

Prediksi Nilai Tukar Petani (NTP) di Indonesia Menggunakan Linear Regression dan Random Forest

Farmers' Terms of Trade (NTP) Prediction in Indonesia Using Linear Regression and Random Forest

Naufal Alif Anargya*¹, Chelsisdeo A.P Sanenek², Egi Rizkiyansyah³, Diah Septiani⁴

^{1,2,3,4}Program Studi Sains Data, Fakultas Informatika, Telkom University Purwokerto
e-mail:

¹anargya@student.telkomuniversity.ac.id, ²chelsisdeo@student.telkomuniversity.ac.id,

³egirizki@student.telkomuniversity.ac.id, ⁴diahseptiani@telkomuniversity.ac.id

Abstrak - Nilai Tukar Petani (NTP) merupakan salah satu indikator penting untuk mengukur tingkat kesejahteraan petani di Indonesia, sehingga kemampuan untuk memprediksi nilai NTP di masa depan sangat bermanfaat bagi pengambilan kebijakan di sektor pertanian. Penelitian ini bertujuan untuk membangun model prediksi NTP per provinsi pada periode 2021–2026 menggunakan algoritma Linear Regression dan Random Forest Regressor. Data yang digunakan berasal dari data NTP bulanan per provinsi yang dipublikasikan oleh Badan Pusat Statistik (BPS), dengan fitur prediktor berupa Tahun, Provinsi, dan Bulan yang telah dikodekan secara numerik. Data dibagi menjadi data latih dan data uji dengan proporsi 80:20, kemudian performa model dievaluasi menggunakan metrik MAE, MSE, RMSE, dan R^2 . Hasil penelitian menunjukkan bahwa model Random Forest memberikan performa yang jauh lebih baik dibandingkan Linear Regression, dengan nilai R^2 sebesar 0,947 dan RMSE sebesar 4,60, dibandingkan Linear Regression yang hanya memperoleh R^2 sebesar 0,062 dan RMSE sebesar 19,38. Hasil ini menunjukkan bahwa hubungan antara fitur Tahun, Provinsi, dan Bulan terhadap NTP bersifat non-linear, sehingga model berbasis ensemble tree seperti Random Forest jauh lebih sesuai digunakan dibandingkan model regresi linear sederhana.

Kata kunci — linear regression; nilai tukar petani; prediksi; random forest; sains data pertanian

Abstract - The Farmers' Terms of Trade (NTP) is an important indicator used to measure the welfare level of farmers in Indonesia, making the ability to predict future NTP values highly valuable for agricultural policy making. This study aimed to build a prediction model for NTP at the provincial level for the period 2021–2026 using the Linear Regression and Random Forest Regressor algorithms. The data used were monthly provincial NTP data published by the Indonesian Central Statistics Agency (BPS), with Year, Province, and Month as encoded predictor features. The data were split into training and testing sets with an 80:20 ratio, and model performance was evaluated using MAE, MSE, RMSE, and R^2 metrics. The results showed that the Random Forest model performed significantly better than Linear Regression, achieving an R^2 of 0.947 and an RMSE of 4.60, compared to Linear Regression which only achieved an R^2 of 0.062 and an RMSE of 19.38. These findings indicate that the relationship between Year, Province, and Month features and NTP is non-linear, making ensemble tree-based models such as Random Forest far more suitable than simple linear regression models.

Keywords — agricultural data science; farmers' terms of trade; linear regression; prediction; random forest

I. PENDAHULUAN

Nilai Tukar Petani (NTP) merupakan salah satu indikator utama yang digunakan pemerintah Indonesia untuk mengukur tingkat kesejahteraan petani. NTP dihitung sebagai rasio antara indeks harga yang diterima petani (It) dengan indeks harga yang dibayar petani (Ib). Apabila nilai NTP berada di atas 100, maka secara umum petani mengalami kondisi yang menguntungkan, karena harga produk pertanian yang dijual relatif lebih tinggi dibandingkan biaya produksi dan kebutuhan konsumsi rumah tangga yang harus dikeluarkan. Sebaliknya, apabila nilai NTP berada di bawah 100, maka kondisi tersebut menunjukkan bahwa petani berada pada posisi yang kurang menguntungkan secara ekonomi.

