Jurnal Informatika dan Teknologi (JICode) Vol.2 No.1 Halaman : 112-121, Februari 2025
ISSN: 3063-346X

PREDIKSI ANGKA HARAPAN HIDUP MENGGUNAKAN *RANDOM FOREST* DAN *XGBOOST REGRESSION*

M. Bagus Prayogi[1] , Fitria Apriani[2] , Nirma[3]
Fakultas Sains dan Teknologi, Universitas Nurul Huda
Email: mhdjesen212@gmail.com[1], fitria@unuha.ac.id[2], nirma@unuha.ac.id[3]

ABSTRAK

Angka harapan hidup mengacu pada estimasi rata-rata durasi kehidupan seseorang sejak kelahirannya. Indikator ini menjadi salah satu komponen penting dalam pengukuran indeks pembangunan manusia (IPM). Peningkatan harapan hidup biasanya berbanding lurus dengan kenaikan nilai IPM. Penelitian ini bertujuan untuk memprediksi tingkat harapan hidup menggunakan 2 algoritma regresi yaitu *Random Forest regression dan XGBoost regression*, serta menganalisis variabel yang paling berpengaruh terhadap harapan hidup. Dataset yang digunakan berasal dari *Global Country Information Dataset* 2023 yang tersedia di platform Kaggle. Berdasarkan hasil analisis, *XGBoost regression* terbukti memiliki performa terbaik dalam melakukan prediksi, sebagaimana ditunjukkan oleh nilai *MAPE* yang lebih rendah sebesar 2.60 dan R^2 yang lebih tinggi sebesar 90.53. Faktor-faktor seperti angka kematian bayi dan rasio kematian ibu ditemukan sebagai prediktor utama, sedangkan pengaruh Indeks Harga Konsumen (CPI) terhadap harapan hidup relatif lebih kecil.

Kata kunci: harapan hidup, *random forest, XGBoost*

1. PENDAHULUAN

Salah satu indikator yang mendorong pertumbuhan ekonomi adalah peningkatan pembangunan sumber daya manusia, yang diukur melalui indeks pembangunan manusia (IPM) [1]. *United Nations Development Program (UNDP)* menerbitkan laporan tentang pembangunan manusia menggunakan ukuran kuantitatif yang disebut Indeks Pembangunan Manusia (IPM). IPM merupakan indeks komposit yang dihitung berdasarkan rata-rata dari tiga indikator utama: Indeks Harapan Hidup, Indeks Pendidikan, dan Indeks Standar Hidup Layak, HDI ini berfungsi sebagai acuan dalam menilai kualitas hidup suatu wilayah, dengan fokus pada tiga komponen utama, yakni dimensi pengetahuan, dimensi kesehatan, dan dimensi kehidupan yang layak[2]. Berdasarkan data dari Badan Pusat Statistik (2014), Indeks Pembangunan Manusia (IPM) dikelompokkan ke dalam empat kategori: rendah jika IPM kurang dari 60, sedang untuk IPM antara 60 dan 70, tinggi jika IPM berada antara 70 dan 80, serta sangat tinggi jika IPM mencapai 80 atau lebih.

Salah satu komponen utama dalam perhitungan Indeks Pembangunan Manusia (IPM) adalah Angka Harapan Hidup (AHH), yang mengacu pada estimasi rata-rata usia yang dapat dicapai seseorang sejak lahir. Angka ini dipengaruhi oleh berbagai faktor penting, seperti tingkat kesehatan, kondisi ekonomi, dan akses terhadap pendidikan. Ketiga faktor tersebut secara signifikan menentukan kualitas hidup dan kesejahteraan masyarakat di suatu negara, sehingga turut memengaruhi AHH secara keseluruhan.[2]. Sebagai indikator kesejahteraan masyarakat, angka harapan hidup juga digunakan untuk mengevaluasi kinerja pemerintah dalam meningkatkan kualitas hidup warganya. Oleh karena itu, penerapan metode *machine learning* menjadi pendekatan inovatif untuk memahami dan meningkatkan angka harapan hidup sebagai ukuran kesejahteraan masyarakat.

