

Laporan Student Performance (Multiple Linear Regression)

Disusun Untuk Memenuhi Tugas Akhir Mata Kuliah Sains Komputasi

Dosen Pengampu :

Farhanna Mar'i, S.Kom., M.Kom.



Disusun oleh :

Daniel Steven Setyawan

24051204042

Program Studi S1 Teknik Informatika

Fakultas Teknik

Universitas Negeri Surabaya

2025

DAFTAR ISI

BAB I LATAR BELAKANG	Error! Bookmark not defined.
BAB II PENJELASAN DATA DAN VISUALISASI	2
BAB III IMPLEMENTASI.....	3
3.1 Hasil Perhitungan Variabel	4
3.2 Perhitungan Mean Squared Error (MSE).....	5
3.3 Hitung Root Mean Squared Error (RMSE)	6
3.4 Mean Absolute Percentage Error (MAPE)	7
3.5 Kode Octave.....	8
BAB IV HASIL ANALISIS DAN KESIMPULAN	3

BAB I

LATAR BELAKANG

Dalam penelitian ini digunakan empat variabel bebas yang meliputi Jam Belajar, Nilai Sebelumnya, Jam Tidur, dan Jumlah Latihan Soal, dengan Performance Index sebagai variabel terikat yang menjadi fokus analisis. Pemilihan keempat variabel bebas tersebut didasarkan pada pertimbangan bahwa aspek-aspek tersebut merupakan representasi dari kebiasaan belajar, kemampuan dasar, kualitas istirahat, serta intensitas persiapan mahasiswa dalam menghadapi kegiatan akademik, sehingga secara teori maupun praktik sering dianggap memiliki hubungan yang erat dengan capaian akademik.

Data yang diperoleh kemudian diolah melalui serangkaian tahapan analisis menggunakan metode Regresi Linier Berganda untuk mengetahui bagaimana masing-masing variabel bebas berkontribusi secara simultan terhadap Performance Index. Dalam proses analisis, digunakan pendekatan matematis melalui Metode Cramer untuk menentukan estimasi nilai koefisien regresi dari setiap variabel bebas. Metode Cramer dipilih karena memberikan prosedur penyelesaian yang sistematis dan memungkinkan peneliti memperoleh nilai koefisien regresi melalui perhitungan determinan matriks, sehingga hasil yang diperoleh bersifat objektif dan dapat diverifikasi kembali melalui tahapan perhitungan manual maupun komputasi.

Setelah nilai koefisien regresi diperoleh, langkah berikutnya adalah melakukan proses visualisasi melalui teknik plotting untuk menggambarkan pola hubungan antara variabel bebas dan variabel terikat serta untuk memverifikasi apakah model yang terbentuk sesuai dengan kecenderungan data yang diamati. Plotting juga dilakukan untuk menilai tingkat linearitas, mendeteksi adanya outlier, serta melihat apakah pola prediksi yang dihasilkan oleh model regresi mencerminkan kenyataan data empiris.

Keseluruhan prosedur analisis ini diharapkan tidak hanya memberikan gambaran kuantitatif mengenai pengaruh masing-masing variabel terhadap Performance Index, tetapi juga membangun model prediksi yang lebih akurat dan dapat digunakan sebagai dasar untuk mengevaluasi efektivitas pola belajar mahasiswa, perencanaan strategi pembelajaran, maupun rekomendasi akademik ke depan. Dengan memadukan pendekatan matematis melalui metode determinan dan analisis grafis melalui plotting, penelitian ini berusaha menghadirkan pemahaman yang lebih komprehensif mengenai dinamika yang memengaruhi performa akademik mahasiswa serta memastikan bahwa model yang dihasilkan memiliki validitas yang kuat baik dari sisi perhitungan maupun visualisasi data.

BAB II

PENJELASAN DATA DAN VISUALISASI

Dataset *Student Performance* ini dirancang untuk menganalisis hubungan antara kebiasaan belajar dan tingkat performa akademik mahasiswa. Tujuannya adalah untuk memprediksi Performance Index (Y) berdasarkan empat faktor utama yang dianggap berpengaruh, yaitu:

1. Jam Belajar (Study Hours / X_1)

Variabel ini menunjukkan berapa lama seorang mahasiswa belajar setiap hari (biasanya dalam satuan jam).

- Semakin banyak jam belajar, biasanya performa akademik cenderung meningkat.
- Namun, jika berlebihan tanpa istirahat cukup, hasilnya bisa menurun.

2. Previous Score (Nilai Sebelumnya / X_2)

Menunjukkan nilai akademik siswa dari ujian atau semester sebelumnya.

- Nilai ini menggambarkan kemampuan dasar siswa.
- Biasanya, siswa dengan nilai sebelumnya tinggi memiliki kecenderungan mempertahankan performa yang baik.

3. Sleep Hours (Jam Tidur / X_3)

Menunjukkan rata-rata jam tidur siswa setiap hari.

- Jam tidur yang ideal berpengaruh terhadap fokus dan daya tangkap.
- Kekurangan tidur dapat menurunkan performa belajar meskipun jam belajar tinggi.

4. Sample Question Papers Practiced (Jumlah Latihan Soal / X_4)

- Variabel ini mencerminkan jumlah latihan soal atau *sample question papers* yang dikerjakan oleh mahasiswa.
- Semakin banyak latihan soal yang dikerjakan, semakin baik pemahaman mahasiswa terhadap pola soal dan konsep materi.
- Latihan soal juga meningkatkan kemampuan problem-solving dan kesiapan menghadapi ujian.

5. Performance Index (Y)

Merupakan variabel dependen atau nilai target yang ingin diprediksi berdasarkan kombinasi keempat faktor sebelumnya (Jam Belajar, Nilai Sebelumnya, Jam Tidur, dan Jumlah Latihan Soal).

