 ISSN:2302-8149
e-ISSN:2540-9719

Perbandingan Metode *Machine Learning* (*Linear Regression, Random Forest,* dan *XGBoost*) dalam memprediksi Kemiskinan di Jawa Tengah Tahun 2024

Comparison of Machine Learning Methods (Linear Regression, Random Forest, and XGBoost) for Predicting Poverty in Central Java in 2024

[1]Zahwa Bunga Putri Pratama, [2]Yani Parti Astuti*
[1,2]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
[1,2]Jl. Imam Bonjol No.207, Semarang, Indonesia
e-mail: [1]zahwabngpr@gmail.com, [2]yanipartiastuti@dsn.dinus.ac.id

(***received***: 3 July 2025, ***revised***: 21 July 2025, ***accepted***: 22 July 2025)

Abstrak

Kemiskinan adalah isu besar yang dihadapi oleh Provinsi Jawa Tengah, dengan tingkat yang mengalami perubahan setiap tahun. Untuk merespons dan menangani masalah ini dengan lebih efektif, dibutuhkan pendekatan berbasis data yang bersifat prediktif. Penelitian ini menerapkan teknik machine learning untuk memprediksi jumlah penduduk miskin pada tahun 2024 di tingkat kabupaten/kota, dengan memanfaatkan data sosial ekonomi antara tahun 2019 hingga 2023 dari Badan Pusat Statistik (BPS). Tujuh indikator digunakan sebagai variabel prediktor, termasuk garis kemiskinan, jumlah serta persentase penduduk yang hidup dalam kemiskinan, tingkat pengangguran terbuka, rata-rata lama pendidikan, indeks pembangunan manusia, dan upah minimum di tingkat kabupaten/kota. Data yang diperoleh dinormalisasi menggunakan StandardScaler dan dibagi menjadi dua bagian, yaitu data pelatihan (80%) dan data pengujian (20%).Penelitian ini membandingkan tiga algoritma regresi, yakni Regresi Linier, Random Forest, dan XGBoost, untuk menilai seberapa efektif model-model tersebut dalam menangani kompleksitas data sosial ekonomi. Hasil analisis menunjukkan bahwa XGBoost menghasilkan kinerja terbaik dengan nilai MAE sebesar 6.665 dan R^2 sebesar 0.978, lebih unggul dibandingkan Random Forest (MAE 9.209; R^2 0.947) dan Regresi Linier (MAE 10.917; R^2 0.896). Perbandingan antara ketiga model ini dilakukan untuk mengisi celah dalam literatur yang berkaitan dengan efektivitas model dalam konteks prediksi kemiskinan lokal. Hasil dari penelitian ini menunjukkan bahwa XGBoost memiliki potensi untuk menjadi instrumen pendukung kebijakan berbasis data, khususnya dalam perencanaan dan usaha pengentasan kemiskinan di tingkat daerah.

Kata kunci: kemiskinan, prediksi, *machine learning,* jawa tengah

Abstract

Poverty is a major issue faced by Central Java Province, with rates fluctuating annually. To respond to and address this challenge more effectively, a predictive, data-driven approach is essential. This study applies machine learning techniques to forecast the number of people living in poverty in 2024 at the district/city level, utilizing socio-economic data from 2019 to 2023 provided by the Central Bureau of Statistics (BPS). Seven indicators are used as predictor variables, including the poverty line, the number and percentage of people living in poverty, the open unemployment rate, average years of schooling, the Human Development Index, and the regional minimum wage. The data were normalized using StandardScaler and split into training (80%) and testing (20%) sets. This study compares three regression algorithms—Linear Regression, Random Forest, and XGBoost—to evaluate their effectiveness in modeling the complexity of socio-economic data. The analysis reveals that XGBoost delivers the best performance, with a Mean Absolute Error (MAE) of 6,665 and an R^2 score of 0.978, outperforming Random Forest (MAE: 9,209; R^2: 0.947) and Linear Regression (MAE: 10,917; R^2: 0.896). By comparing these models, the study addresses a gap in the literature regarding the effectiveness of machine learning models for local-level poverty prediction. The findings suggest

 ISSN:2302-8149
e-ISSN:2540-9719

that XGBoost holds strong potential as a data-driven policy support tool, particularly in poverty alleviation planning and decision-making at the regional level.

