan, maka semakin baik kinerja model prediksi tersebut. Perhitungan MAE dirumuskan sebagai berikut.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{14}$$

Persamaan 12 menunjukkan n : menyatakan jumlah total data pengamatan,  $y_i$  merupakan nilai aktual pada pengamatan ke-i,  $\hat{y}_i$  : nilai hasil prediksi model pada pengamatan ke-i, dan  $|y_i - \hat{y}_i|$  : selisih absolut antara nilai actual dan nilai prediksi. [22]

3) *R2 SCORE*

$R^2$  Score atau koefisien determinasi digunakan untuk mengukur seberapa baik model dalam menjelaskan variasi data aktual. Nilai  $R^2$  menunjukkan proporsi variasi variabel dependen yang dapat dijelaskan oleh model prediksi. [23]

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Persamaan tersebut, n menyatakan jumlah total data observasi yang digunakan dalam evaluasi model. Nilai  $y_i$  merupakan nilai aktual pada data ke-i, sedangkan  $\hat{y}_i$  adalah nilai hasil prediksi model pada data ke-i.  $(\hat{y}_i - y_i)$  sebagai error,  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$  yaitu *Sum of Squared Errors* (SSE).

4) *Uji Statistik Wilcoxon Signed-Rank Test*

Untuk memastikan bahwa perbedaan performa antar model tidak terjadi secara kebetulan, penelitian ini menggunakan uji statistik *Wilcoxon Signed-Rank Test*. Uji ini dipilih karena tidak mengasumsikan distribusi normal pada data, sehingga lebih sesuai untuk membandingkan nilai error prediksi antar model pada data time series. Uji dilakukan terhadap nilai error pada data pengujian antara model Support Vector Regression (SVR), Linear Regression, dan ARIMAX. Hipotesis yang digunakan adalah:

H0: Tidak terdapat perbedaan signifikan antara performa kedua model H1: Terdapat perbedaan signifikan antara performa kedua model Keputusan diambil berdasarkan nilai p-value dengan tingkat signifikansi  $\alpha = 0,05$ . Jika p-value < 0,05, maka H0 ditolak. [24]

**III. HASIL DAN PEMBAHASAN**

*A. Hasil Prediksi Beras*

Tabel berikut menyajikan hasil perbandingan antara nilai aktual dan nilai prediksi harga beras menggunakan metode Support Vector Regression (SVR), Linear Regression, serta model pembanding ARIMAX. Data ditampilkan berdasarkan periode waktu pengamatan dan dikelompokkan berdasarkan komoditas, yaitu beras medium dan beras premium.

Berdasarkan hasil prediksi pada Tabel II, model SVR mampu menghasilkan nilai prediksi yang relatif mendekati nilai aktual pada komoditas beras medium. Selisih antara nilai prediksi dan aktual cenderung kecil dan stabil sepanjang periode pengamatan. Hal ini menunjukkan bahwa SVR

mampu menangkap pola pergerakan harga yang tidak sepenuhnya linear.

TABEL II  
HASIL PREDIKSI BERAS MEDIUM

Komoditas Beras Medium					
Tanggal	Actual	Prediksi SVR	Linear Regres sion	Arima x	Curah Hujan
2025-06-04	13091.419	13071.275	13054.2434	13068.42317	4.2
2025-06-05	13113.08557	13081.33467	13065.02038	13069.37199	4.2
2025-06-06	13158.80014	13098.38266	13078.48474	13073.23483	4.2
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
2025-12-29	13137.92648	13153.17111	13135.36038	13089.13089	12
2025-12-30	13137.92648	13133.17902	13122.87204	13089.13089	12
2025-12-31	13142.68848	13133.14139	13130.96374	13089.13089	12

Sementara itu, model Linear Regression juga mampu mengikuti tren harga beras medium, namun cenderung menghasilkan prediksi yang lebih halus (smooth) dan kurang responsif terhadap fluktuasi harga harian. Model ARIMAX pada beras medium menunjukkan hasil yang kurang optimal, ditandai dengan nilai prediksi yang cenderung konstan pada beberapa periode.

