

## **Klasifikasi Indeks Pembangunan Manusia Menggunakan Model *Stacking Random Forest-XGBoost***

**Tsabita Rosyidah Putri<sup>1\*</sup>, Hanif Ziva Zavira<sup>1</sup>, Niken Sulistyowati<sup>1</sup>, Alfan Rizaldy Pratama<sup>1</sup>**

<sup>1</sup>Sains Data, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

---

Dikirimkan: 20-11-2025

Diterbitkan: 30-11-2025

**Keywords:**

IPM;  
XGBoost;  
Klasifikasi;  
Random Forest;  
Ensemble Learning.

**E-mail Penulis  
korespondensi:**

[22083010012@student.upnja.  
tim.ac.id](mailto:22083010012@student.upnja.tim.ac.id)

**Abstrak.** Pembangunan manusia merupakan indikator kunci keberhasilan pembangunan, di mana Indeks Pembangunan Manusia (IPM) menjadi untuk mengukur kualitas sumber daya manusia dan membedakan kategori wilayah (rendah, sedang, tinggi, sangat tinggi). Penelitian ini bertujuan untuk mengembangkan dan menguji model klasifikasi IPM yang robust dengan memanfaatkan keunggulan ensemble learning melalui teknik Stacking yang mengintegrasikan Random Forest (RF) dan XGBoost. Data yang digunakan adalah IPM Kabupaten/Kota di Provinsi Jawa Timur tahun 2024 (38 sampel dari BPS), yang diproses dengan LabelEncoder dan StandardScaler sebelum dibagi menjadi 75% data latih dan 25% data uji. Model RF dan XGBoost digunakan sebagai base learner yang kemudian digabungkan melalui Stacking. Hasil penelitian menunjukkan bahwa Model Stacking mencapai performa klasifikasi yang sempurna dengan akurasi 100% pada data uji, terkonfirmasi melalui Confusion matrix tanpa adanya kesalahan prediksi. Akurasi ini menyamai RF dan mengungguli XGBoost (90%). Analisis Feature importance secara konsisten mengidentifikasi Rata-rata Lama Sekolah, Indeks Pengetahuan, dan Pengeluaran sebagai variabel paling dominan dalam klasifikasi IPM. Terbukti bahwa model Stacking Ensemble XGBoost-Random Forest terbukti efektif dan sangat andal untuk klasifikasi IPM. Temuan ini menegaskan bahwa faktor pendidikan dan ekonomi adalah kunci utama penentu kategori pembangunan manusia, yang dapat menjadi landasan strategis untuk intervensi kebijakan.

**Abstract.** Human development is a key indicator of development success, where the Human Development Index (HDI) is used to measure the quality of human resources and distinguish between regional categories (low, medium, high, very high). This study aims to develop and test a robust HDI classification model by utilizing the advantages of ensemble learning through the Stacking technique, which integrates Random Forest (RF) and XGBoost. The data used is the HDI of districts/cities in East Java Province in 2024 (38 samples from BPS), which was processed with LabelEncoder and StandardScaler before being divided into 75% training data and 25% test data. The RF and XGBoost models were used as base learners, which were then combined through Stacking. The results showed that the Stacking Model achieved perfect classification performance with 100% accuracy on the test data, confirmed by the Confusion matrix with no prediction errors. This accuracy matched that of RF and surpassed that of XGBoost (90%). Feature importance analysis consistently identified Average Length of Schooling, Knowledge Index, and Expenditure as the most dominant variables in HDI classification. It was proven that the XGBoost-Random Forest Stacking Ensemble model was effective and highly reliable for HDI classification. These findings confirm that educational and economic factors are the main