Keywords: *poverty, prediction, machine learning, central java*

1 Pendahuluan

Kemiskinan masih menjadi tantangan utama dalam pembangunan nasional. Di Provinsi Jawa Tengah, perubahan angka kemiskinan menunjukkan bahwa pendekatan berbasis data sangat penting. Menurut data Badan Pusat Statistik (BPS), angka kemiskinan meningkat dari 10,80% pada tahun 2019 menjadi 11,41% pada 2020 karena dampak pandemi COVID-19, lalu turun drastis menjadi 7,53% pada bulan Maret 2023. Perubahan ini menunjukkan bahwa kebijakan untuk mengurangi kemiskinan perlu didukung oleh sistem prediktif yang mampu merespons perubahan sosial dan ekonomi secara tepat.

Dalam beberapa tahun terakhir, metode prediktif yang berlandaskan machine learning mulai diterapkan untuk mendukung proses pengambilan keputusan dalam kebijakan publik. Berbagai model seperti Regresi Linier, Random Forest, dan XGBoost telah banyak digunakan untuk menganalisis data sosial dan ekonomi. Regresi Linier memiliki keuntungan dalam mempermudah pemahaman hasil, tetapi kurang efisien dalam mengatasi hubungan yang tidak linear. Random Forest dapat menangkap interaksi rumit antar variabel dan telah terbukti berhasil dalam mengklasifikasikan status ekonomi rumah tangga di Cirebon dan Bengkulu [1]. Sementara itu, XGBoost dikenal dengan kinerjanya yang tinggi dalam menangani data besar dan tidak seimbang, seperti pada studi klasifikasi indeks desa dan prediksi GDP [2].

Berbagai penelitian yang ada pada umumnya menunjukkan kemampuan algoritma XGBoost dan Random Forest dalam mengolah data sosial ekonomi. Penelitian lain yang dilakukan oleh peneliti di bidang mitigasi kebakaran hutan menunjukkan bahwa kombinasi antara model XGBoost dan Random Forest juga memiliki efektivitas dalam mengelola data yang besar dan kompleks, seperti dalam memprediksi curah hujan, yang menunjukkan fleksibilitas algoritma ini di berbagai area aplikasi, baik lingkungan maupun sosial [3]. Namun, sebagian besar penelitian masih terfokus pada klasifikasi atau prediksi yang hanya mengandalkan fitur tunggal dan belum melakukan perbandingan secara langsung antara kinerja XGBoost, Random Forest, dan Regresi Linier dalam konteks prediksi jumlah orang miskin secara kuantitatif di tingkat kabupaten atau kota. Inilah yang menjadi celah penelitian yang ingin diatasi oleh studi ini. yang paling efisien dalam meramalkan jumlah penduduk miskin di Provinsi Jawa Tengah berdasarkan data historis dari tahun 2019 hingga 2023, sehingga dapat berfungsi sebagai alat bantu untuk perencanaan kebijakan pengurangan kemiskinan.

Studi-studi tersebut sebagian besar mengevaluasi masing-masing algoritma secara terpisah atau hanya fokus pada klasifikasi, bukan prediksi numerik. Penelitian yang secara langsung membandingkan ketiga metode tersebut dalam konteks prediksi jumlah penduduk miskin di tingkat provinsi dengan pendekatan kuantitatif dan multivariat masih jarang ditemukan. Padahal, pemodelan prediktif seperti ini sangat penting untuk merancang intervensi kebijakan yang lebih akurat dan tepat sasaran.

Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan melakukan evaluasi komparatif terhadap tiga algoritma machine learning, yaitu: (1) Regresi Linier sebagai pendekatan klasik; (2) Random Forest sebagai model berbasis ensemble decision tree; (3) XGBoost sebagai model boosting dengan regularisasi. Kontribusi utama dari studi ini adalah menawarkan rekomendasi algoritma yang paling efektif untuk memperkirakan jumlah penduduk yang hidup dalam kemiskinan di Provinsi Jawa Tengah dengan memanfaatkan data historis dari tahun 2019 hingga 2023, sehingga dapat berfungsi sebagai referensi dalam merumuskan kebijakan untuk menekan angka kemiskinan. Namun demikian, ada beberapa batasan dalam penelitian ini. Salah satu batasan adalah ukuran dataset yang terbatas, yaitu hanya mencakup 35 kabupaten/kota, sehingga jumlah sampel relatif kecil untuk penerapan model machine learning yang kompleks. Selain itu, data yang digunakan bersifat tahunan, sehingga tidak menangkap variasi musiman atau perubahan jangka pendek yang dapat berdampak pada kondisi sosial ekonomi masyarakat. Kedua faktor ini bisa membatasi kemampuan generalisasi model ke konteks atau periode waktu yang berbeda.

