



Implementasi Data Mining Dalam Prediksi Presentase Penduduk Miskin Jawa Tengah Dengan Metode Regresi Linear

Zidan Dwi Anggoro*¹, Fajar Sodik²

¹Sistem Informasi, Fakultas Sains Dan Teknologi, Universitas YPPI Rembang

¹Sistem Informasi, Fakultas Sains Dan Teknologi, Universitas YPPI Rembang

E-mail : zidandwianggoro66@gmail.com*¹, fajaryppi@gmail.com²

Abstract

Poverty is still a crucial problem that affects the quality of life of the community, including in Central Java Province which has diverse socio-economic conditions. This study aims to predict the percentage of poor people in Central Java in 2024 using a linear regression method based on data mining through the RapidMiner application. The data used is secondary data from the Central Statistics Agency (BPS) of Central Java which includes the variables of Adjusted Per Capita Income (PP), Human Development Index (HDI), Poverty Depth Index (P1), and Poverty Severity Index (P2). The research process includes the stages of data selection, pre-processing, transformation, model development, and evaluation. The results of the analysis showed that the HDI, P1, and P2 variables had a significant influence on the percentage of the poor population. The regression model obtained was $Y = 7.198 - 0.069X_2 + 8.861X_3 - 15.423X_4$. Evaluation using Root Mean Square Error (RMSE) yielded a value of 1.125, indicating that the model had a low error rate and was able to produce estimates that were close to actual data. This research is expected to support the formulation of poverty alleviation policies and enrich studies related to the use of data mining in socio-economic analysis.

Keywords: poverty, prediction, linear regression, rapidminer

Abstrak

Kemiskinan masih menjadi permasalahan krusial yang memengaruhi kualitas hidup masyarakat, termasuk di Provinsi Jawa Tengah yang memiliki kondisi sosial-ekonomi beragam. Penelitian ini bertujuan untuk memprediksi persentase penduduk miskin Jawa Tengah tahun 2024 menggunakan metode regresi linear berbasis data mining melalui aplikasi RapidMiner. Data yang



digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS) Jawa Tengah yang mencakup variabel Pendapatan Per Kapita Disesuaikan (PP), Indeks Pembangunan Manusia (IPM), Indeks Kedalaman Kemiskinan (P1), dan Indeks Keparahan Kemiskinan (P2). Proses penelitian meliputi tahap seleksi data, pra-pemrosesan, transformasi, pembangunan model, serta evaluasi. Hasil analisis menunjukkan bahwa variabel IPM, P1, dan P2 memberikan pengaruh signifikan terhadap persentase penduduk miskin. Model regresi yang diperoleh adalah $Y = 7.198 - 0.069X_2 + 8.861X_3 - 15.423X_4$. Evaluasi menggunakan Root Mean Square Error (RMSE) menghasilkan nilai 1.125, yang mengindikasikan bahwa model memiliki tingkat error yang rendah dan mampu menghasilkan estimasi yang mendekati data aktual. Penelitian ini diharapkan dapat mendukung perumusan kebijakan pengentasan kemiskinan serta memperkaya kajian terkait pemanfaatan data mining dalam analisis sosial-ekonomi.

Kata kunci : kemiskinan, prediksi, regresi linier, rapidminer

1. PENDAHULUAN

Kemiskinan merupakan salah satu indikator penting yang mencerminkan tingkat kesejahteraan sosial suatu negara. Tingginya jumlah masyarakat yang tidak mampu memenuhi kebutuhan dasar menunjukkan masih rendahnya kualitas hidup penduduk [1]. Di Indonesia, persoalan ini terus menjadi fokus utama pembangunan, terutama di wilayah-wilayah berpenduduk padat seperti Jawa Tengah dan Jawa Barat yang memberi kontribusi besar terhadap total penduduk miskin nasional. Menurut laporan resmi, pulau Jawa mencatat lebih dari 13 juta penduduk miskin pada 2022, menegaskan urgensi evaluasi dan pemantauan kemiskinan yang lebih akurat untuk mendukung kebijakan pemerintah [1].

Provinsi Jawa Tengah menghadapi persoalan kemiskinan yang cukup kompleks, dipengaruhi oleh beragam aspek seperti pendapatan, pengangguran, sanitasi, dan kesejahteraan sosial. Keterbatasan basis data yang terintegrasi menyebabkan penyaluran bantuan sosial di sejumlah daerah masih belum optimal dan sering kali tidak sesuai kebutuhan masyarakat setempat [2]. Selain itu, berbagai indikator kemiskinan seperti garis kemiskinan dan persentase penduduk miskin terus berubah sepanjang 2020–2024, sehingga diperlukannya proyeksi yang lebih akurat untuk melihat kecenderungan kemiskinan pada tahun 2024 [2]. Kondisi ini menegaskan perlunya model prediksi yang lebih tepat untuk wilayah tersebut.

