

IMPLEMENTASI REGRESI LINEAR BERGANDA UNTUK MEMPREDIKSI TINGKAT KRIMINALITAS DI WILAYAH POLRES DAIRI BERDASARKAN FAKTOR SOSIAL DAN EKONOMI

Fitri Arbaini¹, Hotler Manurung², Hermansyah Sembiring³
Sistem Informasi, STMIK Kaputama, Binjai
E-mail: *fitriarbaini96@gmail.com¹, manurunghotler0@gmail.com²,
hermansyahsembiring240165@gmail.com³

ABSTRAK

Kriminalitas merupakan fenomena sosial yang kompleks dan dipengaruhi oleh berbagai faktor, termasuk kondisi sosial dan ekonomi. Penelitian ini bertujuan untuk mengidentifikasi dan memprediksi tingkat kriminalitas di wilayah Polres Dairi menggunakan metode Regresi Linear Berganda. Dengan menganalisis data kriminalitas dari Polres Dairi, penelitian ini berfokus pada hubungan antara variabel independen (jenis kelamin, pekerjaan, pendidikan terakhir, dan penghasilan per bulan) dan variabel dependen (kategori kriminalitas). Hasil analisis menunjukkan bahwa model regresi yang dibangun memiliki tingkat akurasi yang tinggi dan dapat memberikan gambaran jelas mengenai pengaruh faktor-faktor sosial dan ekonomi terhadap angka kriminalitas. Temuan ini diharapkan dapat menjadi dasar bagi pihak kepolisian dalam merumuskan kebijakan pencegahan kejahatan yang lebih efektif. Kata kunci: Regresi Linear Berganda, Data Mining, Kriminalitas, Faktor Sosial Ekonomi

Kata kunci

Regresi Linear Berganda, Data Mining, Kriminalitas, Faktor Sosial Ekonomi

ABSTRACT

Criminality is a complex social phenomenon influenced by various factors, including social and economic conditions. This study aims to identify and predict the level of criminality in the Dairi Police area using the Multiple Linear Regression method. By analyzing crime data from the Dairi Police, this research focuses on the relationship between independent variables (gender, occupation, last education, and monthly income) and the dependent variable (crime category). The analysis results show that the regression model built has a high level of accuracy and can provide a clear picture of the influence of social and economic factors on crime rates. These findings are expected to serve as a basis for law enforcement agencies in formulating more effective crime prevention policies.

Keywords

Multiple Linear Regression, Data Mining, Criminality, Socio-Economic Factors

1. PENDAHULUAN

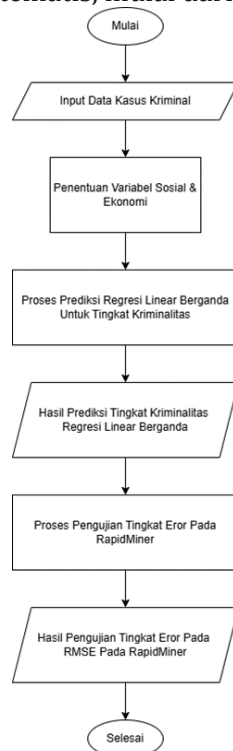
Kriminalitas adalah tindakan yang melanggar norma dan hukum, seperti pencurian, penganiayaan, pembunuhan, dan penipuan (Yuzani et al., 2024). Di Indonesia, angka kriminalitas terus meningkat, dengan The Global Organized Crime Index 2024 menempatkan Indonesia pada peringkat ke-20 di dunia dan ke-2 di ASEAN. Sampai dengan 21 Januari 2025, tercatat 24.327 kasus kriminalitas di seluruh Indonesia (Yonatan, 2025). Peningkatan ini berdampak luas terhadap keamanan dan kesejahteraan masyarakat (Wijoyo et al., n.d.). Faktor sosial-ekonomi seperti kemiskinan, pengangguran, dan rendahnya tingkat pendidikan dianggap berkorelasi langsung dengan tingkat kriminalitas (Rahmah et al., 2024; Duli, 2019).

