

# Estimasi Pertumbuhan Penduduk Jawa Barat Menggunakan Metode Regresi Linear Berganda

Opiana<sup>a1</sup>, Nana Suarna<sup>a2</sup>, Willy Prihartono<sup>b3</sup>

<sup>a</sup>Program Studi Teknik Informatika, STMIK IKMI Cirebon  
Jl. Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon Jawa Barat  
<sup>1</sup>vanopiana@gmail.com  
<sup>2</sup>st\_nana@yahoo.com

<sup>b</sup>Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon  
Jl. Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon Jawa Barat  
<sup>3</sup>willyprihartono@gmail.com

## Abstract

Pertumbuhan penduduk merupakan besaran persentase perubahan jumlah penduduk di suatu wilayah dengan waktu tertentu yang dibandingkan dengan jumlah penduduk pada waktu sebelumnya. Permasalahan yang terjadi saat ini adalah pertumbuhan penduduk yang kian meningkat dari waktu ke waktu, sehingga lembaga sensus sering mengalami kesulitan dalam mengestimasi laju pertumbuhan penduduk tiap tahunnya. Penelitian ini bertujuan untuk mengetahui penerapan data mining dengan metode regresi linear berganda untuk mengestimasi laju pertumbuhan penduduk di Provinsi Jawa Barat. Tahapan yang dilakukan menggunakan proses *KDD (Knowledge Discovery in Database)* yaitu dengan melakukan seleksi data, pra-Proses data, transformasi data, data mining, dan yang terakhir interpretasi dan evaluasi yang akan menghasilkan *output* berupa pengetahuan baru yang dapat memberikan kontribusi yang lebih baik. Hasil prediksi pertumbuhan penduduk di Jawa Barat pada tahun 2023 sebanyak 50.074 juta jiwa atau naik 1.49% dari tahun sebelumnya, pada tahun 2024 sebanyak 50.811 juta jiwa memiliki kenaikan sebesar 1.47% dari tahun sebelumnya. Maka penggunaan metode regresi linear berganda dapat dijadikan referensi dalam estimasi pertumbuhan penduduk di Jawa Barat.

**Keywords:** Pertumbuhan Penduduk, Data Mining, Regresi Linear Berganda

## 1. Introduction

Penduduk ialah sekumpulan warga yang tinggal di wilayah tertentu untuk menetap dengan kebutuhan yang berlaku[1]. Salah satu indikator penting untuk mengukur perkembangan suatu wilayah adalah dari pertumbuhan penduduknya. Pertumbuhan penduduk merupakan besaran persentase jumlah penduduk di suatu wilayah dengan waktu tertentu yang dibandingkan dengan jumlah penduduk pada waktu sebelumnya[2]. Permasalahan yang selama ini terjadi adalah pertumbuhan penduduk yang kian meningkat dari waktu ke waktu[3]. Sehingga, lembaga sensus sering mengalami kendala dalam mengestimasi laju pertumbuhan penduduk tiap tahunnya.

Penelitian sebelumnya yang dilakukan Putri Ardiyanti berjudul "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Denpasar Menggunakan Metode Regresi Linear Berganda" dalam mengaplikasikan metode regresi linear berganda dalam mengestimasi laju pertumbuhan penduduk berhasil menerapkan metode regresi linear berganda untuk memprediksi laju pertumbuhan penduduk di Kota Denpasar dengan mendapatkan hasil bahwa penduduk Denpasar akan mengalami peningkatan pertumbuhan penduduk sebesar 7,5 % pada tahun 2023 dan 12,25% pada tahun 2024[4].

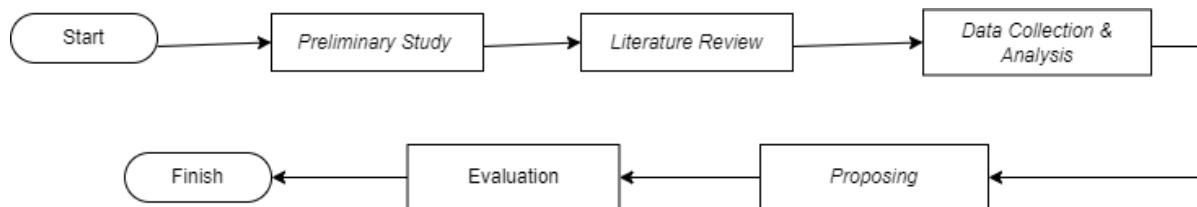
Penelitian kedua yang dilakukan Purwadi dalam penelitian yang berjudul "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linear Berganda Pada BPS Deli Serdang" berhasil mengestimasi pertumbuhan penduduk pada Kabupaten Deli Serdang periode tahun 2018 telah terjadi penambahan penduduk sebanyak 66.243 jiwa dari tahun sebelumnya, dengan menggunakan metode regresi linear berganda ditemukan pola yang berkaitan erat antara atribut jumlah laki-laki dan jumlah perempuan terhadap laju pertumbuhan penduduk[2].