Berbagai penelitian terdahulu menunjukkan bahwa teknik data mining, khususnya regresi linear, dapat digunakan secara efektif untuk memprediksi fenomena sosial. Penelitian mengenai pertumbuhan penduduk, misalnya, berhasil menghasilkan tingkat kesalahan prediksi yang rendah dengan RMSE 0.998 menggunakan regresi linear [3]. Studi lain dalam bidang produksi beras mencatat bahwa metode yang sama menghasilkan nilai *error* minimal, dengan RMSE 285.590 dan *Relative Error* 0.000% [4]. Dalam konteks kemiskinan, regresi linier berganda telah dimanfaatkan untuk memproyeksikan tingkat kemiskinan di Jawa Barat, di mana variabel IPM berperan signifikan dengan korelasi



-0.752 terhadap penduduk miskin [1]. Sementara itu, pemetaan kemiskinan di Jawa Tengah menggunakan K-Means memberikan kualitas klasterisasi yang sangat baik dengan DBI 0.009, sehingga menunjukkan potensi besar pendekatan analitis berbasis data dalam memahami struktur kemiskinan [2]. Selain itu, penelitian lain juga menunjukkan bahwa penerapan berbagai teknik *machine learning* dalam prediksi kemiskinan mampu meningkatkan akurasi hasil prediksi dan mendukung pengambilan keputusan yang lebih tepat sasaran [15].

Walaupun hasil penelitian sebelumnya memberi pondasi teoretis yang kuat, realitas lapangan di Jawa Tengah menunjukkan masih banyak hambatan. Minimnya integrasi data antarindikator kemiskinan membuat proses analisis menjadi tidak efisien dan rentan menghasilkan bias [2]. Selain itu, perubahan cepat kondisi sosial-ekonomi seperti meningkatnya pengangguran dan ketimpangan antarwilayah membuat pemerintah daerah memerlukan instrumen analisis yang lebih adaptif dan berbasis prediksi [1]. Tantangan serupa juga ditemukan pada penelitian terkait pertumbuhan penduduk maupun produksi beras yang menunjukkan ketidaksesuaian antara data historis dengan kondisi aktual di lapangan, sehingga menghambat perumusan kebijakan berbasis data [4][3]. Keadaan ini memperkuat kebutuhan terhadap model prediksi kemiskinan yang lebih akurat.

Hasil kajian menunjukkan bahwa penelitian mengenai kemiskinan di Jawa Tengah masih dominan berfokus pada klasterisasi wilayah tanpa memberikan perkiraan kuantitatif mengenai tingkat kemiskinan pada tahun tertentu. Penelitian yang memanfaatkan regresi linear untuk menghasilkan prediksi numerik terkait persentase penduduk miskin Jawa Tengah, terutama untuk tahun 2024, masih sangat terbatas [2]. Selain itu, meskipun regresi linear sudah diterapkan di beberapa daerah seperti Jawa Barat, model ini belum banyak dimanfaatkan secara khusus untuk konteks Jawa Tengah dengan variabel-variabel yang relevan [1]. Oleh sebab itu, terdapat kekosongan penelitian yang perlu diisi melalui studi prediktif berbasis regresi linear.

Penelitian ini disusun untuk menerapkan metode regresi linear dalam memprediksi persentase penduduk miskin Jawa Tengah tahun 2024 melalui pendekatan data mining. Dengan mengolah data sosial-ekonomi terbaru, model prediksi yang dihasilkan diharapkan mampu memberikan estimasi yang akurat dan dapat dimanfaatkan dalam proses penyusunan kebijakan daerah. Dari sisi akademik, penelitian ini memberikan kontribusi dengan menambah literatur mengenai prediksi kemiskinan berbasis regresi linear di tingkat provinsi. Secara praktis, hasil penelitian dapat membantu pemerintah daerah dalam meningkatkan ketepatan sasaran program penanggulangan kemiskinan sekaligus memperkuat perencanaan pembangunan berbasis data.

