

**LAPORAN PENELITIAN
ANALISIS BIG DATA “HARAPAN HIDUP
LAKI LAKI DAN PEREMPUAN ANTAR PROVINSI DI INDONESIA”
MENGUNAKAN METODE PYSPARK**

Disusun dan Diajukan Untuk Memenuhi Tugas

Mata Kuliah Big Data



Dosen Pengampu :

Dr. Ir. Ananto Tri Sasongko, M.Sc

Di Analisis Oleh :

Kelompok 3

- Raihan Tantowi (312110229)
- Ikram Ramadhan (312110478)
- Nurul fajri (312110506)

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PELITA BANGSA**

Jl. Inspeksi Kalimalang Tegal Danas, Arah Deltamas, Cikarang Selatan, Bekasi Telp. (021)

285118181, 82, 83, 84 Fax (021) 28518181

© 2024

ABSTRAK

Penelitian ini bertujuan menganalisis harapan hidup lelaki dan perempuan antar provinsi di Indonesia dengan menggunakan pendekatan Big Data. Distribusi harapan hidup dieksplorasi untuk mengidentifikasi ketidaksetaraan gender dan faktor-faktor yang memengaruhi variasinya. Analisis mencakup pemahaman pola perubahan harapan hidup dari waktu ke waktu dan identifikasi provinsi-provinsi dengan tren khusus.

Hasil penelitian menunjukkan bahwa terdapat variasi yang signifikan dalam harapan hidup di berbagai wilayah, serta adanya ketidaksetaraan gender yang perlu diperhatikan. Faktor-faktor seperti aksesibilitas layanan kesehatan, pendidikan, dan kondisi ekonomi memainkan peran krusial dalam mempengaruhi harapan hidup.

Implikasi temuan ini memberikan landasan bagi perumusan kebijakan yang lebih efektif untuk meningkatkan kesejahteraan masyarakat secara inklusif. Penelitian ini memberikan kontribusi dalam memahami dinamika harapan hidup di Indonesia dan menyoroti pentingnya tindakan lintas sektor untuk mencapai tujuan pembangunan berkelanjutan.

Bekasi, 14 Januari 2024

Perwakilan Kelompok

DAFTAR ISI

LAPORAN PENELITIAN	i
ABSTRAK	ii
DAFTAR ISI.....	iii
BAB I.....	1
PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	1
C. Tujuan Penelitian	2
BAB II.....	3
TINJAUAN PUSTAKA.....	3
A. Harapan Hidup sebagai Indikator Kesejahteraan.....	3
B. Ketidaksetaraan Gender dalam Harapan Hidup.....	3
C. Big Data dalam Analisis Harapan Hidup	3
D. Variasi Regional dalam Harapan Hidup di Indonesia	3
BAB III.....	4
METODOLOGI PENELITIAN	4
A. Regresi Linier	4
B. Pemilihan Fitur	4
C. Transformasi Fitur.....	4
D. Inisialisasi Model Regresi Linier	5
E. Pembuatan Pipeline.....	5
F. Pemisahan Data Training Dan Testing.....	5
G. Pelatihan Model, Evaluasi Transform dan Evaluasi Kinerja Model	5
H. Menghitung Dan Menampilkan Evaluasi Matriks Lain.....	5
I. Menambahkan Kolom 2023 Dan Membuat Prediksi.....	6
J. Menampilkan Hasil Evaluasi Model Pada Data Pengujian.....	6
BAB IV.....	7
HASIL DAN ANALISA.....	7
A. Harapan Hidup Lelaki Antar Provinsi Indonesia	7
B. Harapan Hidup Perempuan Antar Provinsi Indonesia	11
BAB V	14
KESIMPULAN	14
DAFTAR PUSTAKA	15
PENUTUP.....	16

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Indonesia, sebagai negara kepulauan dengan beragam budaya dan geografi, menyajikan dinamika demografi yang kompleks. Salah satu aspek penting dalam pemahaman kesejahteraan masyarakat adalah harapan hidup, yang mencerminkan tingkat kesehatan dan kualitas hidup penduduk.

Perbedaan harapan hidup antara lelaki dan perempuan serta antarprovinsi menjadi perhatian khusus dalam konteks pembangunan nasional. Faktor-faktor seperti aksesibilitas layanan kesehatan, pendidikan, kondisi ekonomi, dan pola hidup masyarakat dapat memberikan kontribusi signifikan terhadap variasi harapan hidup di berbagai wilayah Indonesia.

Dalam era Big Data, terdapat potensi besar untuk menganalisis dan menginterpretasikan pola-pola kompleks terkait harapan hidup lelaki dan perempuan di setiap provinsi. Data besar ini mencakup informasi yang luas, termasuk data kesehatan, pendidikan, dan ekonomi, yang dapat memberikan wawasan mendalam terhadap faktor-faktor yang mempengaruhi harapan hidup di berbagai daerah.

Dengan menggali informasi dari Big Data, penelitian ini bertujuan untuk menyelidiki perbedaan harapan hidup antar gender dan antarprovinsi di Indonesia. Analisis ini diharapkan dapat memberikan pemahaman yang lebih baik tentang faktor-faktor yang berkontribusi terhadap ketidaksetaraan harapan hidup, sehingga kebijakan yang lebih tepat dapat diimplementasikan untuk meningkatkan kesejahteraan masyarakat Indonesia secara keseluruhan.

Melalui penelitian ini, diharapkan dapat muncul rekomendasi kebijakan yang dapat membantu merumuskan strategi pembangunan nasional yang lebih inklusif dan berkelanjutan, mengurangi disparitas antar wilayah, serta meningkatkan kualitas hidup masyarakat Indonesia secara menyeluruh.