Penelitian terdahulu telah membuktikan efektivitas metode regresi linear berganda dalam memprediksi berbagai fenomena, seperti omset penjualan kamera (Aditya et al., 2023), produksi padi (Om et al., 2024), dan omset penjualan minyak goreng (Fadianty & Sriani, 2024). Regresi linear berganda mampu menganalisis hubungan antara satu variabel terikat dan beberapa variabel bebas (Duli, 2019).

Berdasarkan latar belakang tersebut, penelitian ini berupaya mengimplementasikan metode regresi linear berganda untuk memprediksi tingkat kriminalitas di wilayah Polres Dairi berdasarkan faktor sosial dan ekonomi.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan Data Mining dengan algoritma Regresi Linear Berganda. Prosedur penelitian mengikuti alur sistematis, mulai dari identifikasi masalah hingga evaluasi.



Gambar 1. Metodologi Penelitian

Langkah-langkah yang dilakukan:

- Pengumpulan Data: Mengumpulkan 37 sampel data kasus kriminalitas dari Polres Dairi.
- Pra-pemrosesan Data: Menginisialisasi kriteria status sosial dan ekonomi serta mengkategorikan data kriminalitas ke dalam 6 kategori (Kekerasan Fisik dan Keluarga, Pencurian dan Perampokan, Penipuan dan Pemalsuan, Kejahatan Seksual, Kejahatan Siber dan Informasi, dan Kejahatan Terhadap Ketertiban Umum).
- Perhitungan Regresi Linear Berganda:
- Menghitung nilai rata-rata variabel untuk setiap kategori.
- Menggunakan persamaan regresi linear berganda:
$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$
- Menggunakan metode eliminasi determinan Cramer untuk menemukan koefisien regresi (a,b1,b2,b3,b4).
- Perhitungan Akurasi: Menggunakan rumus Root Mean Square Error (RMSE)

untuk mengevaluasi tingkat kesalahan model.

$$RMSE = \text{SQRT} \left(\sum \frac{|y - y'|}{n} \right) =$$

Di mana:

y' = Hasil Prediksi

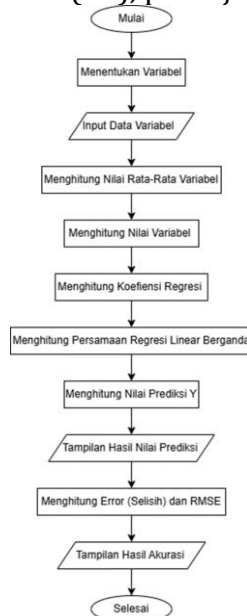
y = Data Aktual

n = Jumlah Data

- h. Perancangan Sistem: Merancang sistem prediksi menggunakan aplikasi RapidMiner, dengan alur proses seperti pada gambar berikut (Setyo & Wardhana, 2019).

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 37 data sampel yang dikategorikan menjadi 6 kategori. Hasil perhitungan manual menunjukkan model regresi sebagai berikut:
jenis kelamin (X1), pendidikan terakhir (X3), pekerjaan (X2), penghasilan per bulan (X4)



Gambar 2. Flowchart Perhitungan Manual

Berdasarkan pada gambar diatas dapat diketahui beberapa tahapan yaitu:

- Mulai
- Menentukan Variabel
Variabel yang digunakan yaitu: Jenis Kelamin, Pekerjaan, Jenis Kriminal, Pendidikan Terakhir, dan Penghasilan/Bulan
- Input Data Variabel
Selanjutnya yaitu input data nilai variabel
- Menghitung Nilai Rata-Rata Variabel
Menghitung nilai rata-rata dari masing-masing variable

$$\text{Rata-rata} = \frac{\text{Jumlah Seluruh Nilai}}{\text{Banyaknya Data}}$$

Kekerasan Fisik dan Keluarga:

$$X1 = \frac{1+1+1+1+1+1+1+1}{8} = \frac{8}{8} = 1$$

$$X2 = \frac{7+7+7+7+4+6+4+4}{8} = \frac{46}{8} = 5,75$$

$$X3 = \frac{4+3+3+4+2+5+4+4}{8} = \frac{29}{8} = 3,625$$

Pencurian dan Perampokan:

$$X1 = \frac{1+0+1+1+0+1}{6} = \frac{4}{6} = 0,666666667$$

$$X2 = \frac{4+7+1+7+2+7}{6} = \frac{28}{6} = 4,666666667$$

$$X3 = \frac{4+4+4+2+4+4}{6} = \frac{22}{6} = 3,666666667$$

$$X4 = \frac{3+2+2+2+2+3+2+3}{8} = \frac{19}{8} = 2,375$$

Penipuan dan Pemalsuan:

$$X1 = \frac{1+1+1+1+1+0+1+0}{8} = \frac{6}{8} = 0,75$$

$$X2 = \frac{7+7+9+10+4+10+8+5}{8} = \frac{60}{8} = 7,5$$

$$X3 = \frac{4+4+5+5+2+5+5+5}{8} = \frac{35}{8} = 4,375$$

$$X4 = \frac{2+2+4+2+2+4+2+2}{8} = \frac{20}{8} = 2,5$$

Kejahatan Siber dan Informasi:
Umum:

$$X1 = \frac{1+1}{2} = \frac{2}{2} = 1$$

$$X2 = \frac{7+7}{2} = \frac{14}{2} = 7$$

$$X3 = \frac{2+5}{2} = \frac{7}{2} = 3,5$$

$$X4 = \frac{3+3}{2} = \frac{6}{2} = 3$$

$$X4 = \frac{2+3+1+2+1+3}{6} = \frac{12}{6} = 2$$

Kejahatan Seksual:

$$X1 = \frac{1+1+1+0+1}{5} = \frac{4}{5} = 0,8$$

$$X2 = \frac{1+1+4+4+3}{5} = \frac{13}{5} = 2,6$$

$$X3 = \frac{4+4+4+3+2}{5} = \frac{17}{5} = 3,4$$

$$X4 = \frac{1+1+3+2+2}{5} = \frac{9}{5} = 1,8$$

Kejahatan Terhadap Ketertiban

$$X1 = \frac{1+1+1+1+1+0+0+1}{8} = \frac{6}{8} = 0,75$$

$$X2 = \frac{4+4+4+4+4+10+2+10}{8} = \frac{42}{8} = 5,25$$

$$X3 = \frac{3+4+2+5+4+5+3+6}{8} = \frac{32}{8} = 4$$

$$X4 = \frac{3+3+2+3+3+4+1+4}{8} = \frac{23}{8} = 2,875$$

e. Hitung Nilai Variabel

f. Menghitung Koefisien Regresi

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 & \sum X_1 X_3 & \sum X_1 X_4 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 & \sum X_2 X_3 & \sum X_2 X_4 \\ \sum X_3 & \sum X_1 X_3 & \sum X_2 X_3 & \sum X_3^2 & \sum X_3 X_4 \\ \sum X_4 & \sum X_1 X_4 & \sum X_2 X_4 & \sum X_3 X_4 & \sum X_4^2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \\ \sum X_3 Y \\ \sum X_4 Y \end{bmatrix}$$

g. Menghitung Persamaan Nilai Regresi Linear Berganda

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$$

h. Menghitung Nilai Prediksi Y

$$Y_1 = -30,5637375 + (10,8054679 \times 1) + (-1,14623271 \times 5,75) + (9,87037744 \times 3,625) + (-1,26928017 \times 2,375) = \mathbf{6,416470134}$$

