

A COMPARISON OF THE PERFORMANCE OF RANDOM FOREST AND EXTRA TREES REGRESSION IN PREDICTING PROVINCIAL GROSS REGIONAL DOMESTIC PRODUCT IN INDONESIA

PERBANDINGAN KINERJA RANDOM FOREST DAN EXTRA TREES REGRESSION DALAM PREDIKSI PRODUK DOMESTIK REGIONAL BRUTO PROVINSI DI INDONESIA

Wiska Radhia Firdaus¹, Rahmandani Bardianto², Aurisa Putri Hawin Yusima³, Fathoni Sabri⁴, Agustin Salsabilla⁵ and Cici Suhaeni⁶

^{1,2,3,4,5,6}Program Studi Statistika dan Sains Data, SSMI, IPB University

[‡]corresponding author: wiskafirdauswiska@apps.ipb.ac.id

Copyright © 2026 Wiska Radhia Firdaus, Rahmandani Bardianto, Aurisa Putri Hawin Yusima, Fathoni Sabri, Agustin Salsabilla and Cici Suhaeni. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Gross Regional Domestic Product (GRDP) is an important indicator for measuring regional economic growth and economic activity. This study aims to compare the performance of Random Forest and Extra Trees Regression methods in predicting provincial GRDP in Indonesia and to identify the dominant factors influencing it. The study used 2025 cross-sectional data from 38 provinces in Indonesia obtained from Statistics Indonesia (BPS). A total of 15 socio-economic predictor variables were employed. The analytical stages included data preprocessing, natural logarithm transformation of GRDP, train-test data splitting, hyperparameter tuning using GridSearch with 5-Fold Cross Validation, and model evaluation using RMSE, MAE, and MAPE. The results indicate that Extra Trees Regression outperformed Random Forest with an RMSE of 0.638, MAE of 0.453, and MAPE of 3.776. Feature importance analysis revealed that domestic investment were the most influential factors affecting GRDP. In addition, the number of healthcare facilities and per capita spending emerged as an important variable in the Extra Trees Regression model. Therefore, Extra Trees Regression is recommended as a more reliable model for predicting provincial GRDP in Indonesia.

Keywords: economic growth, Extra Trees Regression, GRDP, Random Forest.

Abstrak

Produk Domestik Regional Bruto (PDRB) merupakan indikator penting dalam mengukur pertumbuhan dan aktivitas ekonomi daerah. Penelitian ini bertujuan membandingkan kinerja metode *Random Forest* (RF) dan *Extra Trees Regression* (ETR) dalam memprediksi PDRB provinsi di Indonesia serta mengidentifikasi faktor-faktor dominan yang memengaruhinya. Data yang digunakan berupa data *cross section* tahun 2025 dari 38 provinsi di Indonesia yang diperoleh dari Badan Pusat Statistik (BPS). Penelitian menggunakan 15 variabel prediktor sosial-ekonomi. Tahapan analisis meliputi *preprocessing data*, transformasi logaritma natural pada PDRB, pembagian data latih dan data uji, *hyperparameter tuning* menggunakan *GridSearch* dengan *5-Fold Cross Validation*, serta evaluasi model menggunakan RMSE, MAE, dan MAPE. Hasil penelitian menunjukkan bahwa *Extra Trees Regression* memiliki performa lebih baik dibandingkan *Random Forest* dengan nilai RMSE sebesar 0.638, MAE sebesar 0.453, dan MAPE sebesar 3.776. Analisis *feature importance* menunjukkan bahwa penanaman modal dalam negeri merupakan faktor paling dominan dalam memengaruhi PDRB. Selain itu, jumlah sarana kesehatan dan pengeluaran per kapita juga menjadi variabel penting dalam model *Extra Trees Regression*.

Kata Kunci: *Extra Trees Regression*, PDRB, Pertumbuhan Ekonomi, *Random Forest*.

