

KOMPARASI LINEAR REGRESI DAN RANDOM FOREST UNTUK PREDIKSI PENGANGGURAN TERBUKA JAWA TIMUR

Wijayanto¹, Ronny Makhfuddin Akbar², Ahmad Syaifuddin³

^{1,2} Program Studi Informatika Universitas Islam Majapahit; Jl. Raya Jabon No.1, Mojoanyar, Mojokerto 61364, Jawa Timur, Indonesia; Telp. (0321) 399474;

Keywords:

Tingkat Pengangguran Terbuka, *Regresi Linear*, *Random Forest Regression*, Prediksi, *Machine Learning*.

Correspondent Email:

jaayanto440@gmail.com

Abstrak. Tingkat Pengangguran Terbuka (TPT) merupakan indikator penting dalam mengevaluasi kondisi ketenagakerjaan suatu daerah. Prediksi TPT diperlukan untuk mendukung penyusunan kebijakan ketenagakerjaan yang lebih efektif. Penelitian ini bertujuan membandingkan kinerja algoritma Linear Regression dan Random Forest Regression dalam memprediksi TPT di Provinsi Jawa Timur. Dataset yang digunakan berasal dari Badan Pusat Statistik (BPS) periode 2021–2025 yang mencakup 38 kabupaten/kota dengan total 190 data. Variabel prediktor meliputi Produk Domestik Regional Bruto (PDRB), Upah Minimum Regional (UMR), Tingkat Partisipasi Angkatan Kerja (TPAK), inflasi, dan rata-rata lama sekolah. Tahapan penelitian meliputi preprocessing data, standarisasi menggunakan StandardScaler, pembagian data dengan rasio 80:20, pembangunan model, serta evaluasi menggunakan MAE, MSE, RMSE, dan R^2 . Hasil penelitian menunjukkan bahwa Random Forest Regression memiliki performa lebih baik dibandingkan Linear Regression, dengan nilai MAE 0,6768, MSE 0,8211, RMSE 0,9062, dan R^2 0,7295, sedangkan Linear Regression memperoleh MAE 0,8518, MSE 1,2439, RMSE 1,1153, dan R^2 0,5902. Variabel rata-rata lama sekolah menjadi faktor yang paling berpengaruh terhadap prediksi TPT. Dengan demikian, Random Forest Regression lebih efektif untuk memprediksi TPT di Provinsi Jawa Timur.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. The Open Unemployment Rate (OUR) is an important indicator for evaluating regional labor market conditions. Predicting the OUR is essential to support the formulation of more effective employment policies. This study aims to compare the performance of Linear Regression and Random Forest Regression algorithms in predicting the OUR in East Java Province. The dataset was obtained from Statistics Indonesia (BPS) for the 2021–2025 period, covering 38 regencies/cities with a total of 190 observations. The predictor variables include Gross Regional Domestic Product (GRDP), Regional Minimum Wage (RMW), Labor Force Participation Rate (LFPR), inflation, and average years of schooling. The research stages consisted of data preprocessing, standardization using StandardScaler, an 80:20 train-test split, model development, and evaluation using MAE, MSE, RMSE, and R^2 . The results indicate that Random Forest Regression outperformed Linear Regression, achieving an MAE of 0.6768, MSE of 0.8211, RMSE of 0.9062, and R^2 of 0.7295, while Linear Regression obtained an MAE of 0.8518, MSE of 1.2439, RMSE of 1.1153, and R^2 of 0.5902. Feature importance analysis also identified average years of schooling as the most influential predictor. Therefore, Random Forest Regression is more effective for predicting the Open Unemployment Rate in East Java Province.

1. PENDAHULUAN

Tingkat Pengangguran Terbuka (TPT) merupakan indikator makroekonomi yang vital dalam mengukur efektivitas pembangunan ekonomi suatu daerah dalam menyerap angkatan kerja [1]. Sebagai *lagging indicator*, TPT memberikan gambaran komprehensif mengenai sejauh mana kebijakan pembangunan ekonomi mampu menyediakan lapangan pekerjaan yang memadai bagi masyarakat produktif [2].