2. TINJAUAN PUSTAKA

2.1. Kemiskinan

Kemiskinan menggambarkan kondisi ketika seseorang atau sekelompok masyarakat tidak sanggup memenuhi kebutuhan hidup minimum secara layak. Indikator ini menjadi acuan penting untuk menilai tingkat kesejahteraan suatu daerah. Literatur menunjukkan bahwa kemiskinan tidak hanya berakar dari pendapatan rendah, tetapi juga dipengaruhi oleh kualitas pelayanan dasar, pendidikan, serta kemampuan ekonomi masyarakat dalam jangka panjang [1]. Ketimpangan antarwilayah sering kali memperbesar jumlah penduduk miskin, terutama di



provinsi besar seperti Jawa Tengah dan Jawa Barat yang memberi kontribusi signifikan terhadap statistik kemiskinan nasional.

2.2. Faktor-Faktor yang Mempengaruhi Kemiskinan

Tingkat kemiskinan dipengaruhi oleh berbagai faktor multidimensional. Penelitian sebelumnya mengungkap bahwa variabel seperti Indeks Pembangunan Manusia (IPM), pengangguran, pengeluaran rumah tangga, garis kemiskinan, dan kualitas akses layanan sosial menjadi penentu utama dalam analisis kemiskinan. Contohnya, di Jawa Barat variabel IPM memiliki hubungan negatif yang kuat dengan tingkat kemiskinan ($r = -0.752$), menandakan bahwa peningkatan kualitas pembangunan manusia mampu menekan angka kemiskinan secara substansial [1]. Selain itu, hasil klusterisasi kemiskinan di Jawa Tengah menunjukkan adanya ketimpangan antarwilayah yang cukup besar, memperlihatkan variasi kondisi sosial-ekonomi yang memengaruhi kemiskinan di kabupaten/kota [2].

2.3. Data Mining

Data mining merupakan proses penggalian data untuk menemukan pola, hubungan, atau informasi penting dari kumpulan data berukuran besar. Metode ini dapat mengidentifikasi pola tersembunyi yang sulit ditemukan melalui teknik analisis konvensional. Pemanfaatan data mining telah diterapkan secara luas di berbagai sektor, seperti ekonomi, pertanian, dan sosial, karena mampu menghasilkan analisis prediktif yang lebih efektif dan akurat. Dalam proses analisis prediktif, terdapat berbagai algoritma yang dapat digunakan, di antaranya **regresi linear**, **decision tree**, dan Random Forest. Random Forest termasuk metode *ensemble learning* yang bekerja dengan membentuk banyak pohon keputusan, kemudian menggabungkan hasil prediksi dari masing-masing pohon untuk meningkatkan tingkat akurasi model [14]. Salah satu penerapan data mining dapat dilihat pada penelitian prediksi produksi beras yang menghasilkan tingkat kesalahan prediksi relatif rendah, sehingga metode ini dinilai efektif untuk menganalisis data numerik dalam skala besar dan berkelanjutan [4], [8], [9], [13].

2.4. Regresi Linear

Regresi linear merupakan teknik analisis statistik yang digunakan untuk memodelkan hubungan antarvariabel, baik dengan satu variabel bebas maupun beberapa variabel bebas. Metode ini memungkinkan peneliti mengestimasi nilai suatu variabel berdasarkan pola keterkaitan yang terbentuk dari data historis. Pada penelitian pertumbuhan penduduk Cirebon, regresi linear menunjukkan performa yang sangat baik dengan nilai RMSE 0.998, yang berarti tingkat kesalahan prediksi sangat kecil. Penelitian lain mengenai prediksi produksi beras pun memperlihatkan hasil serupa, di mana regresi linear menghasilkan RMSE 285.590 dan Relative Error 0.000%, mengindikasikan tingkat ketelitian prediksi yang tinggi. Hal ini menegaskan bahwa metode regresi linear relevan digunakan untuk memprediksi variabel sosial seperti kemiskinan [10] [11][12]. Persamaan model regresi linier sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Keterangan :

Y = Variabel dependen

β_0 = Konstanta



β_1 = Koefisien determinasi

X = Variabel independen

2.5. Evaluasi Root Mean Square Error

RMSE merupakan salah satu ukuran evaluasi paling umum digunakan dalam memvalidasi model prediksi regresi. RMSE mengukur rata-rata selisih kuadrat antara nilai aktual dan nilai yang diprediksi. Nilai RMSE yang kecil menandakan bahwa perbedaan antara data hasil prediksi dan data aktual semakin kecil, sehingga model dianggap semakin akurat. Dalam penelitian prediksi produksi beras, nilai RMSE sebesar 285.590 dan Relative Error 0.000% menunjukkan bahwa model regresi mampu mengeksekusi prediksi dengan tingkat akurasi yang sangat tinggi [4]. Penggunaan RMSE sangat penting dalam memastikan model prediksi kemiskinan yang dikembangkan benar-benar layak digunakan untuk pengambilan keputusan.