## 1. Pendahuluan

Pembangunan manusia menjadi salah satu unsur yang sangat penting dalam menilai keberhasilan suatu pembangunan, karena pertumbuhan ekonomi yang tinggi saja tidak selalu mampu menjamin peningkatan kesejahteraan masyarakat secara merata [1]. Dalam praktiknya, negara atau wilayah dengan pertumbuhan ekonomi yang pesat masih dapat menghadapi berbagai permasalahan sosial, seperti tingginya tingkat kemiskinan, kesenjangan pendapatan, rendahnya akses terhadap pendidikan dan pelayanan kesehatan, serta kualitas hidup yang kurang memadai bagi sebagian besar penduduk [2]. Oleh karena itu, fokus pada pembangunan manusia, yang mencakup peningkatan kapasitas, kualitas hidup, dan kesempatan masyarakat untuk berpartisipasi dalam proses pembangunan, menjadi indikator kunci untuk menilai seberapa efektif suatu kebijakan pembangunan dalam menciptakan kesejahteraan yang berkelanjutan [3]. Dalam penelitian ini, digunakan data Indeks Pembangunan Manusia (IPM) sebagai objek perbandingan berbagai algoritma klasifikasi. Pemilihan data IPM didasarkan pada dua alasan utama: pertama, IPM merupakan salah satu variabel makroekonomi penting yang mencerminkan kondisi sumber daya manusia di Indonesia; kedua, IPM memiliki kategori klasifikasi yang jelas sehingga cocok digunakan dalam metode *supervised learning*[4]. Berdasarkan ketentuan Badan Pusat Statistik (BPS), klasifikasi IPM dibagi menjadi empat kategori, yaitu: rendah jika  $IPM < 60$ , sedang jika  $60 \leq IPM < 70$ , tinggi jika  $70 \leq IPM < 80$ , dan sangat tinggi jika  $IPM \geq 80$  [5].

Perkembangan terkini dalam *machine learning* menunjukkan bahwa pendekatan *ensemble learning*, khususnya *stacking* atau *hybrid ensemble*, mampu meningkatkan reliabilitas prediksi dengan mengkombinasikan keunggulan masing-masing algoritma [6]. Namun, Sejumlah penelitian sebelumnya umumnya menerapkan metode tunggal seperti Regresi Linier, SVM, *Random Forest*, atau *XGBoost* untuk mengklasifikasikan IPM. Sebagai contoh, penelitian oleh [7] telah menerapkan beberapa algoritma *ensemble learning* seperti *Random Forest*, SVM, dan *AdaBoost* untuk melakukan klasifikasi Indeks Pembangunan Manusia (IPM), yang menunjukkan relevansi metode *ensemble* dalam meningkatkan ketepatan prediksi pada data pembangunan daerah. Penelitian oleh [8] memprediksi jumlah penduduk miskin di Jawa Tengah tahun 2024 menggunakan data sosial ekonomi 2019-2023 dari BPS. Tiga algoritma *machine learning* yaitu *Linear Regression*, *Random Forest*, dan *XGBoost* dibandingkan untuk melihat model yang paling efektif. Hasilnya, *XGBoost* menunjukkan performa terbaik dengan MAE 6.665 dan  $R^2$  0.978 mengungguli dua model lain. Temuan ini menunjukkan bahwa *XGBoost* berpotensi menjadi alat pendukung kebijakan dalam perencanaan pengentasan kemiskinan di tingkat daerah. Selanjutnya, penelitian oleh [9] memprediksi Indeks Pembangunan Manusia (IPM) di wilayah Indonesia bagian Timur menggunakan data komponen IPM seperti Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Pengeluaran per Kapita Disesuaikan (PKD), dan Lama Sekolah (LS). Tiga metode klasifikasi Regresi Logistik, *Decision Tree*, dan *Random Forest* dibandingkan untuk menentukan model terbaik. Hasil menunjukkan bahwa mayoritas daerah (69%) termasuk kategori IPM rendah hingga sedang, sementara 31% berada pada kategori tinggi dan sangat tinggi. Model *Random Forest* memberikan kinerja terbaik dengan akurasi 94,03% dan balanced accuracy 93,33%, serta hanya 4,08% prediksi yang tidak tepat. Variabel UHH terbukti memiliki pengaruh paling signifikan terhadap perubahan nilai IPM di wilayah ini. Hingga saat ini, belum terdapat penelitian yang secara khusus melakukan integrasi antara algoritma *XGBoost* dan *Random Forest* dalam suatu kerangka *ensemble* untuk keperluan klasifikasi Indeks Pembangunan Manusia (IPM) di Indonesia pada tahun 2024. Sebagian besar studi sebelumnya cenderung menggunakan metode tunggal atau membandingkan kinerja masing-masing algoritma secara terpisah, sehingga belum ada pendekatan yang memanfaatkan kekuatan gabungan kedua algoritma tersebut untuk meningkatkan akurasi, stabilitas, dan kemampuan model dalam menangkap pola kompleks dari data IPM terbaru.