TABEL III  
HASIL PREDIKSI BERAS PREMIUM

Komoditas Beras Medium					
Tangga l	Actual	Prediksi SVR	Linear Regressi on	Arimax	Curah Hujan
2025-06-04	14997.14281	14992.44455	15001.21056	14974.9751	4.2
2025-06-05	14998.09538	15002.48085	15009.26001	14964.17509	4.2
2025-06-06	15006.19067	15006.78022	15013.43426	14975.86297	4.2
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
2025-12-29	14681.89543	14732.45984	14769.99258	14952.3517	12
2025-12-30	14681.89543	14723.26156	14763.00932	14952.3517	12
2025-12-31	14679.99071	14723.22591	14759.8903	14952.3517	12

Pada Tabel III untuk komoditas beras premium, terlihat bahwa model SVR menghasilkan prediksi yang lebih mendekati nilai aktual dibandingkan model Linear Regression. Hal ini menunjukkan bahwa meskipun pola harga beras premium relatif stabil, masih terdapat unsur non-linear yang lebih mampu ditangkap oleh model SVR. Sementara itu, model Linear Regression tetap menunjukkan performa yang cukup baik, namun dengan deviasi yang lebih besar dibandingkan SVR. Di sisi lain, model ARIMAX menghasilkan prediksi yang paling jauh dari nilai aktual, yang mengindikasikan bahwa variabel eksternal yang digunakan belum mampu secara optimal menjelaskan variasi harga beras premium.

### B. Hasil Evaluasi Kinerja Model Prediksi yang diuji

Evaluasi kinerja model dilakukan menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *R<sup>2</sup> Score* untuk mengukur tingkat akurasi prediksi terhadap data aktual. MAE dan RMSE digunakan untuk mengukur besar kesalahan prediksi, sedangkan *R<sup>2</sup> Score* menunjukkan kemampuan model dalam menjelaskan variasi data. Nilai MAE dan RMSE dalam penelitian ini diinterpretasikan dalam satuan asli, yaitu Rupiah per kilogram (Rp/kg), sehingga memberikan makna yang lebih praktis. Sebagai contoh, nilai MAE sebesar 15 menunjukkan bahwa rata-rata kesalahan prediksi model adalah sekitar Rp15/kg, yang tergolong kecil dibandingkan dengan rentang harga beras secara keseluruhan.

TABEL IV  
HASIL EVALUASI PREDIKSI

Model	MAE	RMSE	R <sup>2</sup> Score	Komoditas
Naïve Forecasting	10.395018 27	21.336907 57	0.94098637 84	Beras Medium
SVR	26.031438 58	33.568528 33	0.85393239 13	Beras Medium
Linear Regression	45.850747 47	53.263965 92	0.63224692 27	Beras Medium
Arimax	164.90953 77	190.43733 41	- 3.70103518 9	Beras Medium
Naïve Forecasting	15.976326 85	49.225388 86	0.94990565 92	Beras Premium
SVR	28.944398 43	50.064885 53	0.94818245 78	Beras Premium
Linear Regression	42.536324 42	62.740686 49	0.91862164 86	Beras Premium
Arimax	195.37179 76	214.21284 09	0.05136016 705	Beras Premium

Berdasarkan hasil evaluasi pada Tabel IV, kinerja model menunjukkan perbedaan yang cukup signifikan pada masing-

masing komoditas. Pada komoditas beras medium, model Naïve Forecasting menghasilkan performa terbaik dengan nilai MAE sebesar 10,39, RMSE sebesar 21,33, dan R<sup>2</sup> sebesar 0,940. Hal ini menunjukkan adanya ketergantungan kuat terhadap nilai historis, sehingga model sederhana seperti Naïve Forecasting dapat memberikan prediksi yang akurat. Di antara model machine learning yang digunakan, Support Vector Regression (SVR) menunjukkan performa terbaik kedua dengan R<sup>2</sup> sebesar 0,854, diikuti oleh Linear Regression dengan R<sup>2</sup> sebesar 0,632. Sementara itu, ARIMAX menghasilkan performa yang sangat rendah dengan nilai error tinggi dan R<sup>2</sup> negatif, yang menunjukkan ketidakmampuan model dalam menangkap pola data secara efektif. Pada komoditas beras premium, model Naïve Forecasting kembali menunjukkan performa terbaik dengan MAE sebesar 15,98, RMSE sebesar 49,22, dan R<sup>2</sup> sebesar 0,949. Hal ini mengindikasikan bahwa pola harga beras premium juga memiliki ketergantungan temporal yang kuat. Di antara model berbasis machine learning, Support Vector Regression (SVR) menunjukkan performa terbaik dengan MAE sebesar 28,94, RMSE sebesar 50,06, dan R<sup>2</sup> sebesar 0,948, yang sangat mendekati performa Naïve Forecasting. Sementara itu, Linear Regression juga menunjukkan performa yang cukup baik dengan R<sup>2</sup> sebesar 0,918 namun masih berada di bawah SVR. ARIMAX kembali menunjukkan performa yang rendah dengan R<sup>2</sup> sebesar 0,051.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Naïve Forecasting memiliki baseline performa yang sangat kuat pada kedua komoditas, yang mengindikasikan adanya pola ketergantungan temporal yang tinggi dalam data harga beras. Namun demikian, model Support Vector Regression (SVR) mampu memberikan performa yang kompetitif dan stabil pada kedua komoditas dibandingkan model lainnya, terutama dalam menangkap pola non-linear pada data. Sementara itu, Linear Regression menunjukkan performa yang lebih baik pada data dengan karakteristik lebih stabil, khususnya pada beras premium, meskipun masih berada di bawah SVR. Adapun ARIMAX tidak memberikan performa yang optimal pada kedua komoditas, yang menunjukkan bahwa model ini kurang sesuai untuk karakteristik data yang digunakan dalam penelitian ini.