1. Pendahuluan

Produk Domestik Regional Bruto (PDRB) merupakan indikator utama yang mencerminkan total nilai tambah sektor ekonomi di suatu wilayah dan menjadi dasar penyusunan kebijakan serta evaluasi pembangunan daerah (Ria *et al.*, 2022). Selain itu, PDRB juga digunakan pemerintah sebagai dasar penyusunan kebijakan dan evaluasi pembangunan daerah (Gatari *et al.*, 2024). Oleh karena itu, analisis faktor-faktor yang memengaruhi PDRB menjadi penting dalam memahami dinamika pertumbuhan ekonomi regional.

Berbagai faktor ekonomi dan sosial diketahui memengaruhi PDRB. Penanaman modal berkontribusi melalui transfer teknologi, penciptaan lapangan kerja, dan peningkatan investasi (Yuliani *et al.*, 2023). Tingkat Partisipasi Angkatan Kerja (TPAK) juga mendukung pertumbuhan ekonomi melalui ketersediaan tenaga kerja dan potensi pasar (Nazilaturrohman *et al.*, 2024). Sebaliknya, Tingkat Pengangguran Terbuka (TPT) dapat menurunkan output ekonomi sesuai Hukum Okun (Christy A.N & Fahrati Eny, 2024). Selain itu, kualitas sumber daya manusia seperti rata-rata lama sekolah, pengeluaran per kapita, dan Indeks Pembangunan Manusia (IPM)

berperan dalam meningkatkan produktivitas masyarakat (Tarigan *et al.*, 2026). Faktor lain seperti kemiskinan, pariwisata, serta infrastruktur dasar juga memiliki hubungan erat dengan pertumbuhan ekonomi daerah (Permana & Pasaribu, 2023).

Hubungan antara PDRB dan faktor-faktor penentunya cenderung kompleks dan nonlinier sehingga model statistik konvensional sering kali kurang optimal dalam menangkap pola data (Fotopoulos, 2012). Kondisi ini mendorong penggunaan metode *machine learning*, khususnya *Random Forest* (RF) dan *Extra Trees Regression* (ERT), yang dinilai lebih mampu menangani hubungan kompleks antarvariabel. RF bekerja dengan membangun banyak pohon keputusan dari sampel bootstrap (Breiman, 2001), sedangkan ERT menggunakan randomisasi yang lebih tinggi pada proses pemisahan node untuk mengurangi varians antar pohon (Geurts *et al.*, 2006).

Beberapa penelitian menunjukkan bahwa RF dan ERT mampu menghasilkan prediksi ekonomi yang akurat (Mubarak & Septian, 2025). Keunggulan ERT dalam prediksi data kuantitatif di Indonesia juga telah dibuktikan oleh Makhya *et al.* (2022). Namun, studi yang membandingkan kedua metode tersebut dalam prediksi PDRB provinsi di Indonesia masih terbatas. Oleh karena itu, penelitian ini bertujuan menentukan model prediksi terbaik serta mengidentifikasi faktor-faktor dominan yang memengaruhi pertumbuhan ekonomi daerah di Indonesia.

2. Metodologi

Penelitian ini bertujuan untuk membangun model prediksi PDRB provinsi di Indonesia menggunakan metode *Random Forest* dan *Extra Trees Regression*. Kedua metode tersebut merupakan algoritma ensemble berbasis pohon keputusan yang mampu menangani hubungan nonlinier serta interaksi kompleks antar variabel sosial dan ekonomi.

2.1 Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diambil dari Badan Pusat Statistika (BPS). Data sekunder yang digunakan merupakan data *cross section* pada tahun 2025 dengan amatan 38 provinsi yang ada di Indonesia. Peubah-peubah yang digunakan pada penelitian ini disajikan pada Tabel 1.

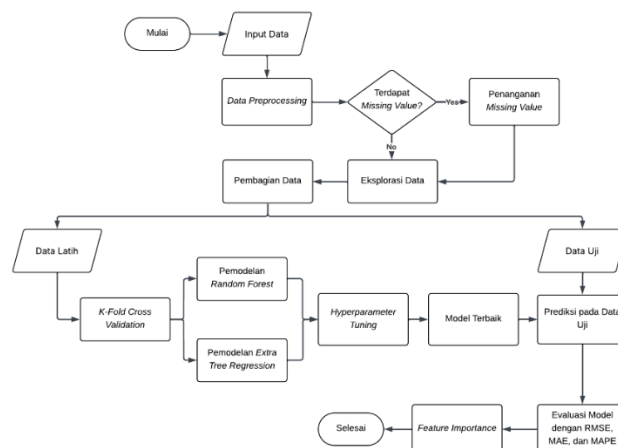
Tabel 1: Daftar peubah-peubah pada penelitian