- Performance Index menggambarkan tingkat performa akademik mahasiswa secara keseluruhan.

- Nilai ini dipengaruhi oleh kebiasaan belajar, kualitas istirahat, kemampuan dasar, dan kesiapan dalam menghadapi soal.

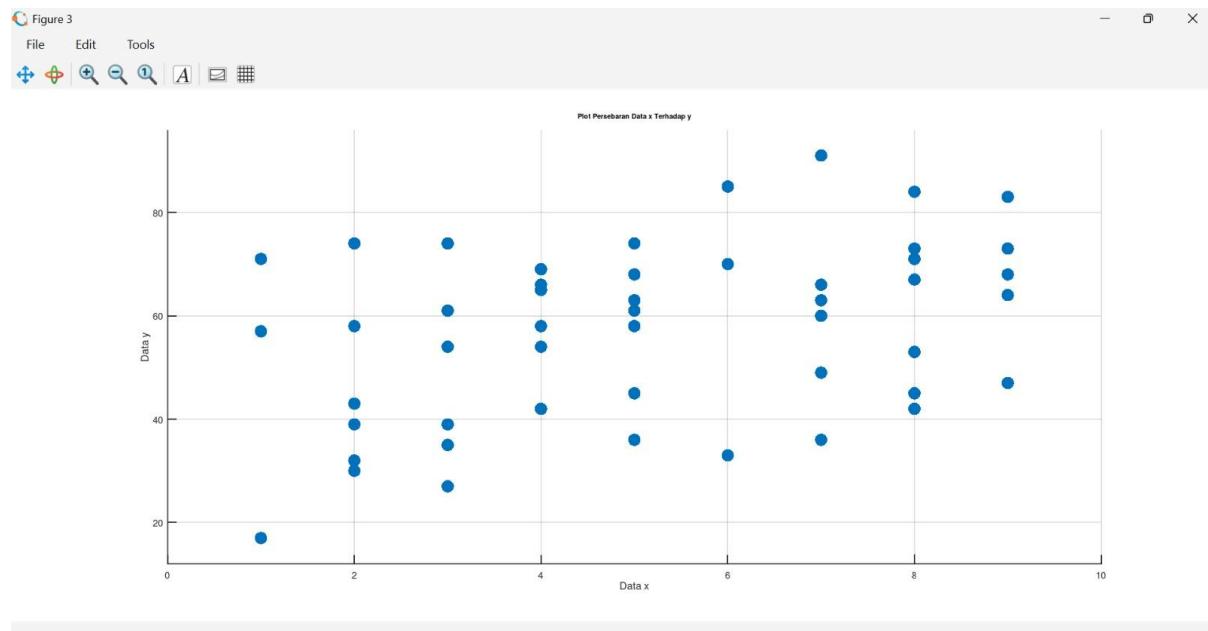
Hours Studied (X)	Previous Scores (x2)	Sleep Hours(X3)	Sample Question(X4)	Performance Index (Y)
7	99	9	1	91
4	82	4	2	65
8	51	7	2	45
5	52	5	2	36
7	75	8	5	66
3	78	9	6	61
7	73	5	6	63
8	45	4	6	42
5	77	8	2	61
4	89	4	0	69
8	91	4	5	84
8	79	6	2	73
3	47	9	2	27
6	47	4	2	33
5	79	7	8	68
2	72	4	3	43
8	73	8	4	67
6	83	7	2	70
2	54	4	9	30
5	75	7	0	63
1	99	4	3	71
6	96	9	0	85
9	74	7	6	73
1	85	5	6	57
3	61	6	3	35
7	62	7	4	49
4	79	8	9	66
9	84	6	6	83
3	94	6	5	74
5	90	4	3	74
3	61	7	3	39
7	44	9	1	36
5	70	6	9	58
9	52	8	1	47
7	67	9	3	60
2	97	9	4	74
4	59	8	3	42
9	72	8	2	68
2	55	4	1	32
9	68	5	3	64

5	62	7	4	45
2	63	6	0	39
4	73	7	0	58
7	46	9	5	36
8	77	6	4	71
3	76	4	3	54
1	43	7	0	17
4	73	4	6	54
2	81	4	3	58
8	61	7	2	53