### C. Hasil Uji Signifikansi Statistik

TABEL V  
HASIL UJI SIGNIFIKASI STATISTIK

Perbandingan	p-value
SVR vs Linear Regression	< 0.001
SVR vs ARIMAX	< 0.001
Linear Regression vs ARIMAX	< 0.001

Hasil uji statistik menunjukkan bahwa seluruh perbandingan model menghasilkan nilai p-value < 0,001. Hal ini menunjukkan bahwa perbedaan kinerja antara model SVR, Linear Regression, dan ARIMAX signifikan secara statistik. Dengan demikian, model dengan nilai error lebih rendah memiliki performa yang lebih baik secara nyata dan bukan disebabkan oleh faktor kebetulan.

### 1) Analisis Matrix

Hasil evaluasi metrik menunjukkan bahwa pada beras medium, SVR menghasilkan error yang lebih rendah dibandingkan Linear Regression dengan MAE sebesar 26,03 dan RMSE sebesar 33,56, sedangkan Linear Regression menghasilkan MAE sebesar 45,85 dan RMSE sebesar 53,26. Hal ini menunjukkan bahwa SVR lebih mampu menangkap pola data yang bersifat non-linear dibandingkan model regresi linear. Pada beras premium, SVR juga menunjukkan performa yang lebih baik dibandingkan Linear Regression dengan MAE sebesar 28,94 dan RMSE sebesar 50,06, sementara Linear Regression menghasilkan MAE sebesar 42,54 dan RMSE sebesar 62,74. Meskipun selisih performa tidak terlalu besar pada metrik  $R^2$ , SVR tetap menunjukkan keunggulan dalam error prediksi. Selain itu, Naïve Forecasting tetap menjadi model dengan performa terbaik pada kedua komoditas, yang menunjukkan bahwa data harga beras memiliki ketergantungan yang kuat terhadap nilai sebelumnya.

### 2) Analisis Kinerja Model SVR

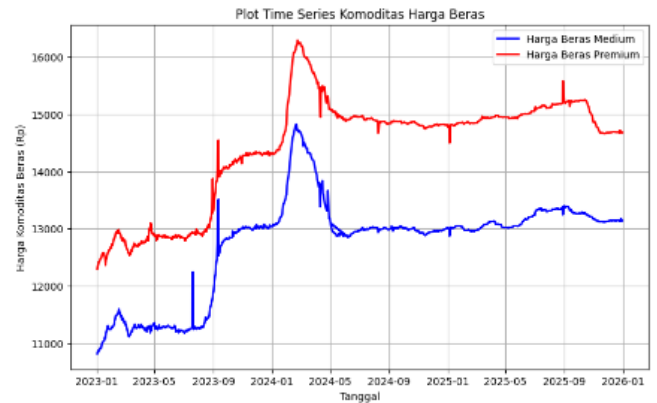
Model SVR menunjukkan performa yang konsisten pada kedua komoditas. Pada beras medium, SVR mampu menangkap pola fluktuasi harga dengan baik, yang ditunjukkan oleh nilai  $R^2$  sebesar 0,853. Pada beras premium, SVR bahkan mendekati performa Naïve Forecasting dengan  $R^2$  sebesar 0,948. Hal ini menunjukkan bahwa SVR efektif dalam memodelkan pola data yang bersifat non-linear maupun semi-linear. Namun, performanya tetap bergantung pada kombinasi parameter seperti C, gamma, dan epsilon yang sangat mempengaruhi kemampuan generalisasi model.