Kode Peubah	Peubah	Satuan
Y	Produk Domestik Regional Bruto	Ribu Rupiah
X_1	Penanaman Modal Asing	Juta US Dollar
X_2	Penanaman Modal Dalam Negeri	Miliar Rupiah
X_3	Tingkat Partisipasi Angkatan Kerja	Persen

Kode Peubah	Peubah	Satuan
X_4	Tingkat Pengangguran Terbuka	Persen
X_5	Rata-Rata Lama Sekolah	Tahun
X_6	Upah Minimum	Rupiah
X_7	Pengeluaran per Kapita per Tahun	Ribu Rupiah
X_8	Indeks Pembangunan Manusia	Persen
X_9	Persentase Penduduk Miskin	Persen
X_{10}	Tingkat Hunian Kamar Hotel	Persen
X_{11}	Kepadatan Penduduk	Jiwa/km ²
X_{12}	Akses terhadap Sanitasi Layak	Persen
X_{13}	Akses terhadap Air Minum Layak	Persen
X_{14}	Jumlah Sarana Kesehatan	Unit

2.2 Metode Penelitian

Penelitian ini menggunakan bahasa pemrograman Python melalui *platform* Google Colab. Alur tahapan penelitian disajikan pada Gambar 1 berikut.



Gambar 1: *Flowchart* Tahapan Analisis

1. Data Preprocessing

Preprocessing data dilakukan melalui standarisasi nama variabel, konversi simbol “-” menjadi *missing value* (NaN), serta perubahan tipe data ke format numerik. Penanganan *missing value* dilakukan menggunakan metode imputasi rata-rata pada variabel yang memiliki nilai hilang. Setelah *preprocessing*, dilakukan analisis deskriptif untuk memahami karakteristik data melalui distribusi variabel, ukuran pemusatan, ukuran penyebaran, dan identifikasi potensi *outlier*.

2. Eksplorasi Data

Pada tahap ini dilakukan pemeriksaan struktur data meliputi informasi tipe data, jumlah observasi, dimensi dataset, serta nama variabel yang digunakan dalam penelitian. Distribusi variabel target PDRB dianalisis menggunakan histogram untuk mengidentifikasi

pola distribusi serta potensi *outlier*. Selanjutnya, dilakukan transformasi logaritma natural pada variabel PDRB guna mengurangi *skewness* dan menstabilkan varians data. Distribusi hasil transformasi kemudian divisualisasikan kembali menggunakan histogram untuk mengevaluasi perubahan distribusi data setelah transformasi.

3. Pembagian Data

Dataset dibagi secara acak menjadi data latih sebesar 80% dan data uji sebesar 20%. Data latih digunakan pada proses pelatihan dan *cross validation* model, sedangkan data uji digunakan untuk evaluasi akhir guna mengukur performa model pada data yang belum pernah dilatih sebelumnya.

4. *K-Fold Cross Validation*

Teknik *K-Fold Cross Validation* diterapkan pada data latih untuk proses pelatihan dan validasi model *Random Forest* dan *Extra Trees Regression*. Metode ini membagi data latih menjadi k *fold* dengan ukuran yang relatif sama. Pada setiap iterasi, satu *fold* berperan sebagai data validasi, sedangkan $k-1$ *fold* lainnya digunakan sebagai data pelatihan. Proses dilakukan berulang hingga seluruh *fold* digunakan sebagai data validasi (Gorriz et al., 2024).

