

# Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode XGBoost

Oleh:

Wildan Kurniawan

Uce Indahyanti

Informatika

Universitas Muhammadiyah Sidoarjo

September, 2023

# Pendahuluan

Angka harapan hidup merupakan rata-rata jumlah tahun kehidupan yang dijalani seseorang yang telah berhasil mencapai umur tertentu. Ada berbagai faktor yang mempengaruhi angka harapan hidup seseorang. Hal ini dapat dimanfaatkan pemerintah untuk melakukan evaluasi kinerja pemerintah dalam meningkatkan kesejahteraan penduduk.

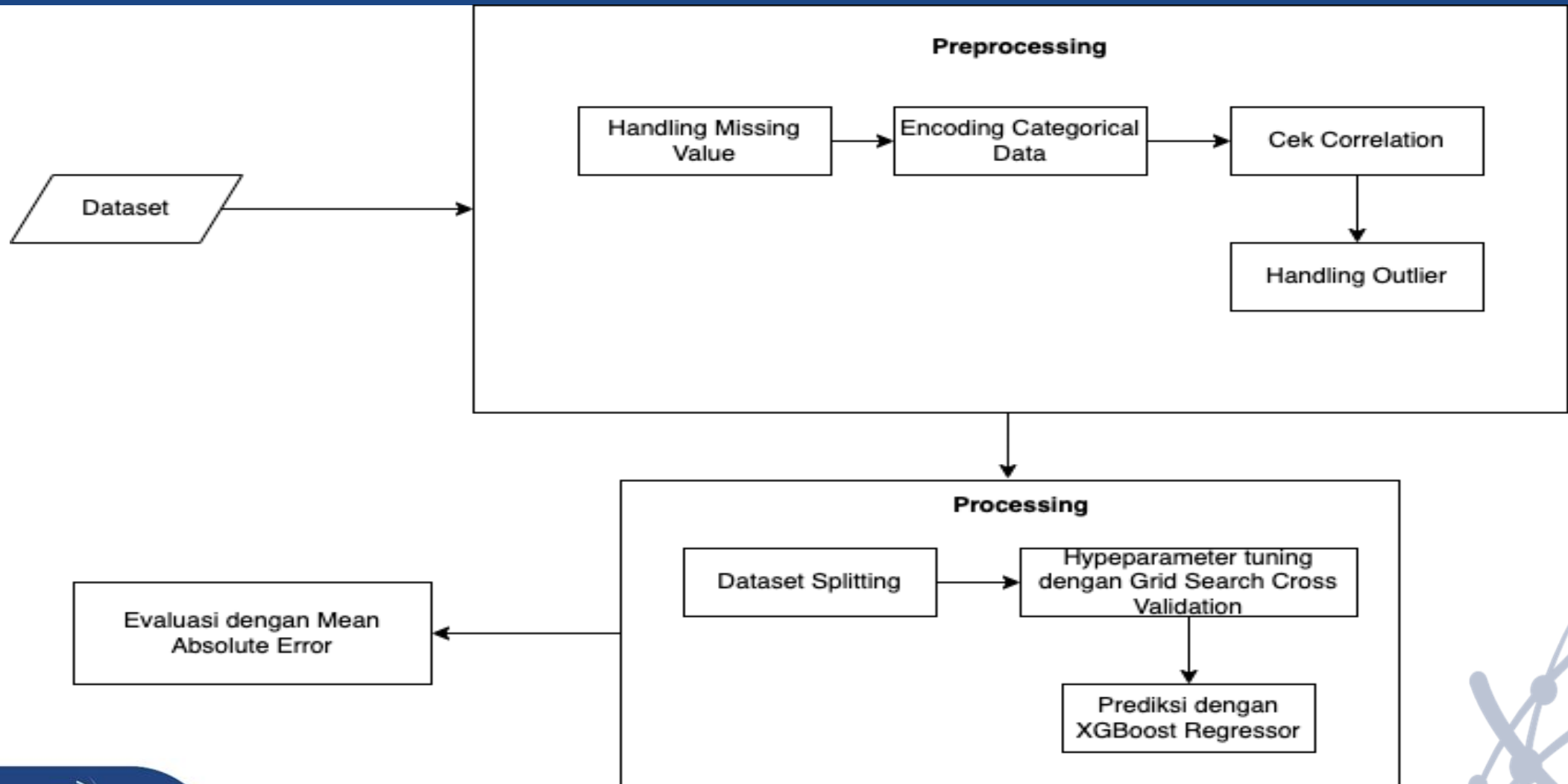
Karena begitu pentingnya pengetahuan tentang usia kehidupan penduduk dunia khususnya di Asia, maka perlu dilakukan prediksi terhadap hal tersebut, agar setiap negara memiliki referensi dan acuan yang jelas untuk menentukan kebijakan ataupun membuat langkah-langkah strategis yang tepat agar usia kehidupan penduduk di Indonesia jangan sampai menurun di masa yang akan datang, bahkan mampu meningkat pada tiap tahunnya.

