

Implementation of the Random Forest algorithm to predict rice needs in DKI Jakarta

Implementasi algoritma Random Forest untuk prediksi kebutuhan beras pada DKI Jakarta

Hadi Santoso^{1*}, Lukman Hakim², Afiyati Afiyati³, Bambang Jokonowo⁴

^{1,2,3} Teknik Informatika, Universitas Mercu Buana, Indonesia

^{1*}hadisantoso@mercubuana.ac.id, ²lukman_hakim@mercubuana.ac.id,

³afiyati.reno@mercubuana.ac.id, ⁴bambang.jokonowo@mercubuana.ac.id

*: Penulis korespondensi (corresponding author)

Informasi Artikel

Received: July 2024

Revised: September 2025

Accepted: October 2025

Published: February 2025

Keywords: predict; rice; random forest
Kata kunci: prediksi; beras; random forest

Abstract

Purpose : to build collaborative partners between government institutions and universities in food processing, especially rice, by predicting rice needs in the DKI Jakarta area.

Design/methodology/approach: The approach in this research uses the Random Forest algorithm which functions to predict rice needs in the DKI Jakarta area.

Results: rice demand prediction application with evaluation values Mean Squared Error 207.86, Mean Absolute Error 9.43, MAPE 0.048, Root Mean Squared Error 14.4, accuracy 0.63

Originality/value/state of the art: research using data from BAPANAS, Cipinang Main Market, with 2 datasets of rice stock, population, year and rice consumption using a random forest algorithm to predict rice needs in the DKI Jakarta area

Abstrak

Tujuan: Membangun mitra kerjasama antara lembaga pemerintah dengan Universitas dalam pengolahan pangan khususnya beras, dengan prediksi kebutuhan beras pada wilayah DKI Jakarta.

Perancangan/metodologi/pendekatan: Pendekatan dalam penelitian ini menggunakan algoritma Random Forest yang berfungsi untuk memprediksi kebutuhan beras pada wilayah DKI Jakarta.

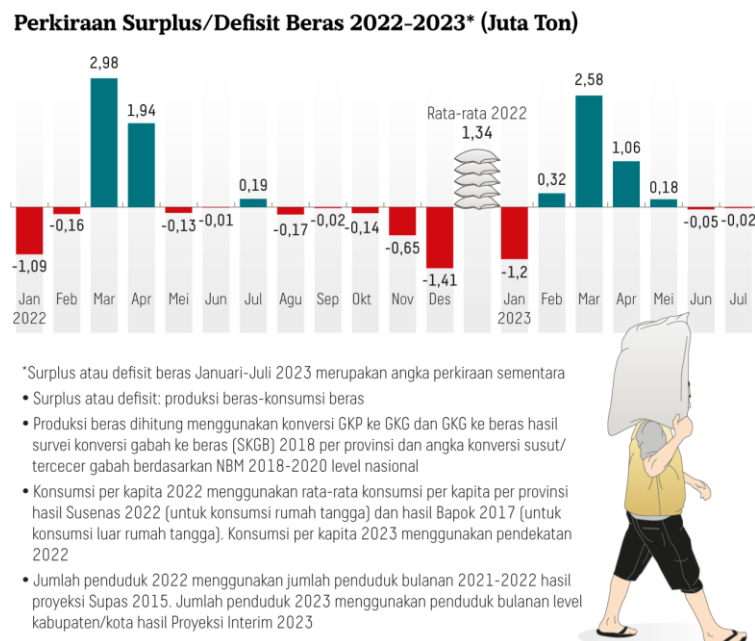
Hasil: aplikasi prediksi kebutuhan beras dengan nilai evaluasi Mean Squared Error 207.86, Mean Absolute Error 9.43, MAPE 0.048, Root Mean Squared Error 14.4, akurasi 0.63

Keaslian/ state of the art: Keaslian : penelitian menggunakan data dari Badan Pangan Nasional (Bapanas), Pasar Induk Cipinang, dengan 2 dataset stok beras, jumlah penduduk,

tahun dan konsumsi beras menggunakan algoritma Random Forest untuk prediksi kebutuhan beras wilayah DKI Jakarta