5. Pemodelan *Random Forest*

Random Forest Regression merupakan metode *ensemble* berbasis pohon keputusan yang bekerja dengan membangun banyak pohon regresi secara acak, kemudian menggabungkan hasil prediksinya untuk meningkatkan akurasi dan stabilitas model (Nugroho, 2025). Metode ini merupakan pengembangan dari *bagging* yang menggunakan *bootstrap sampling* dan pemilihan variabel secara acak pada setiap node sehingga mampu mengurangi varians dan meningkatkan kemampuan generalisasi model (Magalhães et al., 2024). Struktur pohon regresi terdiri atas *root node*, *internal node*, dan *leaf node*, dengan pemilihan *split* terbaik berdasarkan minimisasi galat seperti *mean squared error* (MSE). Proses *Random Forest* dimulai dengan *bootstrap sampling* pada data pelatihan, dilanjutkan dengan pembentukan banyak pohon regresi menggunakan subset variabel acak, kemudian seluruh hasil prediksi digabungkan melalui rata-rata prediksi dari setiap pohon.

6. Pemodelan *Extra Trees Regression*

Extra Trees Regression merupakan metode *ensemble* berbasis pohon keputusan yang membangun banyak pohon secara *top-down* (Mahkya et al., 2022). Berbeda dengan *Random Forest*, metode ini menggunakan seluruh data latih tanpa *bootstrap sampling* dan menentukan *cut-point* secara acak pada setiap node. Pendekatan tersebut bertujuan untuk mengurangi *varians*, meminimalkan bias,

serta meningkatkan efisiensi komputasi (Geurts *et al.*, 2006). Berikut merupakan langkah-langkah yang dilakukan dalam proses *Extra Trees Regression*:

- a) Menentukan peubah respon Y dan peubah penjelas X.
- b) Menentukan jumlah pohon yang akan dibangun, misalnya sebanyak M, serta parameter model, seperti jumlah maksimum fitur, minimum sampel untuk pembelahan node, dan kedalaman maksimum pohon.
- c) Model membangun sebanyak M pohon menggunakan seluruh data latih.
- d) Proses pemisahan node dilakukan dengan memilih K atribut secara acak dari seluruh atribut yang tersedia, kemudian menentukan *cut-point* acak pada masing-masing atribut. Selanjutnya, atribut dan *cut-point* terbaik dipilih sebagai dasar pembagian node.
- e) Proses pembelahan node dihentikan ketika kriteria *stop split* terpenuhi, kemudian node diubah menjadi *leaf* yang menghasilkan nilai prediksi target. Jika belum terpenuhi, proses pembagian node akan terus dilakukan secara berulang.
- f) Mengulang langkah c sampai e hingga diperoleh pohon sebanyak M.
- g) Prediksi akhir diperoleh dengan menghitung rata-rata hasil prediksi dari seluruh pohon yang terbentuk.

7. Tuning Hyperparameter

Hasil prediksi dan kinerja metode *Random Forest* serta *Extra Trees Regression* dipengaruhi oleh *hyperparameter* yang digunakan. Oleh karena itu, dilakukan *hyperparameter tuning* menggunakan metode *GridSearch* untuk meningkatkan performa prediksi model dan mengurangi risiko *overfitting* (Jamshidi *et al.*, 2024). *GridSearch* bekerja dengan menguji setiap kombinasi *hyperparameter* pada setiap *fold* data latih sehingga konfigurasi parameter terbaik dipilih berdasarkan hasil evaluasi empiris (Jamshidi *et al.*, 2024). Parameter yang digunakan dalam model disajikan pada Tabel 2.

Tabel 2: Parameter Model

Hyperparameter		Keterangan
Parameter	Pengaturan	
N_estimator	100, 300, 500, 1000	Maksimal total jumlah pohon regresi yang akan dibangun
Min_sample_split	2, 4, 5, 7	Jumlah sampel paling sedikit untuk bisa membagi sebuah node
Max_features	2, 3, 5, 7, 10, 15	Jumlah batas atribut acak yang akan dipilih dan dievaluasi pada setiap proses pemisahan node

8. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja *Random Forest* dan *Extra Trees Regression* dalam memprediksi nilai PDRB menggunakan data uji. Metrik evaluasi yang digunakan meliputi RMSE, MAE, dan MAPE. Model terbaik ditentukan berdasarkan nilai error terkecil dari masing-masing metrik evaluasi.

