

**RISET INFORMATIKA**  
**PREDIKSI ANGKA HARAPAN HIDUP PENDUDUK MENGGUNAKAN**  
**Metode XGBoost di Jawa Timur**



**Disusun Oleh :**  
M Ryan Nurdiansyah N A (21081010321)

**Dosen Pengampu:**  
Dr. Basuki Rahmat, S.Si., M.T.  
**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"**  
**JAWA TIMUR**  
**2024**

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Angka harapan hidup merupakan salah satu indikator utama dalam menilai kesejahteraan dan kesehatan masyarakat suatu wilayah. Indikator ini mencerminkan kualitas hidup populasi serta efektivitas kebijakan kesehatan yang diterapkan oleh pemerintah. Di Jawa Timur, provinsi dengan populasi besar dan karakteristik sosial-ekonomi yang beragam, prediksi angka harapan hidup sangat penting untuk mendukung perencanaan kebijakan yang berbasis data. Namun, tantangan yang dihadapi adalah keterbatasan metode analisis tradisional dalam menangkap kompleksitas data lokal seperti variasi akses kesehatan, pendidikan, dan ekonomi masyarakat (Smith et al., 2020).

Dalam beberapa tahun terakhir, algoritma machine learning seperti XGBoost telah berkembang pesat dan digunakan secara luas dalam berbagai bidang, termasuk analisis kesehatan masyarakat. Algoritma ini mampu menangani dataset yang kompleks dengan efisiensi tinggi serta memberikan prediksi yang akurat melalui teknik gradient boosting (Chen & Guestrin, 2016). Penelitian menunjukkan bahwa XGBoost dapat mengidentifikasi variabel signifikan secara otomatis, yang menjadikannya alat analisis yang unggul dalam memprediksi indikator kesehatan seperti angka harapan hidup (Rahman et al., 2021).

Selain itu, penelitian di tingkat global telah membuktikan keberhasilan XGBoost dalam prediksi angka harapan hidup di berbagai negara. Misalnya, studi yang diterbitkan oleh Asia Health Organization (2020) menunjukkan bahwa algoritma ini memiliki akurasi yang lebih baik dibandingkan dengan metode tradisional lainnya, terutama dalam konteks prediksi indikator kesehatan masyarakat di Asia. Di sisi lain, penelitian oleh Chen et al. (2019) mengungkapkan bahwa angka harapan hidup dipengaruhi oleh berbagai faktor

sosial-ekonomi, seperti pendapatan per kapita, akses kesehatan, dan tingkat pendidikan.

Di Jawa Timur, kebutuhan untuk menerapkan metode prediksi yang akurat menjadi semakin penting mengingat variasi kondisi sosial-ekonomi antarwilayah. Dengan memanfaatkan metode XGBoost, penelitian ini bertujuan untuk memberikan solusi prediksi yang lebih akurat dan relevan dengan data lokal, sehingga dapat digunakan sebagai dasar dalam merancang kebijakan peningkatan kualitas hidup masyarakat.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang di atas, permasalahan yang dapat dirumuskan adalah sebagai berikut:

1. Bagaimana akurasi metode XGBoost dalam memprediksi angka harapan hidup penduduk di Jawa Timur?
2. Faktor lokal apa saja yang paling berpengaruh terhadap angka harapan hidup di Jawa Timur berdasarkan hasil analisis model XGBoost?
3. Bagaimana model prediksi ini dapat mendukung pengambilan kebijakan yang relevan untuk meningkatkan angka harapan hidup di Jawa Timur?

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah ditetapkan, maka tujuan penelitian ini adalah sebagai berikut:

1. Mengukur akurasi metode XGBoost dalam memprediksi angka harapan hidup di Jawa Timur.
2. Mengidentifikasi faktor lokal utama yang memengaruhi angka harapan hidup.
3. Memberikan rekomendasi berbasis data untuk mendukung kebijakan pembangunan daerah.