Berdasarkan penelitian yang telah dilakukan oleh peneliti sebelumnya yang telah berhasil menerapkan data mining dengan metode regresi linear berganda di daerah lain. Maka, Penelitian ini akan mengestimasi laju Pertumbuhan penduduk di Provinsi Jawa Barat untuk melihat tingkat akurasi dan validitas estimasi laju pertumbuhan penduduk menggunakan pendekatan data mining dengan metode regresi linear berganda. Tujuan melakukan penelitian ini mengetahui penerapan data mining dengan menggunakan metode regresi linear berganda untuk menghasilkan estimasi laju pertumbuhan penduduk di Provinsi Jawa Barat dengan akurat. Selain itu, penelitian ini memiliki tujuan untuk memberikan kontribusi dalam pengembangan model prediksi yang lebih akurat dalam estimasi laju pertumbuhan penduduk.

Metode yang digunakan dalam penelitian ini adalah metode regresi linear berganda. Regresi linear merupakan salah satu cara prediksi menggunakan garis lurus untuk menggambarkan hubungan diantara dua variabel atau lebih[2]. Penggunaan data mining dengan metode regresi linear berganda dapat membantu dalam mengidentifikasi hubungan antar berbagai faktor yang mempengaruhi pertumbuhan penduduk. Tahapan yang dilakukan menggunakan proses KDD (Knowledge Discovery in Database) yaitu dengan melakukan seleksi data, pra-proses data, transformasi data, data mining, dan yang terakhir interpretasi serta evaluasi yang akan menghasilkan output berupa pengetahuan baru yang dapat memberikan kontribusi lebih baik[5].

## 2. Research Methods

Penelitian ini menggunakan metode kuantitatif. Metode penelitian kuantitatif digunakan untuk mengumpulkan data secara sistematis dan mengukur hubungan antar variabel. Pendekatan kuantitatif dianggap efektif dalam memberikan gambaran yang objektif dan dapat diukur terhadap laju pertumbuhan penduduk. Tahapan yang dilakukan selama penelitian ditampilkan dalam *figure 1*.



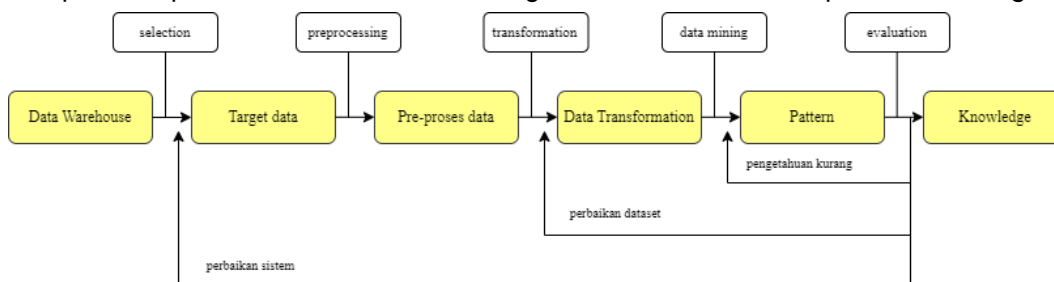
**Figure 1.** Tahapan Metode penelitian

### 2.1. Pengumpulan Data

Penelitian ini menggunakan data sekunder, data sekunder merupakan informasi atau data yang telah dikumpulkan, diolah atau dipublikasikan oleh pihak lain atau lembaga tertentu sebelumnya dengan tujuan yang berbeda dan kemudian digunakan kembali dalam penelitian atau analisis terbaru. Adapun data ini diperoleh dari data publik yang bersumber dari Open Data Jabar <https://opendata.jabarprov.go.id/id/dataset/jumlah-penduduk-berdasarkan-jenis-kelamin-dan-kabupatenkota-di-jawa-barat>.

### 2.2. Teknik Analisis Data

Tahapan analisis data dilakukan dengan metode Knowledge Discovery in Databases (KDD) melibatkan serangkaian langkah sistematis untuk mengekstrak pengetahuan dari data. Berikut tahapan-tahapan dalam analisis data dengan metode KDD ditampilkan dalam *figure 2*.



