

---

---

## PREDIKSI TINGKAT PENGANGGURAN KABUPATEN DAN KOTA DI SUMATERA UTARA MENGGUNAKAN PENDEKATAN DATA MINING DENGAN ALGORITMA LINEAR REGRESION

Tri Yohanes Lumban Gaol<sup>1</sup>, Yogi Anggara Sihaloho<sup>2</sup>, Bayu Prasetyo<sup>3</sup>,  
Saut Parsaoran Tamba<sup>4</sup>

Universitas Prima Indonesia, Medan

Pusat Unggulan IPTEK (PUI-ITIK) Inovasi Teknologi Ilmu Komputer

Email: <sup>1</sup>lumbangaolyohanes5@gmail.com, <sup>2</sup>yogisihaloho16@gmail.com,

<sup>3</sup>bayuprasetyo8018@gmail.com, <sup>4</sup>sautparsaorantamba@unprimdn.ac.id

**Abstract:** *Unemployment Rate is one of the important indicators used to assess the employment and economic conditions of a region. A high unemployment rate can hinder economic growth and reduce public welfare. This study aims to analyze the factors influencing the Unemployment Rate and to develop a prediction model for the Unemployment Rate in regencies/cities of North Sumatra Province using a Data Mining approach with the Multiple Linear Regression algorithm. The data used in this study are secondary data obtained from the Central Statistics Agency (BPS) of North Sumatra Province for the 2019–2023 period, including population, Gross Regional Domestic Product (GRDP), and education level variables with a total of 165 observations. The research stages include data collection, data cleaning, data exploration, splitting the dataset into training and testing data with an 80:20 ratio, model development, and evaluation using MAE, RMSE, and coefficient of determination ( $R^2$ ) metrics. The results show that the education level is a significant variable affecting the Unemployment Rate, indicating the phenomenon of educated unemployment, while population and GRDP do not have a partially significant effect. The resulting model obtained an  $R^2$  value of 30.88%, an MAE of 1.76, and an RMSE of 2.15, indicating that the model has fairly good and stable predictive performance. The findings of this study are expected to support employment policy-making and data-driven regional economic development planning in North Sumatra Province.*

**Keywords:** *Unemployment Rate, Data Mining, Multiple Linear Regression, GRDP, Education Level.*

**Abstrak:** Tingkat Pengangguran merupakan salah satu indikator penting dalam menilai kondisi ketenagakerjaan dan perekonomian suatu daerah. Tingginya tingkat pengangguran dapat menghambat pertumbuhan ekonomi dan menurunkan kesejahteraan masyarakat. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi Tingkat Pengangguran serta membangun model prediksi Tingkat Pengangguran pada kabupaten/kota di Provinsi Sumatera Utara menggunakan pendekatan Data Mining dengan algoritma *Multiple Linear Regression*. Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara periode 2019–2023, meliputi variabel jumlah penduduk, Produk Domestik Regional Bruto (PDRB), dan tingkat pendidikan dengan total 165 *observasi*. Tahapan penelitian meliputi pengumpulan data, pembersihan data, eksplorasi data, pembagian data training dan testing dengan rasio 80:20, pembangunan model, serta evaluasi menggunakan metrik MAE, RMSE, dan koefisien determinasi ( $R^2$ ). Hasil penelitian menunjukkan bahwa tingkat pendidikan merupakan variabel yang berpengaruh signifikan terhadap Tingkat Pengangguran dengan indikasi terjadinya fenomena educated unemployment, sedangkan jumlah penduduk dan PDRB tidak berpengaruh signifikan secara parsial. Model yang dihasilkan memperoleh nilai  $R^2$  sebesar 30,88%, MAE sebesar 1,76, dan RMSE sebesar 2,15, yang menunjukkan kemampuan prediksi yang cukup baik dan stabil. Hasil

penelitian ini diharapkan dapat menjadi pendukung dalam pengambilan kebijakan ketenagakerjaan dan perencanaan pembangunan ekonomi daerah berbasis data di Provinsi Sumatera Utara.