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

Nilai Y_t merepresentasikan data aktual pada periode t , sedangkan y'_t merupakan nilai yang diprediksi pada periode yang sama. Banyaknya data prediksi ditunjukkan oleh n . RapidMiner menyediakan operator khusus untuk menilai performa model regresi linear dengan menghitung Root Mean Square Error (RMSE). Pada metrik ini, sebuah model dikatakan memiliki kualitas yang baik apabila nilai RMSE semakin mendekati angka 0 (nol) [11], [12].

2.6. RapidMiner dalam Analisis Data Mining

RapidMiner merupakan salah satu platform populer dalam data mining yang menyediakan antarmuka visual untuk memudahkan pengguna melakukan proses analisis data, mulai dari preprocessing hingga pemodelan. RapidMiner mendukung berbagai algoritma seperti regresi linear, decision tree, dan clustering. Pada penelitian mengenai prediksi penjualan dan persediaan, RapidMiner digunakan untuk membangun model regresi linear dan berhasil menghasilkan prediksi dengan tingkat akurasi yang baik serta proses analisis yang lebih cepat dan sistematis [5].

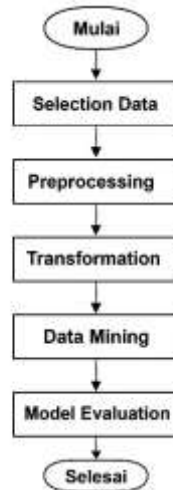
3. METODE PENELITIAN

Pengumpulan data dalam penelitian ini dilakukan dengan mengambil data sekunder dari situs resmi Badan Pusat Statistik (BPS) Provinsi Jawa Tengah melalui alamat <https://jateng.bps.go.id/id>. Data sekunder tersebut merupakan data yang telah tersedia, dikumpulkan, dan diolah oleh lembaga resmi sehingga umumnya telah dipublikasikan dalam bentuk laporan tahunan atau publikasi statistik lainnya [7]. Data yang diperoleh meliputi berbagai indikator sosial dan ekonomi yang relevan dengan kebutuhan analisis penelitian ini (BPS Jawa Tengah, 2024) yang meliputi sebagai berikut :

1. Presentase Penduduk Miskin sebagai variabel yang dipengaruhi atau variabel dependent (Y)
2. Pendapatan Perkapita yang disesuaikan (PP) sebagai variabel yang mempengaruhi (X_1)
3. Indeks Pembangunan Manusia (IPM) sebagai variabel yang mempengaruhi (X_2)



4. Indeks Kedalaman Kemiskinan (P1) sebagai variabel yang mempengaruhi (X_3)
5. Indeks Keparahan Kemiskinan (P2) sebagai variabel yang mempengaruhi (X_4)



Gambar 1. Alur Penelitian

Keterangan dari alur penelitian pada gambar diatas :

1. Selection Data atau tahap seleksi data merupakan proses memilih data yang paling sesuai dan diperlukan untuk penelitian dari keseluruhan kumpulan data yang tersedia. Pemilihan ini dilakukan agar analisis hanya menggunakan data yang relevan sehingga hasil yang diperoleh lebih akurat dan terfokus. Tahap seleksi ini juga dilakukan dalam penelitian terdahulu sebelum membangun model prediksi atau klasterisasi [2].
2. Pra-pemrosesan merupakan kegiatan membersihkan data dengan cara memeriksa konsistensinya, menghapus duplikasi, serta memperbaiki kesalahan atau ketidaksesuaian dalam dataset. Tahap ini penting untuk memastikan kualitas data sebelum digunakan pada proses analisis yang lebih lanjut. Beberapa penelitian sebelumnya juga melakukan pembersihan data untuk memastikan akurasi model prediksi maupun regresi [5].
3. Transformasi data dilakukan dengan mengubah bentuk data sehingga cocok untuk dianalisis menggunakan teknik data mining. Proses ini dapat berupa konversi variabel kategorikal menjadi numerik atau sebaliknya, normalisasi data, atau pengubahan struktur data agar lebih efektif untuk pemodelan. Penelitian mengenai prediksi penjualan menggunakan RapidMiner juga menekankan pentingnya transformasi data agar model dapat berjalan optimal [5].
4. Data mining adalah tahapan untuk menemukan pola, keterkaitan, atau informasi baru yang sebelumnya tidak tampak melalui analisis biasa. Proses ini dilakukan dengan menerapkan metode tertentu seperti regresi, clustering, ataupun teknik prediktif lainnya. Berbagai penelitian sebelumnya telah menunjukkan bahwa data mining mampu menghasilkan pola dan prediksi yang akurat dalam konteks sosial-ekonomi maupun produksi pangan [4].
5. Evaluasi dilakukan untuk menilai sejauh mana model atau pola yang terbentuk dapat memberikan informasi yang benar dan dapat diandalkan. Hasil analisis kemudian disajikan dalam bentuk yang mudah dipahami pengguna. Pada penelitian regresi, proses evaluasi sering menggunakan metrik seperti Root Mean Square Error (RMSE) guna menilai performa model dan memastikan bahwa hasil prediksi mendekati nilai sebenarnya [3].