B. Rumusan Masalah

Dalam konteks Analisis Big Data “Harapan Hidup Lelaki Dan Perempuan Antar Provinsi Di Indonesia”, permasalahan yang akan dijawab melalui penelitian ini dapat dirumuskan sebagai berikut :

1. Bagaimana distribusi harapan hidup lelaki dan perempuan di berbagai provinsi di Indonesia?
2. Apakah terdapat perbedaan signifikan dalam harapan hidup antara lelaki dan perempuan di tingkat nasional dan provinsi?
3. Apa saja faktor-faktor utama yang berkorelasi dengan variasi harapan hidup lelaki dan perempuan di setiap provinsi?
4. Bagaimana pola perubahan harapan hidup lelaki dan perempuan selama periode waktu tertentu di Indonesia?
5. Apakah terdapat provinsi-provinsi tertentu yang menunjukkan tren peningkatan atau penurunan harapan hidup yang perlu mendapatkan perhatian khusus?
6. Bagaimana implikasi hasil analisis Big Data terhadap kebijakan pembangunan nasional yang berfokus pada peningkatan harapan hidup lelaki dan perempuan di Indonesia?

C. Tujuan Penelitian

Tujuan laporan penelitian ini mencakup :

1. Menganalisis Distribusi Harapan Hidup :
Menyelidiki dan menggambarkan distribusi harapan hidup lelaki dan perempuan di setiap provinsi di Indonesia untuk memahami pola geografisnya.
2. Mendeteksi Ketidaksetaraan Gender :
Mengidentifikasi apakah terdapat ketidaksetaraan yang signifikan dalam harapan hidup antara lelaki dan perempuan di tingkat nasional dan provinsi.
3. Mengidentifikasi Faktor-faktor Pengaruh :
Menentukan faktor-faktor utama yang berkorelasi dengan variasi harapan hidup, seperti aksesibilitas layanan kesehatan, pendidikan, dan kondisi ekonomi di setiap wilayah.
4. Menganalisis Perubahan Seiring Waktu :
Menyelidiki pola perubahan harapan hidup lelaki dan perempuan selama beberapa tahun terakhir untuk menangkap tren yang mungkin terjadi.
5. Menyoroti Provinsi dengan Tantangan Khusus :
Mengidentifikasi provinsi-provinsi tertentu yang menunjukkan peningkatan atau penurunan harapan hidup yang memerlukan perhatian khusus dan pemahaman mendalam.
6. Memberikan Rekomendasi Kebijakan :
Merumuskan rekomendasi kebijakan berdasarkan temuan analisis Big Data untuk mendukung upaya meningkatkan harapan hidup secara merata di seluruh Indonesia.

BAB II

TINJAUAN PUSTAKA

A. Harapan Hidup sebagai Indikator Kesejahteraan

Harapan hidup merupakan indikator penting dalam mengukur kesejahteraan masyarakat suatu negara. Keterkaitan erat antara harapan hidup dengan kondisi kesehatan, pendidikan, dan faktor-faktor sosial ekonomi menjadi dasar penting untuk menganalisis keberhasilan pembangunan suatu daerah.

Penelitian terdahulu menunjukkan bahwa harapan hidup yang tinggi seringkali berkorelasi dengan akses yang baik terhadap layanan kesehatan, tingkat pendidikan yang tinggi, dan stabilitas ekonomi. Analisis faktor-faktor ini menjadi landasan untuk memahami perbedaan harapan hidup antara lelaki dan perempuan, serta variasi di antara provinsi-provinsi di Indonesia.

B. Ketidaksetaraan Gender dalam Harapan Hidup

Beberapa studi telah menyoroti ketidaksetaraan gender dalam harapan hidup sebagai permasalahan global. Pertanyaan mendasar mengenai akses terhadap pelayanan kesehatan reproduksi, peran sosial gender, dan faktor-faktor lain yang mempengaruhi harapan hidup perlu dieksplorasi lebih lanjut. Kajian literatur ini akan memberikan landasan konseptual untuk menganalisis perbedaan harapan hidup lelaki dan perempuan di Indonesia.

C. Big Data dalam Analisis Harapan Hidup

Dengan berkembangnya teknologi informasi, analisis Big Data telah menjadi instrumen utama dalam menggali informasi kesehatan dan sosial masyarakat. Penelitian-penelitian sebelumnya telah mengaplikasikan Big Data untuk memahami tren kesehatan masyarakat. Dalam konteks penelitian ini, penggunaan Big Data diharapkan dapat memberikan wawasan yang mendalam terhadap perbedaan harapan hidup di antara provinsi-provinsi di Indonesia.

D. Variasi Regional dalam Harapan Hidup di Indonesia

Konteks geografis, budaya, dan ekonomi yang beragam di Indonesia menciptakan variasi harapan hidup yang signifikan di setiap provinsi. Penelitian regional terdahulu akan memberikan gambaran mendalam tentang faktor-faktor yang mempengaruhi harapan hidup di wilayah-wilayah khusus.

BAB III

METODOLOGI PENELITIAN

A. Regresi Linier

Regresi linier adalah suatu metode statistika yang digunakan untuk memodelkan hubungan linier antara satu atau lebih variabel independen (fitur) dengan variabel dependen (target). Dalam konteks analisis data Kami, regresi linier dapat membantu memahami hubungan antara fitur-fitur tertentu dan variabel target "2022."

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$$

- \hat{y} adalah nilai yang diprediksi.
- w_0 adalah intercept (konstanta) dari model.
- w_1, w_2, \dots, w_n adalah bobot (coefficients) yang sesuai dengan masing-masing fitur.
- x_1, x_2, \dots, x_n adalah nilai dari masing-masing fitur.

Diatas adalah cara bagaimana regresi linear melakukan prediksi pada data kami

B. Pemilihan Fitur

Pada tahap ini, dilakukan pemilihan fitur-fitur yang akan digunakan sebagai variabel input dalam model regresi linier. Fitur-fitur yang dipilih mencakup [1:-1]. Pemilihan fitur dilakukan dengan tujuan meningkatkan keakuratan prediksi dan mengurangi kompleksitas model.

```
# Pemilihan fitur
feature_cols = data.columns[1:-1]
assembler = VectorAssembler(inputCols=feature_cols, outputCol="features")
```

Kami membuat variabel "feature_cols" yang berisi daftar kolom-kolom yang akan diambil sebagai fitur-fitur untuk model. Pilihan ini mengambil semua kolom kecuali kolom pertama dan terakhir. Kolom pertama umumnya adalah kolom ID atau label kategori (seperti nama provinsi), dan kolom terakhir adalah kolom target yang ingin diprediksi.