Gap analisis dalam penelitian membuka peluang untuk memanfaatkan metode yang lebih stabil dan mampu meningkatkan konsistensi prediksi. Teknik *ensemble* seperti *XGBoost* dan *Random Forest* memiliki karakteristik yang saling melengkapi *XGBoost* efektif memperbaiki kesalahan secara bertahap melalui *boosting* [10], sementara *Random Forest* mengurangi varians dan meningkatkan keandalan klasifikasi melalui pembentukan banyak pohon keputusan secara acak [11]. Novelty dari penelitian ini terletak pada pengembangan dan penerapan model *ensemble hybrid XGBoost-Random Forest* yang difokuskan secara khusus untuk klasifikasi tingkat Indeks Pembangunan Manusia (IPM) tahun 2024. Pendekatan ini tidak hanya sekadar membandingkan kinerja masing-masing algoritma secara terpisah, melainkan menghadirkan integrasi dua metode berbasis pohon keputusan yang saling melengkapi, sehingga diharapkan dapat meningkatkan akurasi, kestabilan, dan kemampuan generalisasi model terhadap variasi data antarwilayah. Penelitian ini bertujuan untuk membangun model klasifikasi IPM yang lebih handal dan robust dengan memanfaatkan keunggulan masing-masing algoritma, sekaligus melakukan identifikasi variabel-variabel pembangunan yang memiliki kontribusi paling signifikan terhadap perubahan nilai IPM melalui analisis *feature importance* dari model *ensemble* yang dikembangkan. Dengan demikian, hasil

penelitian ini diharapkan tidak hanya memberikan prediksi yang lebih tepat, tetapi juga menjadi dasar informasi yang dapat mendukung pengambilan keputusan.

## 2. Metode Penelitian

Penelitian ini menggunakan metode kuantitatif berbasis pembelajaran mesin untuk menganalisis Indeks Pembangunan Manusia (IPM) Kabupaten/Kota di Provinsi Jawa Timur tahun 2024. Dataset diperoleh dari Badan Pusat Statistik (BPS) dan terdiri atas 38 sampel, masing-masing mewakili satu kabupaten/kota. Setiap sampel memiliki sejumlah variabel indikator sosial-ekonomi yang seluruhnya direpresentasikan sebagai fitur numerik. Tahapan penelitian dilakukan melalui prosedur berikut.

### 2.1. Pra-pemrosesan Data

Data dibersihkan dari nilai kosong dan disiapkan untuk pemodelan melalui dua teknik standar, yaitu *Encoding* untuk mengkonversi variabel kategorik menjadi bentuk numerik, serta *Z-Score* untuk menormalkan skala seluruh fitur agar memiliki variansi yang seragam. Kedua teknik ini merupakan metode mapan dalam literatur pembelajaran mesin sebagaimana dijelaskan pada rujukan.

### 2.2. Pembagian Data

Dataset dipisahkan menjadi data latih dan data uji menggunakan fungsi *train\_test\_split*. Proporsi yang digunakan adalah 75% data latih dan 25% data uji, dengan pemilihan sampel secara acak untuk menjaga keberagaman dan representativitas distribusi kelas. Pemisahan dilakukan satu kali tanpa replikasi berulang, mengingat ukuran dataset kecil dan mencakup keseluruhan populasi wilayah pengamatan.

### 2.3. Reduksi Dimensi Menggunakan t-SNE

Visualisasi pola dan hubungan antar-wilayah dilakukan menggunakan metode t-Distributed Stochastic Neighbor Embedding (t-SNE) sebagai teknik reduksi dimensi non-linear [12]. Untuk memastikan replikasi, parameter ditetapkan secara eksplisit, yaitu *n\_components* = 2, *perplexity* ditetapkan berdasarkan eksplorasi awal sesuai ukuran sampel, *learning\_rate* disesuaikan untuk menjaga stabilitas proses optimasi.

### 2.4. Pemodelan Klasifikasi

Pemodelan dilakukan menggunakan dua algoritma *supervised learning* yang telah mapan dan banyak digunakan dalam analisis prediktif, yaitu *Random Forest Classifier* dan *XGBoost Classifier*. Kedua model ini dipilih karena mampu menangani data berukuran kecil hingga menengah, memiliki performa prediksi yang stabil, dan tahan terhadap *overfitting* melalui mekanisme pengendalian kompleksitas model.