Provinsi Jawa Timur, sebagai salah satu motor penggerak ekonomi di Indonesia, menghadapi tantangan ketenagakerjaan yang dinamis, di mana pertumbuhan ekonomi tidak selalu berkorelasi positif dengan perluasan kesempatan kerja yang inklusif [3]. Kondisi ini menuntut pemahaman yang mendalam mengenai faktor-faktor determinan yang memengaruhi fluktuasi TPT di wilayah tersebut [4]. Dinamika TPT di Jawa Timur tidak berdiri sendiri, melainkan hasil dari interaksi kompleks variabel-variabel sosioekonomi [5]. Inflasi menjadi penentu stabilitas harga yang berdampak langsung pada daya beli dan keberlangsungan usaha [2]. PDRB per kapita, yang mencerminkan produktivitas agregat daerah, sering kali menunjukkan ketidakselarasan dengan pertumbuhan tenaga kerja (*jobless growth*) [1]. TPAK (Tingkat Partisipasi Angkatan Kerja) juga menjadi elemen kunci yang menentukan besarnya suplai tenaga kerja [3]. UMR (Upah Minimum Regional) memiliki pengaruh yang ambigu; di satu sisi berfungsi sebagai alat kesejahteraan, namun di sisi lain dapat memicu distorsi pada pasar tenaga kerja jika kenaikannya melampaui produktivitas [5]. Rata-rata Lama Sekolah mencerminkan durasi dan tingkat pendidikan angkatan kerja [4]. Meskipun pendidikan yang lebih tinggi diharapkan meningkatkan peluang kerja, ketidaksesuaian antara keterampilan yang dipelajari dan kebutuhan industri (*skill mismatch*) kerap menjadi penyumbang signifikan terhadap angka pengangguran terdidik [4]. Selain itu, kesenjangan ekspektasi upah antara pekerja dan pemberi kerja kerap menyebabkan tenaga kerja memilih untuk tidak terlibat dalam pasar kerja, yang secara akumulatif turut memengaruhi angka pengangguran [6]. Metode analisis statistik

konvensional seperti Regresi Linear sering kali menemui kendala dalam menangkap kompleksitas hubungan *non-linear* antar variabel ekonomi makro [7]. Asumsi linearitas yang kaku dalam model statistik tradisional sering kali gagal merepresentasikan perilaku dinamis pasar tenaga kerja yang cenderung stokastik dan saling berinteraksi [8]. Oleh karena itu, pendekatan melalui *machine learning* menjadi sangat krusial [9]. Algoritma *Random Forest Regression* hadir sebagai solusi melalui mekanisme *ensemble learning* yang mampu mengintegrasikan banyak *decision tree* untuk menghasilkan prediksi yang lebih presisi dan tahan terhadap variasi data [10].

Penelitian ini bertujuan untuk melakukan studi komparatif antara algoritma Regresi Linear dan *Random Forest* dalam memprediksi TPT di Jawa Timur dengan menggunakan dataset ekonomi makro periode 2021-2025 [6]. Dengan membandingkan performa kedua model tersebut, penelitian ini diharapkan mampu memberikan kontribusi ilmiah berupa model prediksi yang lebih akurat serta pemahaman mendalam mengenai variabel mana yang paling berpengaruh dalam fluktuasi TPT [7]. Hasil dari penelitian ini diharapkan dapat menjadi landasan empiris bagi pengambil kebijakan di Jawa Timur dalam merumuskan strategi ketenagakerjaan yang lebih tepat sasaran, adaptif, dan berkelanjutan [1].