C. Transformasi Fitur

Dengan menggunakan "VectorAssembler" bagian kedua, Kami mempersiapkan data dengan cara yang dibutuhkan oleh model, yang umumnya mengharapkan fitur-fitur di dalam satu vektor.

D. Inisialisasi Model Regresi Linier

```
# Inisialisasi model Regresi Linier
lr = LinearRegression(labelCol="2022", featuresCol="features")
# Membuat pipeline
pipeline = Pipeline(stages=[assembler, lr])
```

Inisialisasi model regresi linier dengan kolom target ("2022") dan kolom fitur ("features"). Model ini akan digunakan untuk melakukan prediksi berdasarkan fitur-fitur yang telah dipilih.

E. Pembuatan Pipeline

Pada tahap selanjutnya Kami membuat pipeline yang berisi langkah-langkah transformasi dan model yang akan dieksekusi secara berurutan. Pada kasus ini, pipeline terdiri dari VectorAssembler dan model Linear Regression, guna mengoptimalkan bobot (coefficients) untuk menyesuaikan data pelatihan..

F. Pemisahan Data Training Dan Testing

```
# Pemisahan data menjadi train dan test
train_data, test_data = data.randomSplit([0.8, 0.2], seed=123)
```

Memisahkan data menjadi set pelatihan (80%) dan set pengujian (20%).

G. Pelatihan Model, Evaluasi Transform dan Evaluasi Kinerja Model

```
# Pelatihan model
model = pipeline.fit(train_data)

# Evaluasi model
test_results = model.transform(test_data)

# Evaluasi kinerja model menggunakan Root Mean Squared Error (RMSE)
evaluator = RegressionEvaluator(labelCol="2022", predictionCol="prediction", metricName="rmse")
rmse = evaluator.evaluate(test_results)

# Evaluasi kinerja model menggunakan beberapa metrik
evaluator = RegressionEvaluator(labelCol="2022", predictionCol="prediction")
```

Menggunakan model untuk melakukan prediksi pada data pengujian, Menggunakan RegressionEvaluator untuk mengukur kinerja model menggunakan Root Mean Squared Error (RMSE) beserta evaluasi matrik lainnya pada data pengujian.

H. Menghitung Dan Menampilkan Evaluasi Matriks Lain

```
# Hitung dan tampilkan MSE
mse = evaluator.evaluate(test_results, {evaluator.metricName: "mse"})
print(f"Mean Squared Error (MSE) on test data = {mse}")

# Hitung dan tampilkan MAE
mae = evaluator.evaluate(test_results, {evaluator.metricName: "mae"})
print(f"Mean Absolute Error (MAE) on test data = {mae}")

# Hitung dan tampilkan R2
r2 = evaluator.evaluate(test_results, {evaluator.metricName: "r2"})
print(f"R-squared (R2) on test data = {r2}")
```


Menghitung dan menampilkan Mean Squared Error (MSE), Mean Absolute Error (MAE), dan R-squared (R2) pada data pengujian, dengan metode yang sama pada RMSE.

I. Menambahkan Kolom 2023 Dan Membuat Prediksi

```
# Menambahkan kolom hasil prediksi ke DataFrame
data_with_predictions = model.transform(data).select("province", "prediction")

# Menambahkan kolom "2023" dengan hasil prediksi
data_with_predictions = data_with_predictions.withColumn("2023", data_with_predictions["prediction"])

print(f"Root Mean Squared Error (RMSE) on test data = {rmse}")
```

model.transform(data) digunakan untuk melakukan transformasi pada seluruh DataFrame data menggunakan model yang telah dilatih. Ini menghasilkan DataFrame baru yang telah ditambahkan kolom "prediction" berisi nilai hasil prediksi.

select("province", "prediction") digunakan untuk memilih hanya dua kolom, yaitu "province" (nama provinsi) dan "prediction" (hasil prediksi). DataFrame ini disimpan dalam variabel `data_with_predictions`.

withColumn("2023", data_with_predictions["prediction"]) menambahkan kolom baru dengan nama "2023" ke DataFrame `data_with_predictions`. Nilai dalam kolom "2023" diisi dengan nilai hasil prediksi yang telah disimpan dalam kolom "prediction" sebelumnya.

Sehingga, DataFrame `data_with_predictions` sekarang memiliki dua kolom: "province" dan "2023" yang berisi nilai hasil prediksi untuk tahun 2023.

J. Menampilkan Hasil Evaluasi Model Pada Data Pengujian

```
# Menampilkan hasil evaluasi
test_results.select("Province", "features", "2022", "prediction").show()

# Menampilkan hasil
data_with_predictions.show(35)
```

test_results.select() digunakan untuk memilih beberapa kolom dari DataFrame `test_results` untuk ditampilkan. Kolom-kolom yang dipilih adalah "Province" (nama provinsi), "features" (vektor fitur hasil transformasi), "2022" (nilai sebenarnya untuk tahun 2022), dan "prediction" (hasil prediksi model untuk tahun 2022).