#### 2.4.1. Random Forest

*Random Forest* adalah pengembangan dari metode Decision Tree yang bekerja dengan membangun banyak pohon keputusan sekaligus. Setiap pohon dilatih menggunakan sampel data yang berbeda, dan pemilihan atribut yang digunakan untuk pemecahan node pun dilakukan secara acak dari subset fitur tertentu. Pendekatan ini membuat model lebih tahan terhadap *overfitting* dan mampu meningkatkan akurasi prediksi, karena keputusan akhir diperoleh dari kombinasi berbagai pohon yang dibentuk. Berikut merupakan rumus untuk menghitung mean (rata-rata) populer [13]. Persamaan 1 adalah untuk mencari rata-rata populer:

$$f(x) = \text{rata-rata } (f_1(x), f_2(x), \dots, f_n(x)) \quad (1)$$

Persamaan 1 menjelaskan bahwa  $f(x)$  merupakan nilai prediksi akhir yang dihasilkan oleh model. Variabel  $x$  adalah data input yang dimasukkan ke dalam model. Sementara itu,  $f_1(x), f_2(x), \dots, f_n(x)$  adalah hasil prediksi dari masing-masing pohon keputusan yang membentuk model *ensemble*, seperti pada *Random Forest* atau *Gradient Boosting*. Dengan kata lain, setiap pohon memberikan prediksi terhadap input  $x$ , kemudian seluruh prediksi tersebut digabungkan misalnya dengan cara dirata-ratakan atau dijumlahkan untuk menghasilkan prediksi akhir  $f(x)$  yang lebih stabil dan akurat dibandingkan prediksi dari satu pohon keputusan saja.

#### 2.4.2 XGBoost Classifier

*XGBoost* (Extreme Gradient Boosting) adalah algoritma machine learning yang sangat efisien dan kuat untuk menyelesaikan masalah regresi maupun klasifikasi. Dalam konteks regresi, *XGBoost* digunakan untuk memprediksi nilai kontinu dengan membangun serangkaian pohon keputusan yang dioptimalkan melalui proses boosting. Setiap pohon yang ditambahkan bertujuan untuk memperbaiki kesalahan prediksi dari pohon-pohon sebelumnya. Dengan cara ini, model secara bertahap menjadi lebih akurat dan mampu menghasilkan performa prediksi yang lebih baik secara keseluruhan. Pada metode ini, diperlukan fungsi objektif yang berfungsi untuk

menilai seberapa baik model yang dihasilkan sesuai dengan data pelatihan [14]. Karakteristik utama dari fungsi objektif terdiri dari dua komponen, yaitu nilai kerugian pelatihan dan nilai regularisasi, sebagaimana dijelaskan pada persamaan 2 berikut ini:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (2)$$

Dalam hal ini,  $L$  merepresentasikan fungsi kerugian pada proses pelatihan,  $\Omega$  adalah fungsi regularisasi, dan  $\theta$  menggambarkan parameter-parameter yang digunakan dalam model. Secara umum, bentuk fungsi kerugian pada proses pelatihan dapat dirumuskan seperti pada persamaan 3 berikut.

$$L(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) \quad (3)$$

Di mana  $y_i$  merupakan nilai sebenarnya yang digunakan sebagai acuan kebenaran,  $\hat{y}_i$  adalah nilai prediksi yang dihasilkan oleh model, dan  $n$  menunjukkan jumlah iterasi atau banyaknya data yang digunakan dalam proses pemodelan.

## 2.5. Evaluasi Model

Pada penelitian ini, kinerja model klasifikasi dievaluasi menggunakan data uji dengan menerapkan beberapa metrik yang telah mapan dalam analisis *supervised learning*. Evaluasi dilakukan untuk menilai kemampuan model dalam mengklasifikasikan kategori IPM secara akurat serta mengidentifikasi pola kesalahan prediksi. Tiga metrik utama digunakan dalam penelitian ini, yaitu *Confusion matrix*, *Classification Report (precision, recall, dan F1-score)*, serta tingkat error. Penggunaan ketiga metrik tersebut memungkinkan peneliti memperoleh gambaran komprehensif mengenai performa model, baik secara keseluruhan maupun pada masing-masing kelas.

*Confusion matrix* adalah sebuah tabel yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan prediksi model terhadap nilai sebenarnya [15]. Tabel ini menampilkan jumlah prediksi yang benar dan salah untuk setiap kelas, sehingga memudahkan analisis kesalahan klasifikasi. *Confusion matrix* berbentuk  $2 \times 2$ . Pada Tabel 1 terdapat struktur *confusion matrix* untuk mengetahui klasifikasi biner berikut struktur *confusion matrix*:

Tabel 1. Struktur *Confusion matrix*

		Observasi	
		<i>True</i>	<i>False</i>
<i>Prediction Class</i>	<i>True</i>	<i>True Positif (TP)</i>	<i>False Positive (FP)</i>
	<i>False</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

TP (*True Positive*) merupakan jumlah data yang sebenarnya positif dan berhasil diprediksi sebagai positif oleh model. TN (*True Negative*) adalah jumlah data yang benar-benar negatif dan tepat diprediksi sebagai negatif. FP (*False Positive*) menggambarkan jumlah data yang sebenarnya negatif namun salah diprediksi sebagai positif, yang dikenal sebagai kesalahan tipe I. Sebaliknya, FN (*False Negative*) menunjukkan jumlah data yang sebenarnya positif tetapi keliru diprediksi sebagai negatif, dan kesalahan ini disebut sebagai tipe II. Keempat komponen ini digunakan untuk mengevaluasi kinerja model klasifikasi secara menyeluruh.

