



## **Prediksi Angka Harapan Hidup Menggunakan Algoritma Regresi Linier Berdasarkan Tingkat Kabupaten/Provinsi/Kota**

**Galang Rintang Widya Pratama<sup>1</sup>, Ananto Tri Sasongko<sup>2</sup>**

<sup>1</sup>Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

[galangrwp6@mhs.pelitabangsa.ac.id](mailto:galangrwp6@mhs.pelitabangsa.ac.id), [ananto@pelitabangsa.ac.id](mailto:ananto@pelitabangsa.ac.id)

### **Abstract**

*Life Expectancy is an estimate of the average additional years a person is expected to live. Life Expectancy can also be defined as the average number of years lived by a person after that person reaches their xth birthday. The AHH in Indonesia continues to increase from year to year. Indonesia's AHH in 2022 is 71.9 years, up from 71.6 years in 2021, according to data from the Central Bureau of Statistics. Sukoharjo Regency has the highest AHH in 2022, and Nduga Regency has the lowest. Predicting AHH is one way that can be done to increase AHH evenly. AHH prediction can be used to determine the potential AHH in a region in the future, and the government can take the necessary steps to increase AHH in the region. The data used is AHH data from 2020 to 2022. This data is processed using a linear regression algorithm to predict AHH in 2023. One of the prediction methods that can be used is linear regression. Linear regression is a statistical technique that models the relationship between a dependent variable and one or more independent variables. The data used in this research is the 2020-2022 National Welfare Statistics which is secondary data from the Central Bureau of Statistics.*

**Keywords:** *K Means, Mobile Legends, Clustering*

### **Abstrak**

Angka Harapan Hidup adalah perkiraan rata-rata tambahan umur seseorang yang diharapkan dapat terus hidup. Angka Harapan Hidup juga dapat didefinisikan sebagai rata-rata jumlah tahun yang dijalani oleh seseorang setelah orang tersebut mencapai ulang tahun yang ke-x. AHH di Indonesia terus meningkat dari tahun ke tahun. AHH Indonesia pada tahun 2022 adalah 71,9 tahun, meningkat dari 71,6 tahun pada tahun 2021, menurut data dari Badan Pusat Statistik. Kabupaten Sukoharjo memiliki AHH tertinggi pada tahun 2022, dan Kabupaten Nduga memiliki AHH terendah. Melakukan prediksi AHH adalah salah satu cara yang dapat dilakukan untuk meningkatkan AHH secara merata. Prediksi AHH dapat digunakan untuk mengetahui potensi AHH di suatu wilayah pada masa yang akan datang, dan pemerintah dapat mengambil langkah-langkah yang diperlukan untuk meningkatkan AHH di wilayah tersebut. Data yang digunakan adalah data AHH tahun 2020 hingga tahun 2022. Data ini diolah menggunakan algoritma regresi linier untuk memprediksi AHH pada tahun 2023. Salah satu metode prediksi yang dapat digunakan adalah regresi linier. Regresi linier adalah teknik statistik yang memodelkan hubungan antara variabel terikat dan satu atau lebih variabel bebas. Data yang digunakan dalam penelitian ini adalah Statistik Kesejahteraan Nasional tahun 2020-2022 yang merupakan data sekunder dari Badan Pusat Statistik.

**Kata kunci:** Angka Harapan Hidup, Prediksi, Regresi Linier

### **Pendahuluan**

Angka Harapan Hidup (AHH) adalah perkiraan rata-rata tambahan umur seseorang yang diharapkan dapat terus hidup[1]. Angka Harapan Hidup juga dapat didefinisikan sebagai rata-rata jumlah tahun yang dijalani oleh seseorang setelah orang tersebut mencapai ulang tahun yang ke-x[2]. AHH di Indonesia terus meningkat dari tahun ke tahun. AHH Indonesia pada tahun 2022 adalah 71,9 tahun, meningkat dari 71,6 tahun pada tahun 2021, menurut data dari Badan Pusat Statistik (BPS). Peningkatan AHH ini tidak terjadi karena berbagai upaya pemerintah untuk

mengendalikan penyakit, meningkatkan akses terhadap layanan kesehatan, dan meningkatkan gizi. Meskipun AHH terus meningkat, masih ada perbedaan AHH antar daerah di Indonesia. Kabupaten Sukoharjo memiliki AHH tertinggi pada tahun 2022 (77,86 tahun), dan Kabupaten Nduga memiliki AHH terendah (55,40).