Berdasarkan data WHO, angka harapan hidup di Indonesia pada tahun 2016 tercatat sebesar 69 tahun. Sementara itu, data dari Badan Pusat Statistik (BPS) menunjukkan adanya peningkatan yang signifikan, dengan angka harapan hidup mencapai 71,47 tahun pada tahun 2020. Pada tahun 2023, angka tersebut terus mengalami sedikit kenaikan menjadi 71,5 tahun, mencerminkan tren positif dalam kualitas kesehatan dan kesejahteraan masyarakat Indonesia [3]. Namun demikian, angka harapan hidup Indonesia masih berada sedikit di bawah rata-rata global, yang tercatat sebesar 74,4 tahun. Meskipun demikian,

Indonesia tidak tergolong dalam kategori negara dengan angka harapan hidup terendah. Hal ini menunjukkan bahwa Indonesia memiliki potensi yang cukup besar untuk terus meningkatkan kualitas hidup masyarakatnya. Dengan memperkuat sektor kesehatan, ekonomi, dan pendidikan, Indonesia dapat lebih mendekati atau bahkan melampaui rata-rata global, sekaligus mempersempit kesenjangan dengan negara-negara lain yang memiliki angka harapan hidup lebih tinggi.

Machine learning merupakan cabang ilmu yang mempelajari algoritma dan model statistik yang memungkinkan sistem komputer untuk menyelesaikan tugas tertentu tanpa perlu diprogram secara eksplisit. Konsep ini pertama kali diperkenalkan oleh sejumlah ilmuwan matematika, termasuk Adrien Marie Legendre, Thomas Bayes, dan Andrey Markov pada dekade 1920-an. Mereka mengembangkan teori-teori yang menjadi dasar bagi perkembangan machine learning, yang kini digunakan dalam berbagai aplikasi untuk mengidentifikasi pola dan membuat prediksi berdasarkan data [4]. Definisi *machine learning* (ML) secara umum adalah cabang kecerdasan buatan yang memungkinkan sistem untuk belajar dari data dan membuat keputusan atau prediksi tanpa diprogram secara eksplisit, tujuan dari ML adalah untuk mengembangkan model yang dapat menyelesaikan masalah kompleks secara otomatis, mengidentifikasi pola dalam data, dan meningkatkan kinerjanya seiring waktu, sehingga dapat memberikan solusi adaptif dan efisien dalam berbagai bidang seperti analisis data, prediksi tren, dan automasi[5].

Supervised learning adalah pendekatan dalam machine learning di mana model dilatih menggunakan dataset yang telah diberi label, yaitu data input yang dilengkapi dengan output yang diinginkan, tujuan utama dari metode ini adalah untuk memahami pola atau hubungan antara input dan output, sehingga model dapat menggeneralisasi dan memprediksi output yang tepat untuk data baru yang belum diberi label, dengan pelatihan yang berbasis pada data yang terstruktur ini, model dapat belajar untuk mengenali hubungan yang ada dan membuat prediksi yang akurat ketika dihadapkan dengan data yang serupa di masa depan [6]. Salah satu teknik yang sangat umum digunakan dalam *supervised learning* adalah *regresi*, teknik ini membantu dalam memahami hubungan antara varubel dependen dan varibel independen. Berikut beberapa contoh model dalam teknik regresi yaitu: *random forest, decision tree, regresi linear, XGBoost, adaboost, SVR (Support Vector Regressor) dan gradient boosting.*

Model yang digunakan dalam penelitian ini adalah *Random Forest Regressor* dan *XGBoost Regression* untuk memodelkan angka harapan hidup suatu negara berdasarkan berbagai *prediktor*. *Random Forest* merupakan pengembangan dari metode *Decision Tree* yang menggabungkan beberapa pohon keputusan. dalam Random Forest, setiap pohon keputusan dilatih menggunakan sampel data yang berbeda-beda, dan atribut yang digunakan untuk pemecahan data dipilih secara acak dari subset atribut yang dihasilkan. Pendekatan ini membantu meningkatkan akurasi prediksi dan mengurangi risiko overfitting yang mungkin terjadi pada model *Decision Tree* tunggal, teknik ini memungkinkan model untuk mengurangi overfitting dan meningkatkan akurasi prediksi dengan mengkombinasikan hasil dari berbagai pohon keputusan[7]. Sedangkan *XGBoost (Extreme Gradient Boosting)* adalah algoritma pembelajaran mesin yang efisien dan efektif untuk tugas regresi dan klasifikasi, dalam regresi, *XGBoost* digunakan untuk memprediksi nilai kontinu dengan membangun model berbasis pohon keputusan yang dioptimalkan melalui metode boosting, proses ini melibatkan penambahan pohon keputusan secara bertahap, di mana setiap pohon baru difokuskan untuk memperbaiki kesalahan prediksi yang dihasilkan oleh pohon sebelumnya, sehingga meningkatkan akurasi model secara keseluruhan[8].