**Kata kunci:** Tingkat Pengangguran, *Data Mining*, *Multiple Linear Regression*, PDRB, Pendidikan.

## PENDAHULUAN

Tingkat Pengangguran merupakan salah satu indikator penting dalam menilai kondisi ketenagakerjaan dan stabilitas perekonomian suatu daerah. Tingginya tingkat pengangguran dapat menyebabkan menurunnya kesejahteraan masyarakat, meningkatnya angka kemiskinan, serta menghambat pertumbuhan ekonomi daerah [1], [16]. Permasalahan pengangguran masih menjadi tantangan utama di negara berkembang, termasuk Indonesia, terutama akibat ketidakseimbangan antara jumlah tenaga kerja dan ketersediaan lapangan pekerjaan [3], [17]. Selain itu, peningkatan jumlah penduduk yang tidak diimbangi dengan pertumbuhan kesempatan kerja juga dapat memperbesar angka pengangguran dan kemiskinan [18].

Provinsi Sumatera Utara memiliki kondisi ekonomi yang berbeda pada setiap kabupaten/kota sehingga menyebabkan variasi Tingkat Pengangguran antarwilayah. Faktor-faktor seperti jumlah penduduk, Produk Domestik Regional Bruto (PDRB), tingkat pendidikan, inflasi, dan upah minimum diduga memengaruhi kondisi tersebut [3], [19], [20]. Data Badan Pusat Statistik menunjukkan bahwa tingkat pengangguran di Indonesia mengalami perubahan setiap tahun, terutama pada masa pandemi COVID-19 yang berdampak besar terhadap sektor ekonomi dan ketenagakerjaan [4]. Penelitian mengenai dampak pandemi terhadap pengangguran juga menunjukkan adanya peningkatan ketidakstabilan pasar tenaga kerja selama periode tersebut [1]. Oleh karena itu, diperlukan analisis yang lebih spesifik pada tingkat daerah agar dapat

memahami pola pengangguran secara lebih rinci [12], [13].

Perkembangan teknologi informasi mendorong penggunaan pendekatan berbasis data dalam berbagai bidang, termasuk ekonomi dan ketenagakerjaan. Salah satu pendekatan yang banyak digunakan adalah data mining, yaitu proses pengolahan data untuk menemukan pola dan informasi tersembunyi dari kumpulan data dalam jumlah besar [5], [14]. Pendekatan ini dinilai efektif dalam mendukung proses analisis dan pengambilan keputusan berbasis data, terutama dalam membangun model prediksi ekonomi dan sosial [2], [6].

Dalam penelitian prediktif, algoritma *Multiple Linear Regression* sering digunakan karena mampu menjelaskan hubungan antara variabel dependen dan variabel independen secara matematis dan mudah diinterpretasikan [7], [8]. Metode regresi linear juga memiliki kemampuan yang cukup baik dalam melakukan prediksi pada kasus sosial ekonomi, termasuk prediksi tingkat pengangguran [9]. Penelitian sebelumnya menunjukkan bahwa pendekatan regresi linear mampu menghasilkan model prediksi yang stabil dan memiliki tingkat akurasi yang baik dalam berbagai studi ekonomi dan ketenagakerjaan [10], [11].

Meskipun demikian, sebagian besar penelitian mengenai prediksi pengangguran masih dilakukan pada tingkat nasional atau provinsi, sedangkan penelitian pada tingkat kabupaten/kota masih relatif terbatas [11]. Padahal, analisis pada tingkat wilayah yang lebih rinci sangat dibutuhkan untuk mendukung penyusunan kebijakan ketenagakerjaan yang lebih tepat sasaran [12]. Selain itu, fenomena *educated unemployment* yang menunjukkan tingginya pengangguran

pada kelompok masyarakat berpendidikan juga menjadi perhatian penting dalam pembangunan ketenagakerjaan di Indonesia [13], [19].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun model prediksi Tingkat Pengangguran pada kabupaten/kota di Provinsi Sumatera Utara menggunakan pendekatan data mining dengan algoritma *Multiple Linear Regression*. Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik Provinsi Sumatera Utara periode 2019–2023 yang meliputi variabel jumlah penduduk, PDRB, dan tingkat pendidikan [15]. Model yang dibangun dievaluasi menggunakan *metrik Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan koefisien determinasi ( $R^2$ ) untuk mengetahui tingkat akurasi model [6], [8]. Hasil penelitian diharapkan dapat memberikan informasi yang berguna dalam mendukung pengambilan kebijakan ketenagakerjaan dan perencanaan pembangunan ekonomi daerah berbasis data [5], [14].