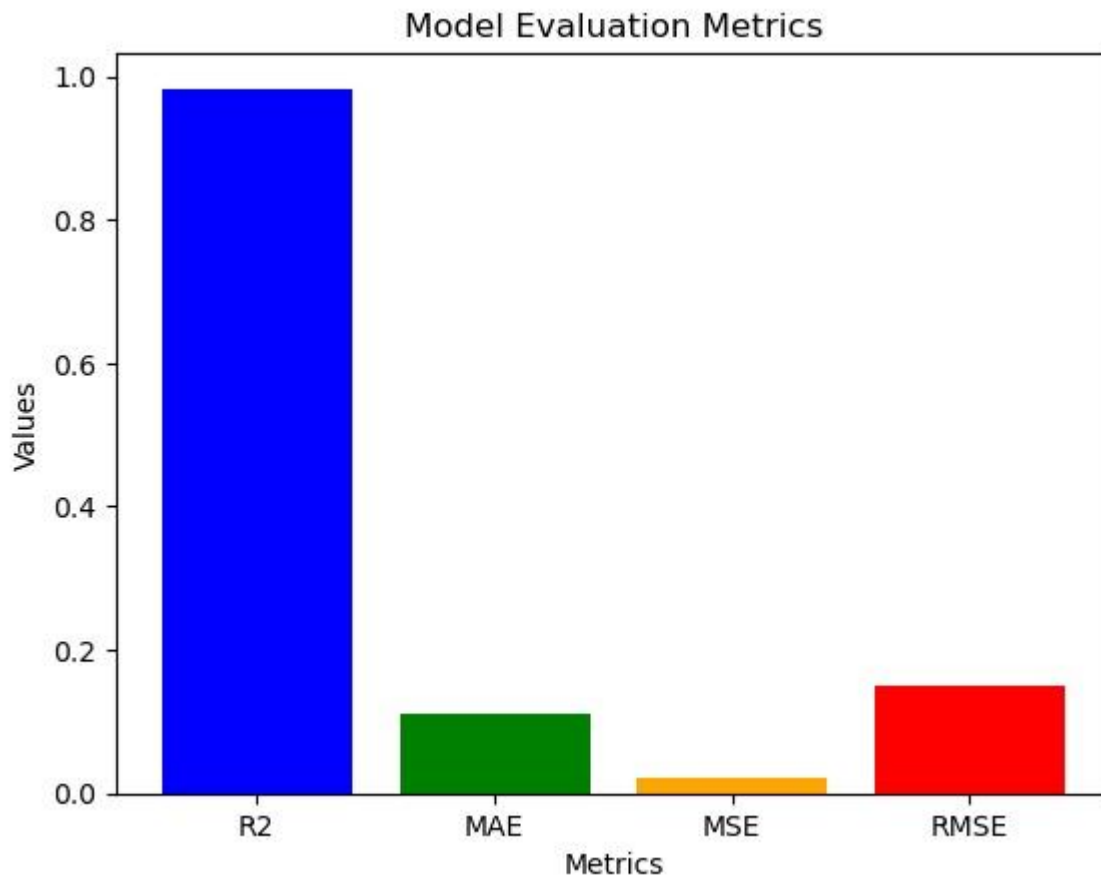
.show() digunakan untuk menampilkan hasil tersebut dalam bentuk tabel pada konsol dan

`data_with_predictions.show(35)` digunakan untuk menampilkan 35 baris pertama dari DataFrame `data_with_predictions`

BAB IV

HASIL DAN ANALISA

A. Harapan Hidup Laki laki Antar Provinsi Indonesia



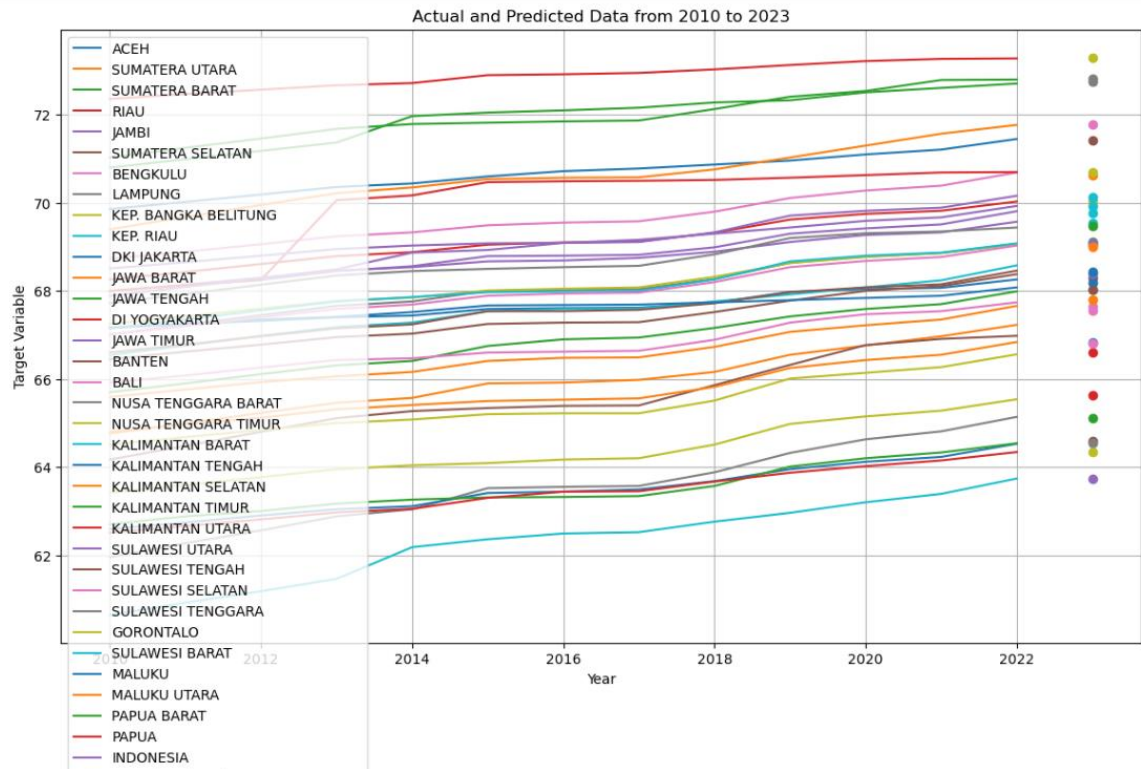
Mean Squared Error (MSE) on test data = 0.022641835428690475
Mean Absolute Error (MAE) on test data = 0.11123236864418118
R-squared (R2) on test data = 0.9830648014810436
Root Mean Squared Error (RMSE) on test data = 0.15047204201674966