Salah satu algoritma machine learning yang dapat digunakan untuk melakukan prediksi terhadap kasus ini adalah XGBoost regressor. algoritma ini merupakan salah satu algoritma machine learning yang populer dan banyak direkomendasikan oleh para juara kompetisi di Kaggle. Oleh karena itu diusulkan penelitian ini yang berjudul “Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode XGBoost”

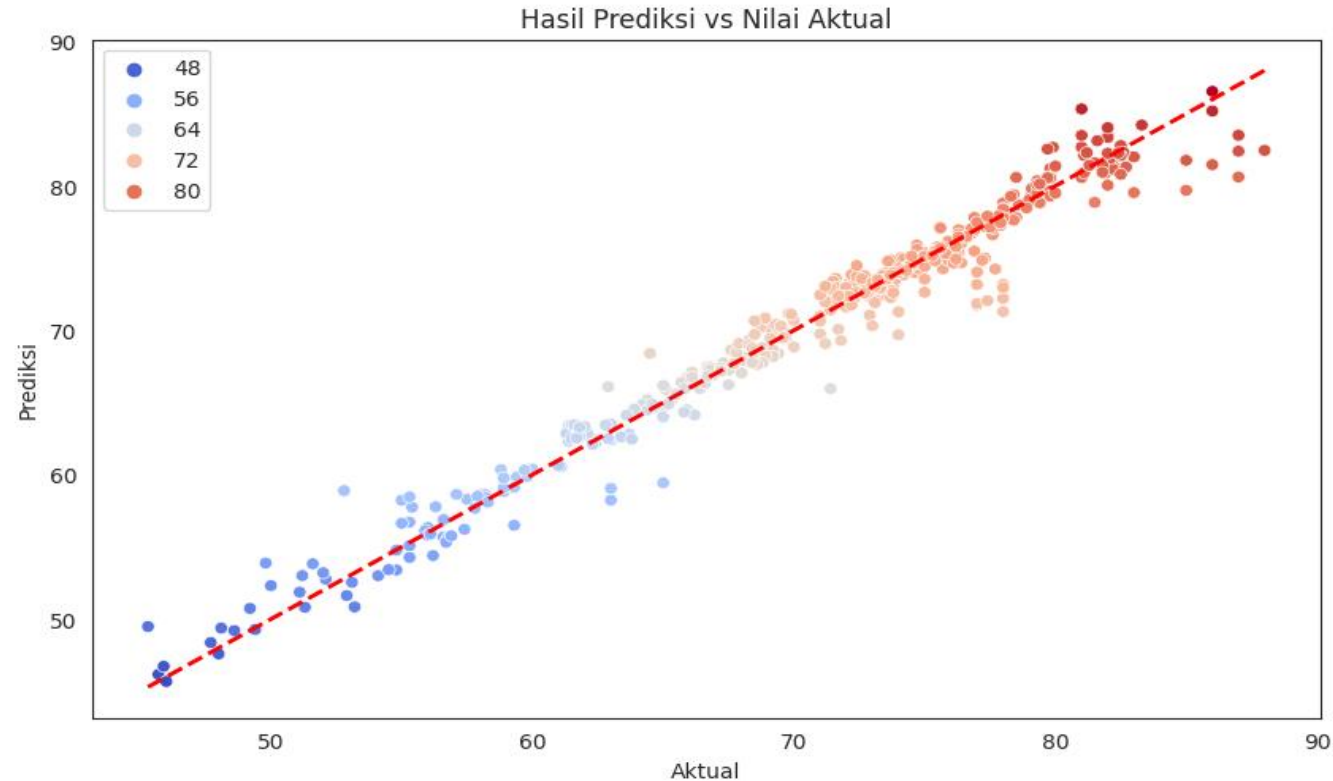
# Pertanyaan Penelitian (Rumusan Masalah)

1. Bagaimana membangun pemodelan prediksi angka harapan hidup penduduk asia menggunakan metode XGBoost ?
2. Bagaimana cara mendapatkan parameter fitting terbaik dengan algoritma XGBoost?
3. Bagaimana melakukan uji performa pemodelan dengan menggunakan confusion matrix?