## METODE

Penelitian ini menggunakan metode kuantitatif dengan pendekatan data mining untuk membangun model prediksi Tingkat Pengangguran pada kabupaten/kota di Provinsi Sumatera Utara. Pendekatan kuantitatif dipilih karena penelitian menggunakan data numerik yang dianalisis secara statistik untuk mengetahui hubungan antar variabel serta menghasilkan model prediksi yang terukur.

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara periode 2019–2023. Dataset terdiri dari 165 observasi yang berasal dari 33 kabupaten/kota selama lima tahun pengamatan. Variabel dependen dalam penelitian ini adalah Tingkat Pengangguran, sedangkan variabel independen meliputi jumlah penduduk

(X1), Produk Domestik Regional Bruto (PDRB) (X2), dan tingkat pendidikan (X3).

Tahapan penelitian diawali dengan pengumpulan data dari publikasi resmi BPS. Selanjutnya dilakukan proses data cleaning untuk memastikan tidak terdapat data duplikat maupun missing value. Setelah proses pembersihan data, dilakukan eksplorasi data menggunakan statistik deskriptif dan analisis korelasi untuk melihat hubungan antarvariabel penelitian.

Tahap berikutnya adalah transformasi data dan pembagian dataset menjadi data training dan data testing dengan rasio 80:20. Sebanyak 132 data digunakan sebagai data training dan 33 data digunakan sebagai data testing. Pembagian data dilakukan secara acak menggunakan random seed 42 agar hasil penelitian dapat direproduksi secara konsisten.

Model prediksi dibangun menggunakan algoritma *Multiple Linear Regression* dengan metode *Ordinary Least Squares (OLS)*. Model regresi digunakan untuk mengetahui pengaruh variabel independen terhadap Tingkat Pengangguran. Bentuk persamaan regresi yang digunakan adalah sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + e$$

Keterangan:

Y = Tingkat Pengangguran

a = konstanta

b1, b2, b3 = koefisien regresi

X1 = jumlah penduduk

X2 = PDRB

X3 = tingkat pendidikan

e = error

Sebelum model diinterpretasikan, dilakukan uji multikolinearitas menggunakan *Variance Inflation Factor (VIF)* untuk memastikan tidak terdapat hubungan yang terlalu tinggi antarvariabel independen. Selanjutnya, performa model dievaluasi menggunakan beberapa metrik evaluasi, yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error*

(RMSE), dan koefisien determinasi ( $R^2$ ). Nilai MAE dan RMSE digunakan untuk mengukur tingkat kesalahan prediksi model, sedangkan  $R^2$  digunakan untuk mengetahui kemampuan model dalam menjelaskan variasi data.

Hasil prediksi kemudian dianalisis untuk melihat tingkat akurasi model serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap Tingkat Pengangguran di kabupaten/kota Provinsi Sumatera Utara.

## HASIL DAN PEMBAHASAN

### Hasil

Dataset terdiri dari 165 observasi (33 Kabupaten/Kota  $\times$  5 tahun, 2019-2023) yang bersumber dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Variabel dependen adalah Tingkat Pengangguran (TP/Y), sedangkan variabel independen meliputi: Jumlah Penduduk (X1), PDRB (X2), dan Tingkat Pendidikan (X3). Setelah proses data cleaning, tidak ditemukan missing value pada seluruh variabel. Dataset dibagi secara acak menjadi 80% data training (132 data) dan 20% data testing (33 data) dengan random seed 42 untuk reproduktibilitas.