\

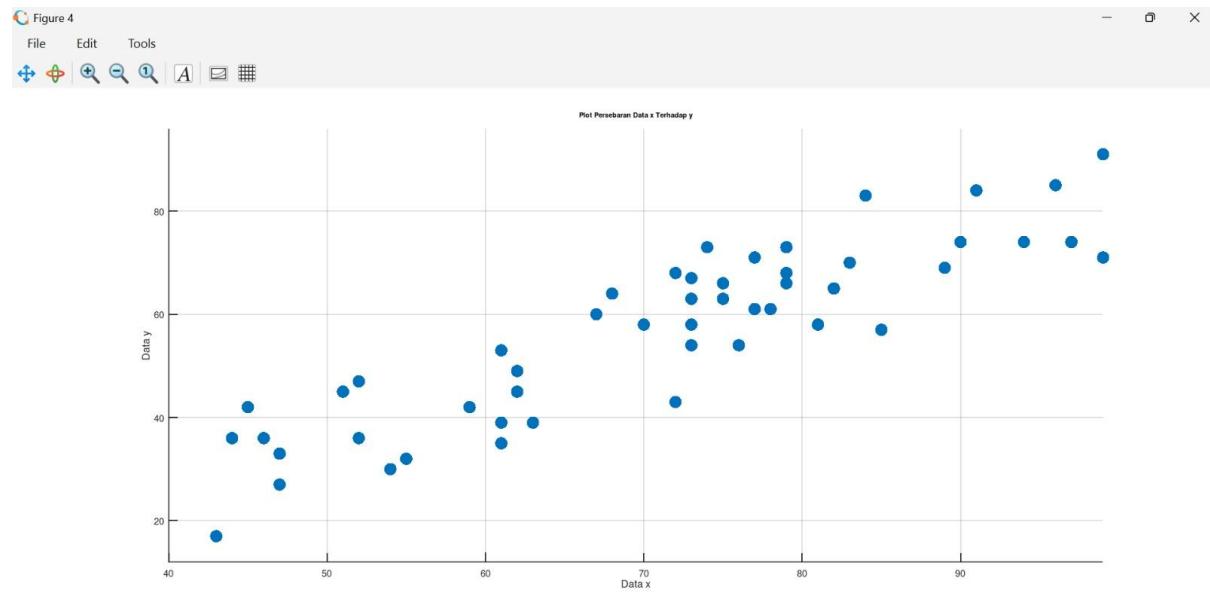
Visualisasi Data menggunakan plotting

1. Gambar 1 – Hours Studied (X1) terhadap Performance Index(Y)



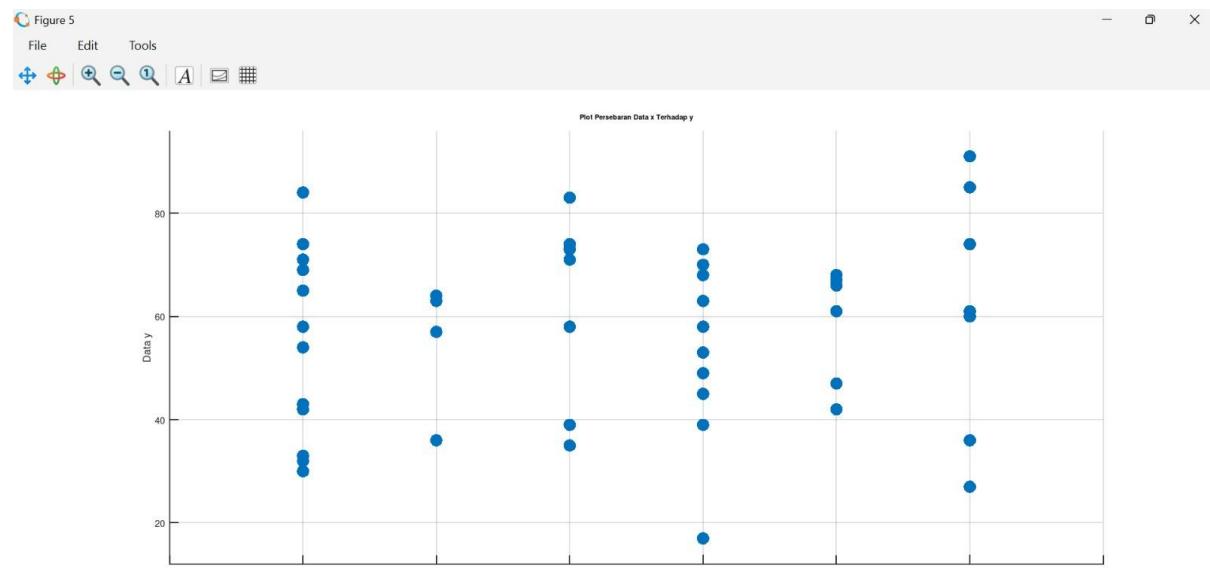
Sebaran titik pada grafik pertama terlihat acak dan tidak membentuk pola tertentu. Hal ini menunjukkan bahwa hubungan antara X_1 dan Y cenderung lemah atau tidak linear, sehingga pengaruhnya perlu diuji lebih lanjut melalui perhitungan regresi.

2. Gambar 2 – Previous Scores (X₂) terhadap Performance Index(Y)



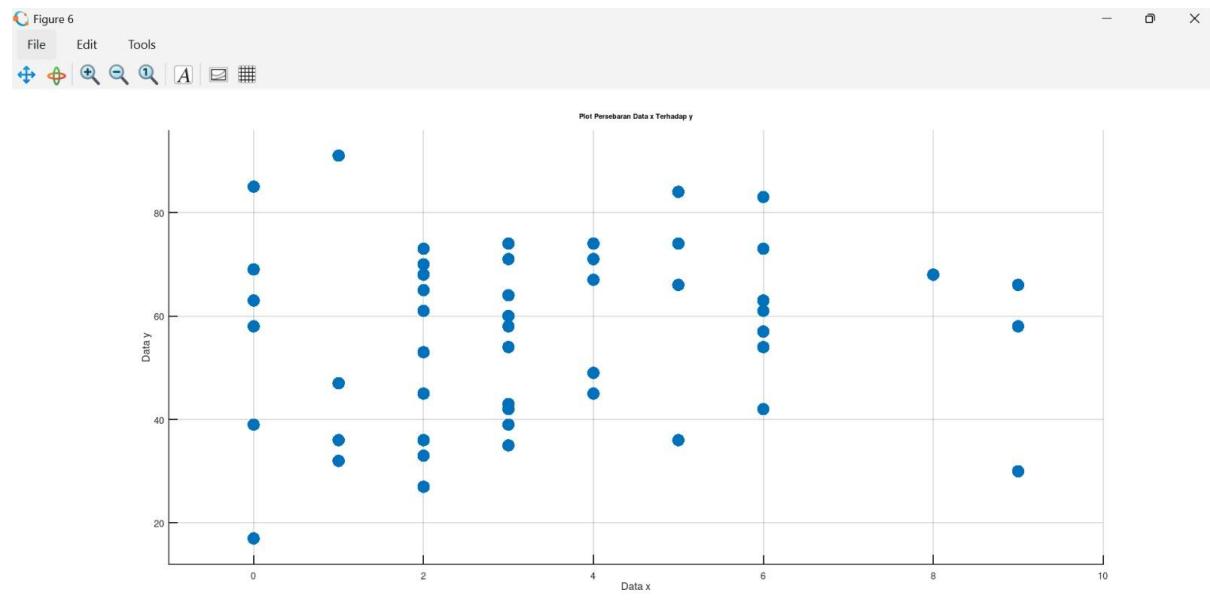
Grafik kedua menunjukkan kecenderungan pola yang lebih teratur, di mana nilai Y meningkat seiring naiknya nilai X₂. Ini mengindikasikan adanya hubungan positif yang lebih kuat dibandingkan variabel lainnya.