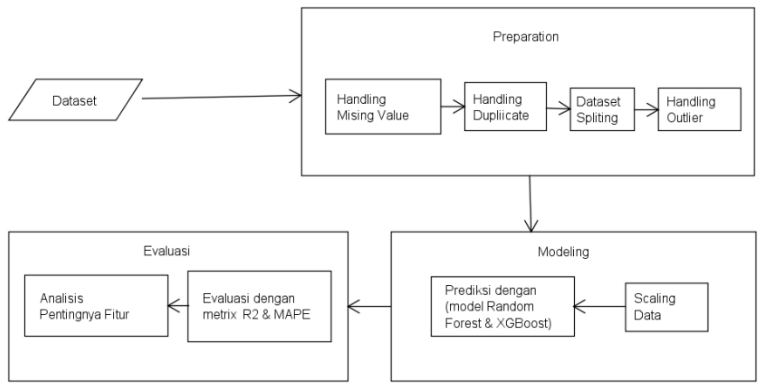
Penelitian terdahulu oleh Samuel Palentino Sinaga dengan judul "Implementasi Jaringan Syaraf Tiruan Resilient Backpropagation dalam Memprediksi Angka Harapan Hidup Masyarakat Sumatera Utara" yang menghasilkan satu model arsitektur terbaik 4-10-1 dengan tingkat keakurasian 88 % dengan 22 epoch[9]. Dari penelitan terdahulu tersebut tujuan dari penelitian ini adalah untuk mencari algoritma yang lebih baik secara performa akurasi.

Penelitian ini bertujuan untuk memprediksi angka harapan hidup dengan menggunakan model yang paling akurat, yaitu melalui penerapan metode *Random Forest* dan *XGBoost Regression*. Setelah menentukan metode yang paling efektif dalam memprediksi angka harapan hidup, tahap selanjutnya adalah mengidentifikasi variabel-variabel penjelas yang memiliki pengaruh paling signifikan terhadap angka harapan hidup. Dengan demikian, penelitian ini tidak hanya berfokus pada akurasi prediksi, tetapi juga memberikan wawasan terkait faktor-faktor yang paling memengaruhi kualitas hidup di suatu negara.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian merangkum langkah-langkah yang diambil dalam suatu studi. Dalam penelitian ini, terdapat empat tahapan utama, yaitu pengumpulan data, persiapan data, pemodelan, dan evaluasi, yang dapat dilihat pada Gambar 1. Setiap tahapan memiliki peran penting dalam memastikan proses penelitian berjalan dengan baik, mulai dari pengumpulan data yang relevan, persiapan data untuk analisis, penerapan model yang tepat, hingga evaluasi hasil yang diperoleh untuk menentukan efektivitas model yang digunakan.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Pada penelitian data yang digunakan adalah dataset *Global Country Information Dataset 2023.* Dataset tersebut di peroleh dari situs Kaggle, dimana terdapat 195 *record* pada data tersebut. Berikut varibel yang digunakan dalam penelitian ini yang dapat dilihat pada tabel 1.

Tabel 1. Variabel

No	Variabel	Deskripsi
y	Harapan Hidup	Rata-rata jumlah tahun yang diharapkan seorang bayi akan hidup sejak lahir.
x_1	Indeks Harga Konsumen (CPI)	Ukuran inflasi dan daya beli.
x_2	Angka Fertilitas	Rata-rata jumlah anak yang dilahirkan oleh seorang wanita sepanjang hidupnya.
x_3	Produk Domestik Bruto (PDB)	Jumlah keseluruhan nilai barang dan jasa yang dihasilkan di suatu negara.
x_4	Angka Kematian Bayi	Jumlah kematian per 1.000 kelahiran hidup sebelum mencapai usia satu tahun.
x_5	Rasio Kematian Ibu	Jumlah kematian ibu per 100.000 kelahiran hidup.
x_6	Upah Minimum	Tingkat upah minimum dalam mata uang lokal.
x_7	Pengeluaran Kesehatan Langsung (%)	Persentase total pengeluaran kesehatan yang dibayar langsung oleh individu.
x_8	Dokter per Seribu Penduduk	Jumlah dokter per seribu orang.
x_9	Populasi	Total populasi suatu negara.
x_{10}	Tingkat Pengangguran	Persentase angkatan kerja yang menganggur.