**Tabel 1 Statistik Deskriptif Variabel Penelitian (N=165)**

Variabel	Mean	Std. Deviasi	Min	Max
TP (%)	4,8367	2,7025	0,18	11,50
Jumlah Penduduk (jiwa)	452.752	510.948	48.935	2.494.512
PDRB (miliar Rp)	17.195	29.111	850	173.445
Tingkat Pendidikan (%)	22,74	6,93	8,45	43,80

Rata-rata TP Sumatera Utara periode 2019-2023 sebesar 4,84% dengan disparitas tinggi (std = 2,70%). PDRB dan Jumlah Penduduk memiliki distribusi

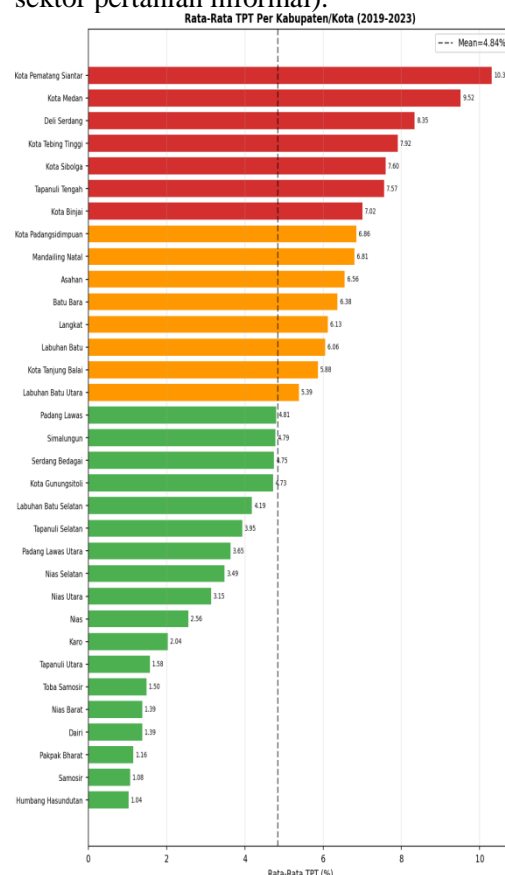
right-skewed akibat dominasi Kota Medan.

**Tabel 1 Rata-Rata Variabel Per Tahun**

Tahun	TP (%)	Penduduk	PDRB (M Rp)	Pendidikan (%)
2019	4,49	441.289	16.550	20,50
2020	5,45	445.562	16.410	21,93
2021	4,93	452.611	16.873	22,44
2022	4,72	458.037	17.637	24,04
2023	4,59	466.262	18.506	24,79

TP melonjak pada 2020 (+21,4%) akibat pandemi COVID-19, kemudian menurun konsisten hingga 2023. PDRB dan pendidikan meningkat setiap tahun.

Daerah perkotaan cenderung TP tinggi (angkatan kerja formal besar, selektivitas tinggi), sedangkan daerah rural/kabupaten TP rendah (dominasi sektor pertanian informal).



**Gambar 1 Rata-Rata TP Per Kabupaten/Kota (2019-2023)**

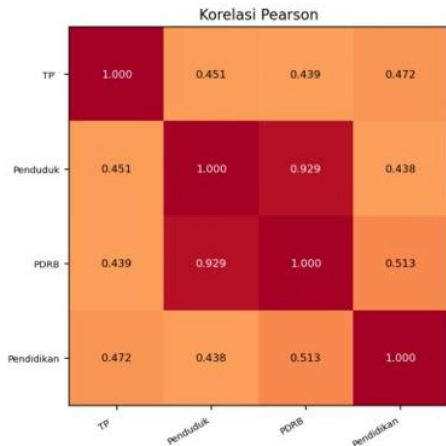
**Tabel 3 TP Tertinggi dan Terendah**

TP Tertinggi	%	TP Terendah	%
Kota	10,32	Humbang	1,04

Pematang Siantar		Hasundutan	
Kota Medan	9,52	Samosir	1,08
Deli Serdang	8,35	Pakpak Bharat	1,16
Kota Tebing Tinggi	7,92	Nias Barat	1,39
Kota Sibolga	7,60	Dairi	1,39

Pendidikan	0,472	0,438	0,513	1,000
------------	-------	-------	-------	-------

Ketiga variabel independen berkorelasi positif sedang (0,44-0,47) terhadap TP. Tingkat Pendidikan memiliki korelasi terkuat (r = 0,472), menunjukkan bahwa daerah dengan pendidikan lebih tinggi cenderung memiliki TP lebih tinggi — sebuah indikasi fenomena *educated unemployment*. Korelasi X1 dan X2 sangat tinggi (r=0,93) mengindikasikan multikolinearitas yang perlu diperhatikan dalam model.