3. Gambar 3 – Sleep Hours(X₃) terhadap Performance Index(Y)



Sebaran titik pada grafik ketiga terlihat sedang—tidak sepenuhnya acak, namun juga tidak membentuk garis tren yang jelas. Ini menunjukkan bahwa hubungan X_3 dengan Y bersifat moderat dan memerlukan analisis lanjutan untuk memastikan pengaruhnya.

4. Gambar 4 – Sample Question Papers Practiced (X_4) terhadap Y



Grafik keempat menunjukkan sebaran titik data yang cukup bervariasi dengan sebagian kecil pola yang tampak meningkat pada nilai tertentu. Secara keseluruhan, pola yang terbentuk tidak terlalu jelas, namun terdapat indikasi bahwa peningkatan nilai X_4 dapat diikuti oleh kenaikan nilai Y pada beberapa kelompok data. Hal ini mengisyaratkan adanya kemungkinan pengaruh dari X_4 terhadap Y , meskipun pengaruh tersebut tidak sekuat variabel X_2 . Penyebaran data yang tidak stabil membuat hubungan linear antara kedua variabel ini tidak terlihat kuat pada visualisasi, sehingga analisis lebih lanjut menggunakan perhitungan regresi sangat diperlukan. Plot ini memberikan gambaran awal bahwa X_4 mungkin berkontribusi, tetapi sifat kontribusinya perlu dibuktikan secara matematis.

BAB III

IMPLEMENTASI

3.1 Hasil Perhitungan Variabel

X1Y	X2Y	X3Y	X4Y	X1X2	X1X3	X1X4	X2X3	X2X4	X3X4
637	9009	819	91	693	63	7	891	99	9
260	5330	260	130	328	16	8	328	164	8
360	2295	315	90	408	56	16	357	102	14
180	1872	180	72	260	25	10	260	104	10
462	4950	528	330	525	56	35	600	375	40
183	4758	549	366	234	27	18	702	468	54
441	4599	315	378	511	35	42	365	438	30
336	1890	168	252	360	32	48	180	270	24
305	4697	488	122	385	40	10	616	154	16
276	6141	276	0	356	16	0	356	0	0
672	7644	336	420	728	32	40	364	455	20
584	5767	438	146	632	48	16	474	158	12
81	1269	243	54	141	27	6	423	94	18
198	1551	132	66	282	24	12	188	94	8
340	5372	476	544	395	35	40	553	632	56
86	3096	172	129	144	8	6	288	216	12
536	4891	536	268	584	64	32	584	292	32
420	5810	490	140	498	42	12	581	166	14
60	1620	120	270	108	8	18	216	486	36
315	4725	441	0	375	35	0	525	0	0
71	7029	284	213	99	4	3	396	297	12
510	8160	765	0	576	54	0	864	0	0
657	5402	511	438	666	63	54	518	444	42
57	4845	285	342	85	5	6	425	510	30
105	2135	210	105	183	18	9	366	183	18
343	3038	343	196	434	49	28	434	248	28
264	5214	528	594	316	32	36	632	711	72
747	6972	498	498	756	54	54	504	504	36
222	6956	444	370	282	18	15	564	470	30
370	6660	296	222	450	20	15	360	270	12
117	2379	273	117	183	21	9	427	183	21
252	1584	324	36	308	63	7	396	44	9
290	4060	348	522	350	30	45	420	630	54
423	2444	376	47	468	72	9	416	52	8
420	4020	540	180	469	63	21	603	201	27
148	7178	666	296	194	18	8	873	388	36
168	2478	336	126	236	32	12	472	177	24
612	4896	544	136	648	72	18	576	144	16
64	1760	128	32	110	8	2	220	55	4
576	4352	320	192	612	45	27	340	204	15

225	2790	315	180	310	35	20	434	248	28
78	2457	234	0	126	12	0	378	0	0
232	4234	406	0	292	28	0	511	0	0
252	1656	324	180	322	63	35	414	230	45
568	5467	426	284	616	48	32	462	308	24
162	4104	216	162	228	12	9	304	228	12
17	731	119	0	43	7	0	301	0	0
216	3942	216	324	292	16	24	292	438	24
116	4698	232	174	162	8	6	324	243	12
424	3233	371	106	488	56	16	427	122	14
15438	212160	18160	9940	18251	1715	896	22504	12299	1066

X1^2	X2^2	X3^2	X4^2
49	9801	81	1
16	6724	16	4
64	2601	49	4
25	2704	25	4
49	5625	64	25
9	6084	81	36
49	5329	25	36
64	2025	16	36
25	5929	64	4
16	7921	16	0
64	8281	16	25
64	6241	36	4
9	2209	81	4
36	2209	16	4
25	6241	49	64
4	5184	16	9
64	5329	64	16
36	6889	49	4
4	2916	16	81
25	5625	49	0
1	9801	16	9
36	9216	81	0
81	5476	49	36
1	7225	25	36
9	3721	36	9
49	3844	49	16
16	6241	64	81
81	7056	36	36
9	8836	36	25
25	8100	16	9
9	3721	49	9
49	1936	81	1