### 3) Analisis Kinerja Model Regresi Linear

Model Linear Regression menunjukkan performa yang lebih terbatas dibandingkan SVR pada kedua komoditas. Pada beras medium, model ini hanya menghasilkan  $R^2$  sebesar 0,632 yang menunjukkan keterbatasan dalam menangkap pola non-linear pada data. Namun, pada beras premium, Linear Regression menunjukkan performa yang lebih baik dengan  $R^2$  sebesar 0,919. Hal ini mengindikasikan bahwa pola harga beras premium cenderung lebih stabil dan mendekati hubungan linear, sehingga lebih sesuai dimodelkan menggunakan regresi linear.

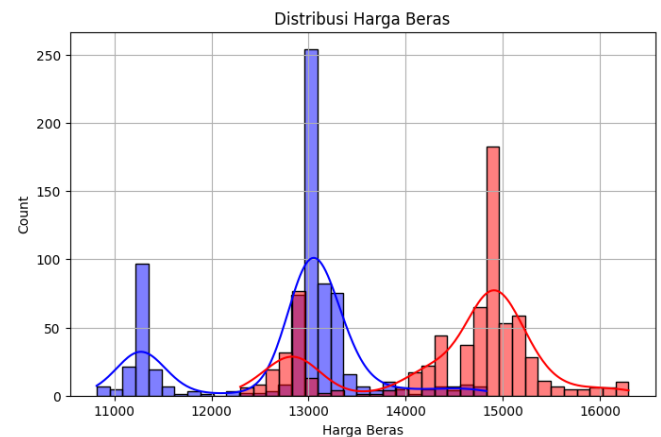
### 4) Analisis Visualisasi Hasil Prediksi

Hasil analisis visual prediksi dapat dilihat pada Gambar 2. Grafik time series menunjukkan pergerakan harga beras medium dan beras premium pada periode 2023–2026. Secara umum, harga beras premium selalu lebih tinggi dibandingkan beras medium. Pada awal periode pengamatan, kedua komoditas menunjukkan fluktuasi yang relatif stabil. Namun, mulai pertengahan 2023 hingga awal 2024 terjadi kenaikan harga yang signifikan, dengan puncak harga pada awal 2024, terutama pada beras premium yang mengalami lonjakan lebih tajam.



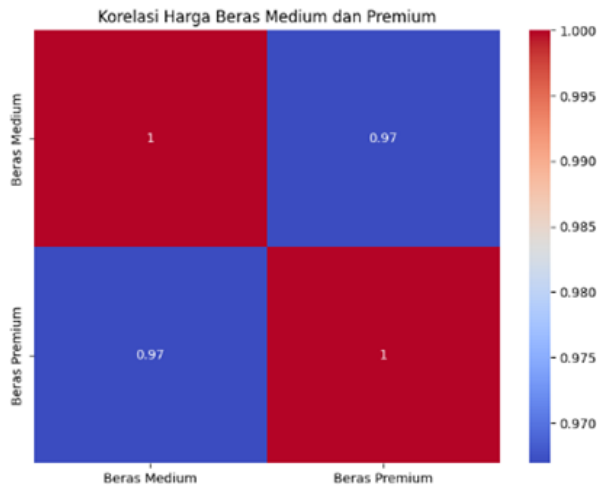
Gambar 2 Grafik Time Series

Setelah periode tersebut, harga kedua komoditas mengalami penurunan dan bergerak lebih stabil hingga akhir periode pengamatan. Secara keseluruhan, beras premium memiliki tingkat volatilitas yang lebih tinggi dibandingkan beras medium, yang mengindikasikan perbedaan karakteristik pola harga antara kedua komoditas.



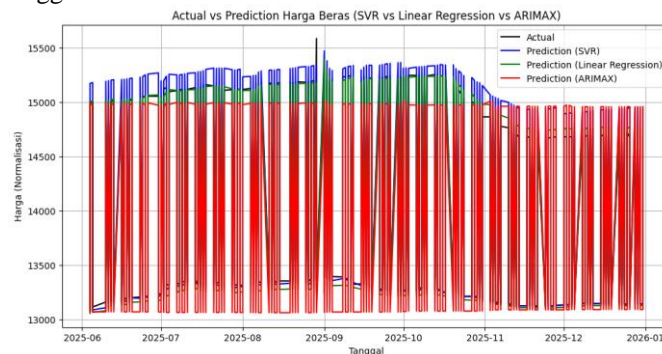
Gambar 3 Grafik Korelasi Harga Beras Medium dan Premium

Grafik heatmap korelasi menunjukkan hubungan antara harga beras medium dan beras premium. Nilai korelasi sebesar 0,97 menandakan adanya hubungan positif yang sangat kuat antara kedua komoditas. Hal ini menunjukkan bahwa pergerakan harga beras medium dan kenaikan atau penurunan harga pada salah satu komoditas akan diikuti oleh komoditas lainnya. Tingginya nilai korelasi ini mengindikasikan bahwa kedua jenis beras dipengaruhi oleh faktor pasar yang serupa, seperti kondisi pasokan, distribusi, dan permintaan.