Gambar 1 Gambaran Umum Data

Tabel 4 Matriks Korelasi Pearson

	TP	Penduduk	PDRB	Pendidikan
TP	1,000	0,451	0,439	0,472
Penduduk	0,451	1,000	<b>0,930</b>	0,438
PDRB	0,439	0,930	1,000	0,513

Tabel 2 Variance Inflation Factor

Variabel	VIF	Keterangan
Jumlah Penduduk (X1)	7,46	Multikolinearitas sedang
PDRB (X2)	8,19	Multikolinearitas sedang
Tingkat Pendidikan (X3)	1,38	Tidak ada masalah

VIF X1 dan X2 di bawah 10 sehingga masih dalam batas toleransi, namun perlu diperhatikan dalam interpretasi koefisien individual. Model diestimasi menggunakan OLS (Normal Equation) pada 132 data training.

Persamaan Regresi:

$$Y = 0,5934 + 0,0000019 X_1 - 0,0000111 X_2 + 0,1598 X_3$$

Tabel 3 Koefisien Regresi dan Uji Signifikansi

Variabel	Koefisien	Std. Error	t-stat	p-value	Sig. (α=0,05)
Konstanta	0,5934	0,7727	0,768	0,4439	Tidak
Jumlah Penduduk (X1)	0,0000019	0,0000011	1,759	0,0809	Tidak
PDRB (X2)	-0,0000111	0,0000202	-0,548	0,5848	Tidak
Tingkat Pendidikan(X3)	0,1598	0,0339	4,712	0,000006	Ya

Dimana dapat dilihat sebagai berikut hasilnya.

Dimana kemudian dapat diringkas hasil model sebagai berikut:

Adjusted R <sup>2</sup>	0,2926
F-statistic	19,06
Prob (F)	0,000000 (Signifikan)

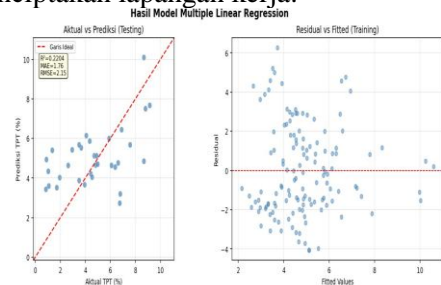
Tabel 7 Ringkasan Model

Metrik	Nilai
R <sup>2</sup>	0,3088

Uji F menghasilkan p-value < 0,001, artinya secara simultan ketiga variabel independen berpengaruh

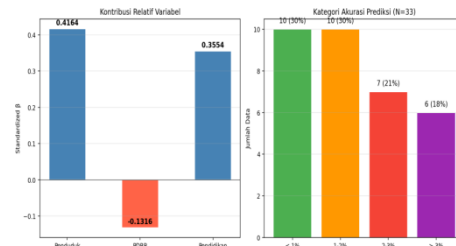
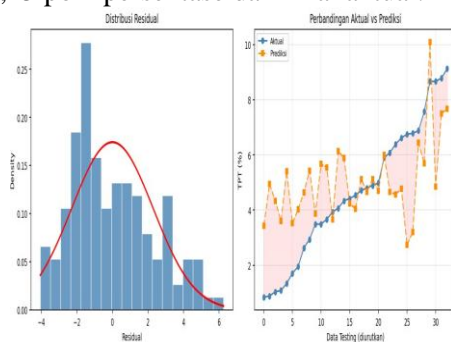
signifikan terhadap TP. Secara parsial (uji-t), hanya Tingkat Pendidikan (X3) yang signifikan pada  $\alpha = 0,05$  dengan koefisien positif 0,1598.