25	4900	36	81
81	2704	64	1
49	4489	81	9
4	9409	81	16
16	3481	64	9
81	5184	64	4
4	3025	16	1
81	4624	25	9
25	3844	49	16
4	3969	36	0
16	5329	49	0
49	2116	81	25
64	5929	36	16
9	5776	16	9
1	1849	49	0
16	5329	16	36
4	6561	16	9
64	3721	49	4
1656	263279	2195	873

Pada tahap ini dilakukan perhitungan seluruh komponen yang diperlukan dalam metode **Cramer** dan penyusunan matriks untuk **Multiple Linear Regression (MLR)**. Perhitungan mencakup hasil perkalian antara setiap variabel bebas (X_1 , X_2 , X_3 , X_4) dengan variabel dependen (Y), perkalian antarvariabel bebas, serta nilai kuadrat masing-masing variabel. Semua nilai tersebut diperlukan untuk membentuk persamaan normal (normal equation) dalam regresi linier berganda dan memastikan proses perhitungan determinan berjalan dengan benar.

1. Perhitungan Perkalian X_1Y , X_2Y , X_3Y , dan X_4Y

Bagian ini menampilkan hasil perkalian langsung antara setiap variabel bebas dengan variabel Y. Tujuan perhitungan ini adalah untuk memperoleh total $\Sigma(X_1Y)$, $\Sigma(X_2Y)$, $\Sigma(X_3Y)$, dan $\Sigma(X_4Y)$, yang nantinya digunakan dalam penyusunan persamaan utama regresi. Nilai-nilai pada kolom menunjukkan kontribusi masing-masing sampel.

2. Perkalian Antar Variabel Bebas (X_1X_2 , X_1X_3 , X_1X_4 , X_2X_3 , X_2X_4 , X_3X_4)

Perhitungan ini diperlukan untuk mengetahui hubungan antarvariabel bebas dalam model. Nilai $\Sigma(X_iX_j)$ menunjukkan tingkat interaksi dan akan dipakai pada bagian koefisien dalam

matriks normal. Hasil pada baris terakhir menjadi nilai dasar dalam matriks utama metode Cramer, semakin besar nilai total, semakin kuat korelasi potensial antarvariabel.

3. Nilai Kuadrat Variabel ($X_1^2, X_2^2, X_3^2, X_4^2$)

Nilai kuadrat digunakan untuk menyusun komponen diagonal pada matriks persamaan regresi linier. $\Sigma(X_1^2)$, $\Sigma(X_2^2)$, $\Sigma(X_3^2)$, dan $\Sigma(X_4^2)$ digunakan pada posisi utama matriks, Nilai total pada baris akhir menjadi nilai kunci dalam proses perhitungan determinan utama.

4. Baris Total

Baris terakhir pada tabel merupakan **Jumlah dari seluruh data**, yaitu:

- $\Sigma X_1 Y = \mathbf{15438}$
- $\Sigma X_2 Y = \mathbf{212160}$
- $\Sigma X_3 Y = \mathbf{18160}$
- $\Sigma X_4 Y = \mathbf{9940}$
- $\Sigma(X_1 X_2) = \mathbf{18251}$
- $\Sigma(X_1 X_3) = \mathbf{1715}$
- $\Sigma(X_1 X_4) = \mathbf{896}$
- $\Sigma(X_2 X_3) = \mathbf{22504}$
- $\Sigma(X_2 X_4) = \mathbf{12299}$
- $\Sigma(X_3 X_4) = \mathbf{1066}$
- $\Sigma(X_1^2) = \mathbf{1656}$
- $\Sigma(X_2^2) = \mathbf{263279}$
- $\Sigma(X_3^2) = \mathbf{2195}$
- $\Sigma(X_4^2) = \mathbf{873}$

Nilai-nilai inilah yang akan kamu gunakan untuk **membentuk matriks koefisien dan matriks hasil**, yang selanjutnya diselesaikan menggunakan **Metode Cramer** untuk mendapatkan nilai koefisien $\beta_0, \beta_1, \beta_2, \beta_3$, dan β_4 .

Pembentukan Matriks (Normal Equation Matrix)

Pada tahap implementasi metode Multiple Linear Regression (MLR) menggunakan pendekatan Cramer's Rule, langkah pertama yang dilakukan adalah membentuk matriks normal (normal equation matrix) berdasarkan hasil pengolahan data. Matriks ini disusun dari berbagai nilai penjumlahan variabel ($\sum X_1$, $\sum X_2$, ...), kuadrat variabel ($\sum X_1^2$, $\sum X_2^2$, ...), serta hasil perkalian silang antarvariabel ($\sum X_1 X_2$, $\sum X_2 X_3$, dan seterusnya). Seluruh nilai tersebut diambil dari tabel hasil komputasi sebelumnya.

Matriks normal untuk model regresi dengan empat variabel bebas (X_1 , X_2 , X_3 , X_4) berbentuk:

n	x1	x2	x3	x4
x1	x_1^2	$x_2 x_1$	$x_3 x_1$	$x_4 x_1$
x2	$x_1 x_2$	x_2^2	$x_3 x_2$	$x_4 x_2$
x3	$x_1 x_3$	$x_2 x_3$	x_3^2	$x_4 x_3$
x4	$x_1 x_4$	$x_2 x_4$	$x_3 x_4$	x_4^2

Setelah seluruh nilai penjumlahan dihitung, matriks tersebut terisi sebagai berikut:

50	260	3545	319	171
260	1656	18251	1715	896
3545	18251	263279	22504	12299
319	1715	22504	2195	1066
171	896	12299	1066	873

Perhitungan Matriks A, A1, A2, A3, A4, dan A5 (Metode Cramer):

Pada metode Cramer's Rule untuk menyelesaikan Multiple Linear Regression, langkah utama yang dilakukan adalah:

1. Membentuk matriks koefisien utama (A).
2. Membentuk matriks pengganti A_1, A_2, A_3, A_4, A_5 dengan mengganti kolom pertama, kedua, ketiga, keempat, dan kelima menggunakan kolom hasil (H).