$$Y_2 = -30,5637375 + (10,8054679 \times 0,666666667) + (-1,14623271 \times 4,666666666) + (9,87037744 \times 3,666666667) + (-1,26928017 \times 2) = \mathbf{4,9436454}$$

$$Y_3 = -30,5637375 + (10,8054679 \times 0,75) + (-1,14623271 \times 7,5) + (9,87037744 \times 4,375) + (-1,26928017 \times 2,5) = \mathbf{8,953318975}$$

i. Tampilan Nilai Hasil Prediksi

j. Menghitung Error (Selisih) dan RMSE

Tahap selanjutnya yaitu menghitung nilai selisih dan Root Mean Square Error (RMSE), dimana semakin kecil nilai RMSE, maka semakin baik model regresi yang diperoleh.

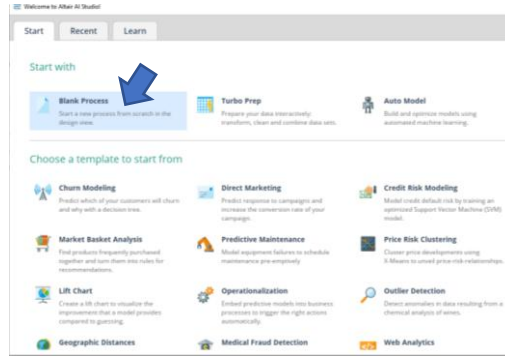
$$\mathbf{RMSE = \sqrt{\left(\sum \frac{|y - y'|}{n} \right)}}$$

k. Tampilan Hasil Akurasi

l. Selesai.

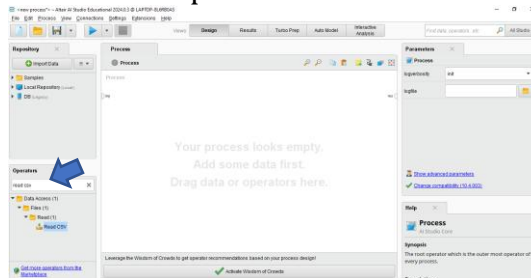
Berikut adalah langkah-langkah implementasi di RapidMiner menggunakan 529 data kasus:

a. Buka aplikasi RapidMiner lalu pilih Blank Proses untuk membuat proses baru.



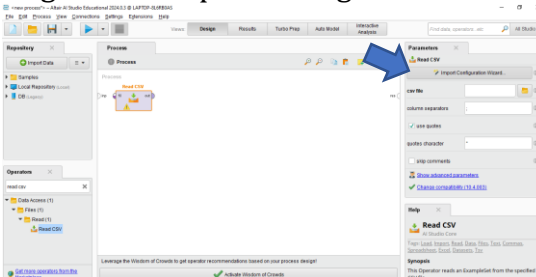
Gambar 3. Proses Baru

b. Tambahkan read csv melalui operator



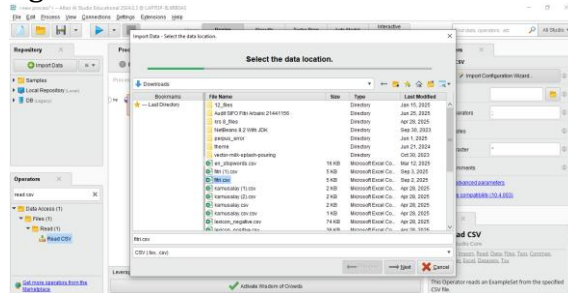
Gambar 4. Mencari Operator Read CSV

c. Masukkan data dengan cara Import Configuration Wizard...