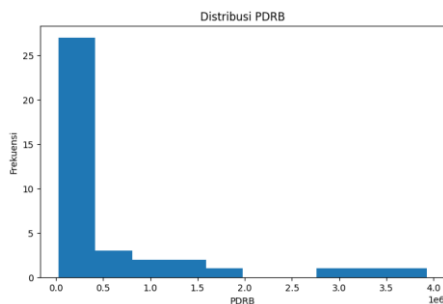
9. Feature Importance

Feature importance adalah salah satu teknik interpretasi model dalam *machine learning* yang digunakan untuk menunjukkan fitur yang paling berperan pada prediksi model berdasarkan kontribusi relatifnya (Saarela & Jauhiainen, 2021). Dalam konteks model berbasis pohon keputusan, *feature importance* umumnya dihitung menggunakan pendekatan *Mean Decrease Impurity* (MDI), yaitu seberapa besar suatu fitur berkontribusi dalam mengurangi ketidakmurnian yang ada pada setiap percabangan pohon keputusan. Semakin tinggi nilainya menunjukkan kontribusi relatif tersebut semakin besar terhadap prediksi model yang digunakan (Chen *et al.*, 2023).

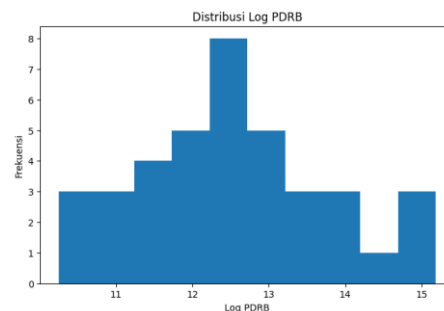
3. Hasil dan Pembahasan

3.1 Eksplorasi Data

Tahap eksplorasi data dilakukan untuk memahami karakteristik distribusi variabel target, yaitu PDRB, sebelum proses pemodelan. Data yang digunakan mencakup 38 provinsi di Indonesia dengan 15 variabel prediktor sosial-ekonomi.



Gambar 2: Distribusi PDRB Provinsi di Indonesia



Gambar 3: Distribusi log PDRB

Distribusi PDRB antar provinsi menunjukkan ketimpangan yang tinggi. Histogram tersebut memperlihatkan sebagian besar provinsi memiliki PDRB relatif rendah, sementara provinsi seperti DKI Jakarta, Jawa Timur, dan Jawa Barat memiliki nilai jauh lebih tinggi sehingga distribusi bersifat menceng ke kanan (*right-skewed*) dan mengandung pencilan. Oleh karena itu, dilakukan transformasi logaritma natural pada PDRB untuk

menghasilkan distribusi yang lebih simetris dan stabil. Hasil transformasi menunjukkan log PDRB memiliki sebaran yang lebih mendekati normal, sehingga lebih sesuai digunakan dalam proses pemodelan.

3.2 Preprocessing

Hasil pemeriksaan data awal menunjukkan terdapat satu nilai hilang pada variabel Penanaman Modal Asing (PMA). Nilai yang hilang tersebut berasal dari salah satu provinsi di wilayah Papua. Penanganan dilakukan melalui imputasi menggunakan rata-rata nilai PMA provinsi-provinsi di Papua yang memiliki data lengkap. Pendekatan ini dipilih karena dianggap lebih representatif dibandingkan menggunakan rata-rata keseluruhan, mengingat karakteristik investasi di wilayah Papua berbeda dengan daerah lainnya. Setelah proses imputasi, tidak ditemukan lagi nilai hilang pada seluruh 17 kolom dataset. Selanjutnya, data dibagi menjadi data latih dan data uji dengan rasio 80:20. Dari total 38 observasi, sebanyak 30 observasi digunakan sebagai data latih dan 8 observasi sebagai data uji.

3.3 Pemodelan

3.3.1 Random Forest

Hyperparameter tuning dilakukan menggunakan metode *grid search* yang dikombinasikan dengan *5-Fold Cross Validation*. Kombinasi *hyperparameter* terbaik dilihat dari nilai RMSE terkecil, yang dapat dilihat pada Tabel 3. Kombinasi *hyperparameter* terbaik yang diperoleh selanjutnya digunakan untuk melatih ulang model pada seluruh data latih. Pemilihan rentang nilai *hyperparameter* bertujuan untuk mengontrol kompleksitas model serta mengurangi risiko *overfitting*. Melalui proses tersebut, diperoleh model *Random Forest* optimal dengan performa prediksi terbaik.