H	Total
2829	Y
15438	$X_1 Y$
212160	$X_2 Y$
18160	$X_3 Y$
9940	$X_4 Y$

3. Menghitung determinan setiap matriks ($\text{DET}(A)$, $\text{DET}(A_1)$, ...).

4. Hasil determinan tersebut nanti digunakan untuk menghitung koefisien regresi

$$B_0, B_1, B_2, B_3, B_4 \text{ menggunakan rumus: } B_i = \frac{\text{DET}(A_i)}{\text{DET}(A)}$$

1. Matriks A (Matriks Koefisien Utama):

Matriks A adalah matriks utama yang berisi:

1. Baris pertama: $n, \Sigma X_1, \Sigma X_2, \Sigma X_3, \Sigma X_4$
2. Baris kedua: $\Sigma X_1, \Sigma X_1^2, \Sigma X_1 X_2, \Sigma X_1 X_3, \Sigma X_1 X_4$
3. Baris ketiga hingga kelima: hasil perkalian silang dan kuadrat antarvariabel

50	260	3545	319	171
260	1656	18251	1715	896
3545	18251	263279	22504	12299
319	1715	22504	2195	1066
171	896	12299	1066	873

Determinannya: $\text{DET}(A) = 7,523,285,062,712.03$

2. Matriks A1 (Kolom 1 diganti H):

2829	260	3545	319	171
15438	1656	18251	1715	896
212160	18251	263279	22504	12299
18160	1715	22504	2195	1066
9940	896	12299	1066	873

Determinannya: $\text{DET}(A_1) = -257,504,197,661,437$

3. Matriks A2 (Kolom 2 diganti H):

50	2829	3545	319	171
260	15438	18251	1715	896
3545	212160	263279	22504	12299
319	18160	22504	2195	1066
171	9940	12299	1066	873

Determinannya: $\text{DET}(A_2) = 21,951,425,932,028$

4. Matriks A3 (Kolom 3 diganti H):

50	260	2829	319	171
260	1656	15438	1715	896
3545	18251	212160	22504	12299
319	1715	18160	2195	1066
171	896	9940	1066	873

Determinannya: $\text{DET}(A_3) = 7,637,048,734,695.01$

5. Matriks A4 (Kolom 4 diganti H):

50	260	3545	2829	171
260	1656	18251	15438	896
3545	18251	263279	212160	12299
319	1715	22504	18160	1066
171	896	12299	9940	873

Determinannya: $\text{DET}(A_4) = 3,228,493,096,044.84$

6. Matriks A5 (Kolom 5 diganti H):

50	260	3545	319	2829
260	1656	18251	1715	15438
3545	18251	263279	22504	212160
319	1715	22504	2195	18160
171	896	12299	1066	9940

Determinannya: $\text{DET}(A_5) = 2,035,003,046,927.98$

Kemudian hitung determinanya menggunakan aturan cramer, dengan rumus:

$$B_i = \frac{\text{DET}(A_i)}{\text{DET}(A)}$$

$$1. A (\text{Intercept}) A = \frac{\text{DET}(A_1)}{\text{DET}(A)} = \frac{-257.504.197.661.437}{7.523.285.062.712,03}$$

$$\text{Hitung: } A = -34,22762736$$

$$2. B1 (\text{Koefisien untuk } X_1 / \text{Jam Belajar}) B1 = \frac{\text{DET}(A_2)}{\text{DET}(A)} = \frac{21.951.425.932.028}{7.523.285.062.712,03}$$

$$\text{Hitung: } B1 = 2,917797976$$

$$3. B2 (\text{Koefisien untuk } X_2 / \text{Previous Score}) B2 = \frac{\text{DET}(A_3)}{\text{DET}(A)} = \frac{7.637.048.734.695,01}{7.523.285.062.712,03}$$

$$\text{Hitung: } B2 = 1,015121542$$

$$4. B3 (\text{Koefisien untuk } X_3 / \text{Jam tidur}) B3 = \frac{\text{DET}(A_4)}{\text{DET}(A)} = \frac{3.228.493.096.044,84}{7.523.285.062.712,03}$$

$$\text{Hitung: } B3 = 0,429133426$$

$$4. B4 (\text{Koefisien untuk } X_4 / \text{Jam tidur}) B4 = \frac{\text{DET}(A_5)}{\text{DET}(A)} = \frac{2.035.003.046.927,98}{7.523.285.062.712,03}$$

$$\text{Hitung: } B4 = 0,270493944$$

Berdasarkan hasil perhitungan determinan matriks menggunakan Metode Cramer, diperoleh koefisien regresi sebagai berikut:

- $A = -34,22762736$

- $B_1 = 2,917797976$ (Jam Belajar / X_1)
- $B_2 = 1,015121542$ (Nilai Sebelumnya / X_2)
- $B_3 = 0,429133426$ (Jam Tidur / X_3)
- $B_4 = 0,270493944$ (Latihan Soal / X_4)

Sehingga diperoleh persamaan regresi:

$$Y = -34,2276 + 2,9178X_1 + 1,0151X_2 + 0,4291X_3 + 0,2705X_4$$

Persamaan ini digunakan untuk memprediksi nilai Performance Index (Y) berdasarkan empat variabel independen.