## **1.4 Manfaat Penelitian**

Penelitian ini memberikan manfaat bagi sejumlah pihak. Adapun manfaat penelitian ini bagi berbagai pihak terkait adalah sebagai berikut:

1. Bagi Akademisi dan Peneliti
  - a. Memberikan kontribusi pada literatur tentang penerapan algoritma XGBoost di bidang kesehatan masyarakat.
  - b. Menjadi dasar bagi penelitian lanjutan yang dapat dikembangkan untuk wilayah atau bidang lain.
2. Bagi Pemerintah Daerah
  - a. Memberikan data berbasis algoritma XGBoost untuk mendukung kebijakan peningkatan kualitas hidup masyarakat di Jawa Timur.
  - b. Membantu perencanaan pembangunan daerah yang lebih efektif dengan memanfaatkan hasil prediksi yang akurat.
3. Bagi Institusi Kesehatan dan Pendidikan
  - a. Mengidentifikasi faktor-faktor utama yang memengaruhi angka harapan hidup untuk mendukung program peningkatan kesehatan dan pendidikan.
  - b. Menyediakan wawasan baru dalam mengintegrasikan data lokal ke dalam kebijakan kesehatan dan pendidikan.
4. Bagi Masyarakat Umum
  - a. Meningkatkan pemahaman masyarakat tentang pentingnya faktor sosial-ekonomi dalam menentukan angka harapan hidup.
  - b. Mendorong kesadaran dan partisipasi masyarakat dalam upaya peningkatan kualitas hidup.

### **1.5 Batasan Masalah**

Adapun batasan masalah ini ditetapkan untuk memperjelas fokus dan arah penelitian sebagai berikut:

1. Penelitian ini hanya mencakup wilayah Provinsi Jawa Timur dan tidak memasukkan data dari provinsi lain di Indonesia.

2. Analisis hanya menggunakan variabel tertentu seperti pendidikan, pendapatan, akses kesehatan, dan infrastruktur lokal, tanpa mempertimbangkan faktor lain seperti budaya atau kebijakan daerah.
3. Penelitian ini menggunakan dataset lokal yang tersedia dan mengabaikan data yang tidak lengkap atau tidak relevan.
4. Model prediksi yang diterapkan hanya menggunakan metode XGBoost tanpa membandingkan secara langsung dengan algoritma machine learning lain seperti Random Forest atau Neural Networks.
5. Data yang digunakan berasal dari periode 5 tahun terakhir dan tidak mencerminkan perubahan data yang terjadi setelahnya.
6. Model prediksi ini tidak mencakup implementasi langsung dalam pengambilan keputusan pemerintah, tetapi hanya menghasilkan rekomendasi berbasis data.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Penelitian Terdahulu**

Penelitian-penelitian yang telah dilakukan oleh peneliti sebelumnya yang berkaitan dengan penelitian yang akan dilakukan adalah sebagai berikut:

1. Penelitian oleh Kurniawan dan Indahyanti (2024).

Penelitian ini menggunakan algoritma XGBoost Regressor untuk memprediksi angka harapan hidup di beberapa negara Asia. Data yang digunakan berasal dari UCI Machine Learning Repository, dengan fokus pada variabel sosial-ekonomi seperti akses kesehatan, tingkat pendidikan, dan pendapatan per kapita. Penelitian menunjukkan bahwa XGBoost memiliki kemampuan tinggi dalam memodelkan hubungan kompleks antar variabel, dengan akurasi mencapai 96,8% dan Mean Absolute Error (MAE) sebesar 0,97.

Temuan penelitian ini menggarisbawahi keunggulan XGBoost dalam memproses dataset berskala besar dan heterogen. Dengan hasil tersebut, penelitian ini memberikan kontribusi penting dalam membuktikan efektivitas machine learning untuk prediksi indikator kesehatan masyarakat. Potensinya sebagai alat pendukung pengambilan kebijakan juga menjadi sorotan utama.