**Figure 2.** Tahapan Proses KDD

#### a. Selection

Seleksi data melibatkan pemilihan dataset dan variabel-variabel yang akan menjadi fokus untuk di analisis. Pemilihan data yang relevan bertujuan untuk memastikan bahwa dataset yang

dipilih tidak hanya sesuai kebutuhan penelitian, tapi juga memiliki potensi untuk menghasilkan informasi berharga.

b. Preprocessing

Tahapan pra pemrosesan data melibatkan rangkaian kegiatan yang bertujuan untuk membersihkan dan merapikan data. Proses ini mencakup penanganan nilai yang hilang, deteksi outlier, normalisasi data, dan kegiatan pra pemrosesan data lainnya. Tujuan utama dari tahapan ini adalah untuk mempersiapkan data dengan cermat dan teliti agar memastikan kualitas dan kebersihan data yang memadai untuk analisis lebih lanjut. Dengan mengatasi nilai yang hilang, mendeteksi anomali, dan normalisasi data, tahapan pra pemrosesan ini bertujuan untuk mengoptimalkan integritas data, sehingga hasil analisis yang dihasilkan dapat diandalkan dan memberikan wawasan yang mendalam.

c. Transformation

Pengubahan data menjadi bentuk yang sesuai atau bermanfaat untuk analisis merupakan suatu proses yang melibatkan sejumlah tindakan, seperti pemfilteran variabel, pembentukan variabel baru, atau konversi format data. Proses ini bertujuan untuk mengoptimalkan data agar dapat diandalkan dalam analisis lebih lanjut.

d. Data Mining

Metode regresi linear berganda digunakan untuk menguji pengaruh lebih dari satu variabel independen terhadap variabel dependen. Bentuk persamaan untuk regresi linear berganda didefinisikan dengan rumus berikut[6]:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Keterangan:

Y = Variabel tak bebas (nilai variabel yang akan diprediksi)

a = Konstanta

$b_1, b_2, \dots, b_n$  = Nilai koefisien regresi

$X_1, X_2, \dots, X_n$  = Variabel bebas

Penelitian ini menggunakan 2 variabel bebas, yaitu  $X_1$  dan  $X_2$ , maka bentuk persamaan regresinya adalah persamaan (2):

$$Y = a + b_1X_1 + b_2X_2 \quad (2)$$

Dalam penelitian ini jumlah penduduk didefinisikan sebagai Y atau variabel dependen, penduduk laki-laki dan penduduk di definisikan sebagai  $X_1$  dan  $X_2$  yaitu sebagai variabel independen. Sehingga apabila dirumuskan maka menjadi persamaan (3):

$$\text{Jumlah Penduduk} = a + b_1 \cdot \text{Penduduk laki - laki} + b_2 \cdot \text{Penduduk Perempuan} \quad (3)$$

e. Evaluation

Proses evaluasi, sebagai tahap integral dalam KDD, mencakup penilaian mendalam terhadap kualitas hasil dari data mining yang telah dilakukan. Fokus utama evaluasi adalah memastikan bahwa model atau pola yang diidentifikasi melalui data mining dapat dianggap sebagai hasil yang dapat diandalkan dan dapat diimplementasikan secara praktis dalam konteks aplikatif. Selain itu, evaluasi ini juga bertujuan untuk memverifikasi bahwa temuan yang dihasilkan sesuai dengan tujuan penelitian yang telah ditetapkan sebelumnya. Oleh sebab itu, untuk mengevaluasi model regresi linear dalam estimasi laju pertumbuhan penduduk digunakan evaluasi metrik *Root Mean Squared Error (RMSE)* dan *R-Squared (R2)*.

RMSE (Root Mean Square Error) adalah akar kuadrat dari Mean Squared Error (MSE) yang dihasilkan. RMSE digunakan karena untuk mengukur seberapa baik model dapat memprediksi nilai respons jumlah penduduk (variabel dependen) berdasarkan Jumlah penduduk laki-laki dan jumlah penduduk perempuan (variabel independen). RMSE mengukur seberapa dekat prediksi model dengan nilai yang sebenarnya[7]. Semakin kecil nilai yang dihasilkan semakin bagus pula hasil yang dihasilkan. Secara sederhana, RMSE merupakan metode untuk menghitung bias dalam model peramalan[8].