Dari *confusion matrix*, berbagai metrik kinerja model klasifikasi dapat dihitung untuk menilai seberapa baik model bekerja. Metrik-metrik ini meliputi akurasi, presisi, *recall* (sensitivitas), dan *F1-score*, yang masing-masing memberikan informasi berbeda tentang performa model. Dengan menggunakan nilai-nilai dalam *confusion matrix*, kita dapat mengetahui proporsi prediksi yang benar, seberapa tepat model dalam memprediksi kelas tertentu, serta kemampuan model dalam mendeteksi semua kasus positif atau negatif.

Akurasi menggambarkan seberapa besar proporsi data yang berhasil diklasifikasikan dengan benar oleh model dibandingkan dengan seluruh data yang diuji. Dengan kata lain, akurasi menunjukkan tingkat keseluruhan ketepatan prediksi model [16]. Nilai akurasi dapat dihitung menggunakan Persamaan 4.

$$Akurasi = \frac{TP}{TP+TN+FP+FN} \times 100\% \quad (4)$$

Metrik presisi menggambarkan seberapa akurat model dalam mengidentifikasi data yang dikategorikan sebagai positif, dengan menekankan proporsi prediksi positif yang benar-benar sesuai dengan kondisi sebenarnya [17]. Berikut perhitungan metrik presisi:

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

*Recall* menggambarkan seberapa banyak data yang benar-benar termasuk dalam kelas positif berhasil terdeteksi dengan tepat oleh model. Dengan kata lain, recall mengukur kemampuan model dalam mengenali seluruh kasus positif yang seharusnya teridentifikasi [18].

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

*F1-score* merupakan rata-rata harmonis dari nilai precision dan recall, sehingga memberikan ukuran kinerja model yang seimbang antara ketepatan prediksi dan kemampuan mendeteksi semua kasus positif [19].

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Metrik error menunjukkan seberapa besar proporsi data yang memang termasuk kategori positif yang berhasil diidentifikasi dengan benar oleh model, sehingga memberikan gambaran tentang kemampuan model dalam mendeteksi semua kasus positif dan mengurangi kemungkinan melewatkannya data positif yang sebenarnya ada [20].

$$Error = \frac{FP}{TP} \times 100\% \quad (8)$$

### 3. Hasil dan Pembahasan

#### 3.1. Hasil Model Dasar dan *Stacking*

Penelitian ini menerapkan metode *ensemble learning* berbasis *stacking* untuk mengklasifikasikan kategori Indeks Pembangunan Manusia (IPM) menggunakan dua model dasar, yaitu *Random Forest* dan *XGBoost*. Proses pelatihan dilakukan menggunakan data latih, sedangkan evaluasi dilakukan pada data uji. Hasil evaluasi performa menunjukkan adanya perbedaan akurasi antara model dasar, serta peningkatan kinerja setelah model digabungkan melalui teknik *stacking*. Akurasi masing-masing model ditampilkan pada Tabel 2.

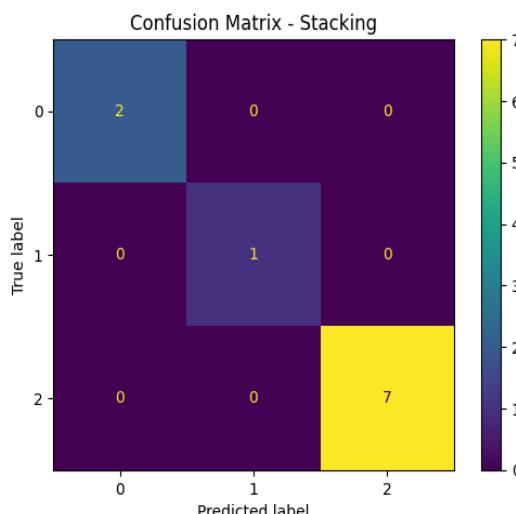
Tabel 2. Akurasi Model Dasar dan Model *Stacking*