2. Penelitian oleh Ghevira Chairunisa, Mohammad K. Najib, Sri Nurdianti, salsabila F. Imni, wardah Sanjawya, Rizka D. Andriani, Henriyansah, Renda S. P. Putri, dan Dhea Eka Putri (2024)

Penelitian ini membandingkan beberapa model regresi, termasuk XGBoost, untuk memprediksi angka harapan hidup menggunakan dataset Global Country Information Dataset 2023. Hasil analisis menunjukkan bahwa model random forest regression memiliki kinerja yang lebih unggul dalam memprediksi hasil, dengan nilai RMSE yang lebih rendah dan nilai  $R^2$  yang lebih tinggi. Kematian bayi dan rasio kematian ibu diidentifikasi sebagai prediktor signifikan di semua model, sedangkan populasi kurang mempengaruhi angka harapan hidup.

3. Penelitian oleh Hamzah Zufarul Furqon, Mochamad, Alfian Rosid, Ade eviyanti, Yunianita Rahmawati, (2024)

Penelitian ini membandingkan algoritma Random Forest dengan metode lain, termasuk XGBoost, untuk memprediksi angka harapan hidup berdasarkan faktor sosial-ekonomi dan kesehatan. Data yang digunakan mencakup variabel seperti rasio kematian bayi, pendapatan per kapita, dan akses layanan kesehatan. Hasil penelitian menunjukkan bahwa Random Forest memiliki kinerja yang lebih baik dalam menangkap pola-pola yang kompleks.

Dengan akurasi yang tinggi, penelitian ini menunjukkan bahwa algoritma machine learning dapat memberikan prediksi yang andal untuk indikator kesehatan masyarakat. Studi ini juga menyoroti pentingnya variabel sosial-ekonomi sebagai faktor utama dalam mempengaruhi angka harapan hidup.

## **2.2 Data Mining**

Data mining adalah proses ekstraksi informasi yang berguna dari dataset besar untuk menemukan pola-pola yang tersembunyi (Han et al., 2012). Proses ini melibatkan berbagai teknik seperti klasifikasi, clustering, regresi, dan asosiasi yang digunakan dalam berbagai bidang termasuk kesehatan masyarakat. Data mining memainkan peran penting dalam

pengolahan data besar, yang sering kali terdiri dari banyak variabel kompleks dan saling terkait.

Dalam konteks prediksi angka harapan hidup, data mining membantu mengidentifikasi variabel signifikan seperti tingkat pendidikan, akses kesehatan, dan pendapatan per kapita. Hal ini mendukung pengambilan keputusan berbasis data, terutama dalam menganalisis dataset lokal seperti di Jawa Timur.

### **2.3 Machine Learning**

Machine learning adalah cabang dari kecerdasan buatan yang memungkinkan sistem untuk belajar dari data tanpa perlu diprogram secara eksplisit (Murphy, 2012). Teknik ini sangat berguna untuk prediksi karena kemampuannya untuk menangkap hubungan kompleks antar variabel. Algoritma seperti Support Vector Machine, Random Forest, dan Gradient Boosting sering digunakan untuk tujuan analisis prediktif.

Dalam penelitian ini, machine learning menjadi pendekatan utama untuk memodelkan angka harapan hidup. Dengan menerapkan algoritma seperti Extreme Gradient Boosting (XGBoost), machine learning mampu menghasilkan prediksi yang lebih akurat dibandingkan metode tradisional seperti regresi linier.

### **2.4 Extreme Gradient Boosting**

Extreme Gradient Boosting (XGBoost) adalah algoritma machine learning berbasis decision tree yang dikembangkan oleh Chen dan Guestrin (2016). Algoritma ini dirancang untuk meningkatkan performa dengan meminimalkan kesalahan prediksi menggunakan metode gradient boosting. XGBoost unggul dalam menangani dataset besar dengan dimensi tinggi dan fitur kompleks.