 ISSN:2302-8149
e-ISSN:2540-9719

2 Tinjauan Literatur

Penelitian tentang kemiskinan telah dilakukan secara luas dengan pendekatan yang beragam. Banyak peneliti menggunakan algoritma machine learning seperti Random Forest, XGBoost, dan regresi linier untuk mengklasifikasikan serta memprediksi tingkat kemiskinan dan indikator sosial ekonomi. Misalnya, [2] menggunakan XGBoost untuk mengklasifikasikan status Indeks Desa Membangun (IDM) di Jawa Barat dan memperoleh akurasi 89%, menunjukkan efektivitas algoritma ini dalam menangani data besar dan tidak seimbang. Hal ini Juga didukung oleh studi [4] serta [5] XGBoost sangat efektif dalam mengklasifikasikan tingkat sosial ekonomi dan mampu menangani data yang tidak seimbang.

Di sisi lain, [6] membandingkan Random Forest dan XGBoost dalam memprediksi Produk Domestik Bruto (GDP), dan menunjukkan bahwa XGBoost memberikan hasil prediksi yang lebih baik setelah dioptimasi dengan Whale Optimization Algorithm. Sedangkan [7] menunjukkan bahwa Random Forest lebih unggul dibandingkan metode lainnya dalam memprediksi angka harapan hidup dengan nilai RMSE lebih rendah dan R^2 lebih tinggi.

Random Forest digunakan secara luas dalam memprediksi kemiskinan dan memberikan hasil yang akurat di berbagai konteks. Random Forest juga menunjukkan kinerja kuat dalam klasifikasi kemiskinan di berbagai wilayah. [8] menggunakan Random Forest dan SMOTE untuk klasifikasi status ekonomi di Kota Cirebon, dengan akurasi hingga 93%. [1] menerapkannya untuk mengidentifikasi karakteristik rumah tangga miskin di Bengkulu menggunakan data SUSENAS, dan menemukan tiga variabel utama yaitu jumlah anggota rumah tangga, pendidikan kepala rumah tangga, dan luas lantai rumah. Sementara itu, [4] menggunakan Random Forest untuk memprediksi tingkat kemiskinan di Kabupaten Bogor berdasarkan data deret waktu, dan berhasil mencapai MAE 0,22% serta R^2 sebesar 0,87. Di sisi lain, [5] menunjukkan fleksibilitas Random Forest dalam domain kesehatan, dengan akurasi tinggi dalam klasifikasi risiko stroke menggunakan data tidak seimbang.

Pendekatan berbasis regresi linier tetap digunakan dalam studi yang mengkaji hubungan antar variabel sosial ekonomi. [9] dan [10] menemukan bahwa inflasi dan pengangguran memiliki hubungan signifikan terhadap kemiskinan, meskipun kontribusi IPM lebih dominan secara statistik. [11] menunjukkan bahwa pertumbuhan ekonomi dan rata-rata lama sekolah memiliki pengaruh negatif signifikan terhadap kemiskinan di Jawa Tengah. [12] bahkan memprediksi tingkat pengangguran dengan regresi linear berganda menggunakan variabel-variabel makroekonomi, meskipun hasil R^2 menunjukkan ruang prediksi yang masih dapat ditingkatkan.

Selain itu, pendekatan hybrid dan spasial juga banyak dikembangkan. [13] mengusulkan model hybrid fuzzy time series dan Holt-Winters untuk memprediksi kemiskinan di Jawa Tengah, dengan tingkat kesalahan prediksi sangat kecil (MAPE < 0,01%). Penelitian [14] menunjukkan bahwa optimalisasi fitur sosial ekonomi dan penggunaan Random Forest dapat meningkatkan prediksi kemiskinan di Sumatera Selatan dengan R^2 sebesar 0,72. Di sisi lain, [15] melakukan tinjauan sistematis dan menemukan bahwa sebagian besar penelitian machine learning berisiko bias tinggi karena permasalahan metodologis seperti ukuran sampel kecil dan pengolahan data hilang yang tidak tepat.