Tabel 3: *Hyperparameter* Random Forest

<i>Hyperparameter</i>	Kombinasi Terbaik
n_estimators	500
max_features	15
min_samples_leaf	4

Proses *tuning* pada model *Random Forest* menghasilkan kombinasi *hyperparameter* optimal yang digunakan pada tahap prediksi. Berdasarkan hasil *5-Fold Cross Validation*, model terbaik dipilih berdasarkan nilai metrik evaluasi yang paling baik. Selanjutnya, model final dievaluasi menggunakan data uji untuk mengukur performa prediksi pada data yang belum pernah dilatih sebelumnya.

3.3.2 Extra Trees Regression

Proses *Grid Search* model *Extra Trees Regression* dilakukan menggunakan ruang pencarian *hyperparameter* yang sama dengan *Random Forest*. Hasil kombinasi *hyperparameter* terbaik untuk *Extra Trees Regression* disajikan pada Tabel 4 dan didasarkan pada nilai RMSE terkecil.

Tabel 4: <i>Hyperparameter Extra Trees Regression</i>	
<i>Hyperparameter</i>	Kombinasi Terbaik
n_estimators	100
max_features	10
min_samples_leaf	2

Melalui evaluasi *5-Fold Cross Validation*, kombinasi *hyperparameter Extra Trees Regression* terbaik yang menghasilkan bias dan *varians* paling seimbang dipilih untuk melatih model akhir. Karena *Extra Trees Regression* memilih *split points* secara acak, optimasi pada *hyperparameter max_features* dan *min_samples_leaf* menjadi sangat krusial untuk menekan *noise* dari data.

3.4 Evaluasi Model

Untuk mengevaluasi kinerja model dalam memprediksi PDRB provinsi, penelitian ini membandingkan dua algoritma *ensemble learning*, yaitu *Random Forest* dan *Extra Trees Regression*. Evaluasi dilakukan menggunakan metrik RMSE, MAE, dan MAPE. Hasil perbandingan performa kedua model disajikan pada Tabel 5.

Tabel 5: Metrik Evaluasi Model			
Model	RMSE	MAE	MAPE
Random Forest	0.687786	0.527837	3.776066
Extra Trees Regression	0.637865	0.453316	3.199110

Karakteristik *Extra Trees Regression* yang menggunakan pemilihan titik pemisahan secara acak terbukti lebih efektif dalam mengurangi *varians* model pada dataset makroekonomi regional. Hal ini tercermin dari nilai RMSE, MAE, dan MAPE yang lebih rendah dibandingkan *Random Forest*. Pendekatan tersebut menghasilkan model yang lebih general dan mampu memberikan prediksi lebih akurat. Oleh karena itu, *Extra Trees Regression* dipilih sebagai model terbaik untuk mengestimasi PDRB provinsi di Indonesia.

3.5 Feature Importance

3.5.1 Random Forest

Analisis *feature importance* dilakukan untuk mengidentifikasi kontribusi masing-masing variabel terhadap prediksi PDRB. Analisis ini membantu mengetahui faktor-faktor utama yang paling memengaruhi tingkat PDRB suatu wilayah. Hasil *feature importance* bagi *Random Forest* disajikan pada Tabel 6.

Tabel 6: Lima peubah paling berkontribusi di model *Random Forest*

Peubah	Nilai
modal_inves_dalam_negeri_milyar_rupiah	0.3535
pengeluaran_per_kapita_ribu	0.0704
jumlah_pelayanan_kesehatan_satuan	0.0641
modal_inves_asing_juta_us_dollar	0.0355
Persentase_penduduk_miskin_persen	0.0251

Berdasarkan hasil analisis *feature importance* pada model *Random Forest*, variabel modal investasi dalam negeri dan pengeluaran per kapita menjadi faktor paling dominan dalam memprediksi PDRB dengan nilai sebesar 0.3535 dan 0.0704. Hasil ini menunjukkan bahwa tingginya investasi domestik dan besarnya pengeluaran per kapita memiliki peran penting dalam mendorong aktivitas ekonomi regional.