Province	features	2022	prediction
BANTEN	[66.6,66.78,66.97...	68.46	68.42772239587062
GORONTALO	[64.51,64.68,64.8...	66.56	66.59812882155047
KALIMANTAN BARAT	[67.15,67.35,67.5...	69.08	69.09460017333653
KALIMANTAN SELATAN	[64.76,64.99,65.2...	67.23	67.5639401808482
KEP. RIAU	[66.54,66.76,66.9...	68.58	68.37860127504918
NUSA TENGGARA TIMUR	[63.42,63.59,63.7...	65.54	65.64075691202898
SULAWESI TENGAH	[64.17,64.48,64.8...	66.98	66.84073807613449
SULAWESI TENGGARA	[67.72,67.93,68.1...	69.44	69.46999239616197
SUMATERA SELATAN	[66.45,66.62,66.7...	68.38	68.17804399713931
SUMATERA UTARA	[65.59,65.76,65.9...	67.66	67.63998905329073

BALI	70.6207559834548	70.6207559834548
NUSA TENGGARA BARAT	65.10727054283855	65.10727054283855
NUSA TENGGARA TIMUR	65.64075691202898	65.64075691202898
KALIMANTAN BARAT	69.09460017333653	69.09460017333653
KALIMANTAN TENGAH	68.02286035690896	68.02286035690896
KALIMANTAN SELATAN	67.5639401808482	67.5639401808482
KALIMANTAN TIMUR	72.82282030530719	72.82282030530719
KALIMANTAN UTARA	70.70226352311273	70.70226352311273
SULAWESI UTARA	70.13291320820703	70.13291320820703
SULAWESI TENGAH	66.84073807613449	66.84073807613449
SULAWESI SELATAN	69.00780246246926	69.00780246246926
SULAWESI TENGGARA	69.46999239616197	69.46999239616197
GORONTALO	66.59812882155047	66.59812882155047
SULAWESI BARAT	63.723097440584425	63.723097440584425
MALUKU	64.58234860463905	64.58234860463905
MALUKU UTARA	66.81630341166611	66.81630341166611
PAPUA BARAT	64.54920768446074	64.54920768446074
PAPUA	64.3484565125731	64.3484565125731
INDONESIA	69.93580402279683	69.93580402279683

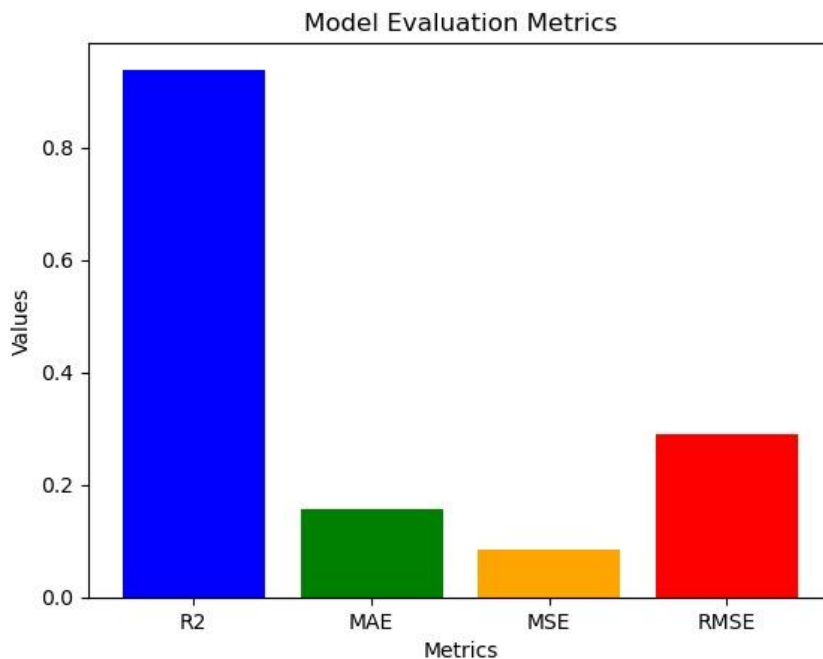
province	prediction	2023
ACEH	68.31001523347358	68.31001523347358
SUMATERA UTARA	67.63998905329073	67.63998905329073
SUMATERA BARAT	68.02002354687043	68.02002354687043
RIAU	70.03905959205902	70.03905959205902
JAMBI	69.54619739930173	69.54619739930173
SUMATERA SELATAN	68.17804399713931	68.17804399713931
BENGKULU	67.79896729407686	67.79896729407686
LAMPUNG	69.10555274807531	69.10555274807531
KEP. BANGKA BELITUNG	69.03820069639075	69.03820069639075
KEP. RIAU	68.37860127504918	68.37860127504918
DKI JAKARTA	71.41880490784827	71.41880490784827
JAWA BARAT	71.78737286611026	71.78737286611026
JAWA TENGAH	72.74533173561208	72.74533173561208
DI YOGYAKARTA	73.30071450532405	73.30071450532405
JAWA TIMUR	69.77785541583746	69.77785541583746
BANTEN	68.42772239587062	68.42772239587062