1. Pendahuluan

Beras merupakan bahan pokok utama masyarakat Indonesia, hampir 90% masyarakat Indonesia mengkonsumsi beras dan beras menjadi komoditas Indonesia yang utama [1], peran pemerintah sangat penting dalam pengendalian atau pengawasan serta pendistribusian beras ke berbagai wilayah Indonesia. Kenaikan beras akibat produksi beras berkurang [2] di pasokan atau tingkat pasar induk. Jakarta memiliki populasi besar dan beragam serta pusat kota yang sangat dinamis untuk konsumsi beras [3]. Berdasarkan data Kompas tahun 2021-2023, rata-rata stok beras melalui pasar induk Cipinang, Jakarta pernah menyentuh 30.000 ton, kini bahkan dibawah 25.000 ton, secara data beras nasional mengalami penurunan akibat penurunan produksi. Selama Juli-September 2023 mengalami defisit beras mencapai 150.000 ton-170.000 ton [4]. Penurunan produksi serta defisitnya bahan pokok ini harus menjadi perhatian penting dalam pengaturan stabilitas nasional, sehingga masyarakat tidak mengalami kesulitan dalam konsumsi beras. Ketersediaan beras pada suatu daerah akan berdampak ketidakstabilan perekonomian bangsa [5]. Berdasarkan hasil kerjasama Badan Pangan Nasional (Bapanas) memberikan data sebanyak 2000 dataset tentang ketersediaan dan kebutuhan beras DKI Jakarta dapat menjadi acuan dalam membangun aplikasi visualisasi prediksi konsumsi beras memberikan kemudahan pola kebutuhan konsumsi beras setiap wilayah DKI Jakarta [6]. Penelitian sebelumnya, Permintaan kebutuhan beras pada masyarakat dipengaruhi harga, rasa, pendapatan dan jumlah tanggungan [7]. Penelitian sebelumnya ketersediaan beras berdasarkan lahan pertanian dan produktivitas beras, semakin berkurangnya beras akibat berkurangnya lahan pertanian [2]. Pada **Gambar 1** menunjukkan perkiraan surplus/defisit beras sebagai berikut:



Gambar 1.Perkiraan Surplus/Defisit Beras

(sumber : Badan Pangan Nasional berdasarkan data KSA BPS April 2023)

Berdasarkan permasalahan tersebut perlu dilakukan pengolahan data kebutuhan pokok seperti beras dengan melakukan prediksi konsumsi beras pada setiap wilayah seperti DKI Jakarta, dengan hal tersebut dapat mengantisipasi kebutuhan beras masa akan datang. Prediksi konsumsi kebutuhan pokok seperti beras sangat penting untuk menyelesaikan permasalahan ketersediaan beras masa akan datang, Penggunaan algoritma *Random Forest Regression* merupakan algoritma dengan tingkat akurasi 80% untuk melakukan prediksi. *Random Forest* adalah metode *machine learning* yang dapat digunakan dalam klasifikasi dan regresi [8].

Beberapa penelitian sebelumnya yang relevan yaitu menurut Nahya Nur, Farid Wajidi, dkk berjudul Implementasi algoritma Random Forest regression untuk memprediksi panen padi di desa Minanga, penelitian ini memprediksi panen padi dengan 7 variabel yaitu luas lahan, jumlah bibit, jenis pupuk, curah hujan, hama dan gulma. Hasil penelitian dengan algoritma Random Forest akurasi 95,11% dan nilai MAPE 4.88% [8].

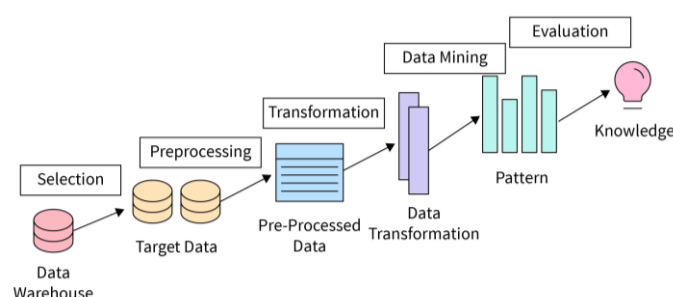
Penelitian selanjutnya yaitu menurut K. Choudhary, W. Shi, dkk berjudul *Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine*, penelitian ini memprediksi panen padi berdasarkan data sentinel-2 citra satelit berdasarkan padi dan bukan padi, Hasil penelitian tingkat akurasi 85% [9].

