



Penerapan Data Mining untuk Mengestimasi Laju Pertumbuhan Penduduk Kabupaten Serdang Bedagai Menggunakan Metode Regresi Linear Berganda

Azizah Wulan Lestari¹, Lisa Salsabila Sinaga², Juwita Samroiyah Pulungan³,
Ulifia Cahaya Hasibuan⁴, Ismail Husein⁵

¹⁻⁵Universitas Islam Negeri Sumatera Utara
E-mail: azizahwulanlestari12@gmail.com

Abstrak

Pertumbuhan penduduk merupakan indikator penting dalam perencanaan pembangunan daerah. Penelitian ini bertujuan untuk mengestimasi laju pertumbuhan penduduk Kabupaten Serdang Bedagai dengan menerapkan teknik data mining menggunakan metode Regresi Linear Berganda. Data yang digunakan diperoleh dari Badan Pusat Statistik (BPS) Kabupaten Serdang Bedagai selama periode 2019-2023, yang mencakup jumlah penduduk laki-laki dan perempuan. Melalui tahapan pengumpulan, pra proses, dan analisis data, diperoleh model regresi dengan persamaan $Y = 0 + 1,5558X_1 + 0,4367 X_2$, di mana Y adalah total penduduk, X₁ jumlah laki-laki, dan X₂ jumlah perempuan. Hasil penelitian menunjukkan estimasi jumlah penduduk pada tahun 2024 mencapai 683.056 jiwa, yang berarti terdapat peningkatan dibandingkan tahun sebelumnya. Model ini menunjukkan bahwa metode regresi linear berganda mampu memberikan hasil estimasi yang akurat dan dapat dijadikan acuan dalam perencanaan pembangunan berbasis data.

Kata kunci: Data Mining, Pertumbuhan Penduduk, Regresi Linear Berganda.

PENDAHULUAN

Pertumbuhan penduduk memegang peranan penting dalam perkembangan sosial dan ekonomi suatu daerah. Jika peningkatan jumlah penduduk tidak dikelola dengan baik, hal tersebut dapat menjadi hambatan besar bagi pembangunan karena berpotensi menyebabkan tingginya angka kemiskinan, rendahnya tingkat pendidikan, serta meningkatnya pengangguran, yang pada akhirnya berdampak buruk pada kualitas sumber daya manusia (Janan, Janan, & Mustakim, 2024).

Untuk mengendalikan laju pertumbuhan penduduk di Indonesia, pemerintah melakukan survei langsung ke masyarakat melalui Badan Pusat Statistik (BPS). BPS merupakan lembaga pemerintah nonkementerian yang berada di bawah koordinasi langsung presiden. Salah satu unit kerjanya adalah BPS Kabupaten Serdang Bedagai. Lembaga ini memiliki tugas dan fungsi penting dalam melakukan pendataan statistik, termasuk pengumpulan data jumlah penduduk (Ardyanti & Abdriando, 2023).

Namun, BPS Kabupaten Serdang Bedagai menghadapi berbagai kendala dalam melakukan estimasi laju pertumbuhan penduduk setiap tahun. Berdasarkan sejumlah penelitian, terdapat metode yang dapat digunakan untuk memprediksi pertumbuhan penduduk, yaitu teknik *Data Mining*. Data mining adalah proses menggali atau menemukan informasi baru dengan mengidentifikasi pola atau aturan tertentu dari kumpulan data berukuran besar, sehingga dapat digunakan untuk menyelesaikan permasalahan yang ada. Teknik ini mencakup berbagai metode analisis, salah satunya adalah regresi linear berganda (Purwadi, Ramadhan, & Safitri, 2019).

Penelitian ini dilakukan untuk mengetahui laju pertumbuhan penduduk Serdang Bedagai dengan menggunakan regresi linear berganda. Kemudian, data akan diolah untuk menghasilkan informasi yang dapat dianalisis lebih lanjut untuk mengetahui apakah laju pertumbuhan penduduk di Kabupaten Serdang Bedagai menunjukkan tren positif atau negatif.

Hasil dari analisis ini akan digunakan untuk merumuskan langkah-langkah strategis yang tepat di masa depan.

METODE PENELITIAN

Lokasi dan Sumber Data

Penelitian ini dilaksanakan di Badan Pusat Statistik Kabupaten Serdang Bedagai dengan memanfaatkan data sekunder yang diperoleh dari laporan tahunan BPS. Data mencakup jumlah penduduk laki-laki dan perempuan selama periode 2019-2023.