Province	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
ACEH	67.17	67.25	67.32	67.4	67.44	67.59	67.61	67.62	67.73	67.98	68.04	68.07	68.26
SUMATERA UTARA	65.59	65.76	65.93	66.06	66.16	66.41	66.48	66.49	66.73	67.07	67.22	67.35	67.66
SUMATERA BARAT	65.7	65.9	66.11	66.31	66.41	66.75	66.9	66.94	67.16	67.42	67.59	67.7	67.99
RIAU	68.26	68.43	68.61	68.79	68.88	69.05	69.1	69.12	69.32	69.62	69.75	69.82	70.03
JAMBI	67.99	68.14	68.3	68.46	68.54	68.67	68.69	68.75	68.89	69.11	69.27	69.33	69.57
SUMATERA SELATAN	66.45	66.62	66.78	66.95	67.03	67.25	67.28	67.29	67.52	67.78	68.0	68.11	68.38
BENGKULU	65.92	66.08	66.26	66.43	66.47	66.6	66.62	66.64	66.89	67.28	67.47	67.54	67.74
LAMPUNG	67.01	67.23	67.45	67.66	67.76	68.0	68.03	68.04	68.27	68.64	68.78	68.86	69.07
KEP. BANGKA BELITUNG	67.25	67.42	67.6	67.77	67.85	68.01	68.05	68.08	68.32	68.63	68.77	68.86	69.06
KEP. RIAU	66.54	66.76	66.98	67.18	67.28	67.54	67.59	67.6	67.76	67.92	68.08	68.24	68.58
DKI JAKARTA	69.86	70.02	70.19	70.36	70.44	70.6	70.72	70.78	70.87	70.96	71.1	71.21	71.45
JAWA BARAT	69.4	69.68	69.95	70.22	70.35	70.54	70.57	70.58	70.76	71.03	71.3	71.57	71.77
JAWA TENGAH	70.8	70.99	71.18	71.37	71.97	72.05	72.1	72.16	72.28	72.33	72.51	72.61	72.71
DI YOGYAKARTA	72.36	72.47	72.57	72.67	72.72	72.9	72.92	72.95	73.03	73.13	73.22	73.27	73.28
JAWA TIMUR	68.0	68.13	68.25	68.45	68.56	68.79	68.8	68.82	68.99	69.3	69.42	69.51	69.81
BANTEN	66.6	66.78	66.97	67.15	67.24	67.54	67.54	67.57	67.72	67.97	68.08	68.15	68.46
BALI	68.73	68.89	69.06	69.24	69.33	69.49	69.55	69.58	69.8	70.11	70.28	70.39	70.69
NUSA TENGGARA BARAT	61.95	62.27	62.56	62.88	63.04	63.52	63.55	63.57	63.88	64.32	64.63	64.81	65.14
NUSA TENGGARA TIMUR	63.42	63.59	63.77	63.95	64.04	64.09	64.17	64.2	64.51	64.98	65.15	65.28	65.54
KALIMANTAN BARAT	67.15	67.35	67.56	67.76	67.86	67.97	67.99	68.0	68.26	68.67	68.8	68.87	69.08
KALIMANTAN TENGAH	67.16	67.25	67.35	67.41	67.52	67.67	67.68	67.69	67.74	67.79	67.84	67.89	68.08
KALIMANTAN SELATAN	64.76	64.99	65.23	65.46	65.57	65.9	65.92	65.98	66.16	66.55	66.76	66.97	67.23
KALIMANTAN TIMUR	71.03	71.25	71.46	71.68	71.79	71.82	71.85	71.87	72.13	72.41	72.54	72.79	72.8
KALIMANTAN UTARA	68.0	68.13	68.25	70.06	70.17	70.47	70.49	70.5	70.52	70.57	70.63	70.69	70.7
SULAWESI UTARA	68.51	68.65	68.8	68.95	69.03	69.08	69.09	69.11	69.33	69.71	69.82	69.89	70.16
SULAWESI TENGAH	64.17	64.48	64.8	65.11	65.27	65.34	65.39	65.4	65.86	66.32	66.77	66.91	66.98
SULAWESI SELATAN	67.01	67.2	67.4	67.59	67.69	67.89	67.94	67.96	68.2	68.54	68.68	68.77	69.03
SULAWESI TENGGARA	67.72	67.93	68.14	68.35	68.45	68.5	68.54	68.57	68.83	69.2	69.31	69.35	69.44
GORONTALO	64.51	64.68	64.84	65.0	65.08	65.2	65.22	65.22	65.51	66.01	66.14	66.27	66.56
SULAWESI BARAT	60.64	60.92	61.18	61.46	62.18	62.36	62.49	62.52	62.76	62.96	63.2	63.39	63.74
MALUKU	62.59	62.74	62.9	63.04	63.11	63.41	63.44	63.49	63.68	63.95	64.12	64.23	64.53
MALUKU UTARA	64.8	64.97	65.12	65.31	65.41	65.5	65.53	65.56	65.82	66.25	66.43	66.55	66.84
PAPUA BARAT	62.7	62.88	63.0	63.17	63.26	63.31	63.32	63.34	63.57	64.01	64.2	64.33	64.54
PAPUA	62.51	62.66	62.81	62.97	63.05	63.3	63.44	63.45	63.67	63.87	64.02	64.15	64.34
INDONESIA	67.89	68.09	68.29	68.49	68.87	68.93	69.09	69.16	69.3	69.44	69.59	69.67	69.93



- ACEH (Predicted)
- SUMATERA UTARA (Predicted)
- SUMATERA BARAT (Predicted)
- RIAU (Predicted)
- JAMBI (Predicted)
- SUMATERA SELATAN (Predicted)
- BENGKULU (Predicted)
- LAMPUNG (Predicted)
- KEP. BANGKA BELITUNG (Predicted)
- KEP. RIAU (Predicted)
- DKI JAKARTA (Predicted)
- JAWA BARAT (Predicted)
- JAWA TENGAH (Predicted)
- DI YOGYAKARTA (Predicted)
- JAWA TIMUR (Predicted)
- BANTEN (Predicted)
- BALI (Predicted)
- NUSA TENGGARA BARAT (Predicted)
- NUSA TENGGARA TIMUR (Predicted)
- KALIMANTAN BARAT (Predicted)
- KALIMANTAN TENGAH (Predicted)
- KALIMANTAN SELATAN (Predicted)
- KALIMANTAN TIMUR (Predicted)
- KALIMANTAN UTARA (Predicted)
- SULAWESI UTARA (Predicted)
- SULAWESI TENGAH (Predicted)
- SULAWESI SELATAN (Predicted)
- SULAWESI TENGGARA (Predicted)
- GORONTALO (Predicted)
- SULAWESI BARAT (Predicted)
- MALUKU (Predicted)
- MALUKU UTARA (Predicted)
- PAPUA BARAT (Predicted)
- PAPUA (Predicted)
- INDONESIA (Predicted)