Penelitian menunjukkan bahwa XGBoost sangat efektif dalam memprediksi indikator kesehatan masyarakat, seperti angka harapan hidup, karena kemampuannya dalam menangkap pola non-linear. Algoritma ini juga memungkinkan identifikasi variabel signifikan secara otomatis,



sehingga mendukung interpretasi hasil yang lebih mendalam (Rahman et al., 2021).

## **2.5 Model Hybrid XGBoost**

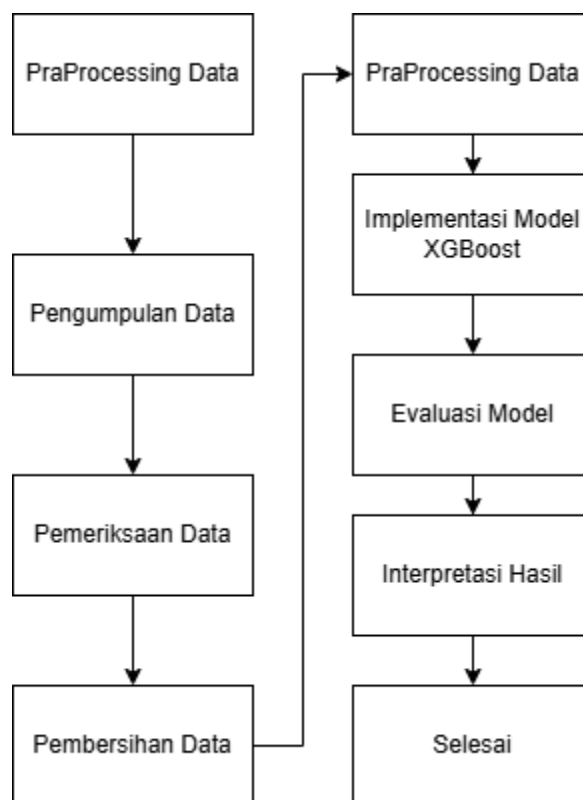
Model hybrid XGBoost adalah pengembangan dari algoritma XGBoost yang dikombinasikan dengan metode lain, seperti Neural Networks atau Long Short-Term Memory (LSTM). Pendekatan ini bertujuan untuk meningkatkan akurasi prediksi dengan mengintegrasikan keunggulan masing-masing metode (Wang et al., 2020). Misalnya, kombinasi XGBoost dengan LSTM telah digunakan untuk analisis deret waktu yang kompleks di bidang kesehatan.

Dalam konteks prediksi angka harapan hidup, model hybrid ini dapat memberikan hasil yang lebih presisi dengan mempertimbangkan faktor temporal dan spasial secara bersamaan. Hal ini sangat relevan untuk data lokal di Jawa Timur, di mana variasi antar wilayah dapat memengaruhi hasil prediksi.

### BAB III: METODE PENELITIAN

#### 3.1 Tahapan Penelitian

Dalam penelitian, untuk mencapai tujuan penelitian secara efektif diperlukan prosedur yang sistematis dalam menjawab rumusan masalah dan mencapai tujuan penelitian, seperti yang telah dijelaskan pada bagian sebelumnya.