3 Data *Preparation*

Persiapan data atau pra-pemrosesan data merupakan tahapan yang mencakup pengumpulan, pengintegrasian, pengorganisasian, dan penataan data agar siap digunakan, tahapan ini bertujuan untuk mengubah data mentah yang sering kali tidak lengkap dan memiliki format yang tidak seragam menjadi format yang lebih terstruktur dan efisien[10]. Tahapan data preparation dalam penelitian ini adalah *handling missing value, handling duplicate value, data spliting* diawal yang bertujuan menghindari kebocoran data dan *handling outlier*.

4 Modeling

Pada tahapan modeling terdapat 2 tahapan yang dilakukan yaitu feature scaling dan pemodelan menggunakan model *Random Forest* dan *XGBoost regression*. *Feature scaling* adalah langkah penting dalam analisis data dan pembelajaran mesin yang bertujuan untuk menyamakan rentang nilai dari berbagai fitur dalam dataset. Proses ini dilakukan untuk menghindari bias, mencegah dominasi fitur tertentu, dan memastikan bahwa semua variabel memiliki rentang nilai yang seimbang, sehingga analisis atau model yang diterapkan dapat bekerja secara optimal[11]. *Random Forest* merupakan pengembangan dari *metode Decision Tree* yang menggabungkan beberapa pohon keputusan. Setiap pohon keputusan dilatih menggunakan sampel data yang berbeda, dan setiap atribut dipecah secara acak pada pohon-pohon yang dipilih berdasarkan subset atribut yang dihasilkan secara acak. Teknik ini memungkinkan model untuk mengurangi overfitting dan meningkatkan akurasi prediksi dengan mengkombinasikan hasil dari berbagai pohon keputusan[7]. Berikut ini rumus untuk mencari rata-rata populer.

$$f(x) = \text{Rata-rata}(f1(x), f2(x), \ldots, fn(x)) \quad (1)$$

Keterangan:

- f(x) : hasil prediksi
- f1(x), f2(x), …, fn(x) : hasil prediksi dari setiap pohon keputusan
- x : input

Sedangkan *XGBoost (Extreme Gradient Boosting)* adalah algoritma pembelajaran mesin yang efisien dan efektif untuk tugas regresi dan klasifikasi, dalam regresi, *XGBoost* digunakan untuk memprediksi nilai kontinu dengan membangun model berbasis pohon keputusan yang dioptimalkan melalui metode boosting, proses ini melibatkan penambahan pohon keputusan secara bertahap, di mana setiap pohon baru difokuskan untuk memperbaiki kesalahan prediksi yang dihasilkan oleh pohon sebelumnya, sehingga meningkatkan akurasi model secara keseluruhan[8].

Pada metode ini, diperlukan fungsi objektif yang berfungsi untuk menilai seberapa baik model yang dihasilkan sesuai dengan data pelatihan. Karakteristik utama dari fungsi objektif terdiri dari dua komponen, yaitu nilai kerugian pelatihan dan nilai regularisasi, sebagaimana dijelaskan pada persamaan berikut ini.

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

Di mana L merupakan fungsi kerugian pelatihan, Ω adalah fungsi regularisasi, dan θ adalah parameter yang terkait dengan model. Secara umum, fungsi kerugian pelatihan dapat dituliskan seperti pada persamaan berikut ini.

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

Dimana *yi* adalah nilai data sebenarnya yang dianggap benar, *ŷi* adalah hasil prediksi yang dihasilkan oleh model, dan n adalah jumlah iterasi dari nilai model.

5 Evaluasi

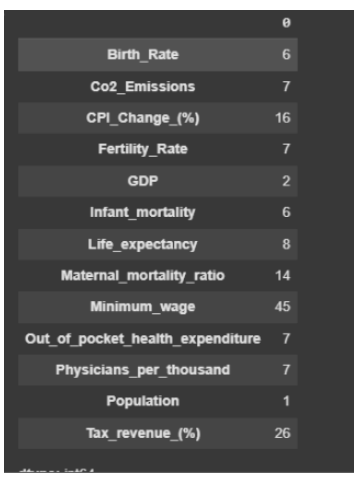
Pada tahapan ini model dievaluasi menggunakan metrik koefisien determinasi ($\boldsymbol{R^2}$) untuk melihat seberapa baik model dalam menjelaslan variabilitas data dan *Mean Absolute Percentage Error* (MAPE) untuk mengukur rata-rata kesalahan prediksi dalam persen[12]. Setelah dievaluasi dilakukan analisi lebih lanjut dengan analisisis pentingnya fitur, untuk mengetahui kontribusi masing masing variabel terhadap prediksi sehingga memberikan wawasan lebih dalam tentang faktor-faktor yang paling memengaruhi hasil model.