Adapun rumus RMSE yang digunakan dirampilkkan pada persamaan (4):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - Y'_t)^2}{n}} \quad (4)$$

Keterangan:

- n = Jumlah Sampel Dalam Data
- Y<sub>t</sub> = Nilai Aktual
- Y'<sub>t</sub> = Nilai Prediksi

### 3. Result and Discussion

Hasil penelitian yang dilakukan dalam pembahasan ini akan menguraikan proses bagaimana estimasi pertumbuhan penduduk di Jawa Barat dengan menggunakan regresi linear berganda.

#### 3.1. Data

Data yang digunakan dalam penelitian ini adalah dataset sebanyak 540 record dan terdiri 9 atribut, dengan rentang waktu data dari tahun 2013 sampai dengan 2022. Dataset bersumber dari Repository Open Data Jabar dengan situs <https://opendata.jabarprov.go.id>. Hasil penelusuran didapatkan data jumlah penduduk berdasarkan kabupaten kota dalam bentuk dokumen soft file format Microsoft Excel. Seperti disajikan pada tampilan tabel 1 berikut ini: (Dataset hanya disajikan sebanyak 10 record, dari dataset keseluruhan yang berjumlah sebanyak 540 record).

**Table 1.** Data Jumlah Penduduk Jawa Barat

	kode_ provinsi	nama_ provinsi	kode_ kabupaten_kota	nama_ kabupaten_kota	jenis_kelamin	jumlah_ penduduk	satuan	Tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	LAKI-LAKI	1930902	JIWA	2013
2	32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUAN	1826962	JIWA	2013
3	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	LAKI-LAKI	1258939	JIWA	2013
4	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	PEREMPUAN	1171101	JIWA	2013
5	32	JAWA BARAT	3203	KABUPATEN CIANJUR	LAKI-LAKI	1154944	JIWA	2013
...	...	...	...	...	...	...	...	...
536	32	JAWA BARAT	3277	KOTA CIMAH	PEREMPUAN	281882	JIWA	2022
537	32	JAWA BARAT	3278	KOTA TASIKMALAYA	LAKI-LAKI	379050	JIWA	2022
538	32	JAWA BARAT	3278	KOTA TASIKMALAYA	PEREMPUAN	367660	JIWA	2022
539	32	JAWA BARAT	3279	KOTA BANJAR	LAKI-LAKI	104346	JIWA	2022
540	32	JAWA BARAT	3279	KOTA BANJAR	PEREMPUAN	102884	JIWA	

#### 3.2. Data Selection

Data yang dipilih pada penelitian ini adalah data jumlah penduduk Jawa Barat dari tahun 2013 sampai 2022. Data ini merupakan data publik, dan memiliki 540 record dan 9 Atribut. Hasil selection dari dataset yang tidak digunakan adalah 5 atribut sehingga hanya akan menggunakan 4 atribut yaitu nama kabupaten kota, jenis kelamin, jumlah penduduk dan tahun. Karena 5 atribut yang tidak digunakan memiliki record data yang sama pada masing-masing atributnya. Atribut sebelum di seleksi ditampilkan pada tabel 2.

**Table 2.** Sebelum Seleksi Atribut

No	Nama Atribut	Data Type
1	Id	integer
2	kode_provinsi	integer
3	nama_provinsi	object
4	kode_kabupaten_kota	integer
5	nama_kabupaten_kota	object
6	jenis_kelamin	object
7	jumlah_penduduk	integer
8	Satuan	object
9	Tahun	integer

Tabel 2 merupakan tabel yang berisi nama atribut beserta tipe datanya sebelum diseleksi. Kemudian atribut yang digunakan untuk penelitian ini setelah hasil diseleksi akan menggunakan 4 atribut seperti ditampilkan pada tabel 3 berikut.

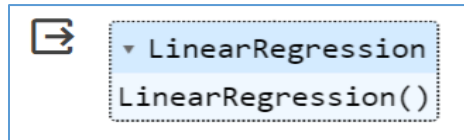
**Table 3.** Hasil Setelah Seleksi Atribut

no	Nama Atribut	Data Type
1	Nama_kabupaten_kota	object
2	Jenis_kelamin	object
3	Jumlah_penduduk	integer
4	tahun	integer