Penelitian menurut Ahmad Zamahzari dan Puryantoro berjudul Forecasting produksi padi dan konsumsi beras di Provinsi Jawa Timur, Hasil penelitian menggunakan ARIMA Box-Jenkins memberikan prediksi mengalami penurunan. Konsumsi beras Provinsi Jawa Timur diprediksi mengalami penurunan di tahun 2022, kemudian secara perlahan diprediksi mengalami kenaikan dari tahun 2023-2027 [10].

Berdasarkan beberapa penelitian, Tujuan penelitian ini adalah melakukan prediksi kebutuhan beras DKI Jakarta berkerjasama dengan Badan Pangan Nasional (Bapanas) dan Pasar Induk Cipinang. Data yang digunakan berasal dari Badan Pangan Nasional (Bapanas) berdasarkan data kebutuhan beras untuk wilayah Indonesia. Dalam penelitian ini menggunakan algoritma Random Forest untuk melakukan prediksi kebutuhan beras wilayah DKI Jakarta.

2. Metode

Penelitian ini menggunakan dataset yang berasal dari Badan Pangan Nasional (Bapanas) untuk mengetahui konsumsi penggunaan beras sebagai kebutuhan pokok pada DKI Jakarta, metode penelitian eksperimental untuk menghasilkan visual data yang akurat, menggunakan metode analisis Knowledge Discovery in database (KDD) dan algoritma Random Forest, dapat dilihat pada **Gambar 2** [11].



Gambar 2. Tahapan penelitian

2.1. Pengumpulan Data

Data diambil dari laporan beras dan konsumsi beras Badan Pangan Nasional (Bapanas) sebanyak 5000 dataset [12], yang diolah disesuaikan berdasarkan tahun, konsumsi penggunaan bahan pokok beras dari setiap wilayah dan tahun. Tahapan seleksi memilih data yang sesuai dengan mengurangi data yang tidak relevan dengan penelitian serta hanya memilih ketersediaan beras untuk wilayah DKI Jakarta saja.

2.2. Preprocessing

Data yang didapatkan di hilangkan atau dibersihkan yang duplikat, serta pemilihan atribut mana yang mempengaruhi prediksi konsumsi bahan pokok beras, dengan program aplikasi *visual code*, bahasa pemrograman Python. Terbagi 3 tahapan dalam preprocessing yaitu *Data Cleaning*, *Data Integration*, *Data Transformasi* [13]. *Data Cleaning* yaitu tahapan menyeleksi data mentah yang diberikan Badan Pangan Nasional (Bapanas) sebanyak 2000 dataset dengan menghilangkan data yang tidak relevan atau tidak diperlukan untuk menghindarkan kesalahan dalam pengolahan data. *Data Integration* yaitu tahapan pengecekan dan menyatukan data dari sumber lain, untuk menghasilkan data yang lebih akurat.

2.3. Transformasi

Penggunaan data yang sudah bersih, dengan transformasi menggunakan operator windowing, menggunakan jendela waktu untuk memastikan secara periode perubahan data. Tahapan ini merubah struktur dan format data yang kompleks menjadi lebih sederhana.

2.4. Data Mining

Pada tahapan ini menggunakan algoritma *Random Forest* untuk prediksi konsumsi bahan pokok beras, pada tahapan ini akan menghasilkan model yaitu prediksi.

2.4.1. Random Forest

Random Forest adalah jenis algoritma supervised learning yang merupakan pengembangan dari *Decision Tree*, *Random Forest* membentuk banyak pohon keputusan yang menentukan prediksi [8]. *Random Forest* metode yang masuk dalam klasifikasi dan regresi [14]. *Random Forest* memiliki kelebihan dari algoritma *Decision Tree* dalam beberapa pengujian tidak mengalami overfitting [15]. Rumus *Random Forest* sebagai berikut [13] :

$$H(x) = \frac{1}{T} \sum_{i=1}^T h_i(x) \quad (1)$$

dimana :

T : jumlah pohon regresi

$H_i(x)$: Keluaran pohon regresi ke- i (h_i) pada sampel x .