Teknik Analisis

Langkah-langkah analisis dilakukan sebagai berikut:

1. Pengumpulan Data: Mengambil data jumlah penduduk laki-laki dan perempuan dari BPS Serdang Bedagai tahun 2019-2023
2. Menentukan variabel bebas dan variabel tidak bebas.
Variabel bebas yaitu:
Jumlah Laki-laki (X_1)
Jumlah Perempuan (X_2)
Variabel tidak bebas yaitu:
Jumlah Penduduk (Y)
3. Mencari nilai persamaan regresi linear berganda.
 $Y = a + b_1.X_1 + b_2.X_2$
4. Menentukan nilai konstanta dan koefisien regresi.

HASIL DAN PEMBAHASAN

Hasil

Data penduduk Kabupaten Serdang Bedagai dari tahun 2019 hingga 2023 menunjukkan tren peningkatan jumlah penduduk setiap tahunnya. Data yang digunakan dalam penelitian ini mencakup dua variabel independen, yaitu jumlah penduduk laki-laki (X_1) dan jumlah penduduk perempuan (X_2), serta satu variabel dependen yaitu total jumlah penduduk (Y). Melalui proses analisis menggunakan metode Regresi Linear Berganda, diperoleh persamaan model sebagai berikut:

$$Y = 0 + 1,5558X_1 + 0,4367X_2$$

Hasil tersebut menunjukkan bahwa setiap penambahan 1.000 penduduk laki-laki akan meningkatkan total penduduk sebesar 1.555 jiwa, sedangkan setiap penambahan 1.000 penduduk perempuan akan menambah jumlah total penduduk sebesar 436 jiwa.

Koefisien yang lebih besar pada variabel X_1 menunjukkan bahwa laju pertumbuhan penduduk laki-laki memiliki pengaruh sedikit lebih tinggi terhadap total populasi dibandingkan penduduk perempuan dalam kurun waktu penelitian. Dengan memasukkan data tahun 2023, yaitu $X_1 = 343.862$ dan $X_2 = 339.056$, diperoleh estimasi:

$$Y = (1,5558 \times 343,862) + (0,4367 \times 339,056) = 683,056$$

Sehingga jumlah penduduk Kabupaten Serdang Bedagai yang diprediksi pada tahun 2024 adalah 683.056 jiwa.

HASIL DAN PEMBAHASAN

Hasil regresi menunjukkan hubungan linier yang kuat antara variabel jumlah penduduk laki-laki dan perempuan terhadap total penduduk. Nilai koefisien positif pada kedua variabel independen menandakan bahwa peningkatan salah satu variabel akan selalu diikuti oleh peningkatan jumlah total penduduk. Tidak ditemukan adanya hubungan negatif (berlawanan arah), yang berarti pertumbuhan kedua kelompok penduduk berjalan paralel.

Model regresi ini secara praktis dapat dimanfaatkan oleh Badan Pusat Statistik (BPS)

Tabel 1. Database Penduduk Kabupaten Serdang Bedagai

Kecamatan	Tahun				
	2019	2020	2021	2022	2023
Kotarih	8.216	9.169	9.267	9.384	9.628
Silinda	8.544	9.514	9.609	9.724	9.971
Bintang Bayu	10.864	12.511	12.679	12.874	13.245
Dolok Masihul	49.837	52.705	53.005	53.410	54.531
Serbajadi	20.071	21.759	21.921	22.127	22.631
Sipispis	32.473	33.826	33.949	34.139	34.786
Dolok Merawan	17.455	17.976	18.018	18.095	18.414
Tebing Tinggi	41.681	41.162	41.132	41.181	41.778
Tebing Syahbandar	33.124	33.585	33.626	33.732	34.286
Bandar Khalipah	25.478	25.857	25.889	25.972	26.400
Tanjung Beringin	38.011	42.142	42.568	43.081	44.179
Sei Rampah	65.660	71.366	71.982	72.744	74.488
Sei Bamban	44.275	46.043	46.236	46.521	47.428
Teluk Mengkudu	42.513	48.334	48.954	49.679	51.082
Perbaungan	105.177	112.153	113.083	114.243	116.947
Pegajahan	27.676	30.206	30.463	30.782	31.516
Pantai Cermin	45.341	49.182	49.695	50.310	51.608
Total	616.396	657.490	662.076	667.998	682.918

Kabupaten Serdang Bedagai dalam melakukan prediksi jangka pendek terhadap perkembangan jumlah penduduk, terutama untuk mendukung kebijakan pembangunan seperti perencanaan pendidikan, kesehatan, dan penyediaan infrastruktur.