Dibutuhkan upaya untuk meningkatkan AHH secara merata di seluruh Indonesia karena perbedaan AHH di antara wilayah ini dapat disebabkan oleh berbagai faktor, seperti pola hidup masyarakat, akses ke layanan kesehatan, dan tingkat pendapatan. Melakukan prediksi AHH adalah salah satu cara yang dapat dilakukan untuk meningkatkan

AHH secara merata. Prediksi AHH dapat digunakan untuk mengetahui potensi AHH di suatu wilayah pada masa yang akan datang, dan pemerintah dapat mengambil langkah-langkah yang diperlukan untuk meningkatkan AHH di wilayah tersebut.

Pada penelitian ini prediksi AHH dilakukan dengan menggunakan algoritma regresi linier berdasarkan tingkat prefektur/provinsi/kota. Data yang digunakan adalah data AHH tahun 2020 hingga tahun 2022. Data ini diolah menggunakan algoritma regresi linier untuk memprediksi AHH pada tahun 2023.

### Metode Penelitian

Menurut Muhammad, dkk(2020) prediksi adalah ukuran memperkirakan kondisi masa depan berdasarkan data masa lalu. Salah satu metode prediksi yang dapat digunakan adalah regresi linier. Metode regresi linier sendiri terdiri dari dua jenis yaitu metode regresi sederhana dan metode regresi linier berganda. Penelitian ini menggunakan metode regresi linier sederhana dengan satu variabel independen[3], [4]. Regresi linier adalah teknik statistik yang memodelkan hubungan antara variabel terikat (dependen, respon, Y) dan satu atau lebih variabel bebas (independen, prediktor, X). Jika hanya ada satu variabel bebas maka disebut regresi linier sederhana. Jika terdapat lebih dari satu variabel bebas disebut regresi linier berganda[5].

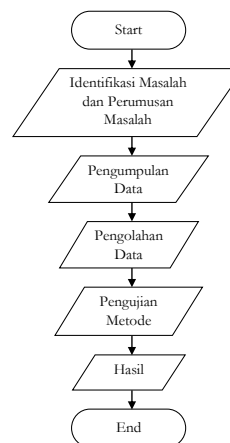
Data yang digunakan dalam penelitian ini adalah Statistik Kesejahteraan Nasional tahun 2020-2022 yang merupakan data sekunder dari Badan Pusat Statistik. Data yang diperkenalkan terdapat 3.288 dataset pada tahun 2020-2022. Data ini disajikan pada tingkat nasional dan regional, sehingga memungkinkan dilakukan perbandingan antar wilayah. Untuk memperjelas maka diberikan simbol  $X$  sebagai variabel bebas dan simbol  $Y$  sebagai variabel terikat.

1. Variabel Bebas  
Variabel yang menyebabkan atau mempengaruhi suatu variabel terikat. Dalam penelitian ini mencakup Provinsi/Kabupaten/Kota[3].
2. Variabel terikat  
Variabel yang dipengaruhi oleh variabel bebas. Angka Harapan Hidup (AHH) memegang peranan penting dalam penelitian ini[3].

Tabel 1. Data Angka Harapan Hidup (AHH)

Provinsi/Kabupaten/Kota (X)	Angka Harapan Hidup (AHH) Menurut Kabupaten/Kota dan Jenis Kelamin (Tahun) (Y)					
	Laki-laki (2020)	Laki-laki (2021)	Laki-laki (2022)	Perempuan n (2020)	Perempuan n (2021)	Perempuan n (2022)
ACEH	68.04	68.07	68.26	71.91	71.94	72.16
Simeulue	63.3	63.33	63.52	67.14	67.15	67.36
Aceh Singkil	65.39	65.4	65.56	69.31	69.31	69.54
Aceh Selatan	62.4	62.46	62.68	66.2	66.26	66.51
Aceh Tenggara	66.11	66.17	66.39	70.06	70.14	70.42
Aceh Timur	66.72	66.76	66.92	70.63	70.65	70.86
Aceh Tengah	66.8	66.8	66.95	70.78	70.78	70.98
Aceh Barat	65.95	65.96	66.13	69.91	69.93	70.14
Aceh Besar	67.8	67.8	67.96	71.69	71.69	71.9

Data yang digunakan dalam penelitian ini berdasarkan data asli angka harapan hidup pada tahun 2020 hingga 2022. Data ini mencakup dua atribut: nama Provinsi/Kabupaten/Kota dan persentase angka harapan hidup[1].