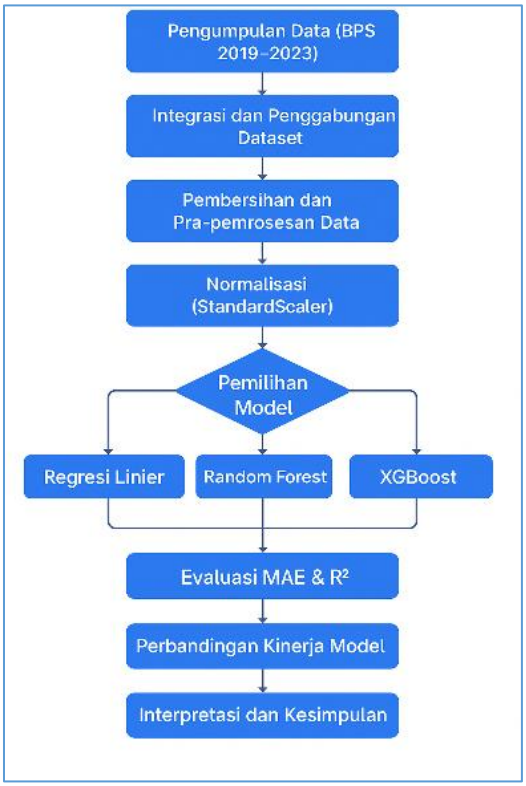
Melalui tinjauan sistematis, [16] menekankan bahwa pemilihan algoritma harus disesuaikan dengan karakteristik data. Linear Regression efektif untuk data dengan pola hubungan linier, sedangkan Random Forest dan KNN lebih unggul untuk data kompleks. Namun, KNN cenderung kurang stabil pada data tidak seimbang.

Berdasarkan literatur yang telah dibahas, tampak bahwa model-model tunggal seperti XGBoost, Random Forest, dan Regresi Linier masing-masing memiliki kelebihan tertentu, demikian pula dengan pendekatan hybrid. Meski demikian, penelitian sebelumnya biasanya hanya fokus pada satu daerah atau satu tahun data saja. Karena itu, penelitian ini bertujuan untuk mengatasi kelemahan tersebut dengan membandingkan ketiga model tersebut dalam memprediksi jumlah penduduk miskin di Provinsi Jawa Tengah secara lebih menyeluruh. Penelitian ini menggunakan data selama beberapa tahun serta indikator sosial ekonomi yang relevan, sehingga diharapkan dapat menghasilkan model prediksi yang lebih praktis dan akurat, yang dapat digunakan dalam merancang kebijakan untuk mengurangi tingkat kemiskinan.

 ISSN:2302-8149
e-ISSN:2540-9719

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode pembelajaran terpandu untuk menunjukkan hubungan antara faktor sosial ekonomi dan jumlah penduduk yang berada dalam kondisi kemiskinan di Provinsi Jawa Tengah. Tujuan utama dari penelitian ini adalah memprediksi jumlah penduduk miskin pada tahun 2024 berdasarkan data dan indikator yang tercatat pada tahun-tahun sebelumnya. Pendekatan ini didukung oleh studi sebelumnya yang menunjukkan efektivitas algoritma supervised learning dalam memodelkan fenomena sosial ekonomi, seperti yang dilakukan oleh [4] menggunakan Random Forest untuk memprediksi kemiskinan di Kabupaten Bogor, dengan hasil R^2 mencapai 0,87.



Gambar 1 Alur penelitian

3.1 Data dan Variabel

Data yang digunakan dalam penelitian ini berasal dari Badan Pusat Statistik (BPS) Provinsi Jawa Tengah, dengan periode waktu mulai tahun 2019 hingga 2023. Jenis data tersebut termasuk data sekunder yang bersifat tahunan dan dibagi berdasarkan kabupaten atau kota di seluruh provinsi tersebut. Peneliti menyampaikan setiap indikator secara manual dari situs resmi BPS, lalu menggabungkannya ke dalam satu file Excel untuk dilakukan analisis secara menyeluruh. Terdapat tujuh indikator utama yang digunakan sebagai fitur (variabel independen), yaitu.: (1) Garis kemiskinan (2019–2023); (2) Jumlah penduduk miskin (2019–2023); (3) Persentase penduduk miskin (2019–2023); (4) Tingkat pengangguran terbuka (2019–2023); (5) Rata-rata lama sekolah (2019–2023); (6) Indeks pembangunan manusia (2019–2023); (7) Upah minimum kabupaten/kota (2019–2023). Semua kolom yang berisi data dari tahun 2019 hingga 2023 digunakan sebagai fitur dalam proses prediksi. Selain itu, variabel yang diteliti adalah jumlah penduduk yang berada dalam kondisi kemiskinan pada tahun 2024, dalam satuan ribuan orang. Data tersebut dimasukkan secara manual oleh peneliti sebagai informasi nyata untuk menguji dan mengevaluasi akurasi prediksi yang telah dibuat.