Nilai NTP di Indonesia menunjukkan variasi yang cukup signifikan antar provinsi maupun antar waktu. Beberapa provinsi seperti Bengkulu dan Riau secara konsisten memiliki rata-rata NTP yang tinggi, sementara provinsi lain berada pada kisaran nilai yang lebih rendah. Selain perbedaan antar wilayah, NTP juga menunjukkan tren peningkatan dari tahun ke tahun, sehingga faktor waktu turut memberikan pengaruh terhadap nilai NTP yang tercatat. Variasi ini menjadikan NTP sebagai data yang menarik untuk dianalisis lebih lanjut menggunakan pendekatan sains data, khususnya untuk membangun model yang dapat memprediksi nilai NTP berdasarkan informasi provinsi, tahun, dan bulan.

Beberapa penelitian terdahulu telah membahas penerapan metode machine learning untuk memprediksi NTP maupun indikator ekonomi pertanian lainnya. Penelitian peramalan NTP di Kalimantan Timur menggunakan pendekatan Neural Network menunjukkan bahwa metode berbasis pembelajaran mesin mampu menangkap pola data NTP dengan cukup baik dibandingkan metode statistik konvensional [1]. Pendekatan serupa juga diterapkan pada data NTP Sumatera Utara menggunakan Recurrent Neural Network dengan optimasi ADAM, yang menunjukkan kemampuan model dalam menangkap pola temporal pada data NTP [2]. Selain pendekatan deep learning, metode hibrida Singular Spectrum Analysis (SSA) dengan SARIMA juga telah digunakan untuk meramalkan NTP nasional dan menunjukkan hasil peramalan yang cukup akurat [3]. Di sisi lain, kajian mengenai faktor-faktor yang memengaruhi NTP di Indonesia menunjukkan bahwa nilai NTP dipengaruhi oleh berbagai faktor ekonomi makro maupun karakteristik wilayah [4].

Pada bidang machine learning untuk regresi, algoritma Random Forest Regression telah banyak digunakan untuk memprediksi nilai-nilai ekonomi maupun hasil pertanian karena kemampuannya dalam menangkap pola data non-linear dan interaksi antar fitur. Penelitian perbandingan antara Linear Regression, Random Forest Regression, dan Gradient Boosted Trees Regression dalam memprediksi harga rumah menunjukkan bahwa model berbasis pohon keputusan secara konsisten memberikan hasil prediksi yang lebih akurat dibandingkan regresi linear sederhana

[5]. Temuan serupa juga diperoleh pada penelitian perbandingan algoritma Linear Regression dan Random Forest untuk memprediksi harga bawang merah, di mana Random Forest menghasilkan nilai error yang lebih kecil dibandingkan Linear Regression [6].

Berdasarkan uraian tersebut, terdapat kesenjangan (gap) berupa belum banyaknya penelitian yang secara khusus membandingkan performa model Linear Regression dan Random Forest Regression pada data NTP multi-provinsi di Indonesia dengan rentang waktu yang relatif panjang (2021–2026). Penelitian ini berkontribusi dengan membangun dan membandingkan kedua model tersebut menggunakan fitur Tahun, Provinsi, dan Bulan, serta mengevaluasi performanya menggunakan metrik MAE, MSE, RMSE, dan R^2 . Hasil dari penelitian ini diharapkan dapat memberikan gambaran mengenai metode prediksi yang paling sesuai untuk data NTP, sekaligus menjadi dasar bagi pengembangan model prediksi NTP yang lebih kompleks pada penelitian selanjutnya.