Gambar 4 Distribusi Harga Beras

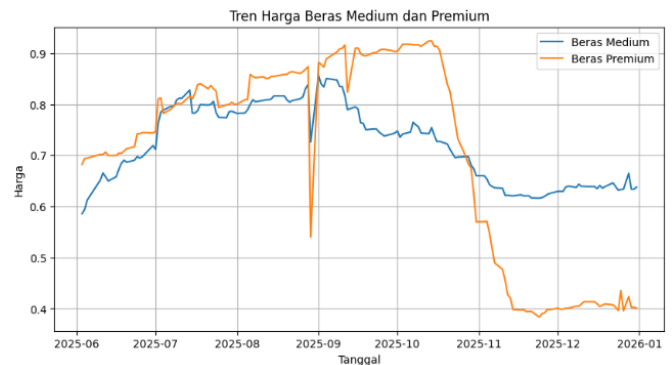
Grafik distribusi menunjukkan sebaran harga beras medium dan beras premium berdasarkan frekuensi kemunculannya. Harga beras medium cenderung terkonsentrasi pada kisaran Rp12.800–13.200, dengan sebaran yang relatif sempit, yang menandakan bahwa harga beras medium lebih stabil. Sebaliknya, distribusi harga beras premium tersebar pada rentang yang lebih luas, yaitu sekitar Rp14.500–15.200, serta menunjukkan variasi yang lebih tinggi.



Gambar 5 Grafik Perbandingan Harga Aktual dan Prediksi

Grafik tersebut menampilkan perbandingan antara harga aktual dengan hasil prediksi dari tiga model, yaitu SVR (biru), Linear Regression (hijau), dan ARIMAX (merah) pada periode Juni 2025 hingga awal 2026. Secara umum, garis aktual (hitam) menunjukkan pola harga yang relatif stabil dengan fluktuasi kecil di kisaran harga tertentu, serta adanya kenaikan hingga mencapai puncak sekitar pertengahan periode sebelum kembali menurun dan stabil di akhir periode. Model SVR (biru) terlihat mampu mengikuti pola pergerakan harga aktual dengan cukup baik. Garis prediksi SVR cenderung mendekati garis aktual, terutama dalam menangkap tren kenaikan dan penurunan harga. Hal ini menunjukkan bahwa SVR memiliki kemampuan yang baik dalam memodelkan pola data yang bersifat non-linear. Sementara itu, Linear Regression (hijau) juga mampu

mengikuti tren umum harga, namun terlihat lebih halus (*smooth*) dan kurang responsif terhadap fluktuasi kecil. Model ini cenderung menghasilkan garis prediksi yang lebih stabil, sehingga pada beberapa titik kurang mampu menangkap perubahan harga secara detail. Berbeda dengan kedua model tersebut, ARIMAX (merah) menunjukkan pola prediksi yang tidak stabil dan cenderung menyimpang jauh dari data aktual. Hal ini terlihat dari banyaknya lonjakan tajam ke bawah yang tidak sesuai dengan pola harga sebenarnya. Kondisi ini mengindikasikan bahwa model ARIMAX tidak mampu memodelkan data dengan baik pada kasus ini.



Gambar 6 Grafik Tren Harga Beras Medium dan Premium

Grafik Tren Harga Beras Medium dan Premium menunjukkan pergerakan harga kedua jenis beras selama periode pengamatan dari sekitar pertengahan tahun 2025 hingga awal tahun 2026. Berdasarkan grafik tersebut, harga beras medium cenderung mengalami kenaikan secara bertahap pada awal periode, kemudian mencapai titik relatif tinggi sekitar bulan Agustus hingga September 2025. Setelah periode tersebut, harga mulai mengalami penurunan secara perlahan hingga sekitar November 2025, kemudian bergerak relatif stabil hingga awal tahun 2026. Dibandingkan dengan beras medium, beras premium menunjukkan fluktuasi harga yang lebih besar. Pada awal periode, harga beras premium berada di atas harga beras medium dan terus meningkat hingga mencapai puncaknya sekitar September hingga Oktober 2025. Namun setelah itu terjadi penurunan yang cukup tajam, terutama pada sekitar November hingga Desember 2025, sebelum akhirnya relatif stabil pada kisaran nilai yang lebih rendah. Secara keseluruhan, grafik ini menunjukkan bahwa harga beras premium memiliki tingkat volatilitas yang lebih tinggi dibandingkan beras medium, sementara beras medium cenderung memiliki pola pergerakan harga yang lebih stabil sepanjang periode pengamatan.