Penelitian ini diawali dengan tahap perumusan masalah, yaitu memprediksi angka harapan hidup di Jawa Timur menggunakan metode XGBoost, yang diawali dengan identifikasi tujuan penelitian dan kesenjangan riset yang ada. Selanjutnya, dilakukan studi literatur untuk memahami penelitian-penelitian sebelumnya, khususnya yang berkaitan dengan algoritma XGBoost dan prediksi angka harapan hidup, guna menentukan inovasi yang dapat diterapkan. Data lokal yang relevan, seperti data demografi, sosial, ekonomi, dan kesehatan masyarakat di Jawa Timur, kemudian dikumpulkan dari berbagai sumber terpercaya. Setelah itu, data tersebut melalui proses pra-pengolahan, termasuk pembersihan data, normalisasi, encoding, dan pembagian dataset menjadi data latih dan data uji untuk memastikan kualitas analisis. Model XGBoost kemudian diterapkan dengan pengaturan parameter yang optimal agar menghasilkan performa prediksi terbaik. Model ini dievaluasi menggunakan metrik seperti Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan R-squared ( $R^2$ ) untuk mengukur akurasi prediksi. Hasil prediksi dan analisis model selanjutnya diinterpretasikan untuk mengidentifikasi faktor-faktor lokal yang paling memengaruhi angka harapan hidup di Jawa Timur. Penelitian ini diakhiri dengan penyusunan laporan yang mencakup kesimpulan dan rekomendasi strategis berbasis data untuk mendukung pengambilan kebijakan yang relevan di daerah tersebut.

### 3.2 Studi Literatur

Studi literatur merupakan langkah awal yang penting dalam penelitian ini, karena memberikan landasan teori dan pemahaman mendalam tentang konteks penelitian serta metodologi yang relevan. Penelitian ini berfokus pada angka harapan hidup sebagai indikator utama kesejahteraan masyarakat, yang dipengaruhi oleh berbagai faktor seperti tingkat pendidikan, kondisi ekonomi, distribusi infrastruktur kesehatan, dan indikator pembangunan lainnya. Selain itu, algoritma XGBoost (Extreme Gradient Boosting) digunakan karena keunggulannya dalam prediksi data berskala besar dengan tingkat akurasi tinggi. Studi literatur

menunjukkan bahwa XGBoost telah banyak digunakan untuk berbagai permasalahan prediksi, namun penerapannya untuk angka harapan hidup, khususnya di tingkat lokal seperti Jawa Timur, masih relatif terbatas. Penelitian sebelumnya cenderung menggunakan dataset umum yang mengabaikan karakteristik lokal, sehingga menghasilkan model dengan akurasi lebih rendah pada wilayah tertentu. Berdasarkan kesenjangan ini, penelitian ini menggunakan data spesifik lokal untuk menghasilkan model prediksi yang lebih relevan. Studi literatur ini juga membantu merancang langkah-langkah penelitian secara terarah, termasuk pengumpulan data, implementasi model, hingga evaluasi kinerja menggunakan metrik seperti Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). Dengan memahami penelitian terdahulu, penelitian ini mampu mengisi kesenjangan yang ada dengan pendekatan yang lebih relevan terhadap kondisi lokal Jawa Timur, sekaligus memberikan kontribusi bagi perencanaan kebijakan berbasis data untuk meningkatkan kualitas hidup masyarakat di wilayah tersebut.

3.3 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini mencakup data lokal Jawa Timur yang mencerminkan faktor-faktor yang memengaruhi angka harapan hidup. Data tersebut diperoleh dari sumber resmi seperti Badan Pusat Statistik (BPS) dan instansi kesehatan setempat. Dataset ini mencakup berbagai variabel yang relevan untuk analisis dan prediksi, termasuk aspek demografi, sosial, ekonomi, dan infrastruktur kesehatan.

Tabel 3.1 Variabel Dataset

Variabel	Keterangan
Angka Harapan Hidup	Rata-Rata Umur Penduduk
Tingkat Kemiskinan	Persentase Penduduk Dibawah Garis Kemiskinan
Tingkat Pendidikan	Rata-Rata Lama Sekolah
Akses Air Bersih	Persentase Penduduk Yang Memiliki Air Bersih

Kepadatan Penduduk

Jumlah Penduduk Per Kilometer  
Persegi

### 3.4 Praprocessing Data

#### 1. Handling Missing Values

Handling missing values adalah langkah awal untuk memastikan dataset siap digunakan dalam analisis. Data yang hilang (missing values) dapat menyebabkan bias atau menurunkan akurasi model prediksi. Oleh karena itu, diperlukan metode untuk menangani data yang hilang dengan hati-hati. Salah satu metode yang umum digunakan adalah menghapus baris atau kolom yang mengandung missing values jika proporsinya sangat kecil ( $<5\%$ ) dan tidak signifikan. Namun, jika data yang hilang memiliki proporsi yang besar atau memuat informasi penting, maka metode imputasi digunakan.