Tujuan dari penelitian ini adalah:

- 1) melakukan eksplorasi data NTP per provinsi di Indonesia periode 2021–2026;
- 2) membangun model prediksi NTP menggunakan algoritma Linear Regression dan Random Forest Regression; serta
- 3) membandingkan performa kedua model berdasarkan metrik evaluasi MAE, MSE, RMSE, dan R^2 guna menentukan model yang paling sesuai untuk memprediksi nilai NTP.

II. TINJAUAN TEORI

Nilai Tukar Petani (NTP) adalah perbandingan antara indeks harga yang diterima petani dengan indeks harga yang dibayar petani, yang digunakan sebagai salah satu proksi tingkat kesejahteraan petani. Data NTP dipublikasikan secara rutin oleh Badan Pusat Statistik (BPS) untuk setiap provinsi di Indonesia setiap bulannya. Faktor-faktor yang memengaruhi NTP di antaranya adalah harga komoditas pertanian, biaya produksi, kondisi geografis wilayah, serta kebijakan ekonomi makro [4].

Linear Regression merupakan algoritma machine learning yang digunakan untuk memodelkan hubungan linear antara satu atau lebih variabel independen (fitur) dengan variabel dependen (target). Model ini bekerja dengan mencari garis atau hyperplane terbaik yang meminimalkan selisih kuadrat antara nilai aktual dan nilai prediksi. Linear Regression memiliki keunggulan berupa interpretasi yang mudah, namun memiliki kelemahan dalam menangkap hubungan non-linear maupun interaksi kompleks antar fitur.

Random Forest Regression merupakan algoritma ensemble learning yang membangun sejumlah pohon keputusan (decision tree) secara acak, kemudian menggabungkan hasil prediksi dari seluruh pohon tersebut (biasanya dengan merata-ratakan) untuk menghasilkan prediksi akhir. Pendekatan ensemble ini membuat Random Forest lebih tahan terhadap overfitting dan mampu menangkap hubungan non-linear maupun interaksi antar fitur yang kompleks, sehingga sering digunakan dalam berbagai kasus prediksi di bidang ekonomi dan pertanian [5][6].

Untuk mengevaluasi performa model regresi, penelitian ini menggunakan empat metrik utama, yaitu Mean Absolute Error (MAE) yang mengukur rata-rata selisih absolut antara nilai aktual dan prediksi, Mean Squared Error (MSE) yang mengukur rata-rata kuadrat dari selisih tersebut, Root Mean Squared Error (RMSE) yang merupakan akar dari MSE dan memiliki satuan yang sama dengan variabel target, serta koefisien determinasi (R^2) yang menunjukkan proporsi variansi data target yang dapat dijelaskan oleh model. Semakin kecil nilai MAE, MSE, dan RMSE, serta semakin mendekati 1 nilai R^2 , maka semakin baik performa model dalam melakukan prediksi.

III. METODE

Data yang digunakan dalam penelitian ini adalah data Nilai Tukar Petani (NTP) menurut provinsi (2018=100) yang dipublikasikan oleh Badan Pusat Statistik (BPS) untuk periode tahun 2021 hingga 2026, dalam format file Excel per tahun. Setiap file tahunan diproses menggunakan fungsi pembersihan data (clean_ntp) yang mengubah struktur data dari format lebar (wide format, kolom per bulan) menjadi format panjang (long format), sehingga setiap baris data merepresentasikan kombinasi Provinsi, Bulan, Tahun, dan nilai NTP. Data dari keenam tahun tersebut kemudian digabungkan menjadi satu dataset utama yang terdiri dari 2.808 baris dan 4 kolom, yaitu Provinsi, Bulan, NTP, dan Tahun.