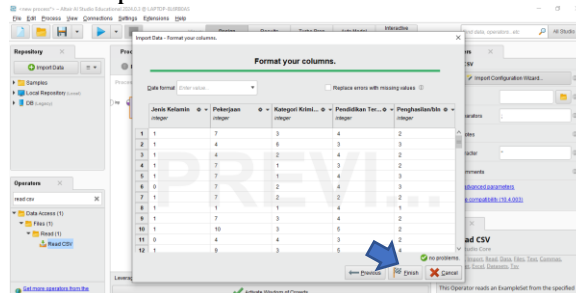
Gambar 5. Memasukkan Data ke Operator Read CSV

d. Pilih data yang ingin dimasukkan



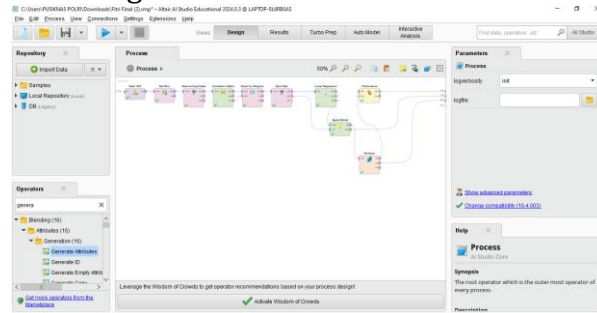
Gambar 6. Pilih Data Yang Ingin di Prediksi

e. Pilih Finish untuk memproses data



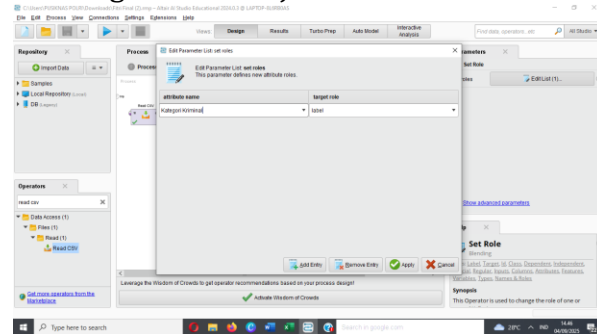
Gambar 7. Data Yang akan Diproses

- f. Tambahkan susunan operator berupa Set Role, Remove Duplication, Correlation Matriks, Select by Weight, Spit Data, Linear Regression, Apply Model, Performance Regression



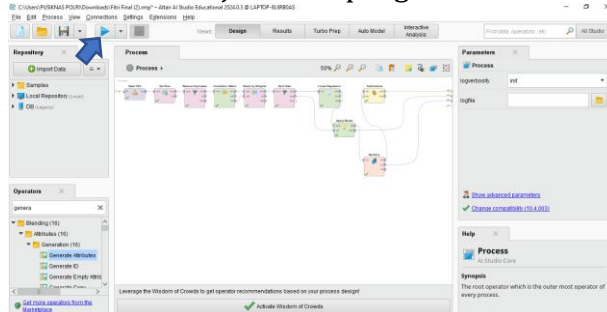
Gambar 8. Menambahkan dan Menyusun Operator Yang Dibutuhkan

- g. Ubah parameters operators Set Role, ubah attribute name menjadi Kategori Kriminal dan ubah target role menjadi label