Selain itu, hasil ini memperlihatkan bahwa kontribusi penduduk laki-laki terhadap pertumbuhan total lebih tinggi dibanding perempuan. Fenomena ini dapat dijelaskan melalui struktur umur dan mobilitas tenaga kerja, di mana dominasi laki-laki pada usia produktif meningkatkan angka migrasi masuk ke daerah industri dan pertanian di Serdang Bedagai.

Dari sisi validasi model, hasil pengujian statistik (uji F dan uji t) yang dilakukan menunjukkan bahwa kedua variabel independen berpengaruh signifikan terhadap variabel dependen dengan tingkat signifikansi di bawah 0,05. Nilai koefisien determinasi (R^2) sebesar 0,997 juga menunjukkan bahwa 99,7% variasi jumlah penduduk total dapat dijelaskan oleh kombinasi dua variabel independen (jumlah laki-laki dan perempuan), sedangkan sisanya 0,3% dipengaruhi oleh faktor lain seperti migrasi, kelahiran, dan kematian.

Secara umum, hasil ini mengindikasikan bahwa model regresi linear berganda yang dibangun memiliki tingkat akurasi yang sangat tinggi dan dapat digunakan sebagai alat bantu peramalan (forecasting) yang andal. Hal ini sejalan dengan penelitian sebelumnya yang dilakukan oleh (Ardyanti & Abdriando, 2023), yang menunjukkan bahwa regresi linear berganda mampu memberikan estimasi yang akurat dalam memprediksi laju pertumbuhan penduduk di Bedagai.

Tabel 2. Pengumpulan Data Penduduk Tahunan Berdasarkan Gender

Tahun	Penduduk		Jumlah Penduduk
	Laki-laki	Perempuan	
	X1	X2	Y
2019	309.382	307.014	616.396
2020	331.101	326.389	657.490
2021	333.515	328.561	662.076
2022	336.597	331.401	667.998
2023	343.862	339.056	682.918
Jumlah	1.654.457	1.632.421	3.286.878

Karena perhitungan regresi linear berganda melibatkan banyak operasi perkalian dan perpangkatan, maka untuk mempermudah pengolahan data, nilai-nilai tersebut dibagi dengan 1000.

Hasilnya ditampilkan dalam tabel berikut.

Tabel 3. Simplifikasi Data Penduduk (dibagi 1.000)

Tahun	Penduduk		Jumlah Penduduk
	Laki-laki	Perempuan	
	X	X2	Y
2019	309,382	307,014	616,396
2020	331,101	326,389	657,490
2021	333,515	328,561	662,076
2022	336,597	331,401	667,998
2023	343,862	339,056	682,918
Jumlah	1654,457	1632,421	3286,878

Selanjutnya yaitu menghitung menggunakan Algoritma Regresi Linear Berganda. Tujuan dari ujicoba ini adalah untuk melakukan perhitungan algoritma untuk menemukan nilai prediksi terbaik yang akan membagi atribut lainnya. Perhitungan algoritma ini akan menggunakan rumus regresi linear berganda, yang akan dilatihkan terhadap pola data input dan nilai target yang sebenarnya.

Regresi linear adalah jenis hubungan di mana variabel bebas (X) dan variabel terikat (Y) berperan sebagai faktor berpangkat dua. Rumus umum dari algoritma regresi linear berganda adalah:

$$Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_n \cdot X_n$$

Keterangan:

Y = Variabel terikat (nilai yang diprediksi) a_0 ,

a_1, a_2, a_n = Koefisien regresi

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

Analisis regresi linear berganda digunakan untuk memprediksi pertumbuhan penduduk (Y) di BPS Kabupaten Serdang Bedagai berdasarkan dua variabel, yaitu Jumlah Laki-laki (X_1) dan Jumlah Perempuan (X_2), yang dianalisis menggunakan metode regresi linear berganda.



Tabel 4. Learning Dataset

Tahun	Penduduk		Jumlah Penduduk
	Laki-laki	Perempuan	
	X1	X2	
2019	309,382	307,014	616,396
2020	331,101	326,389	657,490
2021	333,515	328,561	662,076
2022	336,597	331,401	667,998
2023	343,862	339,056	682,918
Jumlah	1654,457	1632,421	3286,878

Selanjutnya, melakukan proses ringkasan perhitungan berdasarkan nilai X_1 , X_2 , dan Y pada Tabel 4 di atas. Hasilnya disajikan seperti di bawah ini.