Imputasi dilakukan dengan menggantikan nilai yang hilang menggunakan rata-rata (mean) atau nilai tengah (median) dari kelompok data serupa. Misalnya, untuk variabel seperti "angka harapan hidup," nilai yang hilang dapat diisi dengan rata-rata angka harapan hidup di wilayah yang sama. Teknik ini memastikan konsistensi dalam dataset sekaligus meminimalkan kehilangan informasi penting. Dengan menangani missing values secara efektif, dataset menjadi lebih lengkap dan siap digunakan untuk membangun model prediktif.

#### 2. Encoding Data Kategorik

Variabel kategorik, seperti status sosial-ekonomi atau jenis fasilitas kesehatan, harus dikonversi ke format numerik agar dapat diproses oleh algoritma seperti XGBoost. Salah satu metode yang digunakan adalah label encoding, di mana setiap kategori diberi label angka. Sebagai contoh, kategori "Tinggi," "Menengah," dan "Rendah" dalam status sosial-ekonomi dapat diubah menjadi 2, 1,

dan 0. Metode ini sederhana dan efektif untuk variabel kategorik ordinal yang memiliki hubungan urutan.

### 3. Feature Selection

Feature selection adalah proses memilih fitur yang memiliki hubungan signifikan dengan target, yaitu angka harapan hidup. Langkah pertama dalam proses ini adalah melakukan analisis korelasi, seperti Pearson atau Spearman correlation, untuk mengukur hubungan antara fitur dan target. Korelasi positif atau negatif yang signifikan menunjukkan fitur yang relevan untuk dimasukkan dalam model.

### 4. Handling Outliers

Outliers atau data pencilan dapat memengaruhi performa model prediktif secara signifikan, sehingga perlu diidentifikasi dan ditangani dengan hati-hati. Salah satu metode yang digunakan adalah Z-Score, yang mengukur sejauh mana data menyimpang dari rata-rata. Nilai Z-Score di luar rentang  $\pm 3$  dianggap sebagai outlier yang harus ditinjau lebih lanjut.

## 3.5 Implementasi Model

Setelah melalui tahap preprocessing, selanjutnya adalah tahap processing atau pemrosesan data dengan metode XGBoost. Adapun tahapan processing pada penelitian ini adalah sebagai berikut.

### 1. Data Splitting

Tahapan ini melibatkan pembagian dataset menjadi dua bagian utama: data latih (80%) dan data uji (20%). Data latih digunakan untuk melatih model XGBoost, di mana model akan belajar mengenali pola dalam data. Sebaliknya, data uji digunakan untuk mengevaluasi performa model pada data baru yang belum pernah dilihat sebelumnya. Proses ini bertujuan untuk memastikan bahwa model mampu melakukan generalisasi dengan baik dan tidak hanya menghafal data latih (overfitting).

Pembagian dataset dilakukan secara acak menggunakan fungsi bawaan seperti `train_test_split` dari library Scikit-learn, dengan parameter stratifikasi jika dataset memiliki variabel target kategori. Stratifikasi menjaga proporsi variabel target tetap seimbang antara data latih dan data uji. Pembagian ini penting untuk mengukur performa model secara objektif pada data yang mencerminkan skenario di dunia nyata.

## 2. Hyperparameter Tuning dengan Grid Search Cross Validation

Hyperparameter tuning adalah proses penting untuk meningkatkan performa model dengan menemukan kombinasi parameter terbaik. Dalam konteks XGBoost, beberapa hyperparameter yang sering dioptimalkan.