#### IV. DISKUSI

Berdasarkan hasil penelitian, ditemukan bahwa model Naïve Forecasting justru menghasilkan performa terbaik pada kedua komoditas. Hal ini menunjukkan bahwa data harga beras memiliki ketergantungan temporal yang sangat kuat, di mana nilai harga saat ini sangat dipengaruhi oleh nilai

sebelumnya. Kondisi ini mengindikasikan bahwa pola autoregressive pada data sangat dominan, sehingga model sederhana mampu memberikan prediksi yang akurat tanpa memerlukan kompleksitas model yang tinggi.

Pada komoditas beras medium, Support Vector Regression (SVR) menunjukkan performa terbaik di antara model yang diusulkan. Hal ini mengindikasikan bahwa pola harga beras medium memiliki karakteristik yang lebih fluktuatif dan tidak sepenuhnya linear. Kondisi ini dapat dikaitkan dengan dinamika pasar beras medium yang cenderung lebih sensitif terhadap perubahan pasokan, distribusi, serta kondisi produksi, sehingga menyebabkan variasi harga yang lebih tinggi. Oleh karena itu, model non-linear seperti SVR lebih mampu menangkap pola tersebut dibandingkan Linear Regression. Namun demikian, performa SVR masih berada di bawah Naïve Forecasting, yang menunjukkan bahwa ketergantungan terhadap nilai historis tetap menjadi faktor dominan dalam pembentukan harga.

Pada komoditas beras premium, meskipun Linear Regression menunjukkan performa yang cukup baik, SVR tetap menghasilkan nilai error yang lebih rendah. Hal ini menunjukkan bahwa meskipun pola harga beras premium relatif lebih stabil, masih terdapat unsur non-linear yang dapat ditangkap lebih baik oleh SVR. Secara kontekstual, beras premium umumnya memiliki segmentasi pasar yang lebih spesifik dengan tingkat permintaan yang relatif stabil, sehingga pergerakan harga cenderung lebih konsisten dan mendekati pola linear dibandingkan beras medium.

Sementara itu, model ARIMAX tidak menunjukkan performa yang optimal pada kedua komoditas. Hal ini mengindikasikan bahwa variabel eksternal berupa curah hujan belum mampu menjelaskan variasi harga beras secara signifikan dalam penelitian ini. Selain itu, terdapat kemungkinan bahwa faktor lain seperti distribusi, kebijakan pemerintah, kondisi pasar, serta rantai pasok memiliki pengaruh yang lebih dominan terhadap perubahan harga dibandingkan faktor cuaca.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pemilihan model prediksi sangat bergantung pada karakteristik data yang digunakan. Pada kasus ini, model sederhana seperti Naïve Forecasting dan model non-linear seperti SVR lebih sesuai untuk memodelkan data harga beras dibandingkan model statistik seperti ARIMAX. Namun demikian, hasil ini perlu diinterpretasikan secara hati-hati mengingat keterbatasan variabel dan periode data yang digunakan, sehingga belum sepenuhnya merepresentasikan seluruh faktor yang memengaruhi dinamika harga beras di pasar.

## V. KESIMPULAN

Berdasarkan hasil penelitian, kinerja model prediksi harga beras sangat dipengaruhi oleh karakteristik data pada masing-masing komoditas. Penelitian ini telah menerapkan pendekatan time series dengan fitur lag, variabel eksternal curah hujan, serta validasi berbasis waktu untuk membandingkan metode Support Vector Regression (SVR),

Linear Regression (LR), ARIMAX, dan model baseline Naïve Forecasting. Pada komoditas beras medium, model Naïve Forecasting menunjukkan performa terbaik, yang mengindikasikan adanya ketergantungan temporal yang kuat pada data harga. Di antara model yang diusulkan, SVR memberikan hasil yang lebih baik dibandingkan Linear Regression, yang menunjukkan bahwa pola harga beras medium cenderung bersifat non-linear dan fluktuatif. Sementara itu, model ARIMAX belum mampu memberikan performa yang optimal. Pada komoditas beras premium, model Naïve Forecasting kembali menunjukkan performa terbaik, diikuti oleh Linear Regression dan SVR yang menunjukkan kinerja kompetitif. Hal ini mengindikasikan bahwa pola harga beras premium cenderung lebih stabil dan mendekati hubungan linear. Hasil uji statistik menggunakan Wilcoxon Signed-Rank Test menunjukkan bahwa perbedaan performa antar model signifikan secara statistik ( $p\text{-value} < 0,001$ ), sehingga perbedaan kinerja model tidak terjadi secara kebetulan.