B. Harapan Hidup Perempuan Antar Provinsi Indonesia



Mean Squared Error (MSE) on test data = 0.08363924471400459

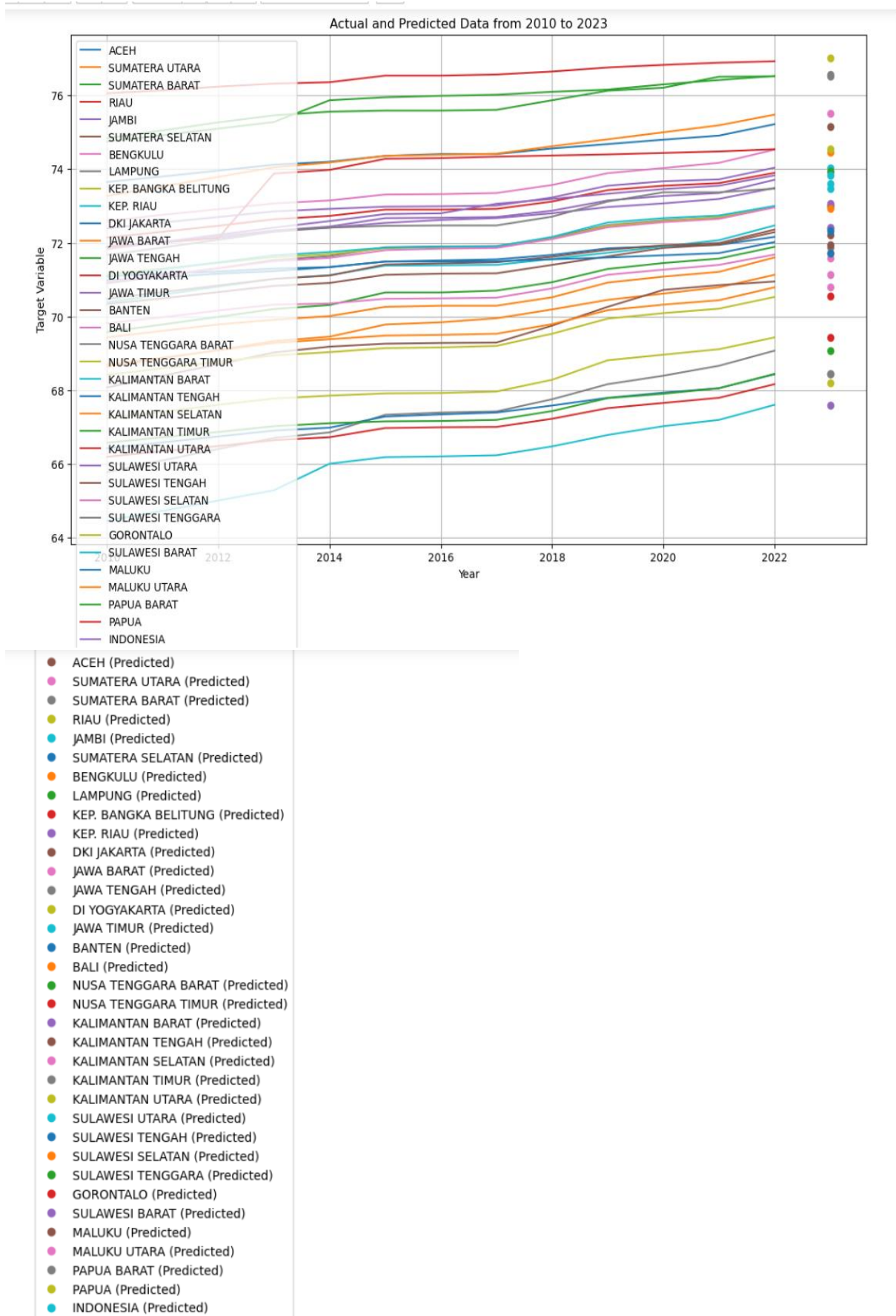
Mean Absolute Error (MAE) on test data = 0.15645140794526782

R-squared (R2) on test data = 0.9385611358843929

Root Mean Squared Error (RMSE) on test data = 0.2892045032740752

Province	features	2022	prediction
BANTEN	[70.48,70.66,70.8...	72.36	72.32002638551693
GORONTALO	[68.41,68.59,68.7...	70.53	70.54570219053427
KALIMANTAN BARAT	[71.07,71.28,71.4...	73.0	73.07031660265739
KALIMANTAN SELATAN	[68.63,68.86,69.0...	71.13	71.13049334044364
KEP. RIAU	[70.39,70.59,70.8...	72.47	72.40435755307357
NUSA TENGGARA TIMUR	[67.24,67.41,67.6...	69.43	69.43886920782968
SULAWESI TENGAH	[68.07,68.39,68.7...	70.95	71.72415292186267
SULAWESI TENGGARA	[71.67,71.87,72.0...	73.47	73.93119036660678
SUMATERA SELATAN	[70.34,70.49,70.6...	72.29	72.40404781608193
SUMATERA UTARA	[69.43,69.61,69.7...	71.6	71.58587442797317