### 3.3. Data Mining

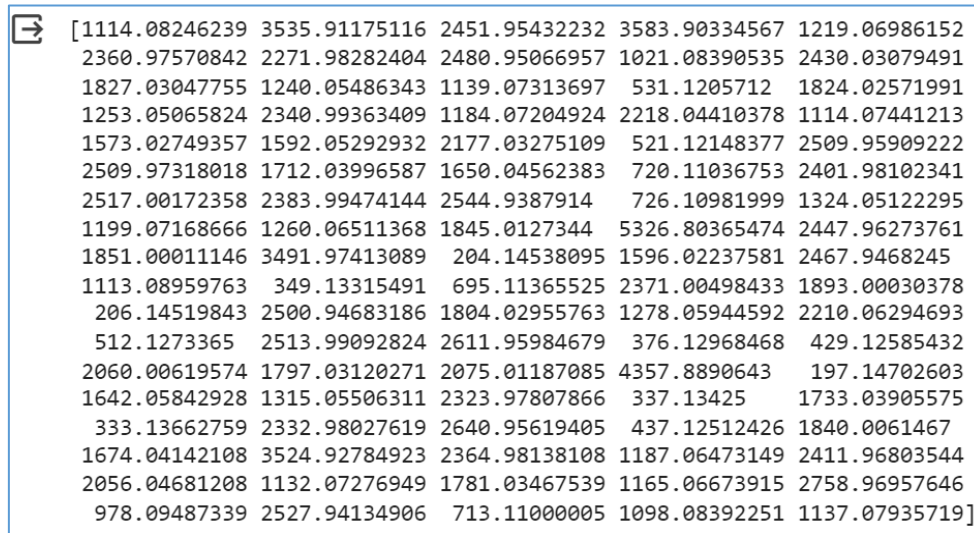
Pada tahapan data mining dilakukan proses perhitungan dengan menggunakan model regresi linear berganda. Model regresi dibuat menggunakan model *linear regression* dari *scikit-learn model*. Model ini dilatih dan diuji menggunakan data training sebesar 70% dan data testing sebesar 30%. Model juga di evaluasi menggunakan evaluasi metrik *root mean squared error (rmse)*, dan *r-squared (r2)*.

Tahap pertama dilakukan pendefinisian atribut penduduk laki-laki dan atribut penduduk perempuan sebagai nilai x (variabel independen) dan jumlah penduduk didefinisikan sebagai nilai y (variabel dependen) untuk digunakan dalam analisis atau pemodelan regresi linear berganda. Setelah variabel didefinisikan selanjutnya data dibagi menjadi 2 bagian yaitu *training set* dan *test set*. Dengan menggunakan fungsi *'train\_test\_split'* dari *scikit learn*. Setelah dilakukan *split data* selanjutnya proses pembuatan model regresi linear dengan menggunakan *library scikit-learn* dan melatihnya dengan data *training*. Model yang telah dibuat digunakan untuk membuat prediksi pada data baru dan untuk mengevaluasi kinerja model pada *data test*.



**Figure 3.** Model Regresi Linear

Pada *Figure 3* menunjukkan bahwa model regresi linear berhasil dibuat. Setelah model berhasil dibuat selanjutnya dilakukan prediksi dengan menggunakan model regresi linear yang telah dilatih pada *data x\_test*. Hasil prediksi nantinya digunakan untuk dibandingkan dengan nilai aktual untuk mengevaluasi kinerja model. Hasil prediksi model ditampilkan pada *figure 4*.



**Figure 4.** Hasil Prediksi Model Regresi Linear

*Figure 4* merupakan hasil prediksi menggunakan model regresi dari data testing (*x\_test*) yang merupakan data latih dari atribut penduduk laki-laki dan penduduk perempuan. Hasil ini nantinya akan dibandingkan dengan nilai data aktual untuk mengevaluasi kinerja model.

### 3.4. Evaluasi

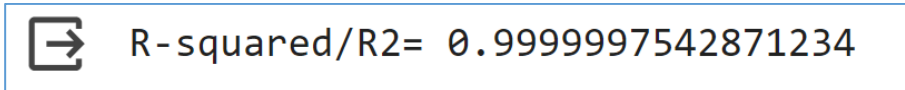
*Root Mean Squared Error (RMSE)* adalah metrik yang digunakan untuk mengukur seberapa baik model regresi dapat memprediksi nilai target jumlah penduduk pada dataset pengujian. RMSE dihitung sebagai akar kuadrat dari Mean Squared Error (MSE). Nilai RMSE yang didapatkan sekitar 0.4685 dalam satuan yang sama dengan variabel dependen menunjukkan bahwa sebaran kesalahan relatif kecil. Hasil nilai RMSE adalah 0.4685329901519414. Semakin kecil nilai RMSE, semakin baik kinerja model regresi. Oleh karena itu, nilai 0.4685329901519414 menunjukkan bahwa model regresi memiliki tingkat kesalahan yang relatif rendah dalam memprediksi nilai target pada dataset pengujian. Disimpulkan bahwa model memiliki kemampuan yang baik dalam melakukan prediksi pada data yang tidak terlihat sebelumnya. Ditampilkan pada *figure 5*.