Artinya setiap kenaikan 1% tingkat pendidikan, TP meningkat 0,16 poin persentase —konsisten dengan fenomena educated unemployment dimana lulusan pendidikan tinggi lebih selektif memilih pekerjaan. PDRB menunjukkan koefisien negatif (-0,0000111), sesuai teori bahwa pertumbuhan ekonomi menurunkan pengangguran, namun pengaruhnya tidak signifikan. Hal ini mengindikasikan adanya jobless growth — pertumbuhan ekonomi yang belum optimal dalam menciptakan lapangan kerja.



**Gambar 2 Hasil Model (Aktual vs Prediksi, Residual, Distribusi Residual)**

Performa yang konsisten antara data training dan testing menunjukkan model tidak mengalami overfitting. MAE testing (1,76) berarti rata-rata kesalahan. Prediksi sebesar  $\pm 1,76$  poin persentase. RMSE testing (2,15) menunjukkan simpangan error sekitar 2,15 poin persentase dari nilai aktual.



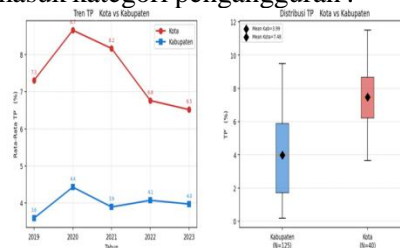
**Gambar 3 Kontribusi Variabel dan Kategori Akurasi**

Dimana dapat diambil Kesimpulan perkategorinya sebagai berikut

**Tabel 8 Kategori Akurasi Prediksi**

Kategori	Error	
< 1% (Akurat)	10	30%
1-2% (Cukup)	9	27%
2-3% (Kurang)	7	21%
> 3% (Buruk)	7	21%

57% data testing memiliki error < 2 poin persentase. Error besar terjadi pada wilayah TP ekstrem rendah (< 2%) dan yang terdampak pandemi berat. Terdapat pola yang jelas: daerah perkotaan memiliki TP lebih tinggi karena angkatan kerja formal yang besar dan selektivitas pekerjaan yang tinggi. Sebaliknya, daerah kabupaten/rural memiliki TP rendah karena dominasi sektor pertanian informal yang menyerap banyak tenaga kerja tanpa termasuk kategori pengangguran.



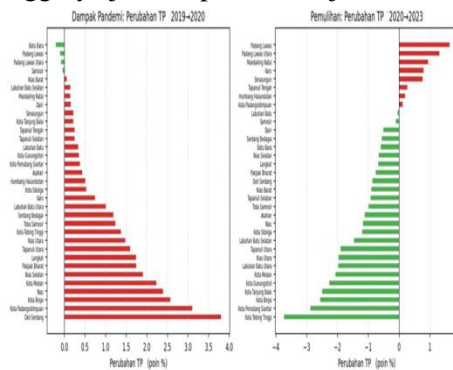
**Gambar 4 Grafik Perbandingan TP Kota vs KAB**

Analisis Kota vs Kabupaten dapat dilihat dari penjelasan berikut dibawah ini **Tabel 4 Statistik TP Berdasarkan Tipe Wilayah**

Tipe	N	Mean TP	Std	Min	Max	Mean Penduduk	Mean PD RB	Mean Pendidikan
Kota	40	7,48	2,05	3,65	11	470	25,14	26,99

ta	%			,5	941	7	%
Ka	12,3,99	2,31	0,18	9,446	14.65	21,38	
b	5	%		50	931	0	%

Terdapat perbedaan yang sangat signifikan antara TP di daerah Kota (rata-rata 7,48%) dan Kabupaten (rata-rata 3,99%) dengan selisih 3,49 poin persentase. Daerah kota juga memiliki rata-rata tingkat pendidikan lebih tinggi (27,0% vs 21,4%), yang konsisten dengan temuan bahwa tingkat pendidikan berkorelasi positif dengan TP (*educated unemployment*). Rata-rata PDRB kota (Rp 25.147 miliar) lebih tinggi dibanding kabupaten (Rp 14.650 miliar), namun tidak mampu menekan TP karena tingginya jumlah pencari kerja formal.