HASIL DAN PEMBAHASAN

4.1. Selection dan Pre-Processing Data

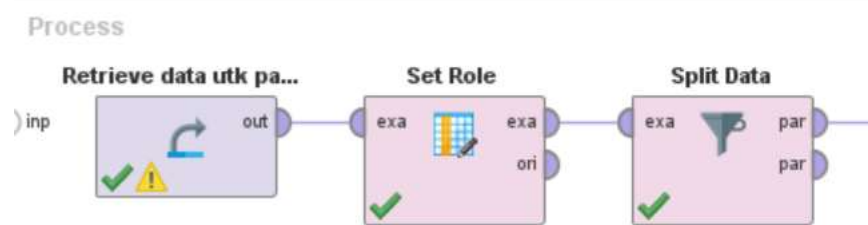
Dalam penelitian ini, dataset yang digunakan meliputi data yang berhubungan tentang Presentase Penduduk Miskin di Jawa Tengah per Kabupaten/Kota tahun 2024

Tabel 1. Dataset Penelitian

Kabupaten/Kota	PP	IPM	Persentase Penduduk		
			Miskin	P1	P2
3301 Kabupaten Cilacap	11.868	72.38	10.68	1.59	0.35
3302 Kabupaten Banyumas	12.96	74.52	11.95	2.09	0.5
3303 Kabupaten Purbalingga	11.343	70.69	14.18	2.1	0.5
3304 Kabupaten Banjarnegara	10.595	69.6	14.71	2.38	0.61
3305 Kabupaten Kebumen	10.176	71.93	15.71	2.35	0.48
3306 Kabupaten Purworejo	11.584	75.11	10.87	1.08	0.18
3307 Kabupaten Wonosobo	11.98	69.82	15.28	2.41	0.6
3308 Kabupaten Magelang	10.926	71.99	10.83	1.23	0.23
3309 Kabupaten Boyolali	14.195	75.96	9.63	1.56	0.4
3310 Kabupaten Klaten	13.513	78.16	12.04	1.46	0.27
3311 Kabupaten Sukoharjo	12.758	79.3	7.47	1.21	0.31
3312 Kabupaten Wonogiri	10.634	72.55	10.71	1.53	0.33
3313 Kabupaten Karanganyar	12.732	78.11	9.59	1.63	0.41
3314 Kabupaten Sragen	13.89	75.53	12.41	2.42	0.67
3315 Kabupaten Grobogan	11.548	72.02	11.43	1.9	0.43
3316 Kabupaten Blora	10.926	71.39	11.42	1.59	0.32
3317 Kabupaten Rembang	11.83	72.53	14.02	2.76	0.84
3318 Kabupaten Pati	11.829	74.1	9.17	1.17	0.25
3319 Kabupaten Kudus	12.533	77.22	7.23	1.02	0.24
3320 Kabupaten Jepara	11.729	74.32	6.09	0.71	0.15
3321 Kabupaten Demak	11.591	74.57	11.89	1.95	0.56
3322 Kabupaten Semarang	13.377	75.67	6.96	0.91	0.18
3323 Kabupaten Temanggung	10.519	71.87	8.67	0.97	0.18
3324 Kabupaten Kendal	13.277	74.34	9.35	0.91	0.15
3325 Kabupaten Batang	10.919	70.74	8.73	1.25	0.27
3326 Kabupaten Pekalongan	11.701	71.84	8.95	1.16	0.23

3327 Kabupaten Pemalang	10.017	68.55	14.92	2.56	0.61
3328 Kabupaten Tegal	10.972	70.77	6.81	0.97	0.22
3329 Kabupaten Brebes	11.389	68.46	15.6	2.5	0.55
3371 Kota Magelang	13.619	82.15	5.94	0.92	0.19
3372 Kota Surakarta	16.291	84.4	8.31	1.14	0.26
3373 Kota Salatiga	17.096	85.72	4.57	0.61	0.11
3374 Kota Semarang	16.99	85.25	4.03	0.59	0.12
3375 Kota Pekalongan	14.547	77.22	6.71	0.84	0.17
3376 Kota Tegal	14.358	77.43	7.64	1.34	0.32

Setelah dataset terkumpul langkah selanjutnya dilakukan *pre-processing* data seperti pada gambar 2 dibawah ini.