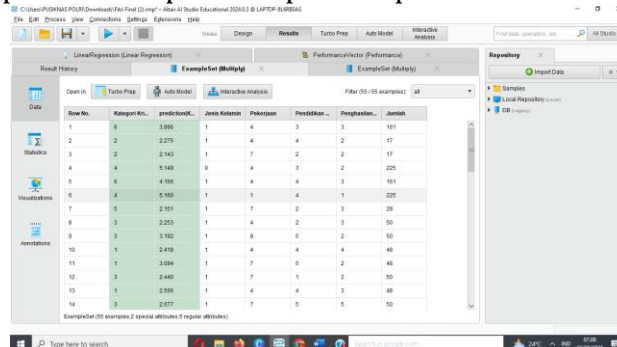
Gambar 9. Menentukan Atribute

- h. Kemudian klik run untuk menjalankan program



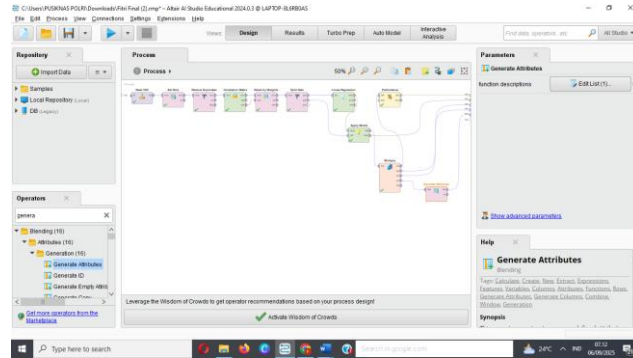
Gambar 10. Menjalankan Program

- i. Berikut merupakan hasil prediksi pada RapidMiner:



Gambar 11. Hasil Prediksi

- j. untuk merubah angka prediksi menjadi hasil bulat tambahkan operator Generate Atributes



Gambar 12. Merubah Hasil Prediksi Menjadi Angka Bulat

k. Hasil prediksi yang telah dibulatkan

The screenshot shows the 'Results' pane of RapidMiner displaying a table of predicted values. The table has columns for 'Row No.', 'Kategori K...', 'prediksiK...', 'Jenis K...', 'Pekerjaan', 'Prediksi...', 'Penghasilan...', and 'Jumlah'. The data is as follows:

Row No.	Kategori K...	prediksiK...	Jenis K...	Pekerjaan	Prediksi...	Penghasilan...	Jumlah
1	6	4	1	4	3	3	161
2	2	2	1	4	4	2	17
3	2	3	1	7	2	2	17
4	4	5	5	4	3	2	225
5	6	4	1	4	4	3	161
6	4	5	1	1	4	1	225
7	5	2	1	7	2	3	28
8	3	2	1	4	2	3	50
9	3	3	1	8	5	2	50
10	1	2	1	4	4	4	48
11	1	3	1	7	1	2	48
12	3	2	1	7	1	2	50
13	1	3	1	4	4	3	48
14	3	3	1	7	5	5	50

Gambar 13. Hasil Prediksi Yang Telah Dibulatkan

l. Berikut merupakan grafik pada Rapidminer:



Gambar 14. Hasil Grafik

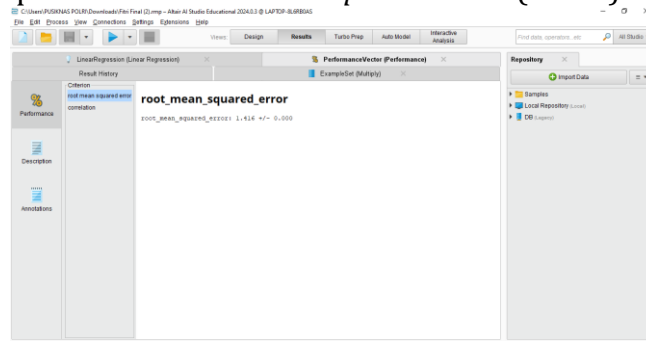
m. Berikut merupakan hasil nilai variabel:

The screenshot shows the 'Results' pane of RapidMiner displaying the results of a linear regression model. The 'LinearRegression' operator is selected, and the 'Description' pane shows the following values:

Variable	Value
Intercept	0.000
Kategori Kriminal	0.000
Pekerjaan	0.000
Prediksi Terakhir	0.000
Penghasilan Terakhir	0.000
Jumlah	0.000

Gambar 15. Hasil Nilai Variabel

n. Berikut merupakan hasil *root mean squared error* (RMSE):



Gambar 16. Hasil Root Mean Squared Error (RMSE)

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa implementasi regresi linear berganda untuk memprediksi tingkat kriminalitas di wilayah Polres Dairi berhasil. Model yang dibangun terbukti memiliki akurasi tinggi (98,6%), dan dapat digunakan untuk memprediksi angka kriminalitas berdasarkan faktor-faktor sosial dan ekonomi. Dimana tingkat kriminalitas paling tinggi terdapat dikategori Kejahatan Seksual yang berjumlah sekitar 262 kasus dan kategori terendah yaitu Pencurian dan Perampokan yang berjumlah sekitar 29 kasus.