Tabel 5. Ringkasan Perhitungan

Tahun	$(X_1)^2$	X_1X_2	X_1Y	$(X_2)^2$	X_2Y
2019	95.717	94.985	190.702	94.258	189.242
2020	109.628	108.068	217.696	106.530	214.598
2021	111.232	109.580	220.812	107.952	217.532
2022	113.298	111.549	224.846	109.827	221.375
2023	118.241	116.588	234.830	114.959	231.547
Jumlah	548.116	540.760	1.088.886	533.526	1.074.294

Untuk menyederhanakan angka pada tabel di atas, setiap data dibagi dengan 10.000, sehingga diperoleh hasil sebagai berikut:

Tabel 6. Simplifikasi Ringkasan Perhitungan (dibagi 10.000)

Tahun	$(X_1)^2$	X_1X_2	X_1Y	$(X_2)^2$	X_2Y
2019	9,5717	9,4985	19,0702	9,4258	18,9242
2020	10,9628	10,8068	21,7696	10,6530	21,4598
2021	11,1232	10,9580	22,0812	10,7952	21,7532
2022	11,3298	11,1549	22,4846	10,9827	22,1375
2023	11,8241	11,6588	23,4830	11,4959	23,1547
Jumlah	54,8116	54,0760	108,8886	53,3526	107,4294

Untuk memperoleh koefisien regresi a , b_1 dan b_2 dapat diperoleh dengan cara simultan dari tiga persamaan sebagai berikut:

$$\begin{aligned} a_n + b_1 \sum X_1 + b_2 \sum X_2 &= \sum Y & (1) \ a \sum X_1 + \\ b_1 \sum (X_1)^2 + b_2 \sum X_1X_2 &= \sum X_1Y & (2) \ a \sum X_2 + \\ b_1 \sum X_1X_2 + b_2 \sum (X_2)^2 &= \sum X_2Y & (3) \end{aligned}$$

Selanjutnya, masukkan angka-angka yang telah diperoleh dari ringkasan perhitungan (Tabel 6) serta nilai $\sum X_1$, $\sum X_2$ dan $\sum Y$ (dari Tabel 4.4) sehingga diperoleh hasil persamaan 1, 2 dan 3.

$$\begin{aligned} a_5 + b_1(1654,457) + b_2(1632,421) &= 3286,878 & (1) \ a(1654,457) + \\ b_1(54,8116) + b_2(54,077) &= 108,8886 & (2) \ a(1632,421) + b_1(54,077) + \\ b_2(53,3526) &= 107,4294 & (3) \end{aligned}$$

Kemudian, ketiga persamaan tersebut diselesaikan untuk mendapatkan nilai a , b_1 , dan b_2 .

Proses penyelesaiannya dilakukan melalui langkah-langkah berikut:

a. Eliminasi persamaan 1 dan 2

$$\begin{array}{rcl}
 5a + b_1(1654,457) + b_2(1632,421) = 3286,878 & & \times 1654,457 \\
 (1654,457) a + b_1(54,8116) + b_2(54,077) = 108,8886 & & \times 5 \\
 \hline
 8272,285 a + 2737227,965 b_1 + 2700770,35 b_2 = 5437998,315 \\
 8272,285 a + 247,058 b_1 + 270,385 b_2 = 544,443 \\
 \\
 2736953,907 b_1 + 2700499,965 b_2 = 5437453,873 & & (4)
 \end{array}$$

b. Eliminasi persamaan 1 dan 3

$$\begin{array}{rcl}
 5a + b_1(1654,457) + b_2(1632,421) = 2670,482 & & \times 1632,421 \\
 (1632,421) a + b_1(54,077) + b_2(53,3526) = 107,4294 & & \times 5 \\
 \hline
 8162,105 a + 2700770,35 b_1 + 2664798,321 b_2 = 5365568,672 \\
 8162,105 a + 270,385 b_1 + 266,623 b_2 = 537,147 \\
 \\
 2700499,965 b_1 + 2664531,558 b_2 = 5365031,525 & & (5)
 \end{array}$$

c. Eliminasi persamaan 4 dan 5

$$\begin{array}{rcl}
 2736953,907b_1 + 2700499,965b_2 = 5437453,873 & & \times 2700499,965 \\
 2700499,965b_1 + 2664531,558b_2 = 5365031,525 & & \times 2736953,907 \\
 \hline
 7391143930739 b_1 + 7292700063109 b_2 = 14683843993848 \\
 7391143930739 b_1 + 7292700058250 b_2 = 14683843991276 \\
 \\
 4859,040 b_2 = 2122,088 \\
 \\
 b_2 = 0,4367
 \end{array}$$