Model	Akurasi
<i>Random Forest</i>	100%
<i>XGBoost</i>	90%
<i>Stacking</i> ( <i>Random Forest</i> + <i>XGBoost</i> )	100%

Hasil menunjukkan bahwa *Random Forest* memberikan akurasi tertinggi, yaitu 100%, sementara *XGBoost* hanya mencapai 90%. Perbedaan ini mengindikasikan bahwa *Random Forest* mampu menangkap pola data secara lebih komprehensif pada dataset ini, sedangkan *XGBoost* masih menghasilkan sejumlah prediksi yang keliru. Namun demikian, ketika kedua model digabungkan melalui arsitektur *stacking*, akurasi yang dihasilkan kembali mencapai 100%, sama seperti performa *Random Forest* sebagai model terbaik.

#### 3.2. Analisis *Confusion matrix - Stacking*

Evaluasi performa model *stacking* dianalisis lebih lanjut menggunakan *confusion matrix* yang disajikan pada Gambar 1.

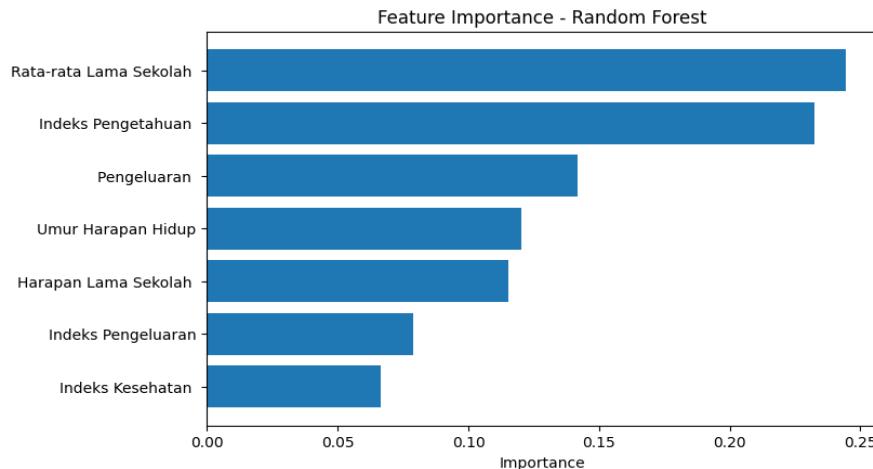


Gambar 1. *Confusion matrix Stacking Model*

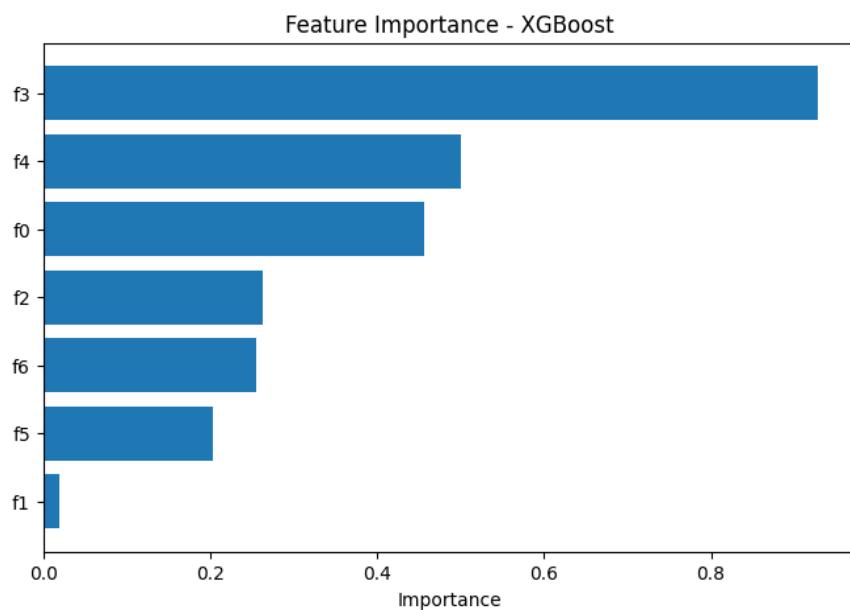
*Confusion matrix* memperlihatkan bahwa model *stacking* mampu mengklasifikasikan seluruh sampel data uji dengan benar pada semua kelas tanpa kesalahan prediksi. Tidak adanya nilai off-diagonal pada *confusion matrix* mengindikasikan bahwa meta-model berhasil mengambil keputusan yang tepat berdasarkan kombinasi prediksi dari kedua model dasar. Hal ini menunjukkan bahwa *stacking* memiliki kemampuan generalisasi yang sangat baik pada dataset IPM yang digunakan dalam penelitian ini.