Secara keseluruhan, penelitian ini menunjukkan bahwa model sederhana seperti Naïve Forecasting masih menjadi baseline yang kuat dalam prediksi data time series harga beras. Selain itu, penggunaan variabel eksternal berupa curah hujan belum memberikan peningkatan akurasi yang signifikan. Penelitian ini memberikan kontribusi awal dalam pengembangan kerangka evaluasi komparatif untuk prediksi harga komoditas pangan, namun hasil yang diperoleh perlu diinterpretasikan secara hati-hati dengan mempertimbangkan keterbatasan data dan model yang digunakan. Penelitian ini memberikan kontribusi dalam analisis komparatif antara metode machine learning dan time series dalam prediksi harga komoditas pangan. Model yang dihasilkan dalam penelitian ini menunjukkan potensi awal sebagai alat bantu dalam pemantauan pergerakan harga beras. Model ini dapat digunakan untuk mengidentifikasi pola perubahan harga berdasarkan data historis. Namun demikian, mengingat keterbatasan data dan variabel yang digunakan, hasil penelitian ini belum dapat digunakan sebagai dasar pengambilan kebijakan secara langsung, melainkan sebagai referensi awal dalam pengembangan sistem prediksi yang lebih komprehensif. Pengembangan lebih lanjut dengan penambahan variabel eksternal yang lebih relevan serta penggunaan data dengan cakupan yang lebih luas diperlukan agar model dapat memberikan dukungan yang lebih kuat dalam konteks pengambilan keputusan dan stabilisasi harga.

## DAFTAR PUSTAKA

- [1] M. Z. Gapari, "Pengaruh kenaikan harga beras terhadap kesejahteraan petani di desa sukaraja," vol. 3, no. April, pp. 14–26, 2021.
- [2] A. D. Lestari, E. Erlikasna, C. Ridho, I. Breta, and M. Daniyal, "Dampak Fluktuasi Harga Beras , Bawang Merah , Cabai Terhadap Inflasi Impact of Price Fluctuations of Rice , Shallots , Chilies on Inflation," pp. 219–226.
- [3] D. Saputra, D. Raihan Trinadi, D. Agustina, and P. Pujiyanto,