**Figure 5.** Hasil RMSE

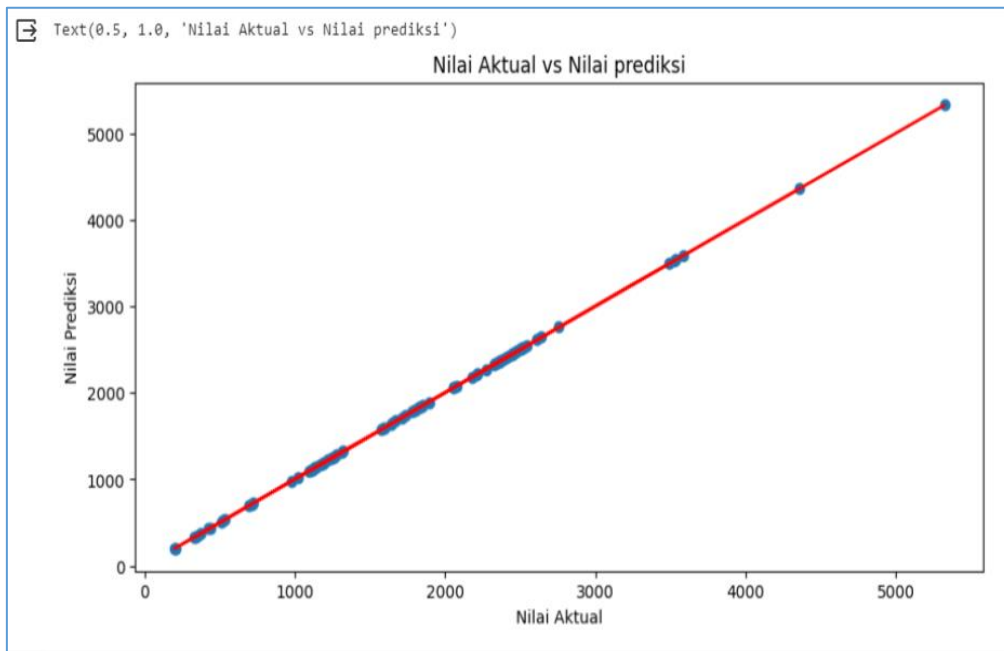
*R-squared (R<sup>2</sup>)* dikenal sebagai koefisien determinasi, adalah metrik evaluasi yang digunakan untuk mengevaluasi seberapa baik model regresi linier cocok dengan data yang diamati. *R-squared* memberikan informasi tentang seberapa baik variabilitas dalam variabel dependen dapat dijelaskan oleh model regresi. Evaluasi model menggunakan *R-squared* memiliki rentang Antara 0 hingga 1, dimana 1 menunjukkan model yang sempurna yang mampu menjelaskan seluruh variabilitas data dan 0 menunjukkan model yang tidak dapat menunjukkan variabilitas sama sekali. Hasil yang didapatkan menunjukkan nilai sekitar 0.99999 menunjukkan bahwa model mampu menjelaskan sebagian besar

variabilitas dalam data jumlah penduduk berdasarkan variabel independen yang digunakan yaitu penduduk laki-laki dan penduduk perempuan. Model regresi linear sangat cocok dengan data dan dapat menjelaskan sebagian besar variabilitas dalam jumlah penduduk. Ditampilkan pada *figure 6*.



**Figure 6.** Hasil R-Squared

Dalam diagram terdapat kemiripan yang sangat tinggi antara nilai aktual dan nilai prediksi, dapat dilihat bahwa sebagian besar titik-titik mendekati garis lurus 45 derajat yang merupakan garis regresi. Garis ini menunjukkan relasi linier antara nilai aktual dan nilai prediksi. Sebaran titik-titik yang rapat menunjukkan bahwa model dengan baik memprediksi nilai-nilai dalam data test. ditampilkan pada *figure 7*.