Tahap praproses data meliputi pembersihan nilai kosong (missing value) menggunakan fungsi dropna(), serta konversi tipe data kolom NTP dari object menjadi numerik agar dapat digunakan dalam proses pemodelan. Setelah proses pembersihan, dilakukan analisis statistik deskriptif yang menunjukkan rata-rata nilai NTP sebesar 115,89 dengan standar deviasi 19,77, nilai minimum 91,46, dan nilai maksimum 211,10.

Pada tahap feature engineering, kolom Provinsi dan Bulan yang bertipe kategorikal dikodekan menjadi nilai numerik menggunakan LabelEncoder, menghasilkan kolom Provinsi_Encoded dan Bulan_Encoded. Fitur yang digunakan sebagai variabel prediktor (X) adalah Tahun, Provinsi_Encoded, dan Bulan_Encoded, sedangkan variabel target (y) adalah nilai NTP. Data kemudian dibagi menjadi data latih (training) dan data uji (testing) dengan proporsi 80:20 menggunakan fungsi train_test_split dengan random_state=42. Untuk model Linear Regression, fitur

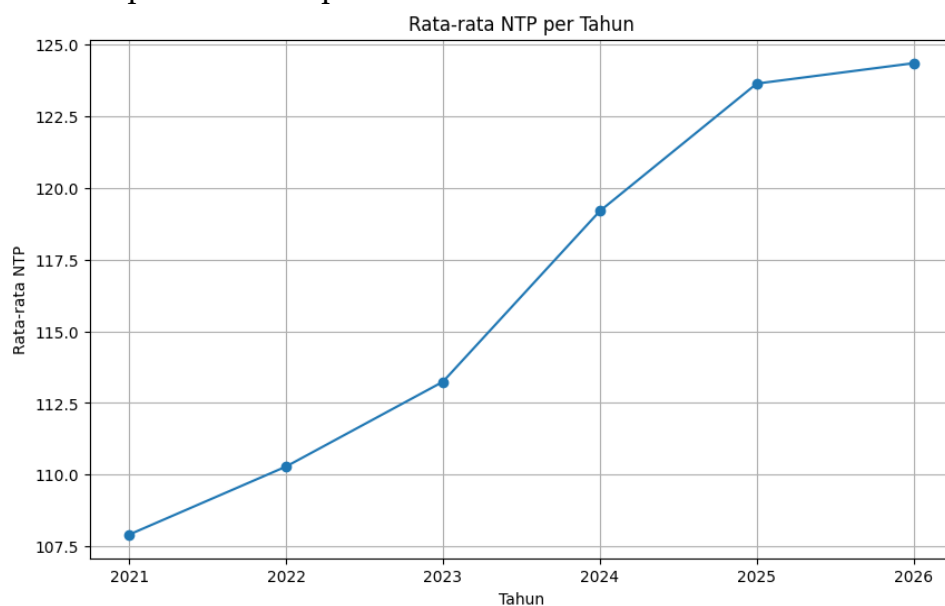
prediktor terlebih dahulu dinormalisasi menggunakan StandardScaler, sedangkan untuk model Random Forest digunakan data tanpa scaling karena algoritma berbasis pohon keputusan tidak sensitif terhadap skala fitur.

Dua model regresi dibangun dan dievaluasi pada penelitian ini, yaitu Linear Regression menggunakan pustaka scikit-learn dengan parameter default, dan Random Forest Regression menggunakan RandomForestRegressor dari scikit-learn. Performa kedua model dievaluasi pada data uji menggunakan empat metrik, yaitu MAE, MSE, RMSE, dan R^2 Score. Selain itu, dilakukan pula analisis feature importance pada model Random Forest untuk mengidentifikasi fitur yang paling berpengaruh terhadap hasil prediksi NTP.