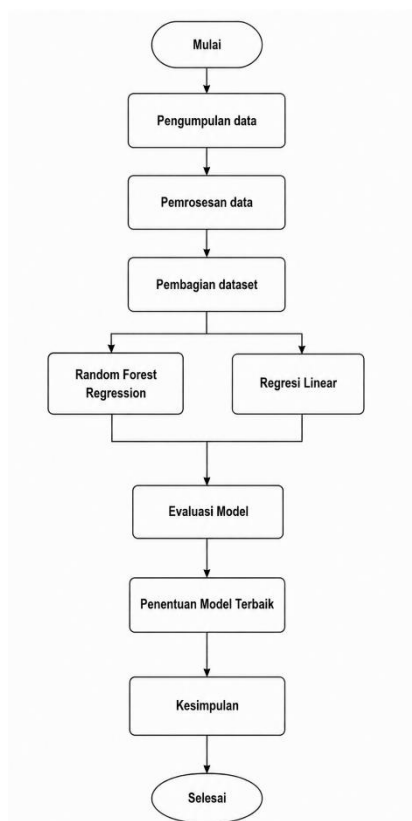
2. TINJAUAN PUSTAKA

Kajian komparatif antara statistik klasik dan *machine learning* menjadi tren riset yang relevan dalam memodelkan data sosial ekonomi [17]. Beberapa penelitian terdahulu mengandalkan model linier dasar untuk skala nasional maupun daerah, seperti penggunaan Regresi Linear untuk prediksi TPT nasional [10] maupun lingkup regional provinsi dan kabupaten [1]. Di sisi lain, algoritma *ensemble* seperti *Random Forest Regression* terbukti efektif memberikan nilai error (MAE dan RMSE) yang lebih kecil serta koefisien determinasi (R^2) yang tinggi pada pemodelan indikator makroekonomi BPS [11], [12]. Studi komparatif berskala makroekonomi juga mengonfirmasi bahwa pendekatan *machine learning* multi-algoritma mampu menangkap hubungan *non-linear* secara optimal, di mana

faktor edukasi atau tingkat pendidikan sering kali muncul sebagai variabel paling dominan [13], [16]. Riset ini mengisi research gap dengan memosisikan diri secara spesifik pada ranah comparative performance akurasi prediksi TPT di 38 kabupaten/kota Jawa Timur [17].

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode komparatif untuk membandingkan performa algoritma *Linear Regression* dan *Random Forest Regression* dalam memprediksi Tingkat Pengangguran Terbuka (TPT) di Provinsi Jawa Timur. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pembagian dataset, pembangunan model, serta evaluasi hasil prediksi. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Penelitian

3.1 Dataset Penelitian

Dataset yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur periode 2021–2025. Dataset terdiri atas 190 data

observasi yang berasal dari 38 kabupaten/kota di Provinsi Jawa Timur. Variabel dependen pada penelitian ini adalah Tingkat Pengangguran Terbuka (TPT), sedangkan variabel independen terdiri atas Produk Domestik Regional Bruto (PDRB), Upah Minimum Regional (UMR), Tingkat Partisipasi Angkatan Kerja (TPAK), inflasi, dan rata-rata lama sekolah. Pemilihan variabel tersebut didasarkan pada penelitian sebelumnya yang menunjukkan bahwa indikator makroekonomi memiliki hubungan dengan kondisi ketenagakerjaan.

3.1.1 Variabel Dataset dan Gambar Isi Dataset

Tabel 1. Variabel Keterangan

Variabel	Keterangan
TPT	Variabel Dependen
PDRB	Variabel Independen
UMR	Variabel Independen
TPAK	Variabel Independen
Inflasi	Variabel Independen
Lama Sekolah	Variabel Independen

	A	B	C	D	E	F	G
1	kab_kota,tahun,tpt,pdrb,umr,tpak,inflasi,Lama_Sekolah						
2	Kabupaten Bangkalan,2021,6.18,27588.1,2004705.75,68.66,1.79,5.96						
3	Kabupaten Banyuwangi,2021,4.75,67081.2,2004705.75,72.32,1.79,7.42						
4	Kabupaten Blitar,2021,5.24,63145.6,2004705.75,69.96,1.79,10.35						
5	Kabupaten Bojonegoro,2021,4.63,78894.2,1954705.75,71.84,1.79,7.38						

Gambar 2. Dataset Penelitian

3.2 Preprocessing Data

Tahap *preprocessing* dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pelatihan model. Proses ini meliputi pengecekan *missing value*, pengecekan data duplikat, analisis statistik deskriptif, analisis korelasi antarvariabel menggunakan matriks korelasi Pearson, serta standardisasi fitur menggunakan metode *StandardScaler*. Standardisasi bertujuan menyamakan skala antarvariabel sehingga proses pelatihan model menjadi lebih optimal, terutama pada algoritma yang sensitif terhadap perbedaan rentang nilai fitur.