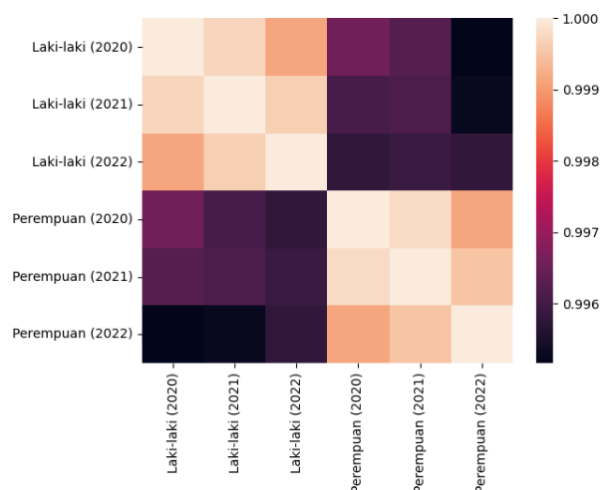


Gambar 1. Bagan Alur

Pada gambar di atas, penelitian diawali dengan identifikasi masalah dan perumusan masalah[6]. Kami mencari referensi dan metode berdasarkan penelitian yang berfokus pada prediksi angka harapan hidup (AHH). Saat kami mengumpulkan data, kami mencari data tersebut dari website resmi Badan Pusat Statistik (BPS) dengan URL <https://www.bps.go.id/id/statistics-table/2/NDU1IzI=/angka-harapan-hidup--ahh--menurut-kabupaten-kota-dan-jenis-kelamin--tahun-.html> dan terlebih dahulu mengolah data tersebut dan membersihkan data tersebut. Data tersebut kemudian dilakukan proses perhitungan dengan menggunakan metode regresi linier sederhana. Ini pertama-tama membagi data menjadi dua bagian: data pengujian (test data) dan data pelatihan (training data)[7]. Hanya dengan begitu Anda akan dapat memprediksi angka harapan hidup (AHH) dengan akurasi dan spesifikasi yang tersedia[6].

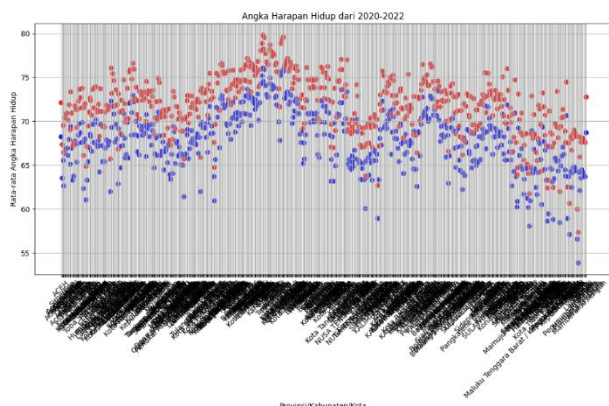
### Hasil dan Pembahasan

#### Hasil Pengolahan Menggunakan Library PySpark



Gambar 2. Visualisasi data menggunakan Seaborn  
Berdasarkan hasil visualisasi data dari Seaborn. Data yang digunakan adalah data AHH tahun 2020 hingga tahun 2022. Data ini diperoleh dari Badan Pusat Statistik (BPS). Angka tersebut menunjukkan korelasi yang sangat kuat antara AHH tahun 2022 dengan AHH tahun 2020 dan 2021. Hal ini ditunjukkan dengan warna gelap pada sel yang menghubungkan kedua variabel tersebut. Visualisasi data ini dapat digunakan untuk menyempurnakan hasil

penelitian yang dilakukan. Gambar 2 menunjukkan bahwa korelasi antara AHH tahun 2022 dengan AHH tahun 2020 dan 2021 sangat tinggi masing-masing sebesar 0,99 dan 0,98. Hal ini menunjukkan bahwa AHH tahun 2022 memiliki hubungan yang sangat kuat dengan AHH tahun 2020 dan 2021. Berdasarkan hasil visualisasi data dapat disimpulkan bahwa AHH tahun 2022 dapat diprediksi dengan sangat akurat berdasarkan AHH tahun 2020 dan 2021. Hal ini menunjukkan bahwa model regresi linier yang digunakan dalam penelitian ini sudah tepat.