1. Perhitungan Nilai Prediksi (Y_pred)

Nilai Y_pred diperoleh dengan memasukkan setiap nilai X ke dalam rumus:

$$Y_{\text{pred}} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

Contoh perhitungannya:

$$Y_{\text{pred}} = -34,2276 + (2,9178X_1) + (1,0151X_2) + (0,4291X_3) + (0,2705X_4)$$

2. Residual (r) dan Residual Kuadrat (r^2)

Residual dihitung sebagai: $r = Y_{\text{actual}} - Y_{\text{pred}}$

Dan residual kuadrat: $r^2 = (Y_{\text{actual}} - Y_{\text{pred}})^2$

3. Total Error ($\sum r^2$)

$$\sum r^2 = 232,4079196$$

Ini merupakan total *Sum of Squared Residuals (SSR)* atau *SSE*, yaitu ukuran akurasi prediksi model. Semakin kecil nilai ini, semakin baik model regresinya. Berikut tabel perhitungan nilai prediksi, residual dan total error.

Y_{pred}	r	r^2	$\frac{y_i - \hat{y}_i}{y_i}$
90,82669	-0,17331	0,030037769	0,001905
62,67056	-2,32944	5,426297016	0,035838
43,88989	-1,11011	1,23234625	0,024669
35,29335	-0,70665	0,499354445	0,019629
65,76414	-0,23586	0,055629212	0,003574
57,56745	-3,43255	11,78241536	0,056271
62,4465	-0,5535	0,306364265	0,008786
36,51176	-5,48824	30,12078323	0,130672

61,95879	0,958789	0,919275722	0,015718
69,50592	0,505915	0,255950491	0,007332
83,20735	-0,79265	0,628293248	0,009436
71,88416	-1,11584	1,245101505	0,015285
26,09868	-0,90132	0,812377951	0,033382
32,70641	-0,29359	0,086197036	0,008897
63,5599	-4,4401	19,71450282	0,065296
46,41325	3,413253	11,65029829	0,079378
66,6517	-0,3483	0,121315373	0,005199
70,53818	0,538182	0,289640379	0,007688
28,14107	-1,85893	3,455637211	0,061964
59,49941	-3,50059	12,2541152	0,055565
70,90374	-0,09626	0,009266567	0,001356
84,59303	-0,40697	0,165625088	0,004788
70,15548	-2,84452	8,091279651	0,038966
57,12117	0,121169	0,014681885	0,002126
39,02298	4,022981	16,1843777	0,114942
52,13843	3,138428	9,849730745	0,06405
61,07123	-4,92877	24,29273643	0,074678
79,87756	-3,12244	9,749603312	0,03762
72,52199	-1,47801	2,184507396	0,019973
73,43884	-0,56116	0,314906138	0,007583
39,45211	0,452115	0,204407631	0,011593
34,72451	-1,27549	1,626881976	0,03543
53,99467	-4,00533	16,04266018	0,069057
48,25194	1,251942	1,567358839	0,026637
58,0723	-1,9277	3,716017141	0,032128
73,93696	-0,06304	0,003974166	0,000852
40,7688	-1,2312	1,515846201	0,029314
68,55437	0,554373	0,307329279	0,008153
29,15619	-2,84381	8,087271735	0,088869
63,20649	-0,79351	0,629663798	0,012399
46,30283	1,302832	1,697371523	0,028952
38,13543	-0,86457	0,747487678	0,022169
54,55137	-3,44863	11,89304127	0,059459

36,75475	0,75475	0,569647937	0,020965
69,85392	-1,14608	1,313509105	0,016142
53,39154	-0,60846	0,370226645	0,011268
15,34433	-1,65567	2,741240135	0,097392
53,26397	-0,73603	0,541738947	0,01363
55,54935	-2,45065	6,005699104	0,042253
54,0411	1,041105	1,083898587	0,019643
		232,4079196	1,658871

3.2 Perhitungan Mean Squared Error (MSE)

MSE dihitung dengan cara membagi total selisih kuadrat antara nilai aktual dan nilai prediksi dengan jumlah data. Dari hasil perhitungan, total error kuadrat sebesar **232,4079**, sehingga diperoleh nilai **MSE = 4,648**.

$$MSE = \frac{R}{n} \quad MSE = \frac{232.4079}{50} = 4.648$$

Nilai ini menunjukkan bahwa rata-rata kesalahan kuadrat model relatif kecil, sehingga model cukup stabil dalam memprediksi nilai Performance Index. Berikut adalah tabel perhitungan validasi untuk regresi polinomial derajat 2 dan derajat 3 untuk menentukan model mana yang memiliki *error* lebih rendah:

3.3 Root Mean Squared Error (RMSE)

RMSE diperoleh dari akar kuadrat nilai MSE. Dengan MSE sebesar 4,648, nilai **RMSE = 2,1559**. RMSE menggunakan satuan yang sama dengan variabel dependen, sehingga dapat diinterpretasikan bahwa rata-rata deviasi nilai prediksi terhadap nilai sebenarnya adalah sekitar **±2,15 poin**.

$$RMSE = \sqrt{MSE} \quad RMSE = \sqrt{4.648} = 2.1559$$

Nilai RMSE yang rendah memperlihatkan bahwa tingkat kesalahan prediksi model berada dalam batas yang wajar.

3.4 Mean Absolute Percentage Error (MAPE)

MAPE dihitung dengan menjumlahkan seluruh nilai kesalahan absolut persentase antara nilai aktual dan prediksi, kemudian dibagi dengan jumlah data. Dari hasil perhitungan, total persentase error sebesar 1,65887 sehingga menghasilkan **MAPE = 0,03318** atau **3,31%**.

$$MAPE = \frac{1}{n} \sum \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad MAPE = \frac{1.65887}{50} = 0.03318 = 3.31\%$$

Nilai MAPE di bawah 10% menunjukkan bahwa akurasi model sangat baik dan model mampu memprediksi nilai Performance Index secara konsisten.