IV. HASIL DAN PEMBAHASAN (11 POINT)

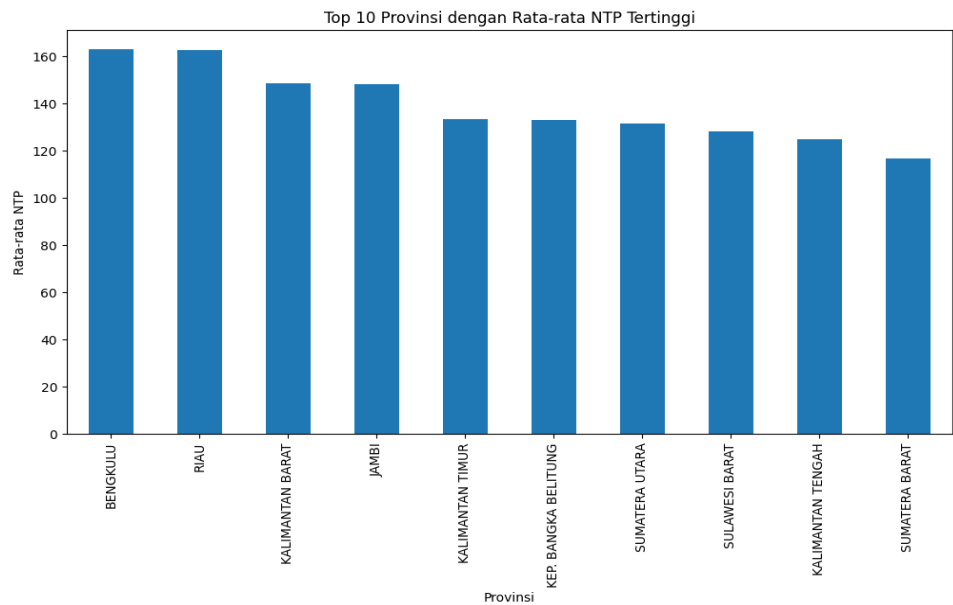
A. Eksplorasi Data (EDA)

Rata-rata nilai NTP mengalami peningkatan dari sekitar 108 pada tahun 2021 menjadi sekitar 124 pada tahun 2026, menunjukkan tren peningkatan kesejahteraan petani dalam periode tersebut.



Gambar 3. 1Rata-rata NTP per Tahun

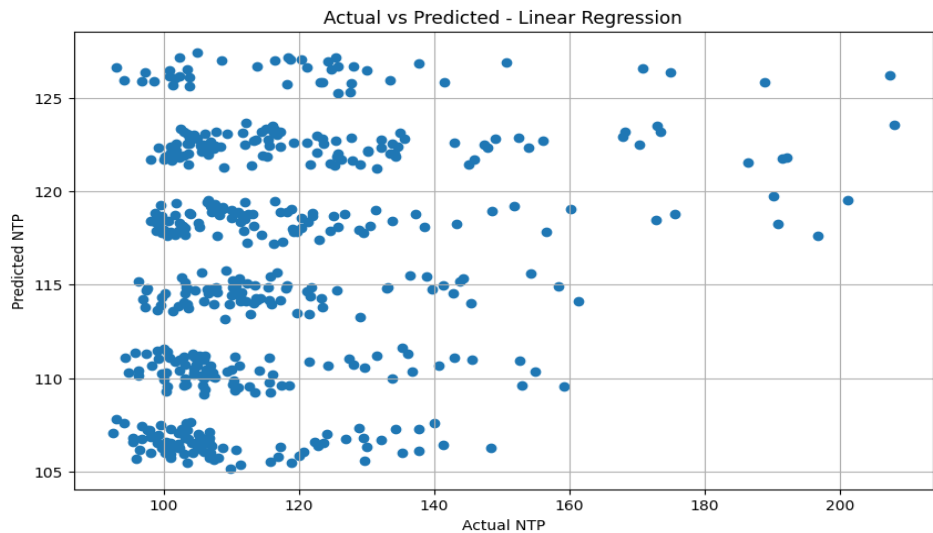
Bengkulu dan Riau memiliki rata-rata NTP tertinggi (sekitar 162), diikuti Kalimantan Barat, Jambi, dan Kalimantan Timur, menunjukkan bahwa faktor wilayah/provinsi sangat memengaruhi nilai NTP — temuan ini akan dikonfirmasi kembali pada analisis feature importance.