Gambar 3. Visualisasi data menggunakan Matplotlib  
Gambar ini menunjukkan AHH di Indonesia terus meningkat dari tahun ke tahun[8]. AHH Indonesia pada tahun 2022 menjadi 71,2 tahun, naik dari 69,5 tahun pada tahun 2021. Gambar 1 juga menunjukkan adanya perbedaan AHH antar wilayah di Indonesia. Pada tahun 2022, nilai AHH tertinggi tercatat di Kabupaten Sukoharjo (77,86 tahun) dan nilai AHH terendah tercatat di Kabupaten Nduga (55,40). Visualisasi data ini dapat digunakan untuk menyempurnakan hasil penelitian yang dilakukan. Gambar 3 menunjukkan bahwa hasil prediksi AHH yang dilakukan oleh algoritma regresi linier konsisten dengan tren data sebenarnya. Berdasarkan hasil visualisasi data dapat disimpulkan bahwa kesenjangan AHH antar wilayah di Indonesia masih perlu diisi. Pemerintah harus mengambil langkah-langkah untuk meningkatkan AHH di daerah yang rendah[9].

Root Mean Squared Error (RMSE): 1.5086904823569852e-11  
R-squared (R2): 1.0

Gambar 4. Hasil visualisasi evaluasi model laki-laki  
Hasil visualisasi evaluasi model laki-laki Gambar 4 menunjukkan hasil visualisasi evaluasi model laki-laki. Data yang digunakan adalah data AHH laki-laki pada tahun 2020 hingga 2022. Seperti yang Anda lihat dari gambar, nilai RMSE (root mean square error) adalah 1,5086904823569852e-11. Nilai RMSE yang kecil menunjukkan bahwa rata-rata selisih antara nilai prediksi dan nilai aktual sangat kecil[10]. Nilai R2 (R2) adalah 1. Nilai R-squared sebesar 1 menunjukkan bahwa model regresi linier dapat menjelaskan 100% variabilitas pada AHH laki-laki. Berdasarkan hasil visualisasi evaluasi model laki-laki, dapat disimpulkan bahwa model regresi linier dapat digunakan untuk memprediksi AHH pada laki-laki dengan sangat akurat. Hal ini ditunjukkan dengan nilai RMSE yang sangat kecil dan nilai R-squared sebesar 1[11].

Root Mean Squared Error (RMSE): 1.0777097012481288e-11  
R-squared (R2): 1.0

Gambar 5. Hasil visualisasi evaluasi model Perempuan  
Data yang digunakan adalah data AHH perempuan pada tahun 2020 hingga 2022. Data ini diperoleh dari Badan Pusat Statistik (BPS). Seperti yang Anda lihat dari gambar, nilai RMSE (root mean square error) adalah 1.0777097012481288e-11. Nilai RMSE yang kecil menunjukkan bahwa rata-rata selisih antara nilai prediksi dan nilai aktual sangat kecil[10].

Nilai R2 (R2) adalah 1. Nilai R-squared sebesar 1 menunjukkan bahwa model regresi linier dapat menjelaskan 100% variabilitas AHH wanita. Berdasarkan hasil visualisasi evaluasi model wanita, dapat disimpulkan bahwa model regresi linier dapat digunakan untuk memprediksi AHH pada wanita dengan sangat akurat. Hal ini ditunjukkan dengan nilai RMSE yang sangat kecil dan nilai R-squared sebesar 1[11].

### Kesimpulan

Dalam penelitian ini, algoritma Regresi Linier Sederhana diterapkan untuk menganalisis Angka Harapan Hidup telah dilakukan. Hasil analisis menunjukkan bahwa Model regresi linier dapat digunakan untuk memprediksi AHH dengan cukup akurat, baik untuk laki-laki maupun perempuan. Kemudian Nilai RMSE yang sangat kecil dan nilai R-squared sebesar 1 menunjukkan bahwa model regresi linier dapat menjelaskan 100% variabilitas AHH dan hasil prediksi AHH yang dilakukan oleh algoritma regresi linier konsisten dengan tren data sebenarnya.