**Gambar 6 Dampak Pandemi**

**Dampak Pandemi COVID-19 (2019 → 2020)**

1. Rata-rata kenaikan TP akibat pandemi: +0,96 poin persentase
2. 29 dari 33 wilayah (88%) mengalami kenaikan TP pada 2020
3. Kenaikan tertinggi: Deli Serdang (+3,81 poin), Kota Medan (+2,24), Kota Tebing Tinggi (+1,38)
4. Hanya 4 wilayah yang TP-nya turun atau stabil: Batu Bara, Padang Lawas, Padang Lawas Utara, Samosir

Daerah perkotaan dan penyangga kota besar (Deli Serdang, Binjai, Medan) paling terdampak karena ketergantungan tinggi pada sektor formal yang rentan terhadap pembatasan aktivitas ekonomi.

**Pemulihan Pasca-Pandemi (2020 → 2023)**

1. Rata-rata penurunan TP sejak puncak pandemi: -0,86 poin persentase
2. 25 dari 33 wilayah (76%) berhasil menurunkan TP dibanding puncak 2020
3. 8 wilayah belum pulih (TP 2023 > TP 2020), antara lain: Karo (+0,80), Padang Lawas Utara (+1,31), Padang Lawas (+1,64), Mandailing Natal (+0,95)
4. Wilayah pemulihan tercepat: Kota Tanjung Balai (-2,50), Kota Tebing Tinggi (-3,74), Labuhan Batu Utara (-1,98)

**Pembahasan**

Berdasarkan hasil analisis, tingkat pendidikan merupakan faktor yang paling dominan dalam memengaruhi tingkat pengangguran (TP). Semakin tinggi tingkat pendidikan, pengangguran cenderung meningkat akibat fenomena *educated unemployment*, yaitu lulusan yang lebih selektif dalam memilih pekerjaan dan adanya ketidaksesuaian antara keterampilan yang dimiliki dengan kebutuhan pasar kerja. Sementara itu, jumlah penduduk berpengaruh positif terhadap TP karena pertumbuhan penduduk meningkatkan jumlah angkatan kerja yang membutuhkan pekerjaan. Di sisi lain, PDRB berpengaruh negatif terhadap TP, namun pengaruhnya relatif lemah karena pertumbuhan ekonomi belum sepenuhnya mampu menciptakan lapangan kerja yang memadai (*jobless growth*). Selain ketiga variabel tersebut, tingkat pengangguran juga dipengaruhi oleh faktor lain seperti struktur ekonomi, investasi, kebijakan ketenagakerjaan, upah minimum, jumlah unit usaha, mobilitas tenaga kerja, dan dampak pandemi.

Penerapan Data Mining dalam penelitian ini dilakukan melalui tahapan pengumpulan data, pembersihan data, eksplorasi data, pembangunan model regresi linear berganda, evaluasi, dan prediksi. Model yang dihasilkan terbukti

signifikan dan mampu menjelaskan hubungan antara variabel-variabel yang diteliti.

Berdasarkan hasil evaluasi, model memiliki nilai  $R^2$  sebesar 30,88% yang menunjukkan kemampuan prediksi pada kategori sedang. Nilai MAE sebesar 1,76% dan RMSE sebesar 2,15% menunjukkan bahwa rata-rata kesalahan prediksi berada pada kisaran  $\pm 2\%$ , sehingga model cukup baik dalam memprediksi tingkat pengangguran, terutama pada wilayah dengan tingkat pengangguran menengah.

Hasil prediksi yang diperoleh dapat dimanfaatkan sebagai dasar pendukung kebijakan ketenagakerjaan, seperti menentukan wilayah prioritas penanganan pengangguran, mengembangkan program yang menyelaraskan pendidikan dengan kebutuhan industri, mendorong kewirausahaan bagi lulusan pendidikan tinggi, mengarahkan pertumbuhan ekonomi ke sektor padat karya, serta menyusun target penurunan pengangguran berdasarkan hasil prediksi yang lebih terukur.

## SIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, Maka Tingkat Pengangguran (TP) di kabupaten/kota Sumatera Utara dipengaruhi oleh jumlah penduduk, PDRB, dan tingkat pendidikan, dengan pendidikan sebagai faktor paling dominan yang menunjukkan fenomena educated unemployment. Penerapan data mining menggunakan algoritma regresi linier melalui serangkaian tahapan menghasilkan model yang cukup akurat, konsisten, dan tidak mengalami overfitting, meskipun masih terdapat faktor lain yang memengaruhi TP seperti struktur ekonomi, investasi, dan upah minimum. Model prediksi ini dapat dimanfaatkan untuk menetapkan wilayah prioritas penanganan pengangguran, memperkuat link- and-match pendidikan dengan kebutuhan industri, mendorong

kewirausahaan bagi lulusan pendidikan tinggi, serta mengarahkan pertumbuhan ekonomi ke sektor padat karya.

## DAFTAR PUSTAKA

- M. Fajar, et al., “Forecasting unemployment during COVID-19,” *Economic Modelling*, 2021.
- R. Bhagat, et al., “Machine learning in economic modeling,” *Procedia Computer Science*, vol. 200, pp. 45–52, 2022.
- A. Mustikasari, “Faktor-faktor pengangguran di Indonesia,” *Economic Development Analysis Journal*, 2025.
- Badan Pusat Statistik, *Statistik Ketenagakerjaan Indonesia*, Jakarta: BPS, 2024.
- R. M. Mustapa, “Data mining for economic analysis,” *Journal of Big Data*, 2025.
- S. Arora, et al., “Machine learning for prediction models,” *Expert Systems with Applications*, 2024.
- K. Kim, “Regression analysis in economic prediction,” *Economic Systems*, 2025.
- W. Badrawani, “Linear regression for forecasting,” *ScienceDirect*, 2025.
- A. S. Singarimbun, “Prediksi tingkat pengangguran menggunakan regresi linear,” *Jurnal Universitas Sumatera Utara*, 2024.
- R. Alauddin, et al., “Analisis pengangguran berbasis regresi,” *Jurnal UIN Sunan Kalijaga*, 2025.
- S. Wahyuddin, “Time series unemployment prediction,” in *Proc. International Conference on Data Science and Analytics*, 2025.
- OECD, *Employment Outlook*, Paris: OECD Publishing, 2022.
- World Bank, *Labor Market Trends Report*, Washington, DC: World Bank, 2023.
- UIN Jakarta, “Data mining dalam ekonomi,” *Jurnal UIN Syarif Hidayatullah*, 2024.

- Badan Pusat Statistik Provinsi Sumatera Utara, Statistik Ketenagakerjaan Daerah, 2024.
- Sari, Y. A. (2021). Pengaruh Upah Minimum Tingkat Pengangguran Terbuka Dan Jumlah Penduduk Terhadap Kemiskinan Di Provinsi Jawa Tengah. *Equilibrium*, 10(2), 12
- Rahayu, D. Y., Ramadani, D. W., Zaldi, N. A., Halimah, N., Agustin, R. Y., & Rustika, T. (2023). Pengaruh Inflasi Terhadap Tingkat Pengangguran Di Indonesia. *Jurnal Pijar*, 1(2), 258-265.
- MUTIARA, R. (2025). PENGARUH POPULASI PENDUDUK, TINGKAT PENGANGGURAN TERBUKA (TPT) DAN KEMISKINAN TERHADAP PERTUMBUHAN EKONOMI DI 11 NEGARA KAWASAN ASEAN DALAM PERSPEKTIF EKONOMI ISLAM TAHUN 2019-2023 (Doctoral dissertation, UIN RADEN INTAN LAMPUNG).
- SHANIA FARZA, S. F. (2021). PENGARUH PERTUMBUHAN EKONOMI, PENDUDUK, PENDIDIKAN TERHADAP TINGKAT PENGANGGURAN TERBUKA DI PROVINSI JAMBI (Doctoral dissertation, Universitas Batanghari).
- Situmorang, F., & Suharianto, J. (2024). Eksplorasi Jejak Tingkat Pengangguran Terbuka, Jumlah Penduduk Miskin, Dan Upah Minimum Regional Terhadap Produk Domestik Regional Bruto Di Provinsi Sumatera Utara 2002-2022. *Journal of Economics and Regional Science*, 4(2), 98-114.