Selanjutnya, dataset dibagi menjadi 80% data pelatihan (*training set*) dan 20% data pengujian (*testing set*) menggunakan metode *train-test split*. Data pelatihan digunakan untuk membangun model, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan model dalam memprediksi data yang belum pernah dipelajari sebelumnya.

3.3 Pembangunan Model

Penelitian ini membandingkan dua algoritma regresi, yaitu *Linear Regression* dan *Random Forest Regression*.

A. Linear Regression

Linear Regression digunakan sebagai model dasar (*baseline*) untuk mempelajari hubungan linear antara variabel independen dengan Tingkat Pengangguran Terbuka. Persamaan umum *Linear Regression* ditunjukkan pada Persamaan (3.1).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \dots (3.1)$$

di mana:

Y = Tingkat Pengangguran Terbuka,

β_0 = konstanta,

β_i = koefisien regresi,

X_i = variabel independen,

ε = galat (error).

B. Random Forest

Random Forest Regression merupakan metode *ensemble learning* yang membangun sejumlah *decision tree*. Prediksi akhir diperoleh dari rata-rata seluruh hasil prediksi setiap pohon keputusan sehingga menghasilkan model yang lebih stabil dan memiliki kemampuan generalisasi yang lebih baik.

Secara matematis, prediksi Random Forest dirumuskan sebagai:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x) \dots \dots \dots (3.2)$$

dengan:

$\hat{y}$ = hasil prediksi,

T = jumlah pohon keputusan,

$f_t(x)$ = hasil prediksi pohon ke- i .

3.4 Evaluasi Model

Kinerja kedua model dievaluasi menggunakan empat metrik regresi, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2). Keempat metrik tersebut digunakan karena mampu memberikan gambaran mengenai tingkat kesalahan prediksi serta kemampuan model dalam menjelaskan variasi data.

Persamaan MAE ditunjukkan pada persamaan (3.3)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \dots \dots \dots (3.3)$$

Persamaan MSE ditunjukkan pada persamaan (3.4)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots \dots \dots (3.4)$$

Persamaan RMSE ditunjukkan pada persamaan (3.5)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \dots \dots \dots (3.5)$$

Persamaan R^2 ditunjukkan pada persamaan (3.6)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots \dots \dots (3.6)$$

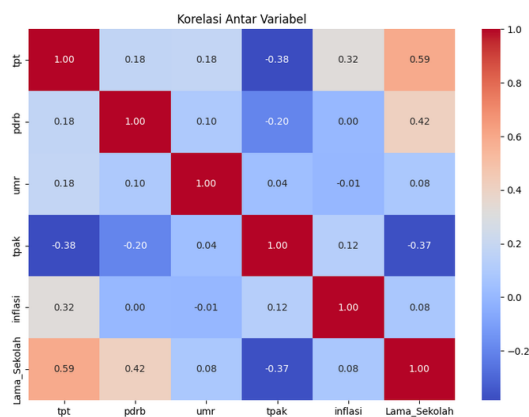
4 HASIL DAN PEMBAHASAN

4.1 Hasil Preprocessing Data

Sebelum dilakukan proses pembangunan model, dataset terlebih dahulu melalui tahap *preprocessing* untuk memastikan kualitas data yang digunakan. Berdasarkan hasil pemeriksaan, dataset tidak memiliki *missing value* maupun data duplikat sehingga seluruh data dapat digunakan pada proses pelatihan

model. Selanjutnya dilakukan standardisasi menggunakan metode *StandardScaler* agar setiap variabel memiliki skala yang seragam. Tahap ini bertujuan meningkatkan stabilitas proses pelatihan model, khususnya pada algoritma *Linear Regression*.

Analisis korelasi juga dilakukan untuk mengetahui hubungan antarvariabel penelitian. Hasil korelasi ditunjukkan pada Gambar 3.