2.4.2. Mean Absolute Error

Mean Absolute Error (MAE) adalah merupakan metrik yang dapat mengukur kesalahan absolut suatu nilai rata-rata antara nilai prediksi dan nilai sebenarnya [13]:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (2)$$

dimana :

P_i dan O_i : nilai perkiraan dan pengamatan

n : jumlah sampel

2.4.3. Mean Squared Error

Mean Squared Error (MSE) merupakan metrik yang dapat mengukur selisih nilai rata-rata yang diperoleh dari jumlah kesalahan kuadrat dibagi dengan jumlah total nilai prediksi, sehingga metrik ini sangat cocok digunakan pada pembobotan yang lebih besar. nilai untuk nilai kesalahan yang lebih besar.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (3)$$

dimana :

n : jumlah sampel

\hat{Y}_i : nilai prediksi

Y_i : nilai observasi

2.5. Evaluasi

Tahapan evaluasi menggunakan rumus RMSE, MAE pada model yang dihasilkan. Evaluasi ini memberi gambaran apakah prediksi yang dilakukan menggunakan algoritma *Random Forest* memberikan hasil prediksi konsumsi bahan pokok beras secara periodik (bulan atau Tahunan) yang cukup signifikan tingkat kebenarannya. Rumus RMSE (*Root Mean Squared Error*) dan MAPE (*Mean Absolute Percentage Error*) sebagai berikut [8]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - Y_n)^2} \quad (4)$$

Semakin rendah nilai MAPE semakin baik model tersebut dalam melakukan prediksi. Rumus akurasi sebagai berikut :

$$Akurasi = 100\% - MAPE \quad (5)$$

Rumus nilai MAPE sebagai berikut :

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - Y_n|}{Y_i} \times 100\% \quad (6)$$

di mana:

Y_i : nilai Aktual

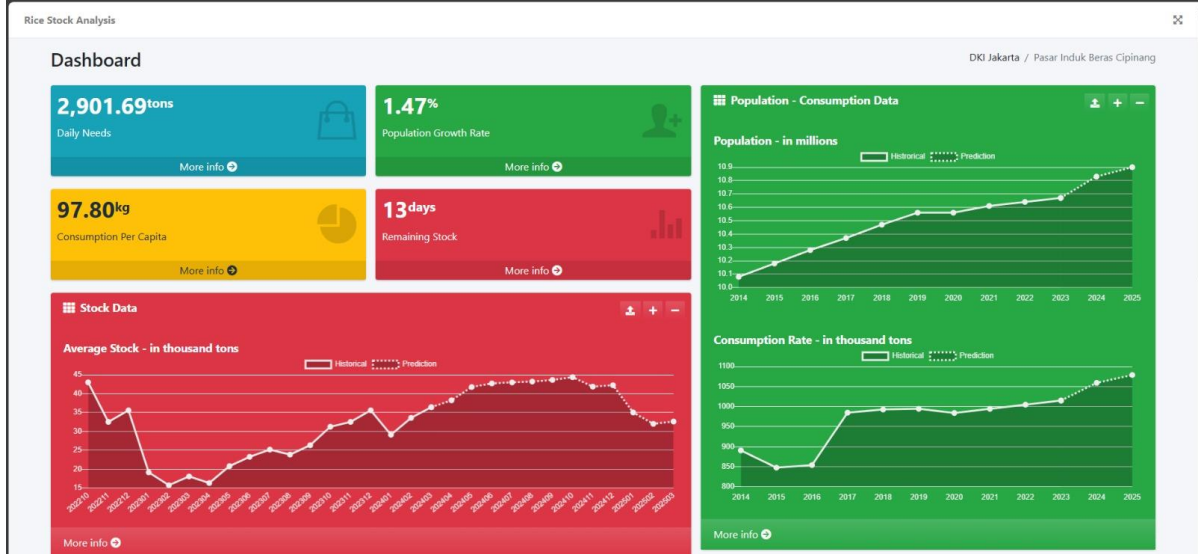
Y_n : nilai Prediksi

N : jumlah dataset sampel

3. Hasil dan Pembahasan

Penelitian menghasil aplikasi web yang dapat digunakan untuk mempermudah dalam pengolahan prediksi kebutuhan beras khususnya wilayah DKI Jakarta. Penelitian ini menggunakan algoritma *Random Forest* yang digunakan untuk melakukan prediksi kebutuhan beras. **Gambar 3** Halaman web untuk visualisasi prediksi kebutuhan beras sebagai berikut :