Kesimpulan ini mempunyai implikasi penting bagi pemerintah dan masyarakat. Bagi pemerintah, kesimpulan ini menunjukkan bahwa proyeksi AHH dapat digunakan untuk mengetahui potensi AHH di masa depan di suatu wilayah. Apabila pemerintah mengetahui potensi AHH di suatu daerah, maka pemerintah dapat mengambil langkah-langkah yang diperlukan untuk meningkatkan AHH di daerah tersebut. Bagi masyarakat umum, kesimpulan ini menunjukkan bahwa AHH di Indonesia terus meningkat dari tahun ke tahun. Namun masih terdapat perbedaan AHH antar wilayah di Indonesia. Hal ini menandakan kesenjangan AHH antar wilayah di Indonesia masih perlu ditutup. Masyarakat dapat berperan aktif dalam meningkatkan AHH di wilayahnya dengan menerapkan pola hidup sehat dan menjaga lingkungan.

### Daftar Pustaka

- [1] Y. Pratiwi dan M. Mulyawan, "Implementasi Algoritma K-Means untuk Menentukan Angka Harapan Hidup berdasarkan Tingkat Provinsi," *Blend Sains Jurnal Teknik*, vol. 1, no. 4, hlm. 284–294, 2023, doi: 10.56211/blendsains.v1i4.233.
- [2] F. Faturahman, "Angka Harapan Hidup Sebagai Intervening Anggaran Kesehatan Masyarakat Dan Sumber Daya Manusia Kesehatan Terhadap Kemiskinan Di Kota Jambi," *ILTIZAM Journal of Shariah*

- Economic Research*, vol. 3, no. 1, hlm. 24, 2019, doi: 10.30631/iltizam.v3i1.269.
- [3] N. Almutazah, N. Azizah, Y. L. Putri, dan D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana," *Jurnal Ilmiah Matematika Dan Terapan*, vol. 18, no. 1, hlm. 31–40, 2021, doi: 10.22487/2540766x.2021.v18.i1.15465.
- [4] A. M. Siregar dan T. A. Hasan, "Aplikasi Linier Regresi dengan Algoritma Jaringan Syaraf Tiruan untuk Sentimen Analisis," *Techno Xplore : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 2, hlm. 43–51, 2018, doi: 10.36805/technoxplore.v3i2.800.
- [5] D. Kurniawan, "Regresi Linier," *Statistic*, hlm. 1–6, 2008.
- [6] N. Riskya dan S. Yuliana, "Penerapan Data Mining Untuk Prediksi Perilaku Pelanggan Menggunakan Multiple Linear Regression," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, 2023, doi: 10.23960/jitet.v11i3.3194.
- [7] Harsiti, Z. Muttaqin, dan E. Srihartini, "Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet," *JSiI (Jurnal Sistem Informasi)*, vol. 9, no. 1, hlm. 12–16, 2022, doi: 10.30656/jsii.v9i1.4426.
- [8] E. Dewi, S. Mulyani, I. Septianingrum, N. Nurjanah, dan R. Rahmawati, "Prediksi Curah Hujan Di Kabupaten Majalengka Dengan Menggunakan Algoritma Regresi," *e-Jurnal JUSITI (Jurnal Sistem Informasi dan Teknologi Informasi)*, vol. 8–1, no. 1, hlm. 67–77, 2019, doi: 10.36774/jusiti.v8i1.602.
- [9] R. Hasanah, S. Syaparuddin, dan R. Rosmeli, "Pengaruh angka harapan hidup, rata-rata lama sekolah dan pengeluaran perkapita terhadap tingkat kemiskinan pada Kabupaten /Kota di Provinsi Jambi," *e-Jurnal Perspektif Ekonomi dan Pembangunan Daerah*, vol. 10, no. 3, hlm. 223–232, 2021, doi: 10.22437/pdpd.v10i3.16253.
- [10] C. Wahyuni, "Penerapan Analisis Jalur (Path Analysis) dalam Menentukan Faktor-faktor yang Mempengaruhi Angka Harapan Hidup di Wilayah Indonesia Bagian Tengah," *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 1, hlm. 74–83, 2023, doi: 10.21009/jsa.07107.
- [11] Wufron, "Analisis Regresi Linier dengan IBM SPSS Statistics," *Universitas Garut*, vol. 1, no. March 2020, hlm. 0–10, 2020, [Daring]. Tersedia pada: <https://doi.org/10.31219%2Fosf.io%2Ffwex8>