Gambar 3. Heatmap Korelasi

Berdasarkan Gambar 3, terlihat bahwa setiap variabel memiliki tingkat hubungan yang berbeda terhadap Tingkat Pengangguran Terbuka (TPT). Variabel Lama Sekolah menunjukkan hubungan yang relatif lebih kuat dibandingkan variabel lainnya. Sementara itu, variabel Inflasi, PDRB, UMR, dan TPAK memiliki tingkat korelasi yang bervariasi sehingga seluruh variabel tetap digunakan dalam proses pembangunan model.

4.2 Hasil Pengujian

Model yang dibangun pada penelitian ini terdiri atas *Linear Regression* dan *Random Forest Regression*. Kedua model dilatih menggunakan data latih sebesar 80% dan diuji menggunakan data uji sebesar 20%. Evaluasi dilakukan menggunakan empat metrik, yaitu MAE, MSE, RMSE, dan koefisien determinasi (R^2).

Hasil evaluasi kedua model ditunjukkan pada Tabel 2 berikut.

Tabel 2. Hasil Evaluasi Model

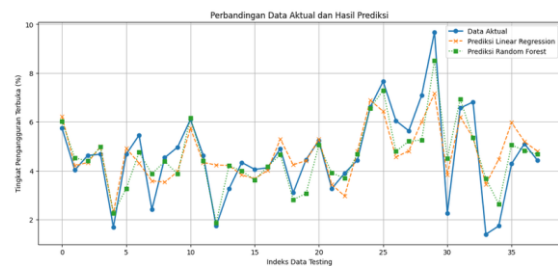
Metrik	Linear Regresi	Random Forest
MAE	0,8518	0,6768

MSE	1,2439	0,8211
RMSE	1,1153	0,9062
R^2	0,5902	0,7295
RANKING	2	1

Berdasarkan Tabel 2, *Random Forest Regression* menghasilkan nilai MAE, MSE, dan RMSE yang lebih rendah dibandingkan *Linear Regression*. Selain itu, nilai koefisien determinasi (R^2) *Random Forest Regression* lebih tinggi sehingga model mampu menjelaskan variasi data dengan lebih baik.

4.3 Analisa Perbandingan Prediksi

Selain menggunakan metrik evaluasi, kualitas model juga dianalisis melalui perbandingan antara nilai aktual dan nilai prediksi. Visualisasi hasil prediksi kedua algoritma ditunjukkan pada Gambar 4.

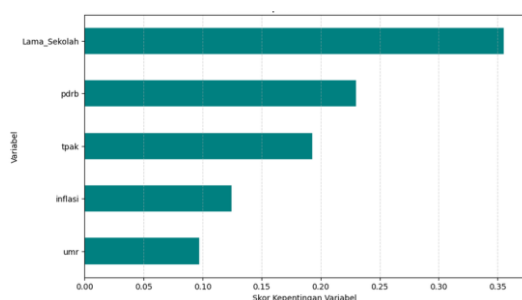


Gambar 4. Perbandingan Nilai Aktual dan Hasil Prediksi

Berdasarkan Gambar 4, hasil prediksi *Random Forest Regression* cenderung mengikuti pola data aktual dengan lebih baik dibandingkan *Linear Regression*. Prediksi yang dihasilkan memiliki penyimpangan yang lebih kecil sehingga mampu menggambarkan perubahan nilai Tingkat Pengangguran Terbuka secara lebih konsisten. Sebaliknya, *Linear Regression* menunjukkan beberapa penyimpangan pada titik-titik tertentu karena model hanya mampu mempelajari hubungan linear antarvariabel.

4.4 Analisis Feature Importance

Salah satu keunggulan *Random Forest Regression* adalah kemampuannya dalam mengukur tingkat kontribusi setiap variabel terhadap hasil prediksi melalui analisis *feature importance*. Hasil analisis ditunjukkan pada Gambar 5.