3.1. Tampilan Aplikasi Prediksi kebutuhan beras



Gambar 3. Halaman web prediksi kebutuhan beras DKI Jakarta

Aplikasi prediksi kebutuhan beras untuk wilayah DKI Jakarta, mengacu pada jumlah penduduk dan konsumsi masyarakat, sehingga pertumbuhan penduduk mempengaruhi kebutuhan konsumsi beras.

3.2. Preprocessing data

Tahapan 1, pengumpulan dataset pada Tabel 1, perkiraan ketersediaan dan kebutuhan beras yang perlu dilakukan cleaning menghilangkan beberapa atribut yang tidak dibutuhkan. Untuk dataset yang digunakan ada 2 file yaitu data penduduk, tingkat konsumsi dan rerata stok bulan di ambil dari data pasar induk sebagai berikut:

Tabel 1. Dataset ketersediaan dan kebutuhan beras

Perkiraan Ketersediaan						Perkiraan Kebutuhan			
Bulan	Stok Awal (Ton)	Produk si GKG (Ton)	Produk si (ton)	Baran g masu k (ton)	Barang keluar(to n)	Total ketersedia an (ton)	Kebutuh an RT (Ton)	Kebutuh an non RT (Ton)	Total Kebutuh an (Ton)
Januari 2020	340.4	0	0	213.6	0	532.654,3	60.796,9	1.114,83	62.631,8
Februari 2020	728.6	0	0	324.5	0	964.756.3	60.173,8	1.003,38	70.237,2
Maret 2020	114.6	0	0	112.3	0	214.788.3	70.335,0	1.116,97	72.562,0
April 2020	159.0	0	0	125.7	0	254.544,6	80.696,9	1.014,83	81.711,8
Mei 2020	202.2	0	0	112.3	0	354.442,6	78.093,8	1.073,38	79.237,2
Juni 2020	202.3	0	0	102.2	0	354.775,6	80.775,0	1.716,97	82.992,0
...

Perkiraan Ketersediaan						Perkiraan Kebutuhan			
Oktober 2023	496.9	0	0	164.1	0	661.119,5	80.696,9	2.214,83	82.911,8
	74,67			44,86		4	9		2
				4					
Novemb er 2023	578.4	0	0	141.4	0	719.851,6	78.093,8	2.143,38	80.237,2
	14,32			37,30		2	6		5
				7					
Desembe r 2023	639.8	0	0	125.1	0	764.924,4	80.775,0	2.216,97	82.992,0
	14,32			10,17		9	8	9	65
				5					

Berdasarkan **Tabel 1**, dilakukan cleansing dataset berdasarkan kebutuhan dalam penelitian untuk prediksi kebutuhan beras didapatkan pada **Tabel 2** sebagai berikut:

Tabel 2. Hasil Cleansing stok beras

Bulan	Tahun	Stok Avg (ton)
5	2020	1.716.97
6	2020	1.475.32
7	2020	2.124.34
...
11	2023	2.143.38
12	2023	2.216.97

Dataset populasi dan konsumsi beras pada Provinsi DKI Jakarta, menentukan prediksi kebutuhan beras masa akan datang terdapat pada **Tabel 3** sebagai berikut:

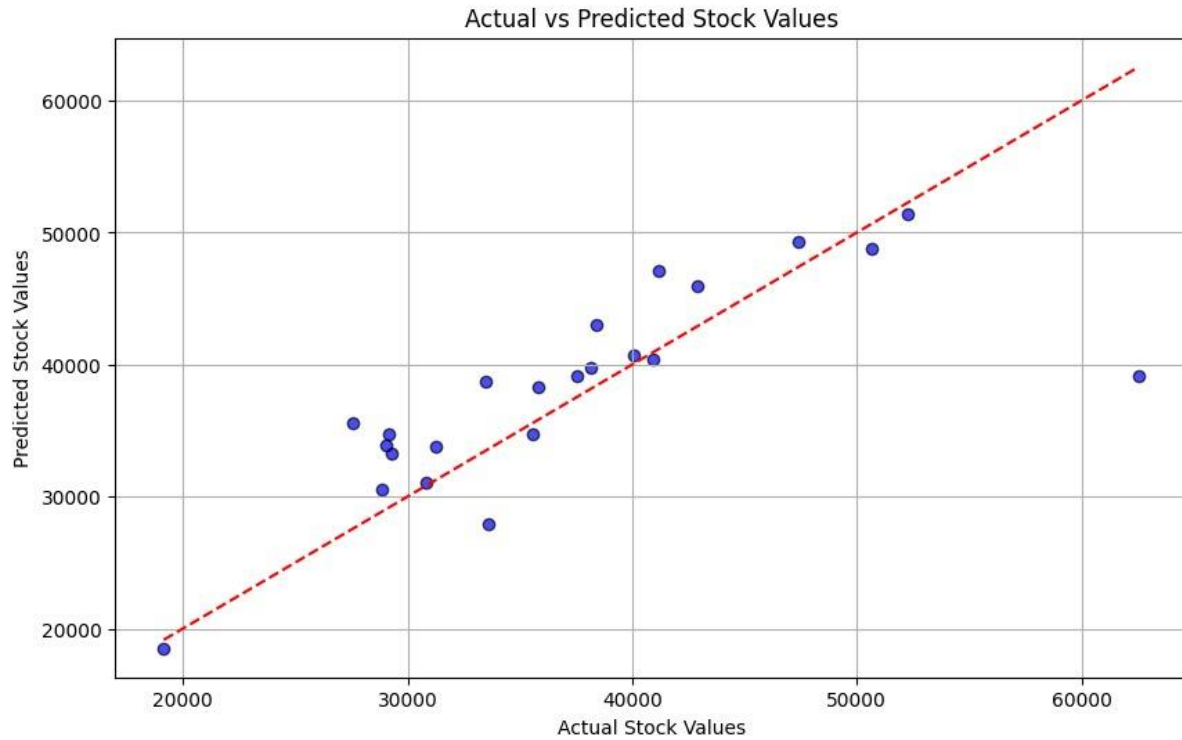
Tabel 3. Dataset Populasi dan konsumsi DKI Jakarta

Tahun	Populasi	Konsumsi (ton)
2013	9.969.948	851.433,55
2014	10.075.310.8	890.657.40
2015	10.177.924	847.821,06
...
2021	10.609.681	994.339,30
2022	10.640.007	1.004.714,48
2023	10.672.100	1.015.089,18

Stok beras dan konsumsi menentukan prediksi kebutuhan beras, jumlah konsumsi beras pada masyarakat atau daya beli, meningkatkan kebutuhan masa yang akan datang, dengan pola setiap bulan dan tahun akan menentukan prediksi tahun depan kebutuhan beras.

3.3. Evaluasi Model

Evaluasi merupakan tahapan untuk mengetahui sejauh mana kinerja model yang digunakan dengan data testing. Pengujian suatu algoritma Random Forest dikatakan baik dengan menguji nilai RMSE, Akurasi dan MAE. Pada **Gambar 4** dan **Tabel 4** merupakan nilai RMSE dari pengujian sebagai berikut:



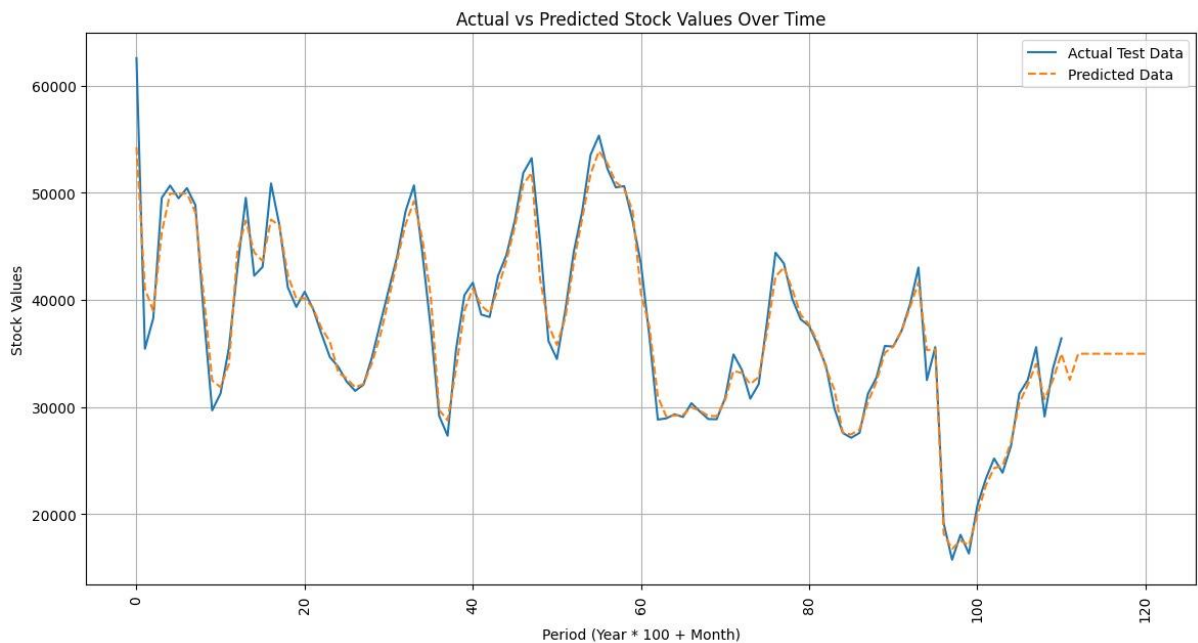
Gambar 4. Prediksi aktual dan stok kebutuhan beras

Berdasarkan nilai evaluasi nilai MAE, RMSE dan akurasi pada pengujian data pembelajaran sebagai berikut :

Tabel 4. Evaluasi pengujian pada algoritma *Random Forest*

Evaluasi	Hasil
Mean Squared Error (MSE)	207.86
Mean Absolute Error (MAE)	9.43
MAPE	0.048
Root Mean Squared Error (RMSE)	14.4
Akurasi	0.63

Berdasarkan hasil evaluasi *Mean Squared Error* 207.86 masih dianggap baik tergantung skala yang digunakan, nilai *Mean Absolute Error* pada pengujian 9.43 masih kurang untuk dengan skala 0-5 masih dianggap baik, untuk *Root Mean Squared Error* 14.4 masih tergolong kurang baik dengan skala 0-10 maksimal dikatakan baik. MAPE dengan nilai 0.048 sangat baik secara nilai aktual, sedangkan untuk tingkat akurasi 63% masih dianggap kurang baik. Penerapan algoritma *Random Forest* untuk penerapan time series dan prediksi masih belum baik secara evaluasi. Hasil ujicoba prediksi mutli tahun pada algoritma Random Forest, pada **Gambar 5**, sebagai berikut.



Gambar 5. Hasil prediksi kebutuhan beras 2 tahun

Gambar 5, menunjukkan kinerja algoritma *Random forest* untuk menampilkan prediksi *time series* tidak cukup berhasil, pola prediksi cenderung mengacu pada tahun sebelumnya. Kegagalan untuk penerapan *time series* pada algoritma *Random Forest*, sebagai metode ensemble dari pohon keputusan, tidak dirancang untuk menangkap keterkaitan temporal ini secara langsung.

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian dan pengujian aplikasi prediksi kebutuhan beras menggunakan algoritma *Random Forest* dengan variabel bulan, tahun, stok dan populasi, konsumsi masyarakat menjadi acuan dalam menentukan prediksi kebutuhan beras masa akan datang dengan nilai akurasi 63%, MSE 207.86, MAE 9.43, MAPE 0.048 dan RMSE 14.4 dengan rata-rata nilai masih kurang baik dalam penerapan prediksi kebutuhan beras. Untuk aplikasi dapat memprediksi kebutuhan beras secara konstan mengacu pada tahun sebelumnya.