3.5.2 Extra Trees Regression

Selain pada *Random Forest*, analisis *feature importance* juga dilakukan menggunakan *Extra Trees Regression* untuk menguji konsistensi pengaruh setiap variabel prediktor. Lima variabel dengan nilai *feature importance* tertinggi pada model ini disajikan pada Tabel 7.

Tabel 7: Lima peubah paling berkontribusi di model *Extra Trees Regression*

Peubah	Nilai
modal_inves_dalam_negeri_milyar_rupiah	0.2947
jumlah_pelayanan_kesehatan_satuan	0.1749
pengeluaran_per_kapita_ribu	0.0489
Indeks_pembangunan_manusia_persen	0.0448
Persentase_penduduk_miskin_persen	0.0425

Pada model *Extra Trees Regression*, penanaman modal dalam negeri menjadi variabel paling berpengaruh, diikuti oleh jumlah pelayanan Kesehatan pada urutan kedua. Selain itu, pengeluaran per kapita juga

muncul sebagai faktor penting ketiga, yang menunjukkan adanya keterkaitan antara kemampuan daya beli masyarakat dan produktivitas ekonomi daerah.

4. Simpulan dan Saran

Penelitian ini membandingkan kinerja *Random Forest* dan *Extra Trees Regression* dalam memprediksi PDRB provinsi di Indonesia menggunakan 15 variabel prediktor sosial-ekonomi dari 38 provinsi. Hasil evaluasi menunjukkan bahwa *Extra Trees Regression* memiliki performa lebih baik dibandingkan *Random Forest* dengan RMSE 0.638, MAE 0.453, dan MAPE 3.199, sedangkan *Random Forest* menghasilkan RMSE 0.688, MAE 0.528, dan MAPE 3.776. Analisis *feature importance* menunjukkan bahwa penanaman modal dalam negeri merupakan variabel paling dominan pada kedua model, sementara jumlah sarana Kesehatan dan pengeluaran per kapita juga menjadi faktor penting pada model *Extra Trees Regression*. Dengan demikian, *Extra Trees Regression* direkomendasikan sebagai model yang lebih andal untuk estimasi PDRB provinsi di Indonesia. Penelitian selanjutnya disarankan untuk menggunakan data panel multitahun serta mempertimbangkan perbandingan dengan algoritma *ensemble* lainnya seperti XGBoost dan LightGBM untuk memperkuat validasi pemilihan model terbaik.

Daftar Pustaka

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45:5–32. <https://doi.org/https://doi.org/10.1023/A:1010933404324>
- Chen, J., Ooi, L. Q. R., Tan, T. W. K., Zhang, S., Li, J., Asplund, C. L., Eickhoff, S. B., Bzdok, D., Holmes, A. J., & Yeo, B. T. T. (2023). Relationship between prediction accuracy and feature importance reliability: An empirical and theoretical study. *NeuroImage*, 274. <https://doi.org/10.1016/j.neuroimage.2023.120115>
- Christy A.N, & Fahrati Eny. (2024). Analisis Pengaruh PDRB, Angkatan Kerja, dan Upah Minimum Terhadap Penyerapan Tenaga Kerja di Provinsi Kalimantan Selatan Tahun 2010-2022. *JIEP: Jurnal Ilmu Ekonomi dan Pembangunan*, 7(2):411–420. <https://doi.org/https://doi.org/10.20527/jiep.v7i2.309>
- Gatari, A. P., Dwi Asmara, G., & Khasanah, U. (2024). Analisis Produk Domestik Regional Bruto (PDRB) Terhadap Sektor Unggulan Guna Meningkatkan Pembangunan Ekonomi Kabupaten Madiun. *MULTIPLE: Journal of Global and Multidisciplinary*, 2(3):1470–1484. <https://journal.institercom-edu.org/index.php/multiple>

- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1):3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Gorritz, J. M., Clemente, R. M., Segovia, F., Ramirez, J., Ortiz, A., & Suckling, J. (2026). Is K-fold cross validation the best model selection method for Machine Learning? *Information Fusion*, 135. <https://doi.org/https://doi.org/10.1016/j.inffus.2026.104404>
- Jamshidi, E., Yusup, Y., Hooy, C., Kamaruddin, M., Hassan, H.M., Muhammad, S.A., Shafri, H.Z.M., Then, K.H., Norizan, M.S., Tan, C.C. (2024). Predicting Oil Palm Yield using a Comprehensive Agronomy Dataset and 17 Machine Learning and Deep Learning Models. *Ecological Informatics*, 81. <https://doi.org/10.1016/j.ecoinf.2024.102595>
- Magalhães, B., Bento, P., Pombo, J., Calado, M. do R., & Mariano, S. (2024). Short-Term Load Forecasting Based on Optimized Random Forest and Optimal Feature Selection. *Energies*, 17(8). <https://doi.org/10.3390/en17081926>
- Mahkya, D. Al, Notodiputro, K. A., & Sartono, B. (2022). Extra Trees Method for Stock Price Forecasting with Rolling Origin Accuracy Evaluation. *MEDIA STATISTIKA*, 15(1):36–47. <https://doi.org/10.14710/medstat.15.1.36-47>
- Mubarak, M. H., & Septian, F. (2025). Prediksi GDP dengan RF dan CGBoost Berdasarkan Aspek Sosia, Ekonomi, dan Lingkungan. *4TH MDP STUDENT CONFERENCE (MSC) 2025*, (1):338–346. <https://doi.org/https://doi.org/10.35957/mdp-sc.v4i1.11206>
- Nazilaturrohman, Nugroho, R. Y. Y., Wijanarko, A., Ahmad, & Setiawan, D. Y. (2024). Pengaruh Pertumbuhan Dan Kepadatan Penduduk, Teknologi, Dan Industri Manufaktur Terhadap Perekonomian Daerah. *Buletin Ekonomika Pembangunan*, 5(2):289–301. <https://doi.org/10.21107/bep.v5i2.26360>
- Nugroho, M. W. (2025). Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi. *Jurnal JTik: Jurnal Teknologi Informasi Dan Komunikasi*, 9(4):1562–1571. <https://doi.org/https://doi.org/10.35870/jtik.v9i4.4236>
- Permana, H., & Pasaribu, E. (2023). Pengaruh Inflasi, IPM, UMP, dan PDRB Terhadap Kemiskinan di Pulau Sumatera. *JIMEA: Jurnal Ilmiah*

- Manajemen, Ekonomi, dan Akutansi*, 7(3):1113–1132.
<https://doi.org/https://doi.org/10.31955/mea.v7i3.3516>
- Ria, R., Hasibuan, A., Kartika, A., Suwito, F. A., & Agustin, L. (2022). Pengaruh Produk Domestik Regional Bruto (PDRB) terhadap Tingkat Kemiskinan Kota Medan. *Reslaj: Religion Education Social Laa Roiba Journal*, 4(3):683–693.
<https://doi.org/10.47476/reslaj.v4i3.887>
- Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3(2).
<https://doi.org/10.1007/s42452-021-04148-9>
- Tarigan, S. A. B., Lopian, A. L. Ch. P., & Lumi, A. N. M. (2026). Pengaruh Pertumbuhan Ekonomi, Pengangguran, Indeks Pembangunan Manusia dan Penduduk Terhadap Kemsikinan di Kabupaten Karo Provinsi Sumatera Utara. *Jurnal Berkala Ilmiah Efisiensi*, 26(3).
- Yuliani, N. M., Fuadi, A. B., Arkan, M. N., & Helmi, S. G. Y. (2023). Pengaruh PMA dan PMDN Terhadap Pertumbuhan Ekonomi 34 Provinsi di Indonesia The Influence of FDI and PMDN on Economic Growth in 34 Provinces Indonesia. *JEMeS: Jurnal Ekonomi Manajemen Dan Sosial*, 6(2):3–50. <https://doi.org/https://doi.org/10.56071/jemes.v6i2.612>