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah dataset *Global Country Information Dataset 2023* yang diperoleh dari situs Kaggle, dengan total 195 record. Dari data tersebut hanya diambil beberapa variabel yang relevan dengan angka harapa hidup, berikut variabel independen yang digunakan: Indeks Harga Konsumen (CPI), Angka Fertilitas, Produk Domestik Bruto (PDB), Angka Kematian Bayi, Rasio Kematian Ibu, Upah Minimum, Pengeluaran Kesehatan Langsung (%), Dokter per Seribu Penduduk, Populasi, Tingkat Pengangguran. Setelah mendapatkan varibel yang akan dianalisi selanjutnya masuk ke tahapan selanjutnya yaitu *data preparation* atau proses pra-persiapan sebelum modeling. Berikut ini tahapan dari *data preparation* yang dilakukan:

3.1 Cek dan *Handling Missing Value*

Dalam tahapan ini didapatkan bahwa setiap variabel memiliki missing value yang dapat dilihat pada gambar berikut ini:



Gambar 2. *Missing Value*

Untuk menangani missing value tersebut, diterapkan metode imputasi dengan mengisi nilai yang kosong menggunakan nilai median dari masing-masing variabel. Pendekatan ini dipilih karena median lebih tahan terhadap outlier, sehingga dapat menjaga kestabilan distribusi data.

3.2 Cek dan *Handling Duplicate Value*

Pada tahap ini, data diperiksa secara menyeluruh untuk memastikan tidak adanya duplikasi yang dapat memengaruhi kualitas analisis. Setelah proses pengecekan selesai, hasilnya menunjukkan bahwa data bebas dari duplikasi. Dengan demikian, data dianggap valid dan siap digunakan untuk melanjutkan ke tahap analisis berikutnya.

3.3 *Spliting Data*

Data spliting atau pemisahan data adalah teknik untuk memisahkan data menjadi beberapa subset, umumnya data di split menjadi data pelatihan dan data testing. Tujuan dari pemisahan ini adalah untuk untuk melatih, menguji, dan mengevaluasi model yang dibangun berdasarkan data tersebut. Pada tahapan ini data displit dengan rasio 80:20 yaitu 80% data pelatihan dan 20% data testing dengan random state 42, yang dapat dilihat pada gambar berikut ini.

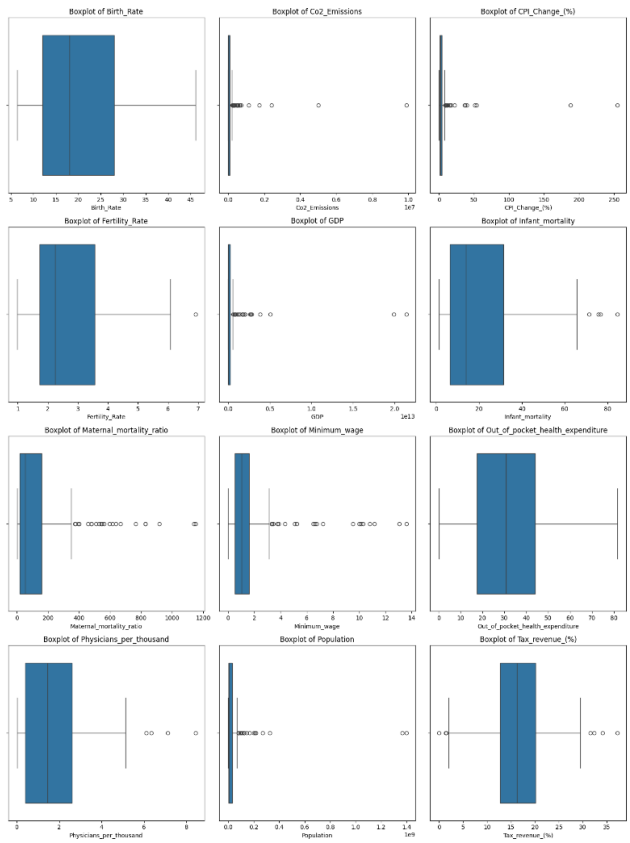
Spliting Data

```
[ ]  # train-test split
     X = df.drop(columns=['Life_expectancy'])
     y = df.Life_expectancy
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 3. Spliting Data

3.4 Cek dan *Handling Outlier*

Tahapan ini merupakan tahapan terakhir dari pra persiapan data untuk melihat apakah data yang kita gunakan memiliki outlier atau data yang memiliki nilai yang jauh berbeda dari sebagian besar data lainnya dalam dataset, sehingga dapat memengaruhi hasil analisis dan interpretasi. Dalam tahapan ini digunakan visualisasi box plot untuk melihat outlier pada data, berikut ini visualisasi dari boxplotnya.



Gambar 4. Boxplot

Berdasarkan hasil visualisasi, ditemukan bahwa sebagian besar variabel memiliki outlier. Untuk menangani hal ini, dilakukan pemeriksaan mendalam, dan penanganan outlier diterapkan menggunakan metode IQR dengan menyesuaikan batas minimum dan maksimum data. Metode ini dipilih karena kami berupaya menjaga integritas data dengan tetap mempertahankan nilai-nilai penting tanpa menghapusnya.

Setelah melalui tahap data preparation dan spliting data tahapan selanjutnya adalah feature enggeneering yaitu dengan menscaling data. Scaling data adalah proses transformasi nilai data ke dalam skala tertentu tanpa mengubah distribusi relatif antar nilai, tujuannya adalah untuk memastikan bahwa semua fitur memiliki rentang nilai yang serupa[13]. Pada penelitian ini jenis scaler yang digunakan adalah MinMaxScaler, MinMaxScaler adalah teknik scaling data yang mengubah nilai fitur ke dalam rentang tertentu, biasanya antara 0 dan 1. Berikut ini adalah hasil dari scaling.