**Figure 7.** Visualisasi Regresi

Untuk melihat selisih antara nilai aktual dan nilai prediksi dibuat satu dataframe baru untuk memberikan gambaran lebih rinci tentang bagaimana model melakukan prediksi pada data test. Proses ditampilkan pada *figure 8*.

	nilai aktual	nilai prediksi	selisih
64	1114.0	1114.574542	0.574542
135	3536.0	3535.937879	-0.062121
153	2452.0	2452.191901	0.191901
189	3583.0	3582.820903	-0.179097
253	1218.0	1218.206907	0.206907
...	...	...	...
21	978.0	978.063250	0.063250
234	2528.0	2527.803497	-0.196503
161	714.0	713.665385	-0.334615
16	1098.0	1098.512846	0.512846
118	1137.0	1137.573014	0.573014

90 rows x 3 columns

**Figure 8.** Hasil Selisish Nilai Aktual Dan Prediksi

Pada Figure 6 dapat dilihat bahwa atribut "nilai aktual" berisi nilai aktual dari variabel dependen ( $y_{test}$ ), atribut "nilai prediksi" berisi nilai yang diprediksi oleh model regresi ( $y_{pred}$ ) dan atribut "selisih" merupakan perbedaan antara nilai prediksi dan nilai aktual ( $\text{nilai prediksi} - \text{nilai aktual}$ ). Dalam gambar terlihat adanya nilai selisih positif yang berarti model memprediksi nilai yang lebih tinggi daripada nilai aktual dan terdapat nilai selisih negatif yang berarti model memprediksi nilai yang lebih rendah daripada nilai aktual. Dari hasil selisih yang didapat terlihat bahwa selisih dari nilai prediksi dan nilai aktual memiliki nilai yang sangat kecil sehingga dapat disimpulkan bahwa model yang dibuat cenderung memberikan prediksi yang baik.

### 3.5. Prediksi 5 Tahun Berikutnya

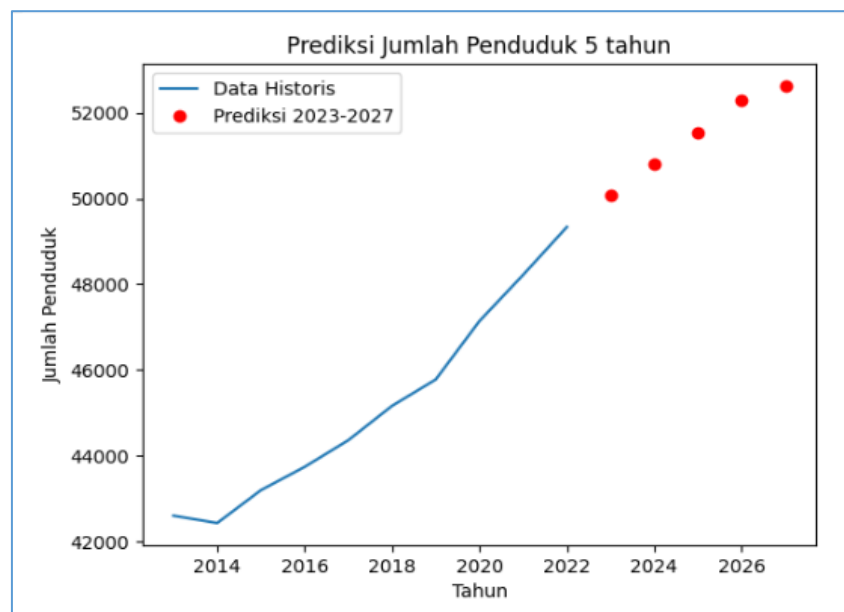
Pada hasil menunjukkan bahwa pada tahun 2023 jumlah penduduk diperkirakan sekitar 50.074 juta. Tahun 2024 berjumlah sekitar 50.811 juta. Pada tahun 2025 diperkirakan berjumlah 51.548 juta. Kemudian pada tahun 2026 sekitar 52.285 juta dan pada tahun 2027 jumlah penduduk sekitar 52.633 juta jiwa.. Ditampilkan pada *figure 9*.

	Tahun	Penduduk_Laki-laki	Penduduk_Perempuan	Jumlah_Penduduk_Prediksi
0	2023	25334.0	24744.0	50074.0
1	2024	25681.0	25134.0	50811.0
2	2025	26029.0	25524.0	51548.0
3	2026	26376.0	25914.0	52285.0
4	2027	26723.0	25914.0	52633.0

**Figure 9.** Hasil Prediksi

Untuk melihat bagaimana hasil prediksi jumlah berperilaku terhadap data historis maka dibuat tampilan visualisasi tren data historis. Ditampilkan pada *figure 10*.





**Figure 10.** Hasil Visualisasi Data Prediksi

Figure 8 menunjukkan bahwa garis biru pada plot (label 'Data Historis') mewakili data historis jumlah penduduk pada tahun tahun sebelumnya. Titik merah pada plot (label 'Prediksi 2023-2027') menunjukkan prediksi jumlah penduduk untuk tahun 2023-2027 berdasarkan hasil dari penerapan model regresi yang telah dibuat. Pada grafik titik merah menunjukkan bahwa model regresi memprediksi peningkatan jumlah penduduk dari tahun ke tahun dalam rentang waktu 2023-2027.