3.5 Kode Octave

```
x = [ ...
7 4 8 5 7 3 7 8 5 4 ...
8 8 3 6 5 2 8 6 2 5 ...
1 6 9 1 3 7 4 9 3 5 ...
3 7 5 9 7 2 4 9 2 9 ...
5 2 4 7 8 3 1 4 2 8 ...
];
y = [ ...
91 65 45 36 66 61 63 42 61 69 ...
84 73 27 33 68 43 67 70 30 63 ...
71 85 73 57 35 49 66 83 74 74 ...
39 36 58 47 60 74 42 68 32 64 ...
45 39 58 36 71 54 17 54 58 53 ...
];
figure;
scatter(x, y, 'filled');
xlabel("Data x");
ylabel("Data y");
title("Plot Persebaran Data x Terhadap y");
% Set rentang sumbu X menjadi 15 sampai 30
xlim([-1 10]);
% Y tetap otomatis atau bisa ditambah padding
ylim([min(y)-5, max(y)+5]);
grid on;
```

Penjelasan Kode Plotting Pada Octave

- `x = [...`
- `7 4 8 5 7 3 7 8 5 4 ...`
- `8 8 3 6 5 2 8 6 2 5 ...`
- `1 6 9 1 3 7 4 9 3 5 ...`
- `3 7 5 9 7 2 4 9 2 9 ...`
- `5 2 4 7 8 3 1 4 2 8 ...`
- `];`

Bagian ini berfungsi untuk memasukkan data variabel x ke dalam bentuk array. Data ini terdiri dari 50 nilai yang mewakili faktor atau variabel independen yang akan dianalisis hubungannya terhadap variabel y.

- `y = [...`
- `91 65 45 36 66 61 63 42 61 69 ...`
- `84 73 27 33 68 43 67 70 30 63 ...`
- `71 85 73 57 35 49 66 83 74 74 ...`
- `39 36 58 47 60 74 42 68 32 64 ...`
- `45 39 58 36 71 54 17 54 58 53 ...`
- `];`

Bagian ini berfungsi untuk memasukkan data variabel **y** sebagai *nilai target* atau *variabel dependen*. Terdiri dari 50 nilai yang sebarannya akan divisualisasikan terhadap variabel **x**.

- `figure;`

Perintah ini digunakan untuk membuka atau membuat jendela grafik baru sehingga plot tidak tercampur dengan grafik lain.

- `scatter(x, y, 'filled');`

`scatter(x, y)` digunakan untuk membuat diagram persebaran antara data **x** dan **y**.

Argumen 'filled' memberi efek *mengisi warna* pada titik-titik plot agar lebih jelas dan mudah dilihat.

- `xlabel("Data x");`
- `ylabel("Data y");`
- `title("Plot Persebaran Data x Terhadap y");`

`xlabel()` memberikan label pada sumbu **X**, `ylabel()` memberikan label pada sumbu **Y**, `title()` memberikan judul pada grafik, sehingga pembaca mengetahui hubungan apa yang sedang divisualisasikan.

- `xlim([-1 10]);`

Perintah `xlim()` berfungsi untuk menentukan batas minimum dan maksimum pada sumbu X. Dalam code ini, rentang ditetapkan antara -1 sampai 10, agar plot terlihat lebih proporsional dan tidak terlalu mepet dengan batas grafik.

- `ylim([min(y)-5, max(y)+5]);`

`min(y)` mengambil nilai terkecil dari data **y**, `max(y)` mengambil nilai terbesar, Tambahan -5 dan +5 memberikan *padding* sehingga titik tidak menempel ke tepi grafik.

- `grid on;`

Perintah ini menampilkan garis grid pada grafik sehingga posisi setiap titik lebih mudah diamati dan dibandingkan, terutama ketika menjelaskan hubungan antar data.

BAB 4

HASIL ANALISIS DAN KESIMPULAN

Berdasarkan hasil analisis dan perhitungan Multiple Linear Regression menggunakan Metode Cramer, dapat disimpulkan bahwa keempat variabel independen, yaitu Jam Belajar, Nilai Sebelumnya, Jam Tidur, dan Jumlah Latihan Soal, secara simultan berpengaruh terhadap Performance Index mahasiswa.

Nilai Sebelumnya (X_2) merupakan variabel dengan pengaruh paling kuat dan konsisten terhadap performa akademik, diikuti oleh Jam Belajar (X_1). Hal ini menunjukkan bahwa kemampuan dasar dan konsistensi belajar memainkan peran utama dalam pencapaian akademik. Jam Tidur (X_3) dan Latihan Soal (X_4) juga memberikan kontribusi positif, meskipun dengan pengaruh yang lebih moderat.

Model regresi yang dihasilkan memiliki tingkat error yang rendah dengan nilai RMSE sebesar $\pm 2,15$ dan MAPE sebesar 3,31%, yang menandakan bahwa model memiliki tingkat akurasi yang baik dalam memprediksi Performance Index. Oleh karena itu, model ini dapat digunakan sebagai dasar untuk mengevaluasi pola belajar mahasiswa serta membantu dalam perencanaan strategi pembelajaran yang lebih efektif.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan matematis melalui Metode Cramer yang dipadukan dengan analisis visualisasi data mampu menghasilkan model prediksi yang valid dan reliabel. Model ini diharapkan dapat memberikan kontribusi dalam pengambilan keputusan akademik dan pengembangan sistem evaluasi performa mahasiswa di masa mendatang.

DAFTAR PUSTAKA

Nikhil7280. (n.d.). (2025). *Student performance: Multiple linear regression* [Data set]. Kaggle.
<https://www.kaggle.com/datasets/nikhil7280/student-performance-multiple-linear-regression>