Gambar 5. Feature Importance

Berdasarkan Gambar 5, variabel Lama Sekolah memiliki nilai *feature importance* tertinggi dibandingkan variabel lainnya. Hasil ini menunjukkan bahwa tingkat pendidikan memberikan kontribusi paling besar terhadap perubahan Tingkat Pengangguran Terbuka di Provinsi Jawa Timur. Selanjutnya, Urutan kontribusi dari yang tertinggi setelah Lama Sekolah adalah PDRB, disusul TPAK, Inflasi, dan yang paling rendah adalah UMR.

Temuan tersebut menunjukkan bahwa peningkatan kualitas sumber daya manusia melalui pendidikan berpotensi memberikan dampak terhadap penurunan tingkat pengangguran. Sementara itu, faktor ekonomi seperti PDRB, UMR, dan inflasi tetap berperan dalam memengaruhi kondisi pasar tenaga kerja meskipun kontribusinya tidak sebesar variabel pendidikan.

4.5 Analisis Hasil Penelitian

Berdasarkan seluruh hasil pengujian, *Random Forest Regression* memberikan performa yang lebih baik dibandingkan *Linear Regression* dalam memprediksi Tingkat Pengangguran Terbuka di Provinsi Jawa Timur. Hal ini ditunjukkan oleh nilai MAE, MSE, dan RMSE yang lebih rendah serta nilai R^2 yang lebih tinggi.

Perbedaan performa tersebut terjadi karena *Linear Regression* hanya memodelkan hubungan linear antara variabel independen dan variabel dependen. Sementara itu, *Random Forest Regression* mampu mempelajari hubungan yang lebih kompleks melalui kombinasi sejumlah *decision tree*. Karakteristik tersebut memungkinkan Random Forest menghasilkan prediksi yang lebih stabil terhadap data ekonomi yang memiliki pola *non-linear*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Informatika Fakultas Teknik Universitas Islam Majapahit atas dukungan selama pelaksanaan penelitian. Penulis juga menyampaikan apresiasi kepada Badan Pusat Statistik (BPS) yang telah menyediakan data penelitian sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] M. R. Fahlevi and A. S. R. Anshari, "Analisis Pengaruh PDRB, Upah Minimum Regional, dan Tingkat Pendidikan Terhadap Tingkat Pengangguran Terbuka di Provinsi Jawa Timur," *Jurnal Ekonomi dan Pembangunan Regional*, vol. 5, no. 2, pp. 112–125, Jul. 2022.
- [2] S. Handayani and W. Utami, "Dinamika Inflasi, Pertumbuhan Ekonomi, dan Hubungannya dengan Tingkat Pengangguran Terbuka di Indonesia," *Jurnal Ekonomi Makro Indonesia*, vol. 10, no. 1, pp. 45–58, Jan. 2023.
- [3] R. D. Saputra and N. K. S. Astuti, "Pengaruh Tingkat Partisipasi Angkatan Kerja (TPAK) dan Rata-rata Lama Sekolah Terhadap Penyerapan Tenaga Kerja Pemuda," *Jurnal Ketenagakerjaan dan Kebijakan Publik*, vol. 3, no. 3, pp. 201–214, Oct. 2023.
- [4] T. H. Purwanto, "Analisis Kesenjangan Keterampilan (*Skill Mismatch*) dan Dampaknya Terhadap Tingkat Pengangguran Terdidik di Jawa Timur," *Jurnal Pendidikan Ekonomi dan Bisnis*, vol. 12, no. 1, pp. 33–47, Feb. 2024.
- [5] A. Nurhayati and F. Rosyidi, "Dilema Upah Minimum dan Efeknya Terhadap Elastisitas Kesempatan Kerja pada Sektor Manufaktur di Kabupaten/Kota," *Jurnal Ilmiah Riset Ekonomi*, vol. 9, no. 4, pp. 310–322, Dec. 2024.
- [6] H. Setiawan, "Evaluasi Capaian Indikator Ketenagakerjaan Jawa Timur Pasca-Pandemi Berdasarkan Data Badan Pusat Statistik," *Jurnal Statistik Daerah*, vol. 14, no. 2, pp. 88–99, Aug. 2025.
- [7] B. Juanda and J. M. Syarif, "Model Regresi Data Panel untuk Analisis Faktor-Faktor yang Mempengaruhi Pengangguran Tingkat Regional," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 15, no. 1, pp. 15–28, Mar. 2023.
- [8] D. Prasetyo and R. Wardani, "Penerapan Metode Ordinary Least Squares (OLS) dalam Memprediksi Pertumbuhan Ekonomi Daerah Berbasis Komponen PDRB," *Jurnal Statistika dan Matematika Terapan*, vol. 5, no. 2, pp. 77–89, Sep. 2023.
- [9] K. A. Nugroho, "Uji Asumsi Klasik dan Limitasi Model Linear dalam Kompleksitas Data Sosial