### 3.3. Analisis Feature importance Model Dasar

Selain itu, interpretasi fitur dilakukan melalui analisis *feature importance* dari model dasar, yaitu *Random Forest* dan *XGBoost*. Hasil visualisasi ditunjukkan pada Gambar 2 dan Gambar 3.



Gambar 2. *Feature importance Random Forest*



Gambar 3. *Feature importance XGBoost*

Pada model *Random Forest*, fitur dengan kontribusi terbesar adalah Rata-rata Lama Sekolah, diikuti oleh Indeks Pengetahuan dan Pengeluaran. Ketiga variabel ini memiliki nilai importance tertinggi, sehingga dapat disimpulkan bahwa faktor pendidikan dan aspek ekonomi sangat dominan dalam proses klasifikasi kategori IPM. Sementara itu, pada model *XGBoost*, fitur dengan pengaruh terbesar adalah f3, yaitu fitur yang mewakili variabel Pengeluaran (berdasarkan urutan fitur pada dataset). Fitur penting berikutnya adalah f4 dan f0, yang berhubungan dengan variabel pendidikan, seperti Indeks Pengetahuan dan Rata-rata Lama Sekolah.

Konsistensi fitur penting pada kedua model dasar, khususnya variabel yang berkaitan dengan pendidikan dan pengeluaran, memperkuat bahwa kedua aspek tersebut memiliki peran signifikan dalam membedakan kategori

IPM. Hal ini mengindikasikan bahwa daerah dengan pendidikan lebih baik dan tingkat pengeluaran yang lebih tinggi cenderung memiliki kategori IPM yang lebih tinggi.

Secara keseluruhan, hasil penelitian ini membuktikan bahwa metode *stacking* efektif digunakan pada kasus klasifikasi IPM. Meskipun *XGBoost* menunjukkan performa yang lebih rendah secara individual, keberadaannya tetap memberikan keragaman prediktif dalam *ensemble*. Pada akhirnya, *stacking* mampu mengoptimalkan keputusan meta-model dan menghasilkan performa prediksi yang sempurna dengan akurasi **100%** pada data uji.

#### 4. Kesimpulan

Penelitian ini membuktikan bahwa metode *Stacking Ensemble Learning* yang menggabungkan *Random Forest* dan *XGBoost* sangat efektif dalam mengklasifikasikan kategori Indeks Pembangunan Manusia (IPM) Kabupaten/Kota di Jawa Timur. Model *Stacking* mencapai akurasi sempurna sebesar 100% pada data uji, menyamai performa *Random Forest* dan melampaui *XGBoost* yang memperoleh akurasi 90%. Hasil *confusion matrix* menunjukkan tidak ada kesalahan prediksi, menandakan kemampuan generalisasi model yang sangat baik. Analisis fitur penting dari kedua model dasar juga menegaskan bahwa variabel pendidikan dan pengeluaran seperti Rata-rata Lama Sekolah, Indeks Pengetahuan, serta Pengeluaran merupakan faktor paling dominan dalam menentukan kategori IPM. Temuan ini menunjukkan bahwa peningkatan kualitas pendidikan dan daya beli masyarakat memiliki pengaruh besar terhadap pembangunan manusia di suatu wilayah. Ke depannya, penelitian dapat diperkuat dengan penggunaan data *time-series*, validasi silang, serta analisis lebih mendalam terhadap kontribusi masing-masing model dasar dalam mekanisme *Stacking* untuk memastikan kestabilan dan ketahanan model pada data yang lebih luas.