Gambar 3. 2 Top 10 Provinsi dengan Rata-rata NTP Tertinggi

B. Hasil Model Linear Regression

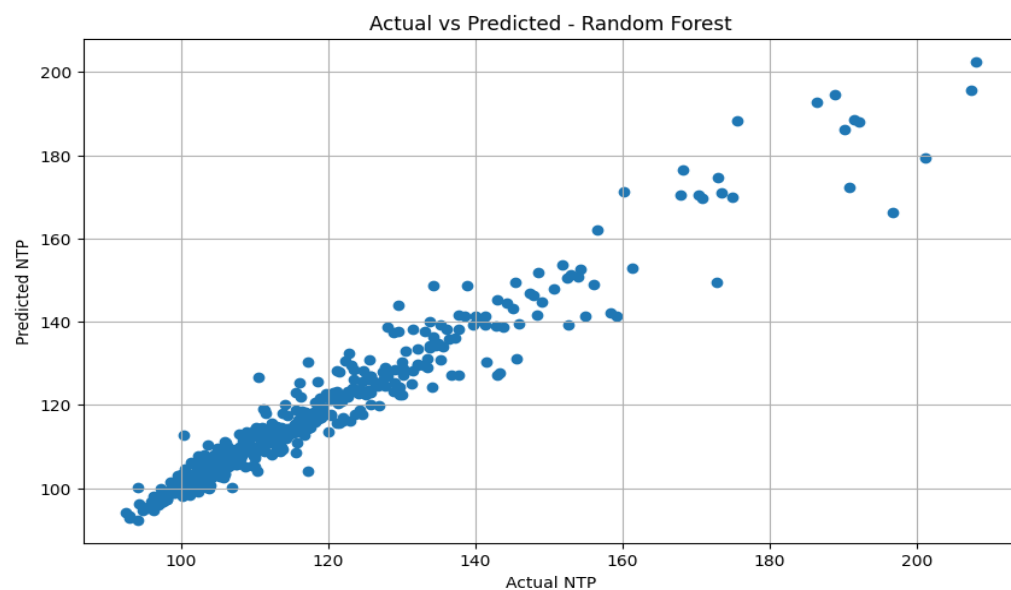
Model Linear Regression menunjukkan performa yang kurang baik, dengan R^2 hanya 0,062 dan RMSE sebesar 19,38. Hal ini terlihat dari scatter plot di mana titik-titik prediksi tidak mengikuti garis diagonal ideal. Prediksi model hanya berkisar 105–127, sedangkan nilai aktual memiliki rentang hingga di atas 200, sehingga saat nilai aktual tinggi, model tetap memprediksi rendah dan menghasilkan error besar. Hal ini menunjukkan Linear Regression mengalami underfitting karena tidak mampu menangkap hubungan non-linear antara fitur Tahun, Provinsi, dan Bulan terhadap NTP.