```
     CPI_Change_(%)  Fertility_Rate       GDP  Infant_mortality  \
5          0.131148        0.189449  0.002857          0.053284
135        0.417288        0.483939  0.042371          0.541721
122        0.298063        0.114420  1.000000          0.028122
167        1.000000        0.270106  0.006695          0.229417
85         0.047690        0.082532  1.000000          0.005920

     Maternal_mortality_ratio  Minimum_wage  Out_of_pocket_health_expenditure  \
5                    0.109141      0.936993                          0.296069
135                  0.390177      0.355624                          0.068796
122                  0.008186      1.000000                          0.148649
167                  0.321965      0.320062                          0.121622
85                   0.008186      1.000000                          0.158477

     Physicians_per_thousand  Population
5                   0.466383    0.001129
135                 0.008511    0.113708
122                 0.611064    0.224702
167                 0.202553    0.007410
85                  0.406809    1.000000
```

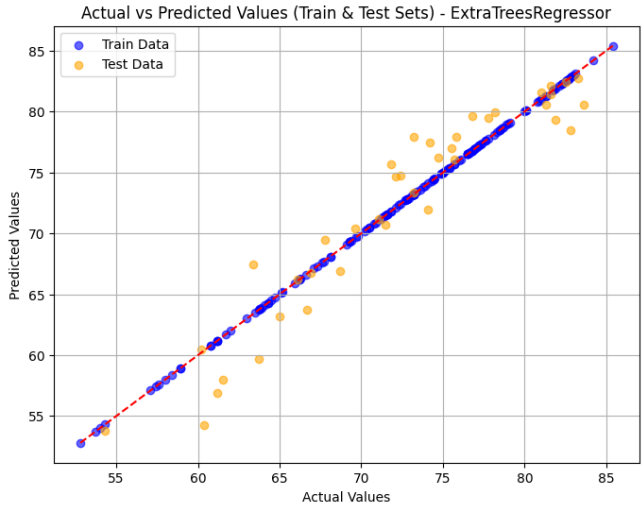
Gambar 5. Scaling Data

Setelah data terscaling di rentang 0 – 1, data tersebut siap memasuki tahap modeling. Dimana modeling adalah proses menciptakan representasi realitas yang disederhanakan dalam bentuk konstruk matematis atau komputasi untuk menganalisis hubungan antar variabel, membuat prediksi, dan menghasilkan wawasan[14]. Pada tahapan ini kami akan membandingkan performa dari model *Random Forest* dan *XGBoost Regression*, setelah dilakukan pelatihan terhadap model berikut ini adalah rangkuman dari performa masing masing model yang dikukur dengan matrik koefisien determinasi (R^2) dan *Mean Absolute Percentage Error* (MAPE). Berikut ini tabel dari evaluasi model.

Tabel 2. Evaluasi

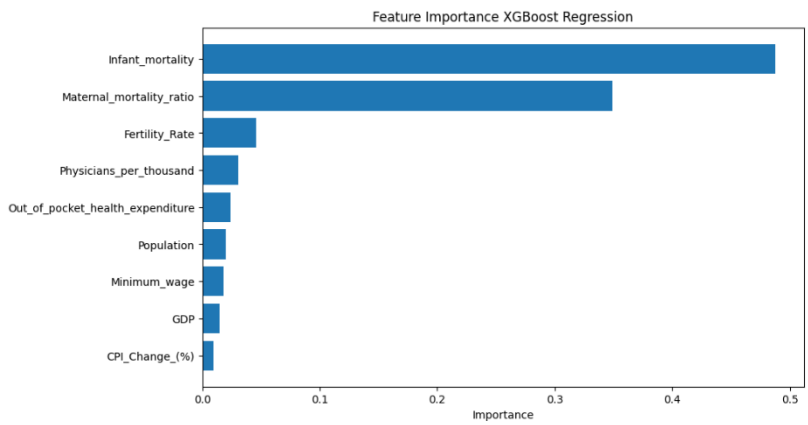
No	Model	R^2	MAPE
1	Random Forest	87.091775	3.009974
2	XGBoost Regression	90.531463	2.607853

Berdasarkan tabel di atas, dapat disimpulkan bahwa model *XGBoost Regression* menunjukkan kinerja yang sangat optimal dalam menjelaskan variabilitas data. Selain itu, model ini memiliki tingkat kesalahan (error) yang lebih rendah dibandingkan dengan model Random Forest, sehingga menjadikannya pilihan yang lebih unggul untuk analisis data dalam konteks ini yang akan di evaluasi lebih lanjut. Selain dievaluasi menggunakan *Mean Absolute Percentage Error* (MAPE) dan koefisien determinasi (R^2), kinerja model regresi juga dianalisis melalui scatter plot yang membandingkan nilai aktual dengan prediksi, seperti yang terlihat pada Gambar 6. Dalam scatter plot ini, garis putus-putus merah menggambarkan garis identitas, yang menunjukkan posisi prediksi yang sempurna—di mana nilai prediksi sama persis dengan nilai sebenarnya. Titik-titik yang mendekati garis identitas ini menunjukkan prediksi yang akurat, sementara titik yang jauh dari garis tersebut menunjukkan prediksi yang kurang tepat. Evaluasi ini memberikan pandangan visual mengenai sejauh mana model dapat menghasilkan prediksi yang akurat secara keseluruhan.