- “Perbandingan Metode Random Forest, Linear Regression, SVM Untuk Memprediksi Harga Beras Premium,” *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 154–160, 2025, doi: 10.47701/yqx4ss53.
- [4] Aviva Pradasyah and A. Baita, “Comparative Study of Support Vector Regression and Long Short-Term Memory for Stock Price Prediction,” *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1301–1311, 2025, doi: 10.30871/jaic.v9i4.9425.
- [5] A. A. Suyono and R. Arief, “Prediksi Indeks Harga Konsumen Komoditas Makanan di Kota Surabaya menggunakan Support Vector Regression,” 2022, doi: 10.47002/metik.v6i1.339.
- [6] C. A. Bahri and K. D. Tania, “Perbandingan Kinerja LSTM , Random Forest , dan SVR Berbasis Knowledge Discovery untuk Prediksi Harga Beras Sumatera Selatan,” vol. 12, no. 5, pp. 721–732, 2025, doi: 10.30865/jurikom.v12i5.9140.
- [7] A. T. Widiyatmoko, S. Butsianto, and A. Nugroho, “Application of Machine Learning for Premium Rice Price Increase Prediction Using Linear Regression Algorithm Penerapan Machine Learning untuk Prediksi Kenaikan Harga Beras Premium Menggunakan Algoritma Regresi Linear,” vol. 5, no. July, pp. 1125–1132, 2025.
- [8] V. Arinal and M. Azhari, “Penerapan Regresi Linear untuk Prediksi Harga Beras di Indonesia,” *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 341–346, 2023, [Online]. Available: <http://ejournal.sisfokomtek.org/index.php/saintek/article/download/1417/1109>
- [9] N. Niako, J. D. Melgarejo, G. E. Maestre, and K. P. Vatcheva, *Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with ARIMA and LSTM*, vol. 24, no. 1, 2024. doi: 10.1186/s12874-024-02448-3.
- [10] Nimatul Mamuriyah, Richard, and Haeruddin, “Implementation Mean Imputation and Outlier Detection for Loan Prediction Using the Random Forest Algorithm,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 4, pp. 937–944, 2025, doi: 10.33480/jitk.v10i4.6437.
- [11] Shabila Ocktavia and Wahyu T Atmojo, “Analisis Segmentasi Pelanggan Pembiayaan Berdasarkan Demografi Untuk Memprediksi Tingkat Kredit Menggunakan Algoritma K-Means,” *J. Inform. Teknol. dan Sains*, vol. Vol 7, pp. 394–402, 2025.
- [12] G. Brandon Abelio Ogaden *et al.*, “Perbandingan RFE dan SelectKbest untuk Klasifikasi Penyakit Diabetes dengan Random Forest,” *Jnatia*, vol. 3, no. 3, p. 2025, 2025.
- [13] K. Hidayat, W. Witanti, E. Ramadhan, F. Sains Dan Informatika, and U. Jenderal Achmad Yani, “Analisis Tren dan Prediksi Penjualan Restoran Menggunakan Model Time Series Prophet,” *Metik J.*, vol. 9, p. 2025, 2025, doi: 10.47002/metik.v9i2.1101.
- [14] M. Khairani and I. Zufria, “Klasifikasi Tingkatan Perokok dengan Analisis Data Survei Masyarakat menggunakan Algoritma K-Means dan XGBoost,” *J. Edukasi dan Penelit. Inform.*, vol. 11, no. 2, pp. 230–241, 2025.
- [15] D. N. Handayani and S. Qutub, “Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan,” vol. 4, no. 2, pp. 406–412, 2025.
- [16] M. Y. T. Sulistyono, E. S. Pane, E. M. Yuniarno, and M. H. Purnono, “Hybrid Significant Stroke Feature: A Novel Stroke Feature Analysis Approach for Stroke Severity Classification of EEG Signals Based on Time Domain, Frequency Domain, and Signal Decomposition Domain,” *Int. J. Intell. Eng. Syst.*, vol. 17, no. 6, pp. 1241–1267, 2024, doi: 10.22266/ijies2024.1231.91.
- [17] M. A. Adam, H. Y. Chrisnanto, and R. F. Umbara, “Prediksi Pola Keuangan pada Pasar Saham Bursa Efek Indonesia Menggunakan Algoritma Support Vector Machine For Regression (SVR),” *J. Edukasi dan Penelit. Inform.*, vol. 10, no. 3, pp. 476–484, 2024.
- [18] A. N. Maharadja, I. Maulana, and B. A. Dermawan, “Penerapan Metode Regresi Linear Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi,” *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 95–102, 2021, doi: 10.30871/jaic.v5i1.3184.
- [19] Muhammad Iqbal, Rozzi Kesuma Dinata, and Rizki Suwanda, “Comparison of Linear Regression and Polynomial Regression for Predicting Rice Prices in Lhokseumawe City,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 3, pp. 387–394, 2025, doi: 10.32736/sisfokom.v14i3.2396.
- [20] F. Riestiansyah, D. Damayanti, M. Reswara, and R. Susetyoko, “Perbandingan metode ARIMA dan ARIMAX dalam Memprediksi Jumlah Wisatawan Nusantara di Pulau Bali,” vol. 7, no. 2, 2022.
- [21] J. Halif, D. Wahiddin, I. Sanjaya, and S. Faisal, “Model Regresi Linear Berganda untuk Prediksi Tingkat Pengangguran di Provinsi Jawa Barat,” *J. Algoritm.*, vol. 22, no. 1, pp. 324–335, 2025, doi: 10.33364/algoritma/v.22-1.2312.
- [22] D. I. Mulyana and Marjuki, “Optimasi Prediksi Harga Uang Vaname Dengan Metode Rmse Dan Mae Dalam Algoritma Regresi Linear,” *J. Ilm. Betrik*, vol. 13, no. 1, pp. 50–58, 2022, doi: 10.36050/betrik.v13i1.439.
- [23] R. Linear and J. Zakaria, “Metode Regresi Linear Sederhana,” vol. 7, no. 4, pp. 2884–2888, 2023.
- [24] G. A. Mukhbita, I. Masti, N. I. Fazila, R. Wahyuni, and Y. Wulan, “Perbandingan Nilai Uts Dan Uas Matematika Siswa Kelas Iv Sdn Rawa Badak Utara 23 Menggunakan Uji Wilcoxon Signed Rank Test,” vol. 4, no. 2, pp. 83–87, 2025.