Penelitian serupa oleh [13] Juga memanfaatkan data garis kemiskinan dan jumlah penduduk yang miskin dalam merancang model prediktif hybrid untuk wilayah Jawa Tengah, yang telah terbukti memberikan tingkat akurasi yang tinggi. Namun, penelitian oleh [15] perlu diberi peringatan mengenai kemungkinan besar terjadinya bias dalam model prediksi machine learning apabila tidak didukung oleh validasi dan pengelolaan data yang memadai.

 ISSN:2302-8149
e-ISSN:2540-9719

3.2 Pra-pemrosesan Data

Sebelum proses pemrosesan dimulai, dilakukan langkah awal untuk memastikan data tetap konsisten dan berkualitas. Nilai-nilai angka dalam data yang semula menggunakan tanda titik sebagai pemisah ribuan terlebih dahulu diubah dengan cara menghilangkan titik dan mengganti koma menjadi tanda desimal. Setelah proses ini selesai, semua data dikonversi menjadi bentuk numerik. Baris-baris yang mengandung nilai kosong (NaN) akibat kesalahan format kemudian dihapus.

Setelah pembersihan data selesai, semua fitur dinormalisasi menggunakan StandardScaler agar setiap variabel berada dalam skala yang seragam. Selanjutnya, dataset dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji, dengan parameter random_state=42 agar hasil pelatihan model dapat direplikasi. Teknik ini juga banyak digunakan dalam studi sebelumnya, seperti pada penelitian oleh [14] yang menerapkan normalisasi dan pengurangan dimensi untuk meningkatkan hasil prediksi kemiskinan di wilayah Sumatera Selatan.

3.3 Model dan Implementasi

Penelitian ini menggunakan tiga algoritma regresi untuk memodelkan dan membandingkan performa prediksi:

1. Regresi Linier: digunakan sebagai model dasar karena kesederhanaannya dan kemudahan dalam interpretasi.
2. Random Forest Regressor: model ensemble berbasis pohon keputusan yang mampu menangani data non-linear dan kompleks. Diterapkan dengan n_estimators=100 dan random_state=42.
3. XGBoost Regressor: metode boosting yang membangun model secara bertahap untuk meminimalkan error, dengan n_estimators=100.

Seluruh model dikembangkan menggunakan Python versi 3. 10 di platform Google Colaboratory, dengan menggunakan pustaka pandas untuk memproses data awal, scikit-learn untuk melakukan Regresi Linier dan Random Forest, serta xgboost untuk membuat model XGBoost. Platform ini juga digunakan oleh ((M. Husni Mubarok, Firza Septian) dalam studi prediksi PDB berbasis algoritma RF dan XGBoost.

3.4 Teknik Evaluasi

Evaluasi model dilakukan menggunakan dua metrik utama: (1) Mean Absolute Error (MAE): mengukur rata-rata kesalahan absolut antara hasil prediksi dan data aktual, dengan satuan yang sesuai dengan target (ribu jiwa). (2) R-squared (R^2): menunjukkan proporsi variansi data target yang dapat dijelaskan oleh model; nilai mendekati 1 menandakan prediksi yang baik. Evaluasi serupa digunakan oleh [7] dalam penelitian prediksi angka harapan hidup, dan oleh [13] Untuk mengevaluasi model hybrid kemiskinan, di mana R kuadrat dan MAE digunakan sebagai indikator utama kinerja tahun 2024 di tingkat kabupaten dan kota di Provinsi Jawa Tengah.

4 Hasil dan Pembahasan

Tabel 1 menampilkan hasil evaluasi dari tiga model yang digunakan dalam penelitian ini, yaitu Regresi Linier, Random Forest Regressor, dan XGBoost Regressor. Evaluasi dilakukan dengan menggunakan dua metrik, yaitu Mean Absolute Error (MAE) dan R-squared (R^2).