Grid Search Cross Validation adalah metode yang digunakan untuk mencoba berbagai kombinasi parameter secara sistematis. Proses ini melibatkan pembagian data latih menjadi beberapa lipatan (fold) untuk validasi silang. Model dilatih pada sebagian data dan divalidasi pada bagian lainnya, sehingga performanya dapat dievaluasi secara menyeluruh. Dengan menggunakan fungsi seperti `GridSearchCV` dari Scikit-learn, peneliti dapat menentukan grid parameter yang ingin diuji dan secara otomatis mengevaluasi kombinasi terbaik berdasarkan metrik tertentu, seperti Mean Absolute Error (MAE) atau Root Mean Squared Error (RMSE).

Setelah menemukan parameter terbaik, model dilatih ulang menggunakan parameter ini pada seluruh data latih, dan performanya dievaluasi menggunakan data uji. Proses tuning ini memastikan bahwa model bekerja secara optimal dengan hasil prediksi yang lebih akurat dan andal. Hyperparameter tuning menjadi tahap kunci dalam menghasilkan model XGBoost yang unggul dalam memprediksi angka harapan hidup.

### 3.6 Evaluasi Model

Tahap evaluasi dalam konteks machine learning adalah proses mengukur kinerja dan keefektifan model yang telah dibangun. Evaluasi model penting untuk memahami sejauh mana model tersebut mampu melakukan prediksi yang akurat dan berguna dalam praktiknya. Peneliti menggunakan Mean Absolute Error (MAE)

dimana metode ini digunakan untuk mengukur sejauh mana model prediksi numerik (seperti XGBoost Regressor) mendekati nilai sebenarnya dalam data pengujian. MAE menghitung rata-rata dari selisih absolut antara prediksi model dan nilai sebenarnya, selain itu penulis juga menggunakan Root Mean Squared Error (RMSE) digunakan untuk mengukur kesalahan prediksi dengan lebih sensitif terhadap outlier, memberikan gambaran yang realistis terhadap kesalahan besar. R-squared ( $R^2$ ) digunakan untuk menilai sejauh mana model dapat menjelaskan varians dalam data target, dengan nilai mendekati 1 menunjukkan kemampuan model yang sangat baik dalam menangkap pola data. Selain itu, akurasi diukur untuk mengevaluasi tingkat kesesuaian hasil prediksi dengan nilai aktual dalam margin tertentu. Kombinasi metrik ini memberikan gambaran menyeluruh tentang seberapa baik model XGBoost bekerja dalam memprediksi angka harapan hidup berdasarkan data yang digunakan.



## DAFTAR PUSTAKA

- Chairunisa, G., Najib, M. K., Nurdianti, S., Imni, S. F., Sanjaya, W., Andriani, R. D., Henriyansah, Putri, R. S. P., & Ekaputri, D. (2024). Life Expectancy Prediction Using Decision Tree, Random Forest, Gradient Boosting, and XGBoost Regressions. *Jurnal Sintak*, 2(2), 71–82. <https://doi.org/10.62375/jsintak.v2i2.249>
- Furqon, H. Z., Rosid, M. A., Eviyanti, A., & Rahmawati, Y. (2024). *Prediction of Life Expectancy Based on Socioeconomic and Health Factors Using Random Forest Regressor Algorithm [ Prediksi Angka Harapan Hidup Berdasarkan Faktor Sosioekonomi Dan Kesehatan Menggunakan Algoritma Random Forest Regressor ]*. 1–10.
- Kurniawan, W., & Indahyanti, U. (2024). Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode XGBoost. *Indonesian Journal of Applied Technology*, 1(2), 18. <https://doi.org/10.47134/ijat.v1i2.3045>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Rahman, M., et al. (2021). Predictive Modeling in Public Health Using XGBoost. *Journal of Data Science and Applications*, 34(2), 45–58.
- Wang, S., et al. (2020). Hybrid Models Combining XGBoost and Neural Networks for Time Series Prediction. *Applied Soft Computing*, 89, 106102.