Daftar Pustaka

- [1] H. Santoso, Hadi;Hakim, Lukman;Afiyati; Magdalena, "Jurnal abdidias," *Sos. Dampak Kenaikan Beras dengan Prediksi Kebutuhan Beras Masy. di Pasar Induk Cipinang dengan Kerjasama Badan Pangan Nas.*, vol. 5, no. 2, pp. 90–96, 2024.
- [2] M. F. F. Mardianto, E. Tjahjono, M. Rifada, A. Herawanto, A. L. Putra, and K. A. Utama, "The Prediction of Rice Production in Indonesia Provinces for Developing Sustainable Agriculture," *J. First Int. Conf. Food Agric.*, vol., no., pp. 325–333, 2018.
- [3] Hakim, Lukman;Santoso, Hadi; yusuf, M "Jurnal abdidias," *Sos. Peran Teknol. Artif. Intel. untuk Klasifikasi Status Sos. Masy. DKI Jakarta*, vol. 5, no. 3, pp. 97–102, 2024.

- [4] M. P. J. J, “Pasokan Berkurang, Pasar Beras Tertekan,” Kompas. Accessed: Nov. 23, 2023. [Online]. Available: <https://www.kompas.id/baca/ekonomi/2023/08/13/pasokan-berkurang-pasar-beras-tertekan>
- [5] Y. Yulida and M. A. Karim, “Prediction of rice consumption in South Kalimantan by considering population growth rate,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 758, no. 1, 2021, doi: 10.1088/1755-1315/758/1/012022.
- [6] Badan Pangan Nasional, “Dataset Permintaan beras,” 2023.
- [7] and W. Wachdijono, Dwi Purnomo, “Rice Demand and Its Predictions as a Basis for Realizing Food Security Programs in Higher Education Institutions,” *1st Int. Agric. Nutraceutical, and Food Science 2022 “Praising The Trop. Resour. Glorif. Biodiversity Potential of Nusantara” November 9-10th*, no. 5, pp. 103–117, 2022, [Online]. Available: <https://jurnal.untidar.ac.id/index.php/icanfs/article/view/6640>
- [8] T. Online, N. Nur, and F. Wajidi, “Jurnal Politeknik Caltex Riau Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga,” *J. Komput. Terap.*, vol. 9, no. 1, pp. 58–64, 2023.
- [9] K. Choudhary, W. Shi, Y. Dong, and R. Paringer, “Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine,” *Adv. Sp. Res.*, vol. 70, no. 8, pp. 2443–2457, 2022, doi: 10.1016/j.asr.2022.06.073.
- [10] P. Padi, “Forecasting padi dan konsumsi Beras Jawa Timur,” vol. 20, 2023.
- [11] M. Adjie Setyadj, A. Faqih, and Y. Arie Wijaya, “Peramalan Harga Komoditas Beras Di Kalimantan Timur Menggunakan Algoritma Neural Network,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 320–324, 2023, doi: 10.36040/jati.v7i1.6327.
- [12] BPN, “Laporan Beras DKI Jakarta,” Jakarta, 2024.
- [13] E. Pramudianzah, Y. S. Triana, and R. Budiarto, “Analysis of Waste Transportation Volume in Jakarta Province using Linear Regression and Random Forest Regression,” *ACM Int. Conf. Proceeding Ser.*, vol. i, 2022, doi: 10.1145/3557738.3557876.
- [14] T. A. U. Azmi, L. Hakim, D. C. R. Novitasari, and W. D. U. D. Utami, “Application Random Forest Method for Sentiment Analysis in Jamsostek Mobile Review,” *Telematika*, vol. 20, no. 1, p. 117, 2023, doi: 10.31315/telematika.v20i1.8868.
- [15] D. S. B M, D. C. N K, D. P. T, and D. R. R, “Rice and Wheat Yield Prediction in India Using Decision Tree and Random Forest,” *Comput. Intell. Mach. Learn.*, vol. 3, no. 2, pp. 1–8, 2022, doi: 10.36647/ciml/03.02.a001.