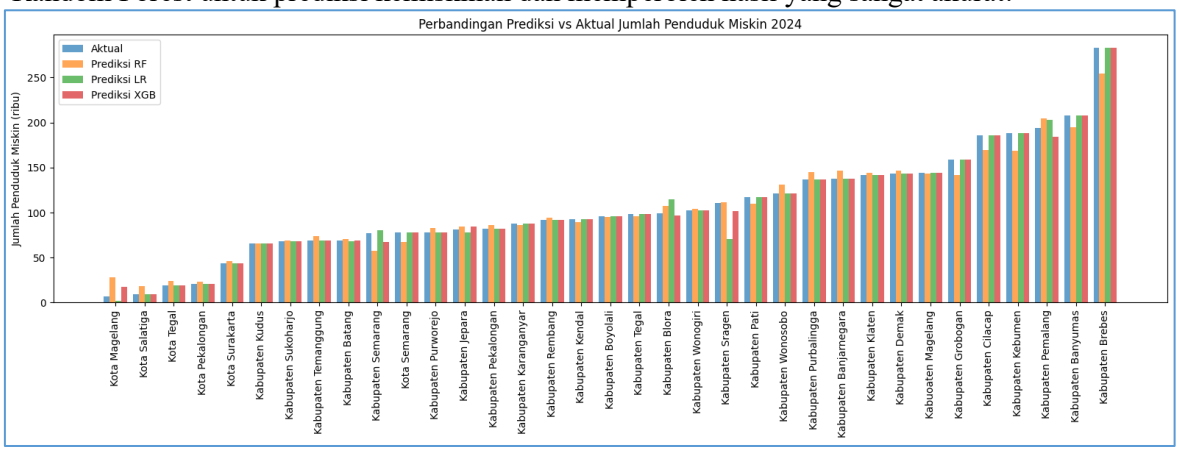
Tabel 1. Hasil evaluasi model

Model	MAE	R^2
Regresi Linier	10.917	0.896
Random Forest	9.209	0.947
XGBoost	6.665	0.978

Model XGBoost menunjukkan performa terbaik dalam meramalkan jumlah penduduk miskin pada tahun 2024, dengan penurunan MAE sebesar 38,94% dibandingkan Regresi Linier, serta 27,6% lebih baik dibandingkan Random Forest Regressor. Hasil ini sejalan dengan penelitian oleh (M. Husni Mubarok, Firza Septian) membuktikan keunggulan XGBoost dalam menangani variabel sosial ekonomi yang kompleks dan beragam. Selain itu, [2] juga menunjukkan bahwa XGBoost efektif dalam klasifikasi sosial ekonomi desa dengan tingkat akurasi tinggi.

 ISSN:2302-8149
e-ISSN:2540-9719

Kelemahan utama dari Regresi Linier adalah ketidakmampuannya dalam menghadapi pola data yang tidak linier serta rentan terhadap pengaruh outlier, seperti yang ditunjukkan dalam studi sistematis oleh Borianto dan [16] yang menyatakan bahwa Regresi Linier hanya cocok untuk hubungan data yang linier dan sederhana. Di sisi lain, Random Forest lebih mampu mengatasi kompleksitas dan ketidaklinieran dalam data, meskipun kinerjanya masih sedikit kalah dibandingkan XGBoost.Hal ini juga didukung oleh penelitian oleh [4] yang menggunakan Random Forest untuk prediksi kemiskinan dan memperoleh hasil yang sangat akurat.



Gambar 2 Grafik perbandingan nilai aktual dan prediksi tahun 2024

Sebagai ilustrasi hasil prediksi secara spesifik, Menampilkan nilai aktual dan prediksi dari model untuk beberapa kabupaten/kota dengan akurasi tinggi:

Tabel 2 Contoh hasil prediksi model XGBoost

Kabupaten/Kota	Nilai Aktual (ribu)	Prediksi (ribu)	Selisih Absolut (ribu)
Kota Semarang	77.79	77.79	0
Kabupatenn Banyumas	207.78	207.78	0
Kabupaten Cilacap	186.08	186.07	0.01
Kabupaten Brebes	283.28	283.28	0

Dari Tabel 2 di atas, terlihat bahwa model XGBoost mampu menghasilkan prediksi yang sangat dekat dengan nilai aktual. Hal ini memperkuat temuan bahwa model ini memberikan performa terbaik, sejalan dengan studi oleh [7] yang membandingkan berbagai model regresi dan menemukan bahwa Random Forest dan XGBoost memiliki performa prediktif paling unggul dalam konteks indikator social. Temuan ini juga didukung oleh [13] Model hybrid prediksi kemiskinan berhasil dibangun dan mencapai akurasi yang sangat tinggi melalui kombinasi teknik statistik dan machine learning.