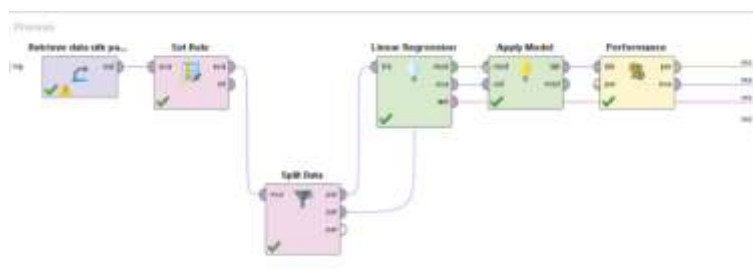


Gambar 2. *PreProcessing* Data Presentase Penduduk Miskin

Langkah awal dalam proses ini adalah memilih atribut-atribut yang relevan dari dataset, yaitu kolom PP, IPM, P1 dan P2. Selanjutnya, operator Set Role digunakan untuk menetapkan Presentase Penduduk Miskin sebagai label, sementara kolom-kolom lainnya ditetapkan sebagai *regular attributes*. Setelah peran masing-masing kolom ditentukan, data kemudian dibagi menjadi dua bagian menggunakan operator Split Data, dengan proporsi 80% sebagai data training dan 20% sebagai data testing. Bagian data testing inilah yang selanjutnya diproses untuk menghasilkan nilai prediksi.

4.2. Data Mining

Proses data mining diterapkan pada data yang sudah melalui tahap pra-pemrosesan. Tahapan dimulai dengan preprocessing, kemudian dilanjutkan dengan penerapan algoritma Linear Regression untuk melakukan pemodelan. Alur proses tersebut ditampilkan pada Gambar 3.





Gambar 3. Proses Data Mining

LinearRegression

```
- 0.069 * IPM  
+ 8.861 * P1  
- 15.423 * P2  
+ 7.198
```

Gambar 4. Hasil Uji Persamaan Regresi

Merujuk pada tabel yang ditampilkan pada Gambar 4, hasil menunjukkan bahwa IPM, P1 dan P2 merupakan variabel yang berpengaruh terhadap Presentase Penduduk Miskin menurut pemodelan Regresi Linier. Hasil tersebut diuji menggunakan dataset training sehingga membentuk model persamaan regresi sebagai berikut :

$$Y = 7.198 - 0.069X_2 + 8.861X_3 - 15.423X_4$$

Dimana :

- Y : Presentase Penduduk Miskin
X₂ : Indeks Pembangunan Manusia (IPM)
X₃ : Indeks Kedalaman Kemiskinan (P1)
X₄ : Indeks Keparahan Kemiskinan (P2)

Berdasarkan persamaan regresi tersebut maka dapat di interpretasikan sebagai berikut:

- 1) Konstanta $\beta_0 = 7.198$

Jika IPM (X₂), P1 (X₃) dan P2 (X₄) bernilai 0 (nol) maka Presentase Penduduk Miskin di Jawa Tengah pada tahun 2024 adalah sebesar 7.198.

- 2) Koefisien $\beta_2 = - 0.069$

Jika IPM (X₂) mengalami peningkatan 1% maka Presentase Penduduk Miskin di Jawa Tengah pada tahun 2024 menurun sebesar 0.069. dan jika IPM (X₂) mengalami penurunan 1% maka Presentase Penduduk Miskin di Jawa Tengah pada tahun 2024 meningkat sebesar 0.069. Hal tersebut menunjukkan bahwa nilai koefisien regresi $\beta_2 = - 0.069$ adalah korelasi negatif sehingga hubungannya tidak searah.

- 3) Koefisien $\beta_3 = 8.861$

Jika P1 (X₃) mengalami peningkatan 1% maka Presentase Penduduk Miskin di Jawa Tengah pada tahun 2024 meningkat sebesar 8.861. dan jika P1 (X₃) mengalami penurunan 1% maka Presentase Penduduk Miskin di Jawa Tengah pada tahun 2024 menurun sebesar 8.861. Hal tersebut menunjukkan bahwa nilai koefisien regresi $\beta_3 = 8.861$ adalah korelasi positif sehingga hubungannya searah.