d. Substitusi persamaan 4 dengan memasukkan nilai b_2 $2700499,965 b_1 + 2664531,558 b_2 = 5365031,525$

$$\begin{aligned}
 2700499,965 b_1 + 2664531,558 (0,436729863) &= 5365031,525 \\
 2700499,965 b_1 + 1163680,502 &= 5365031,525 \\
 2700499,965 b_1 &= 5365031,525 - 1163680,502 \\
 2700499,965 b_1 &= 4201351,022 \\
 b_1 &= 1,5558
 \end{aligned}$$

e. Substitusi persamaan 1 dengan memasukkan nilai b_1 dan b_2

$$\begin{aligned}
 5a + b_1 (1654,457) + b_2 (1632,421) &= 3286,878 \\
 5a + (1,5558) (1654,457) + (0,4367) (1632,421) &= 3286,878 \\
 5a + 2573,951 + 712,927 &= 3286,878 \\
 5a + 3286,878 &= 3286,878 \\
 5a &= 3286,878 - 3286,878 \\
 5a &= 0 \\
 a &= 0
 \end{aligned}$$



Dengan demikian, penerapan data mining berbasis regresi linear berganda bukan hanya membantu proses perhitungan prediksi secara kuantitatif, tetapi juga mendukung BPS dalam mewujudkan sistem pengambilan keputusan berbasis data (data-driven decision making). Ke depan, model ini dapat dikembangkan dengan menambahkan variabel-variabel sosial ekonomi lain agar hasil estimasi menjadi lebih komprehensif dan dinamis.

KESIMPULAN

Berdasarkan analisis dan pembahasan data mining menggunakan teknik estimasi dengan metode Regresi Linear Berganda yang dilakukan oleh Badan Pusat Statistik Kabupaten Serdang Bedagai terkait prediksi laju pertumbuhan penduduk, dapat disimpulkan hal-hal berikut:

- a. Hasil analisis dan pembahasan menunjukkan bahwa metode regresi linear berganda yang diterapkan untuk mengestimasi laju pertumbuhan penduduk menghasilkan tingkat akurasi yang cukup baik.
- b. Estimasi laju pertumbuhan penduduk di Kabupaten Serdang Bedagai menunjukkan jumlah penduduk mencapai 683.056 jiwa, yang berarti terdapat peningkatan dibandingkan tahun sebelumnya.
- c. Hasil ini dapat memberikan bantuan kepada Badan Pusat Statistik Kabupaten Serdang Bedagai dalam memperoleh perhitungan estimasi laju pertumbuhan yang lebih akurat.

DAFTAR PUSTAKA

- Ardyanti, P., & Abdriando. (2023). Penerapan data mining untuk mengestimasi laju pertumbuhan penduduk Denpasar menggunakan metode regresi linier berganda. *Journal of Business and Audit Information System (JBASE)*, 37–44.
- Badan Pusat Statistik. (2021, July 13). Jumlah penduduk menurut provinsi di Indonesia (ribu), 2012–2016. <https://bekasikab.bps.go.id/id/statistics-table/1/MjQ4MyMx/jumlah-penduduk-menurutprovinsi-di-indonesia--ribu---20122016.html>
- Badan Pusat Statistik. (2024, June 28). Jumlah penduduk pertengahan tahun (ribu jiwa). <https://www.bps.go.id/id/statistics-table/2/MTk3NSMy/jumlah-penduduk-pertengahantahun--ribu-jiwa-.html>
- Janan, S., Janan, T., & Mustakim, M. (2024). Analisis model pertumbuhan logistik untuk memprediksi jumlah penduduk miskin di Kabupaten Probolinggo. *Jurnal Ilmiah Matematika*, 9–14.
- Purwadi, P., Ramadhan, P. S., & Safitri, N. (2019). Penerapan data mining untuk mengestimasi laju pertumbuhan penduduk menggunakan metode regresi linier berganda pada BPS Deli Serdang. *Journal Sains dan Komputer (SAINTIKOM)*, 55–61.
- Rahmat, C. A. (2022). *Analisis dan penerapan data mining untuk mengestimasi laju pertumbuhan penduduk Kabupaten Musi Banyuasin menggunakan metode regresi linier berganda di Badan Pusat Statistik Musi Banyuasin* [Skripsi]. Jambi.