5 Kesimpulan

Hasil evaluasi performa model menunjukkan bahwa XGBoost memiliki kinerja terbaik di antara kedua metode lainnya, dengan nilai Mean Absolute Error (MAE) sebesar 6. 665 dan koefisien determinasi (R^2) sebesar 0. 978. Ini menunjukkan bahwa model XGBoost dapat menjelaskan sekitar 97,8% variasi pada data target, mencerminkan kemampuannya untuk menangani pola yang non-linier dan kompleks dalam data sosial ekonomi. Di sisi lain, model Regresi Linier menghasilkan MAE sebesar 10. 917 dan R^2 sebesar 0. 896, menandakan bahwa model ini kurang mampu menangkap kompleksitas data yang tidak bersifat linier. Sementara itu, Random Forest menunjukkan kinerja yang cukup baik dengan MAE sebesar 9. 209 dan R^2 sebesar 0. 947, tetapi masih kalah dibandingkan dengan kinerja XGBoost. Temuan ini membuktikan bahwa metode boosting seperti XGBoost lebih efektif dalam meramalkan data sosial ekonomi yang kompleks dan tidak linier, yang juga sejalan dengan penelitian sebelumnya yang menyoroti keunggulan model boosting dalam berbagai domain prediksi. Selain itu, penelitian ini juga memberikan sumbangan untuk mengisi kekurangan dalam

 ISSN:2302-8149
e-ISSN:2540-9719

literatur, khususnya yang terkait dengan pemanfaatan data terbuka daerah untuk mengembangkan model prediktif kemiskinan yang aplikatif dan berdasarkan data aktual. Keberhasilan dalam menggunakan data BPS daerah juga menunjukkan bahwa sumber data terbuka nasional dapat dimanfaatkan secara strategis dalam mendukung pembangunan yang berbasis data. Selain itu, hasil penelitian ini juga memiliki potensi untuk dikembangkan menjadi sistem pendukung keputusan berbasis data untuk pemerintah daerah. Sistem ini bisa membantu dalam merancang dan mengevaluasi kebijakan pengentasan kemiskinan dengan cara yang lebih efektif dan responsif terhadap perubahan kondisi sosial ekonomi. Integrasi hasil prediksi dengan kebijakan berbasis bukti akan sangat penting untuk meningkatkan efisiensi program pemerintah serta memastikan alokasi sumber daya yang lebih tepat sasaran. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dari segi metodologi, tetapi juga potensi aplikasi dalam mendukung tata kelola pemerintahan yang lebih berbasis data.

Ucapan Terima Kasih

Saya ingin mengucapkan rasa terima kasih yang tulus kepada Badan Pusat Statistik (BPS) Provinsi Jawa Tengah atas kesediaan menyediakan data yang sangat berharga dan menjadi dasar utama dalam penelitian ini. Tanpa adanya data yang lengkap dan terbuka, penelitian ini tidak akan dapat berjalan secara optimal.

Referensi

[1] W. Agwil, D. Agustina, H. Fransiska, and N. Hidayati, “Klasifikasi Karakteristik Kemiskinan di Provinsi Bengkulu Tahun 2020 menggunakan Metode Pohon Klasifikasi Gabungan”, doi: 10.34123/jurnalasks.v14i2.348.

[2] S. A. Latfalia and R. R. Marliana, “Klasifikasi Status Indeks Desa membangun Jawa Barat menggunakan *Algoritma XGBoost*,” *Jurnal Riset Statistika*, pp. 75–82, Dec. 2024, doi: 10.29313/jrs.v4i2.5011.

[3] H. Hardianto, Y. N. Kunang, E. S. Negara, and T. Sutabri, “Model Data Mining dalam menganalisis Faktor-Faktor yang mempengaruhi Kemiskinan di Sumatera Selatan,” 2025, doi: 10.36040/jati.v9i2.12989.