province	prediction	2023
ACEH	72.21713765181346	72.21713765181346
SUMATERA UTARA	71.58587442797317	71.58587442797317
SUMATERA BARAT	71.88414785186544	71.88414785186544
RIAU	73.89863104376617	73.89863104376617
JAMBI	73.4625898471867	73.4625898471867
SUMATERA SELATAN	72.40404781608193	72.40404781608193
BENGKULU	71.7235511373566	71.7235511373566
LAMPUNG	72.96574189735713	72.96574189735713
KEP. BANGKA BELITUNG	73.0131639803405	73.0131639803405
KEP. RIAU	72.40435755307357	72.40435755307357
DKI JAKARTA	75.15179066719489	75.15179066719489
JAWA BARAT	75.51658767408206	75.51658767408206
JAWA TENGAH	76.56267351446206	76.56267351446206
DI YOGYAKARTA	77.02294750939176	77.02294750939176
JAWA TIMUR	73.61269062810923	73.61269062810923
BANTEN	72.32002638551693	72.32002638551693
BALI	74.46887321930217	74.46887321930217
NUSA TENGGARA BARAT	69.07702280470913	69.07702280470913
NUSA TENGGARA TIMUR	69.43886920782968	69.43886920782968
KALIMANTAN BARAT	73.07031660265739	73.07031660265739
KALIMANTAN TENGAH	71.95449124537842	71.95449124537842
KALIMANTAN SELATAN	71.13049334044364	71.13049334044364
KALIMANTAN TIMUR	76.51484154749944	76.51484154749944
KALIMANTAN UTARA	74.54424344562405	74.54424344562405
SULAWESI UTARA	74.04246014148288	74.04246014148288
SULAWESI TENGAH	71.72415292186267	71.72415292186267
SULAWESI SELATAN	72.94106446252783	72.94106446252783
SULAWESI TENGGARA	73.93119036660678	73.93119036660678
GORONTALO	70.54570219053427	70.54570219053427
SULAWESI BARAT	67.59316231437539	67.59316231437539
MALUKU	68.45170253063355	68.45170253063355
MALUKU UTARA	70.79377303174961	70.79377303174961
PAPUA BARAT	68.43248414251619	68.43248414251619
PAPUA	68.19491824325716	68.19491824325716
INDONESIA	73.81930946802193	73.81930946802193



BAB V

KESIMPULAN

Melalui penelitian ini, kami dapat menyimpulkan bahwa harapan hidup lelaki dan perempuan di Indonesia menunjukkan variasi yang signifikan antar provinsi. Distribusi harapan hidup dipengaruhi oleh sejumlah faktor, termasuk aksesibilitas layanan kesehatan, pendidikan, dan kondisi ekonomi di setiap wilayah. Analisis Big Data membantu mengidentifikasi ketidaksetaraan gender dalam harapan hidup, menyoroti pentingnya mengatasi disparitas hak dan akses bagi semua individu.

Pola perubahan harapan hidup dari waktu ke waktu memberikan pemahaman yang lebih baik tentang tren kesejahteraan masyarakat. Provinsi-provinsi tertentu menunjukkan peningkatan atau penurunan yang perlu mendapatkan perhatian khusus dalam perumusan kebijakan pembangunan.

Hasil penelitian ini memberikan wawasan penting bagi pembuat kebijakan dalam merancang strategi pembangunan yang lebih inklusif dan berkelanjutan. Implikasi temuan dapat membantu meningkatkan efektivitas kebijakan kesehatan, pendidikan, dan pembangunan ekonomi di berbagai wilayah Indonesia.

Namun, penting untuk diingat bahwa penelitian ini memiliki keterbatasan, seperti keterbatasan data dan asumsi yang digunakan. Oleh karena itu, perlu dilakukan penelitian lebih lanjut untuk memperdalam pemahaman tentang faktor-faktor yang mempengaruhi harapan hidup.

Dengan demikian, kesimpulan penelitian ini menggarisbawahi pentingnya kerjasama lintas sektor dan tingkat untuk mencapai tujuan pembangunan berkelanjutan, dengan fokus pada peningkatan kesejahteraan masyarakat dan pengurangan ketidaksetaraan di seluruh Indonesia.

DAFTAR PUSTAKA

Link Git Hub : <https://github.com/lkramRamadhan08/UASBigDataKelompok3.git>

Link Data Set : <https://www.bps.go.id/id>

Referensi : <https://chat.openai.com/>

PENUTUP

Penelitian ini memberikan kontribusi signifikan dalam pemahaman perbedaan harapan hidup lelaki dan perempuan antar provinsi di Indonesia. Dengan menggali data melalui pendekatan Big Data, kami berhasil mengidentifikasi pola distribusi, ketidaksetaraan gender, dan faktor-faktor krusial yang mempengaruhi harapan hidup di berbagai wilayah.

Analisis ini mengonfirmasi bahwa harapan hidup tidak hanya mencerminkan kondisi kesehatan, tetapi juga dipengaruhi oleh faktor-faktor seperti pendidikan, aksesibilitas layanan kesehatan, dan kondisi ekonomi. Temuan ketidaksetaraan gender menjadi sorotan utama, memperlihatkan bahwa perlu upaya lebih lanjut untuk memastikan pemerataan hak dan akses bagi semua individu.

Selain itu, pemahaman tentang perubahan harapan hidup dari waktu ke waktu dan provinsi-provinsi yang memerlukan perhatian khusus memberikan landasan bagi formulasi kebijakan yang lebih terarah. Implikasi hasil penelitian ini dapat memberikan panduan bagi pembuat kebijakan dalam merancang strategi pembangunan yang lebih inklusif dan berkelanjutan.

Kami menyadari bahwa penelitian ini memiliki batasan tertentu, seperti keterbatasan data dan asumsi yang digunakan dalam analisis. Oleh karena itu, kami mendorong penelitian lebih lanjut untuk menyempurnakan dan memperluas pemahaman terkait harapan hidup di Indonesia.

Dengan harapan bahwa temuan ini dapat memberikan kontribusi positif bagi perumusan kebijakan yang lebih efektif, serta memberikan inspirasi bagi penelitian selanjutnya dalam memahami dinamika kesejahteraan masyarakat Indonesia. Terima kasih atas dukungan dan partisipasi dari semua pihak yang telah terlibat dalam penelitian ini.