Gambar 3. 3 Actual vs Predicted - Linear Regression

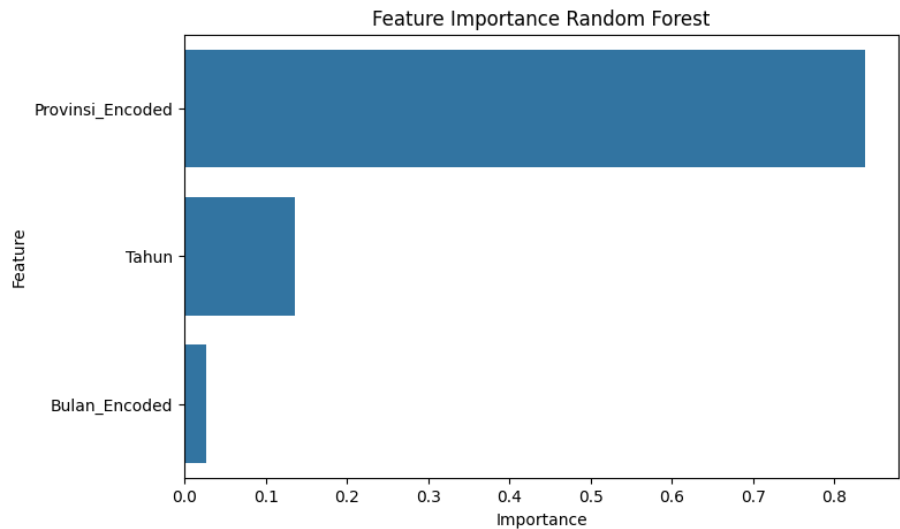
C. Hasil Model Random Forest

Model Random Forest menunjukkan performa sangat baik dengan MAE = 2,82, RMSE = 4,60, dan $R^2 = 0,947$, yang berarti model mampu menjelaskan 94,7% variasi nilai NTP. Pada scatter plot, sebagian besar titik mengikuti pola diagonal dengan rapat pada rentang NTP 90–150. Namun pada rentang NTP di atas 150, titik mulai menyebar dan cenderung berada di bawah diagonal (underestimate), mengindikasikan model kurang optimal pada nilai NTP ekstrem tinggi yang jarang muncul di data training.



Gambar 3. 4 Actual vs Predicted - Random Forest

Hasil feature importance menunjukkan Provinsi_Encoded mendominasi (~0,83), menegaskan bahwa perbedaan karakteristik antar provinsi adalah faktor paling menentukan dalam prediksi NTP — sejalan dengan temuan EDA pada Gambar 2. Fitur Tahun berkontribusi sekitar 0,13, mencerminkan tren NTP antar tahun, sedangkan Bulan_Encoded hanya 0,03, menunjukkan fluktuasi musiman bulanan tidak signifikan.



Gambar 3. 5 Feature Importance Random Forest

D. Perbandingan Hasil Evaluasi Model

Tabel 3. 1 Perbandingan Hasil Evaluasi Model

Model	MAE	MSE	RMSE	R ² Score
Linear Regression	13,66	375,75	19,38	0,062
Random Forest	2,82	21,16	4,60	0,947
SVR	11,81	422,25	20,55	-0,054
XGBoost	2,79	17,27	4,16	0,957

Berdasarkan hasil evaluasi model pada Tabel 1, XGBoost menunjukkan performa terbaik dengan nilai MAE sebesar 2,79, MSE sebesar 17,27, RMSE sebesar 4,16, dan R² Score sebesar 0,957, yang menandakan bahwa model mampu menjelaskan sekitar 95,7% variasi data NTP. Random Forest juga memberikan hasil yang sangat baik dengan R² sebesar 0,947, meskipun sedikit lebih rendah dibandingkan XGBoost. Sementara itu, Linear Regression hanya memperoleh R² sebesar 0,062, yang menunjukkan kemampuan prediksi yang rendah karena hubungan linear sederhana tidak mampu merepresentasikan pola data secara optimal. Di sisi lain, Support Vector Regression (SVR) menghasilkan R² negatif (-0,054), yang mengindikasikan bahwa performa model lebih buruk dibandingkan prediksi menggunakan rata-rata data. Secara keseluruhan, hasil ini menunjukkan bahwa model berbasis ensemble, khususnya XGBoost dan Random Forest, lebih efektif dalam menangkap pola kompleks pada data NTP dibandingkan model regresi konvensional.

V. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun dan membandingkan model prediksi Nilai Tukar Petani (NTP) di Indonesia menggunakan algoritma Linear Regression dan Random Forest Regression berdasarkan fitur Tahun, Provinsi, dan Bulan. Hasil evaluasi menunjukkan bahwa model Random Forest memberikan performa yang jauh lebih baik dibandingkan Linear Regression, dengan nilai R^2 sebesar 0,947 dan RMSE sebesar 4,60, sedangkan Linear Regression hanya memperoleh R^2 sebesar 0,062 dan RMSE sebesar 19,38, sehingga dapat disimpulkan bahwa hubungan antara fitur yang digunakan dengan nilai NTP bersifat non-linear dan lebih sesuai dimodelkan menggunakan pendekatan ensemble tree seperti Random Forest. Hasil analisis feature importance juga menunjukkan bahwa faktor provinsi merupakan penentu utama nilai NTP, diikuti oleh faktor tahun, sementara faktor bulan memberikan pengaruh yang relatif kecil. Meskipun performa model Random Forest sudah sangat baik, penelitian ini masih memiliki keterbatasan, di antaranya penggunaan pembagian data latih dan uji secara acak yang berpotensi menimbulkan kebocoran informasi pada data time series, serta belum digunakannya fitur tambahan yang relevan seperti harga komoditas pertanian atau kondisi cuaca. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan pembagian data berbasis waktu (time-based split), menambahkan fitur lag NTP periode sebelumnya, serta mengeksplorasi algoritma lain seperti Support Vector Regression dan XGBoost untuk memperoleh model prediksi NTP yang lebih akurat dan robust.

VI. DAFTAR PUSTAKA

- [1] "Peramalan Nilai Tukar Petani Kalimantan Timur Menggunakan Metode Neural Network," *Indones. J. Appl. Stat. Data Sci.*, vol. 2, no. 1, hal. 1–11, 2025.
- [2] A. Damanik, O. Harefa, N. S. U. Purba, dan J. T. Hardinata, "Prediksi RNN Terhadap Nilai Tukar Petani Sumatera Utara Menggunakan ADAM," *J. Komput. Teknol. Inf. Sist. Komput.*, vol. 4, no. 3, hal. 2125–2134, 2026.
- [3] O. A. Setyaningrum, E. Zukhronah, dan S. S. Handajani, "Peramalan Nilai Tukar Petani (NTP) Di Indonesia Menggunakan Metode Hibrida Singular Spectrum Analysis (SSA)-SARIMA," *Pros. Semin. Nas. Pendidik. Mat. (SENPIKA)*, vol. 1, hal. 254–266, Des 2023.
- [4] S. S. Aulia, D. S. Rimbodo, dan M. G. Wibowo, "Faktor-faktor yang Mempengaruhi Nilai Tukar Petani (NTP) di Indonesia," *J. Econ. Bus. Aseanomics*, vol. 6, no. 1, hal. 44–59, 2021.
- [5] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *J. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, hal. 58–64, Jul 2023.
- [6] M. A. Pratama, M. Munawaroh, dan W. J. Pranoto, "Perbandingan Performa Algoritma Linear Regresi dan Random Forest untuk Prediksi Harga Bawang Merah di Kota Samarinda," *TEKTONIK J. Ilmu Tek.*, vol. 1, no. 2, hal. 172–182, 2024.
- [7] "Implementasi Algoritma Support Vector Regression dalam Memprediksi Harga Sayuran Komoditas," *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 5, Okt 2024.
- [8] "Penerapan Support Vector Regression Kernel Linier dalam Meramalkan Harga CPO Indonesia," *Equator J. Math. Stat. Sci.*, 2024.
- [9] "Prediksi Angka Harapan Hidup Menggunakan Random Forest dan XGBoost Regression," *JICode J. Inform. Komput.*, vol. 2, no. 1, hal. 112–121, 2025.
- [10] "Perbandingan Kinerja Metode Regresi Linear dan Extreme Gradient Boosting (XGBoost) dalam Prediksi Harga Saham BUMN Sektor Perbankan," *Trigonometri J. Mat.*, vol. 3, no. 1, hal. 27–35, 2026.