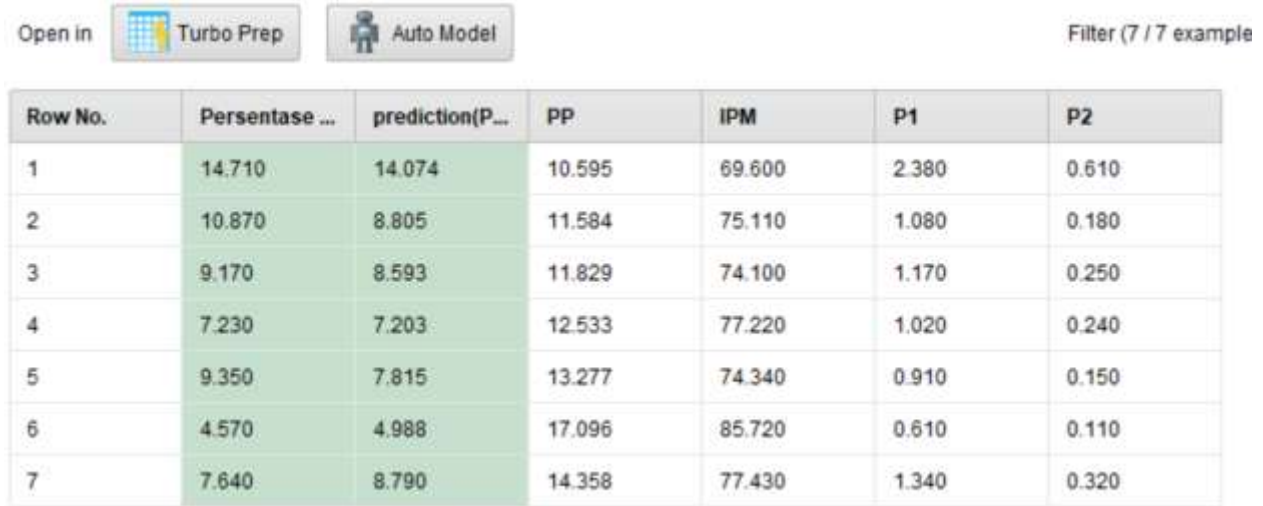
- 4) Koefisien $\beta_4 = - 15.423$

Jika PP (X₄) mengalami peningkatan 1% maka Presentase Penduduk Miskin di Jawa Tengah pada tahun 2024 menurun sebesar 15.423. dan jika PP (X₄) mengalami penurunan 1% maka Presentase Penduduk Miskin di Jawa



Tengah pada tahun 2024 meningkat sebesar 15.423. Hal tersebut menunjukkan bahwa nilai koefisien regresi $\beta_4 = -15.423$ adalah korelasi negatif sehingga hubungannya tidak searah.

Hasil prediksi Presentase Penduduk Miskin dapat dilihat pada gambar 5. Hasil tersebut menunjukkan antara nilai Presentase Penduduk Miskin asli dan nilai prediksi.



Row No.	Persentase ...	prediction(P...	PP	IPM	P1	P2
1	14.710	14.074	10.595	69.600	2.380	0.610
2	10.870	8.805	11.584	75.110	1.080	0.180
3	9.170	8.593	11.829	74.100	1.170	0.250
4	7.230	7.203	12.533	77.220	1.020	0.240
5	9.350	7.815	13.277	74.340	0.910	0.150
6	4.570	4.988	17.096	85.720	0.610	0.110
7	7.640	8.790	14.358	77.430	1.340	0.320

Gambar 5. Hasil Prediksi Presentase Penduduk Miskin

4.3. Evaluasi Model

Setelah model berhasil disusun, tahap berikutnya adalah melakukan evaluasi untuk menilai seberapa baik model tersebut bekerja. Pengujian performa dilakukan menggunakan operator Performance, dan hasil evaluasi tersebut ditampilkan pada Gambar 6.



Gambar 6. Hasil Root Mean Squared Error

Berdasarkan hasil evaluasi, diperoleh nilai RMSE sebesar 1.125 dan . Nilai-nilai yang relatif rendah dan dekat dengan angka nol tersebut menunjukkan bahwa kemampuan model dalam menghasilkan estimasi cukup baik dan mendekati data sebenarnya. Temuan ini sejalan dengan studi yang dijelaskan oleh (Saputra, S. Sumarjono; & M. A. S., 2022), yang menyatakan bahwa kinerja regresi linear dapat dinilai melalui perhitungan root mean square error, di mana suatu model prediksi dianggap optimal apabila nilai RMSE mendekati 0 (nol).



KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan mengenai prediksi persentase penduduk miskin di Provinsi Jawa Tengah tahun 2024 menggunakan metode regresi linear, diperoleh bahwa variabel IPM, Indeks Kedalaman Kemiskinan (P1), dan Indeks Keparahan Kemiskinan (P2) merupakan faktor yang berpengaruh terhadap tingkat kemiskinan di wilayah tersebut. Model regresi yang terbentuk, yaitu $Y = 7.198 - 0.069X_2 + 8.861X_3 - 15.423X_4$, menunjukkan adanya hubungan linier antara variabel-variabel independen dan persentase penduduk miskin. Hasil evaluasi model menggunakan RMSE menghasilkan nilai 1.125, yang termasuk rendah dan mengindikasikan bahwa model memiliki kemampuan prediksi yang cukup akurat serta mendekati nilai sebenarnya. Secara keseluruhan, penerapan teknik data mining melalui aplikasi RapidMiner terbukti efektif dalam mengolah dan memodelkan data sosial-ekonomi, sehingga dapat dijadikan sebagai salah satu pendekatan untuk mendukung proses perumusan kebijakan pengentasan kemiskinan. Penelitian ini memberikan kontribusi baik secara akademik maupun praktis, dengan menghadirkan model prediksi yang dapat dimanfaatkan oleh pemerintah daerah dalam menyusun strategi pembangunan yang lebih tepat sasaran.

REFERENSI

- [1] A. Sazwati, et al., “Penerapan Data Mining untuk Mengestimasi Persentase Penduduk Miskin di Jawa Barat Menggunakan Regresi Linier Berganda,” *Jurnal Informatika*, vol. 8, no. 1, pp. 1103–1108, 2024.
- [2] M. M. E. I. Tawakal and A. M. Majid, “Analisis Tingkat Kemiskinan dengan Algoritma K-Means,” *Jurnal Data Science*, vol. 7, no. 1, pp. 112–119, 2025.
- [3] A. Nur, A. S. Sunge, and A. T. Zy, “Prediksi Pertumbuhan Penduduk dengan Metode Regresi Linear,” *Jurnal Teknologi Informasi*, vol. 11, no. 2, 2023.
- [4] A. I. P. Adinata and I. Ali, “Penerapan Data Mining dalam Prediksi Produksi Beras Menggunakan Metode Regresi Linear,” *Jurnal Komputasi*, vol. 8, no. 2, pp. 2020–2026, 2024.
- [5] A. H. Pradhana, M. Irfa, A. Ali, A. Ristyawan, and E. Daniati, “Penerapan Regresi Linear Menggunakan RapidMiner untuk Memprediksi Penjualan dan Persediaan,” *Jurnal Sistem Informasi*, vol. 8, pp. 291–297, 2024.
- [6] S. Saputra, Sumarjono, and M. A. S., “Penerapan Data Mining untuk Prediksi Kinerja Menggunakan Evaluasi RMSE,” *Jurnal Informatika*, 2022.
- [7] Badan Pusat Statistik Provinsi Jawa Tengah, “Profil Kemiskinan di Jawa Tengah Tahun 2024,” 2025.
- [8] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. Cambridge, MA: Morgan Kaufmann, 2022.
- [9] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, 2nd ed. New York: Springer, 2021.
- [10] M. Sholeh, et al., “Penerapan Data Mining dengan Metode Regresi Linear Menggunakan RapidMiner,” *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 8, no. 2, 2023.
- [11] A. Septa, et al., “Analisis Variabel Kemiskinan di Indonesia Menggunakan Regresi Linear dan K-Means,” *Jurnal Profit*, 2024.
- [12] E. S. H. Hasibuan, et al., “Pengaruh Jumlah Penduduk dan UMP terhadap Tingkat Kemiskinan,” *JPTAM*, vol. 8, no. 3, 2024.
- [13] A. F. Abdillah, et al., “Pengaruh Pengeluaran Per Kapita terhadap Kemiskinan,” *Jurnal Sistem Informasi*, 2025.



- [14] I. P. Sofyani, et al., “Analisis Faktor-Faktor yang Mempengaruhi Kemiskinan Menggunakan Regresi Linear,” Jurnal APOKSI, 2024.
- [15] A. N. A. Mokoagow, et al., “Pemodelan Tingkat Kemiskinan Menggunakan Ridge dan Lasso Regression,” Prosiding Saintek UT, 2024.
- [16] F. H. Muzaki and W. J. Pranoto, “Analisis Regresi Linear dalam Data Mining untuk Prediksi,” Jurnal Tektonik, 2023.