Gambar 6. Scatter Plot Prediksi

Analisis pentingnya fitur dalam model *XGBoost Regression* menunjukkan pola unik dalam mengevaluasi dan memanfaatkan informasi. Kematian bayi muncul sebagai variabel yang paling berpengaruh terhadap prediksi, menunjukkan bahwa variabel ini adalah indikator utama dalam dataset yang dianalisis. Selain itu, rasio kematian ibu juga secara konsisten dianggap penting, meskipun pengaruhnya tidak sebesar kematian bayi. Model *XGBoost Regression* juga memberikan distribusi yang lebih seimbang dalam feature importance, di mana fitur seperti Angka Fertilitas, jumlah dokter per seribu penduduk, dan Pengeluaran Kesehatan Langsung (%) turut berperan signifikan dalam prediksi. Analisis ini menunjukkan bahwa meskipun beberapa variabel dianggap penting secara universal, model *XGBoost* memiliki kemampuan untuk mengaitkan bobot yang lebih proporsional pada berbagai variabel, mencerminkan cara algoritma ini memproses informasi secara efektif untuk menghasilkan prediksi yang akurat.



Gambar 7. Feature Importance

4. KESIMPULAN

Secara keseluruhan, model *XGBoost Regression* menunjukkan kinerja yang lebih baik dalam memprediksi hasil dibandingkan dengan model lainnya. Hal ini tercermin dari nilai *Mean Absolute Percentage Error* (MAPE) yang lebih rendah dan koefisien determinasi (R^2) yang lebih tinggi, yang menunjukkan kemampuan model ini untuk menghasilkan prediksi yang lebih akurat dan dapat diandalkan. Analisis pentingnya fitur juga mengungkapkan bahwa kematian bayi dan rasio kematian ibu secara konsisten menjadi prediktor yang signifikan dalam model. Di sisi lain, variabel Indeks Harga Konsumen (CPI) memberikan kontribusi yang lebih kecil dalam memprediksi angka harapan hidup, mengindikasikan bahwa faktor ini kurang berpengaruh dibandingkan dengan variabel lainnya.

DAFTAR PUSTAKA

[1] A. A. Kania Azzahra and D. Soebagyo, "Analisis Determinan Pertumbuhan Ekonomi di Indonesia Tahun 2000-2021," *Action Res. Lit.*, vol. 8, no. 3, pp. 340–345, 2024, doi: 10.46799/arl.v8i3.279.

[2] M. D. D. Akasumbawa, A. Adim, and M. G. Wibowo, "Faktor-Faktor yang Memengaruhi Pertumbuhan Ekonomi di Negara dengan Jumlah Penduduk Terbesar di Dunia (Studi pada Negara China, India, Indonesia, Pakistan dan Amerika Serikat)," *Welf. J. Ilmu Ekon.*, vol. 2, no. 1, pp. 67–74, 2021, doi: 10.37058/wlfr.v2i1.2611.

[3] R. Aprilia and S. F. Nurhayati, "Analisis Faktor-Faktor Yang Berpengaruh Terhadap Angka Harapan Hidup Di Kabupaten/Kota Se Jawa Tengah Tahun 2020-2022 Analysis of Factors That Influence Life Expectancy Rate in Districts/Cities of Central Java in 2020-2022," *J. Inov. Pembang.*, vol. 12, no. 2, 2024.