Variabel Jenis Kelamin dan Penghasilan/bln menunjukkan pengaruh negatif terhadap tingkat kriminalitas, dimana Jenis Kelamin laki-laki memiliki tingkat kriminalitas yang lebih tinggi dibandingkan dengan perempuan, Demikian pula, penghasilan/bln menunjukkan makin tinggi penghasilan seseorang maka, tingkat kriminalitas menurun.

5. DAFTAR PUSTAKA

- Aditya, A., Auliasari, K., & Pranoto, Y. A. (2023). Metode Regresi Linier Berganda untuk Prediksi Omset Penyewaan Kamera di Joe Kamera. Institut Teknologi Nasional Malang.
- APLIKASINYA Tim Penulis, D., Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Agnes Manuhutu, M., Ridwan, M., Khrisna Wardhani, A., Alim, S., Romli, I., Religia, Y., Tri Octafian, D., Utan Sufandi, U., & Ernawati, I. (2021). DATA MINING. www.penerbitwidina.com
- Aziz, M., & Fatchiyatur, M. (2022). Sistem Prediksi Planned Maintenance Studi Kasus di PT. XYZ. SUBMIT (Jurnal Ilmiah Teknologi Informasi dan Sains), 2(2), 68–74.
- Duli, N. (2019). Metodologi Penelitian Kuantitatif. Deepublish.
- Fadianty, R., & Sriani, S. (2024). Penerapan Data Mining dengan Algoritma Regresi Linear Berganda Untuk Memprediksi Omset Penjualan Minyak Goreng. Building of Informatics, Technology and Science (BITS), 6(2), 1191–1200.
- M. K., & Maysanjaya, I. M. D. (2024). BUKU AJAR DATA MINING (Efitra, Ed.). PT. Sonpedia Publishing Indonesia. https://www.researchgate.net/publication/377415198_BUKU_AJAR_DATA_MINING
- Oktafiani, J. (2018). Tinjauan Pustaka: Pengertian Implementasi. Sereal Untuk, 51(1), 51.
- Om, K., et al. (2024). Implementasi Algoritma Regresi Linear Berganda untuk Prediksi Produksi Padi. Jurnal Ilmu Pertanian.

- Queyrut, S., Bromberg, Y.-D., & Schiavoni, V. (2022). *Pelta*. 12(1), 37–43.
<https://doi.org/10.1145/3565010.3569064>
- Rahmah, A., et al. (2024). Analisis Hubungan Faktor Sosial-Ekonomi terhadap Kriminalitas di Indonesia. *Jurnal Sosial Humaniora*.
- Setyo, W. N., & Wardhana, S. (2019). Implementasi Data Mining Pada Penjualan Produk Di Cv Cahaya Setya Menggunakan Algoritma Fp-Growth. *Petir*, 12(1), 54–63.
- Siswo, A., Syahra, Y., & Yetri, M. (2022). Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda. *STMIK Triguna Dharma*.
- Wijoyo, A., Saputra, A. Y., Ristanti, S., Sya'ban, R., Amalia, M., & Febriansyah, R. (n.d.). Pembelajaran Machine Learning. Retrieved July24,2025 [tautan mencurigakan telah dihapus]
- Yonatan, A. Z. (2025). Angka Kriminalitas Indonesia Tertinggi Ke-2 di ASEAN. *GoodStats*.
- Yuzani, A. D., et al. (2024). Pengaruh Faktor Sosial Ekonomi terhadap Tingkat Kriminalitas: Analisis di Kota Tanjungpinang. *Jurnal Ilmu Hukum*, 1(4), 165–173.