#### 4. Conclusion

Berdasarkan hasil penelitian menggunakan regresi linear berganda untuk memprediksi jumlah penduduk di provinsi di Provinsi Jawa Barat dapat disimpulkan bahwa:

Hasil proses prediksi menggunakan metode regresi linear berganda untuk memprediksi jumlah penduduk di Jawa Barat menggunakan model regresi dapat bekerja dengan baik. Model prediksi yang di evaluasi menggunakan berbagai evaluasi metrik menunjukkan bahwa model prediksi yang dibuat memiliki akurasi yang baik dalam memprediksi jumlah penduduk berdasarkan penduduk laki-laki dan penduduk perempuan.

Hasil akurasi model untuk prediksi laju pertumbuhan penduduk di Jawa Barat menunjukkan tingkat akurasi yang cukup baik. Hasil evaluasi metrik Root Mean Squared Error (RMSE) menunjukkan nilai 0.4685. Dari evaluasi metrik yang telah dilakukan menunjukkan bahwa hasil nilai yang didapatkan semuanya memiliki nilai yang kecil sehingga dapat disimpulkan bahwa model regresi bekerja dengan baik dan memiliki tingkat kesalahan prediksi atau error kecil sehingga dapat memprediksi jumlah penduduk dengan akurat. Selain itu hasil evaluasi menggunakan R-Squared menunjukkan nilai 0.9999, hasil yang didapat mendekati nilai 1 yang menunjukkan bahwa model regresi dapat membaca variabilitas data sehingga dapat mendukung kesimpulan bahwa model yang digunakan memang memiliki tingkat akurasi yang sangat baik.

Dari hasil penelitian didapatkan hasil prediksi jumlah penduduk untuk 5 (lima) tahun berikutnya yaitu pada tahun 2023 jumlah penduduk diperkirakan sekitar 50.074 juta jiwa atau naik 1.49% dari tahun sebelumnya yang hanya berjumlah 49.339 juta jiwa pada tahun 2022. Pada tahun berikutnya juga diprediksi terjadi kenaikan sebesar 1.47% pada tahun 2024 sehingga jumlah penduduk diperkirakan akan sebanyak 50.811 juta jiwa. Kemudian pada tahun 2025 diprediksi terjadi kenaikan sebesar 1.45% pada tahun ini jumlah penduduk diprediksi sebanyak 51.548 juta jiwa. Tahun 2026 juga diprediksi mengalami peningkatan jumlah penduduk sebesar 1.43% sehingga berjumlah 52.285 juta jiwa dan pada tahun 2027 diperkirakan akan mengalami sedikit kenaikan sekitar 0.66% sehingga pada tahun 2027 jumlah penduduk di Jawa Barat akan sebanyak 52.633 juta jiwa

## References

- [1] I. Indriani, D. Siregar, and A. P. Windarto, "Penerapan Metode Linear Regression dalam Mengestimasi Jumlah Penduduk," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1112, 2022, doi: 10.30865/jurikom.v9i4.4676.
- [2] P. Purwadi, P. S. Ramadhan, and N. Safitri, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 55, 2019, doi: 10.53513/jis.v18i1.104.
- [3] E. D. Sri Mulyani *et al.*, "Estimasi Pertumbuhan Penduduk Di Kabupaten Tasikmalaya Menggunakan Metode Regresi Linear Berganda," *Infosys (Information Syst. J.)*, vol. 6, no. 1, p. 1, 2021, doi: 10.22303/infosys.6.1.2021.1-11.
- [4] A. A. A. P. Ardyanti and A. Abdriando, "Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Denpasar Menggunakan Metode Regresi Linier Berganda," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 6, no. 1, pp. 37–44, 2023, doi: 10.30813/jbase.v6i1.4317.
- [5] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021, doi: 10.47065/jimat.v1i2.104.
- [6] I. B. M. Swarbawa, I. G. Arta Wibawa, and I. K. Gede Suhartana, "Prediksi Hasil Panen Padi Di Kabupaten Jembrana Dengan Metode Linear Regression," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 11, no. 3, p. 671, 2023, doi: 10.24843/jlk.2023.v11.i03.p24.
- [7] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 1, pp. 8–17, 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- [8] N. Litha and T. Hasanuddin, "Analisis Performa Metode Moving Average Model untuk Prediksi Jumlah Penderita Covid-19," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 87–95, 2020, [Online]. Available: <https://jurnal.yoctobrain.org/index.php/ijodas/article/view/19>.