- Ekonomi Makro," *Indonesian Journal of Applied Statistics*, vol. 7, no. 2, pp. 142–155, Nov. 2024.
- [10] M. I. Fadilah and S. Sulastri, "Pemodelan Regresi Linear Berganda Untuk Proyeksi Angkatan Kerja di Sektor Industri Strategis," *Jurnal Analisis Data Kuantitatif*, vol. 4, no. 1, pp. 22–34, Jan. 2025.
- [11] L. A. Pratama and Y. S. Wijaya, "Penerapan Algoritma Random Forest Regression dalam Memprediksi Indikator Kemiskinan Tingkat Kabupaten," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 4, pp. 789–798, Aug. 2023.
- [12] N. W. Utami, I. G. A. P. R. Agung, and I. M. S. Putra, "Prediksi Tingkat Pengangguran Menggunakan Metode *Ensemble Learning* Berbasis *Decision Tree*," *Jurnal Ilmu Komputer dan Informatika*, vol. 9, no. 2, pp. 130–141, Nov. 2023.
- [13] R. Kusuma and A. B. Santoso, "Analisis Feature Importance pada Algoritma Pohon Keputusan untuk Identifikasi Variabel Dominan Sosial Ekonomi," *Jurnal Sistem Informasi dan Sains Data*, vol. 2, no. 1, pp. 55–66, Jan. 2024.
- [14] F. A. Ramadhan, "Optimasi Hyperparameter Random Forest Regression Menggunakan Grid Search untuk Prediksi Indikator Makro Ekonomi," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 3, pp. 412–420, Jun. 2024.
- [15] S. Mulyani and R. Hartono, "Komparasi Algoritma Supervised Learning Berbasis Pohon untuk Estimasi Tingkat Kesejahteraan Masyarakat," *Jurnal Informatika dan Komputasi*, vol. 16, no. 2, pp. 105–116, Aug. 2025.
- [16] A. R. Hakim and E. N. Fatimah, "Studi Komparatif Akurasi Model Statistik Konvensional dan *Machine Learning* pada Data Deret Waktu Ekonomi," *Jurnal Jatisi (Jurnal Teknologi Informasi dan Sistem Informasi)*, vol. 10, no. 3, pp. 512–526, Sep. 2023.
- [17] W. Budiharto and J. Purwanto, "Analisis Perbandingan Regresi Linear Berganda dan Random Forest dalam Pemetaan Ketimpangan Pendapatan Daerah," *Jurnal JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 2, pp. 189–199, Apr. 2024.
- [18] M. H. Asy'ari and T. Baryanto, "Mengatasi Hubungan Non-Linear pada Data Publikasi BPS: Pendekatan Regresi Parametrik versus Non-Parametrik *Machine Learning*," *Jurnal Komputasi Statistik*, vol. 6, no. 1, pp. 44–56, Jan. 2025.
- [19] G. D. Permadi and I. M. Sukarsa, "Evaluasi Robustness Metrik MAE, MSE, dan R-Square dalam Penilaian Model Prediksi Sosial," *Journal of Computer Science and Informatics Engineering*, vol. 5, no. 2, pp. 95–107, May 2025.
- [20] H. Kurniawan and S. Wahyuni, "Keunggulan Metode *Ensemble* Terhadap Algoritma Linear Tunggal dalam Prediksi Indikator Ketenagakerjaan Nasional," *Jurnal Riset Komputer (JURIKOM)*, vol. 13, no. 1, pp. 67–79, Feb. 2026.