[4] F. Baharuddin and A. Tjahyanto, "Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Peningkatan Dataset," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 1, pp. 25–31, 2022, doi: 10.32736/sisfokom.v11i1.1337.

[5] Wijoyo A, Saputra A, Ristanti S, Sya'ban S, Amalia M, and Febriansyah R, "Pembelajaran Machine Learning," *OKTAL (Jurnal Ilmu Komput. dan Sci.*, vol. 3, no. 2, pp. 375–380, 2024, [Online]. Available: https://journal.mediapublikasi.id/index.php/oktal/article/view/2305

[6] J. C. Mestika, M. O. Selan, and M. I. Qadafi, "Menjelajahi Teknik-Teknik Supervised Learning untuk Pemodelan Prediktif Menggunakan Python," *Bul. Ilm. Ilmu Komput. dan Multimed.*, vol. 1, no. 1, pp. 216–219, 2023, [Online]. Available: https://jurnalmahasiswa.com/index.php/biikma/article/view/101

[7] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.

[8] R. Siringoringo, R. Perangin Angin, and B. Rumahorbo, "Model Klasifikasi Genetic-Xgboost Dengan T-Distributed Stochastic Neighbor Embedding Pada Peramalan Pasar," *J. TIMES*, vol. 11, no. 1, pp. 30–36, 2022, doi: 10.51351/jtm.11.1.2022672.

[9] S. P. Sinaga, A. Wanto, and S. Solikhun, "Implementasi Jaringan Syaraf Tiruan Resilient Backpropagation dalam Memprediksi Angka Harapan Hidup Masyarakat Sumatera Utara," *J. Infomedia*, vol. 4, no. 2, p. 81, 2020, doi: 10.30811/jim.v4i2.1573.

[10] K. R. Putra, S. Umaroh, N. Fitrianti, and S. Nugraha, "RESULTANT: Data Preparation Techniques to Improve XGBoost Algorithm Performance," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 8, no. 1, pp. 42–51, 2023, [Online]. Available: https://doi.org/10.26760/mindjournal.v8i1.42-51

[11] M. Radhi, S. Hamonangan Sinurat, D. Ryan Hamonangan Sitompul, E. Indra, and S. Informasi, "Prediksi water quality index (wqi) menggunakan algoritma regressi dengan hyper-parameter tuning," *J. Sist. Inf. dan Ilmu Komput. Prima)*, vol. 5, no. 1, pp. 44–50, 2021.

[12] R. C. Rohmana, D. Triwanti, P. R. Setiyaningrum, T. Perminyakan, T. Perminyakan, and T. Perminyakan, "Penerapan Machine Learning dalam Penentuan Porositas Batuan : Studi Kasus Menggunakan Regresi Linier Berganda dan Regresi KNN pada Data Log Sumur Application of Machine Learning in Rock Porosity Determination : Case Study Using Multiple Linear Regression ," vol. 13, no. 02, pp. 42–50, 2024.

[13] M. J. Vikri *et al.*, "RICE QUALITY IDENTIFICATION FOR INDONESIAN FOOD STANDARDS BASED ON ELECTRONIC NOSE BERDASARKAN STANDAR PANGAN INDONESIA BERBASIS," pp. 49–60, 2025.

[14] L. Selviana, W. Afgani, and R. A. Siroj, "Correlational Research," *Innov. J. Soc. Sci. Res.*, vol. 4, pp. 5118–5128, 2024, [Online]. Available: https://j-innovative.org/index.php/Innovative