[4] D. N. Handayani and S. Qutub, “Penerapan *Random Forest* untuk Prediksi dan Analisis Kemiskinan,” *RIGGS: Journal of Artificial Intelligence and Digital Business*, Vol. 4, No. 2, pp. 405–412, May 2025, doi: 10.31004/riggs.v4i2.512.

[5] D. A. Ubaid and M. I. F. Rachmad, “Klasifikasi Risiko Strok menggunakan *Algoritma Random Forest* dengan Teknik *Knowledge Discovery in Database*,” *Journal Computer Science*, Vol. 4, No. 1, 2025, [Online]. Available: https://www.kaggle.com/datasets/mahatiratushe, doi: 10.31294/5tbkg781

[6] M. H. Mubarok and F. Septian “Prediksi GDP dengan RF dan *XGBoost* berdasarkan Aspek Sosial, Ekonomi, dan Lingkungan,” doi: 10.35957/mdp-sc.v4i1.11206.

[7] G. Chairunisa, M. K. Najib, S. Nurdiati, S. F. Imni, W. Sanjaya, R. D. Andriani, H. Henriyansah, R. S. P. Putri, and D. Ekaputri, “*Life Expectancy Prediction using Decision Tree, Random Forest, Gradient Boosting, and XGBoost Regressions*,” *Jurnal Sintak*, Vol. 2, No. 2, 2024, doi: 10.62375/jsintak.v2i2.249

[8] N. N. Sholihah and A. Hermawan, “*Implementation of Random Forest and Smote Methods for Economic Status Classification in Cirebon City*,” *Jurnal Teknik Informatika (Jutif)*, Vol. 4, No. 6, pp. 1387–1397, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1135.

[9] R. P. P. Sinurat, “Analisis Faktor-Faktor Penyebab Kemiskinan sebagai Upaya Penanggulanagan Kemiskinan di Indonesia,” *Jurnal Registratie*, Vol. 5, No. 2, pp. 87–103, Dec. 2023, doi: 10.33701/jurnalregistratie.v5i2.3554.

[10] L. Priseptian and W. P. Primandhana,“Analisis Faktor-Faktor yang mempengaruhi Kemiskinan,” *Forum Ekonomi*, Vol. 24, No. 1, pp. 45–53, 2022, [Online]. Available: http://journal.feb.unmul.ac.id/index.php/FORUMEKONOMI, doi: 10.33701/jurnalregistratie.v5i2.3554.

[11] M. N. Faritz, “Pengaruh Pertumbuhan Ekonomi dan Rata-Rata Lama Sekolah terhadap Kemiskinan di Provinsi Jawa Tengah Ady Soejoto.”, doi: 10.26740/jupe.v8n1.p15-21

 ISSN:2302-8149
e-ISSN:2540-9719

[12] J. Halif, D. Wahiddin, I. Sanjaya, and S. Faisal, "Model Regresi Linear Berganda untuk Prediksi Tingkat Pengangguran di Provinsi Jawa Barat," *Jurnal Algoritma*, Vol. 22, No. 1, pp. 324–335, May 2025, doi: 10.33364/algoritma/v.22-1.2312.

[13] T. W. A. Putra, S. Solikhin, and M. Z. Abdillah, "Model *Hybrid* untuk Prediksi Jumlah Penduduk yang Hidup dalam Kemiskinan," *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 10, No. 6, pp. 1253–1264, Dec. 2023, doi: 10.25126/jtiik.2023107484.

[14] A. Heryati and T. S. Saputra, "*Optimizing Socioeconomic Features for Poverty Prediction in South Sumatera,*" *TIERS Information Technology Journal*, Vol. 6, No. 1, pp. 16–32, 2025, doi: 10.38043/tiers.v6i1.6244.

[15] C. L. A. Navarro *et al.*, "*Risk of Bias in Studies on Prediction Models Developed using Supervised Machine Learning Techniques: Systematic Review*," Oct. 20, 2021, *BMJ Publishing Group*. doi: 10.1136/bmj.n2281.

[16] Borito, B. Hendrik, "Infomatika Tinjauan Sistematis Metode Linear *Regression*, *K-Nearest Neighbor* dan *Random Forest* untuk Prediksi Tingkat Kemiskinan" *Jurnal Informatika, Manajemen dan Komputer*, Vol. 16, No. 2, 2024, doi: 10.36723/juri.v16i2.720