#### Daftar Rujukan

- [1] S. Y. Kusumastuti, L. Judijanto, T. Kunawangsih, and S. Sa'dianoor, *Ekonomi Pembangunan Lanjutan*. PT. Sonpedia Publishing Indonesia, 2025.
- [2] M. P. Hababil, M. K. Firdaus, N. Nazhmi, M. D. Hamdani, M. R. Alghifary, and A. Fadilla, “Analisis pengaruh pemerataan ekonomi dalam upaya menghapus ketimpangan sosial-ekonomi antar masyarakat,” *J. Macroecon. Soc. Dev.*, vol. 1, no. 4, pp. 1–9, 2024.
- [3] N. Ramadhani and Y. P. Utomo, “Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia (IPM) di Jawa Timur Tahun 2010-2020,” *J. Bisnis Dan Manaj.*, vol. 3, no. 2, pp. 269–276, 2023.
- [4] R. Ningsih, “Pengaruh Pertumbuhan Ekonomi, Upah Minimum dan Indeks Pembangunan Manusia Terhadap Jumlah Penduduk Miskin di Provinsi Lampung,” IAIN Metro, 2025.
- [5] “indeks-pembangunan-manusia-provinsi-jawa-timur-2024.” [Online]. Available: <https://jatim.bps.go.id/publication/2025/05/27/47fde052cb353c601c21c209/indeks-pembangunan-manusia-provinsi-jawa-timur-2024.html>
- [6] C. Supriyanto and A. Marjuni, “Peningkatan keberagaman data untuk klasifikasi penyakit diabetes berbasis *stacking ensemble learning*,” *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 1–10, 2025.
- [7] R. I. Arumnisa and A. W. Wijayanto, “Comparison of *Ensemble Learning* Method: *Random Forest*, Support Vector Machine, AdaBoost for Classification Human Development Index (HDI),” *SISTEMASI*, vol. 12, no. 1, pp. 206–218, 2023.
- [8] Z. B. P. Pratama and Y. P. Astuti, “Comparison of *Machine learning* Methods (Linear Regression, *Random Forest*, and *XGBoost*) for Predicting Poverty in Central Java in 2024,” *Sist. J. Sist. Inf.*, vol. 14, no. 5, pp. 2492–2499, 2025.
- [9] A. Arisandi and S. Syarifuddin, “Memprediksi Indeks Pembangunan Manusia di Wilayah Indonesia Bagian Timur Menggunakan *Random Forest Classification*,” *J. Math. Theory Appl.*, pp. 1–6, 2023.
- [10] A. Syaifuddin and T. Prabowo, “Optimasi Akurasi Model Prediksi Magnitudo Gempa Bumi dengan Integrasi Clustering DBSCAN pada *Ensemble Learning* (*Random Forest* & *XGBoost*),” 2024.
- [11] M. N. Cholil, “Implementasi algoritma *Random Forest* pada model klasifikasi sekuen DNA manusia dalam penyakit diabetes,” Universitas Islam Negeri Maulana Malik Ibrahim, 2025.
- [12] P. Zhang and X. Chen, “An *Unsupervised learning* Approach for Coal Spontaneous Combustion Warning Level Classification Using t-SNE and k-Means Clustering,” *Appl. Sci.*, vol. 15, no. 7, p. 3756, 2025.
- [13] M. B. Prayogi and F. Apriani, “Prediksi Angka Harapan Hidup Menggunakan *Random Forest* dan *XGBoost Regression*,” *JICode J. Inform. Dan Komput.*, vol. 2, no. 1, pp. 112–121, 2025.
- [14] J. M. A. S. Dachi and P. Sitompul, “Analisis Perbandingan Algoritma *XGBoost* dan Algoritma *Random Forest Ensemble Learning* pada Klasifikasi Keputusan Kredit,” *J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam*, vol. 2, no. 2, pp. 87–103, 2023.
- [15] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-label *confusion matrix*,” *Ieee Access*, vol. 10, pp. 19083–19095, 2022.
- [16] H. Zhang, Y. Wang, F. Dayoub, and N. Sunderhauf, “Varifocalnet: An iou-aware dense object detector,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8514–8523.
- [17] M. El-Geneidy, H. El-Din Moustafa, H. Khater, S. Abd-Elsamee, and S. A. Gamel, “Advanced real-time detection of acute ischemic stroke using YOLOv12, YOLOv11, and YOLO-NAS: a comparative study for multi-class classification,” *Sci. Rep.*, vol. 15, no. 1, p. 32546, 2025.
- [18] S. Saidah, Y. N. Fuadah, F. Alia, N. Ibrahim, R. Magdalena, and S. Rizal, “Facial skin type classification based on microscopic images using convolutional neural network (CNN),” in *Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics: ICEBEHI 2020, 8-9 October, Surabaya, Indonesia*, Springer, 2021, pp. 75–83.
- [19] N. M. Alahdal, F. Abukhodair, L. H. Meftah, and A. Cherif, “Real-time object detection in autonomous vehicles with YOLO,” *Procedia Comput. Sci.*, vol. 246, pp. 2792–2801, 2024.
- [20] G. Canbek, T. Taskaya Temizel, and S. Sagiroglu, “PToPI: A comprehensive review, analysis, and knowledge representation of binary classification performance measures/metrics,” *SN Comput. Sci.*, vol. 4, no. 1, p. 13, 2022.