# Metode



# Hasil



Gambar di atas merupakan visualisasi dari uji coba model yang telah dibangun. Garis lurus putus-putus (---) menunjukkan usia aktual sedangkan scatter plot menunjukkan prediksi usia. Dari gambar tersebut, model yang telah dibangun berhasil melakukan prediksi dengan baik dimana persebaran scatter plot mendekati garis.

# Pembahasan

**Table Hasil Hyperparameter Tuning**

Hyperparameter	Value
Eta	0.3
Learning_rate	0.1
Max_depth	12
Min_child_weight	8
Reg_alpha	1
Reg_lambda	0.1

Dari parameter terbaik yang dihasilkan oleh grid search cross validation yang terdapat pada table di atas, selanjutnya dihasilkan skor train sebesar 99,6% dan skor test sebesar 96,8%. Selain itu, untuk skor mean absolute error yang dihasilkan adalah 0.97, jika skor MAE semakin mendekati 0 maka model yang telah dibangun semakin bagus.

# Temuan Penting Penelitian

Temuan penting dalam penelitian ini adalah metode machine learning yang salah satunya adalah XGBoost dapat menghasilkan nilai akurasi yang tinggi jika dikombinasikan dengan grid search cross validation.

# Manfaat Penelitian

Memberikan kontribusi bagi upaya pengembangan aplikasi yang dapat membantu pemerintah untuk meningkatkan harapan hidup di negara tersebut.



# Referensi

- [1] T. Afriliansyah and Z. Zulfahmi, "Prediksi Angka Harapan Hidup Masyarakat Aceh dengan Model Terbaik Algoritma Cyclical Order," *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, vol. 2, pp. 441–449, 2020.
- [2] *Cost and affordability of healthy diets across and within countries*. 2020. doi: 10.4060/cb2431en.
- [3] P. Parulian et al., "Analysis of Sequential Order Incremental Methods in Predicting the Number of Victims Affected by Disasters," in *Journal of Physics: Conference Series*, 2019. doi: 10.1088/1742-6596/1255/1/012033.
- [4] A. Wanto et al., "Forecasting the Export and Import Volume of Crude Oil, Oil Products and Gas Using ANN," in *Journal of Physics: Conference Series*, 2019. doi: 10.1088/1742-6596/1255/1/012016.
- [5] C. Wang, C. Deng, and S. Wang, "Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost," *Pattern Recognit Lett*, vol. 136, 2020, doi: 10.1016/j.patrec.2020.05.035.
- [6] S. P. Sinaga, A. Wanto, and S. Solikhun, "Implementasi Jaringan Syaraf Tiruan Resilient Backpropagation dalam Memprediksi Angka Harapan Hidup Masyarakat Sumatera Utara," *Jurnal Infomedia*, vol. 4, no. 2, 2020, doi: 10.30811/jim.v4i2.1573.
- [7] P. R. , , D. A. Sihombing, S. Suryadiningrat, and Y. P. A. C. Yuda, . "Identifikasi Data Outlier (Pencilan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya," *Jurnal Ekonomi dan Statistik Indonesia*, pp. 307–316, 2022.
- [8] C. Nkikabahizi, W. Cheruiyot, and A. Kibe, "Chaining Zscore and feature scaling methods to improve neural networks for classification[Formula presented]," *Appl Soft Comput*, vol. 123, 2022, doi: 10.1016/j.asoc.2022.108908.
- [9] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis," *Informatics*, vol. 8, no. 4, 2021, doi: 10.3390/informatics8040079.
- [10] D. A. Anggoro and N. A. Afdallah, "Grid Search CV Implementation in Random Forest Algorithm to Improve Accuracy of Breast Cancer Data," *Int J Adv Sci Eng Inf Technol*, vol. 12, no. 2, 2022, doi: 10.18517/ijaseit.12.2.15487.
- [11] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Inf Sci (N Y)*, vol. 585, 2022, doi: 10.1016/j.ins.2021.11.036.

