



Kelompok 5

LINEAR DAN POLYNOMIAL REGRESSION

Sadinal Mufti (2208107010007)
M. Agradika Ridhal Eljatin (2208107010020)
Jihan Nabilah (2208107010035)
Firjatullah Afny Abus (2208107010059)
Athar Rayyan Muhammad (2208107010074)



Latar Belakang

Dalam era digital, analisis data sangat penting untuk pengambilan keputusan. Salah satu metode yang umum digunakan adalah regresi, baik linear maupun polinomial, untuk memahami hubungan antar variabel dan membuat prediksi.

Pada tugas ini, kami menganalisis dataset performa akademik siswa yang mencakup kebiasaan belajar, kehadiran, nilai sebelumnya, kondisi sosial ekonomi, dan akses belajar. Dengan regresi, kami ingin mengidentifikasi faktor yang mempengaruhi keberhasilan akademik dan memberikan wawasan bagi pendidik dalam merancang strategi pembelajaran yang efektif.





Deskripsi Dataset

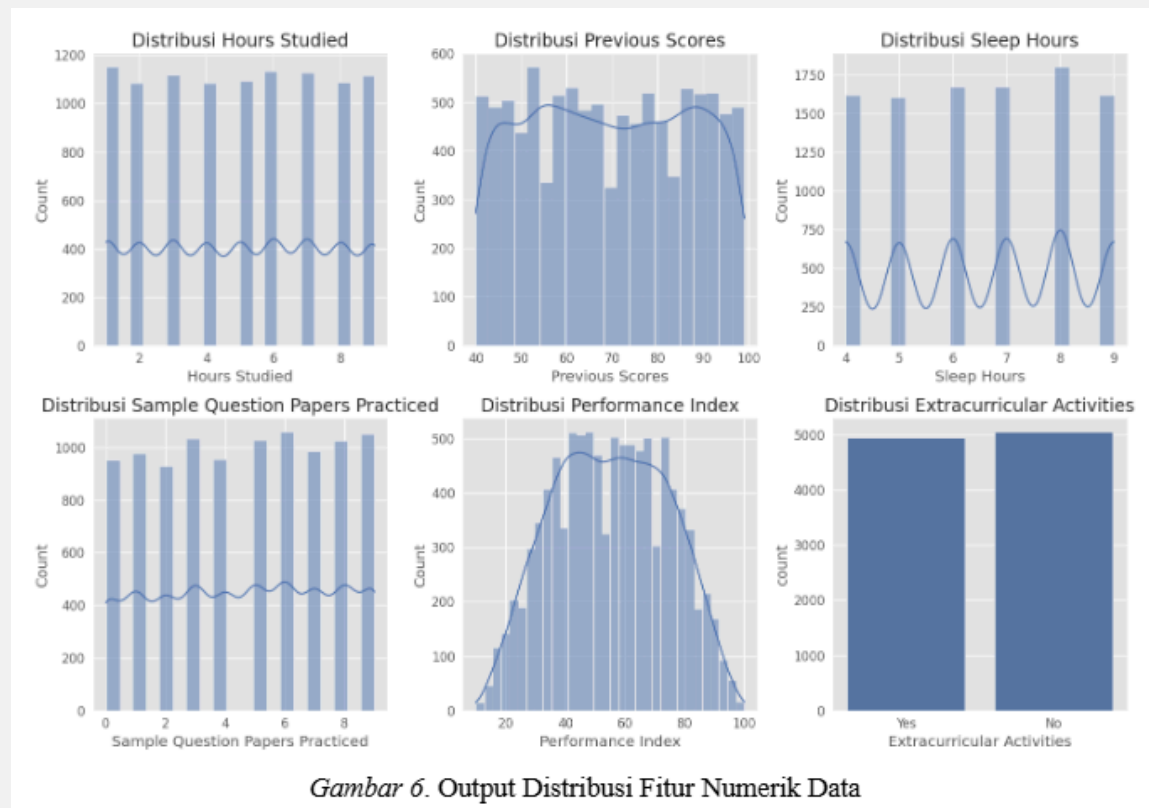
- Nama: Student Performance (Multiple Linear Regression)
- Sumber: [Kaggle](#)
- Jumlah Data: 10.000 sampel, 6 fitur
- Format: CSV
- Label: Performance Index (10–100)

Fitur Utama

- Hours Studied: Jam belajar siswa
- Previous Scores: Nilai ujian sebelumnya
- Extracurricular Activities: Keikutsertaan dalam kegiatan ekstrakurikuler
- Sleep Hours: Rata-rata jam tidur harian
- Sample Question Papers Practiced: Jumlah soal latihan yang dikerjakan

Analisis Awal Data

- Distribusi Fitur Numerik dan Analisis



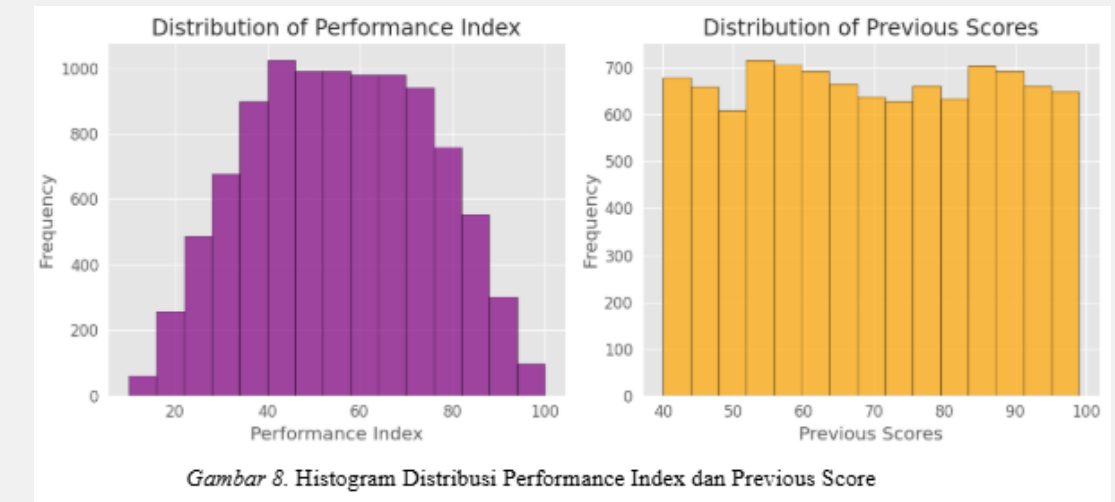
Dapat dilihat bahwa sebagian besar data terdistribusi normal atau seragam, dengan sedikit kecenderungan ke kanan pada Previous Scores. Tidak ada outlier yang signifikan dalam grafik ini, menunjukkan data relatif bersih.

- Perbandingan Variabel Dependen dan Independen

Previous Scores	Performance Index
0	694457.0
	552248.0

Gambar 7. Hasil Perbandingan Kedua Variabel

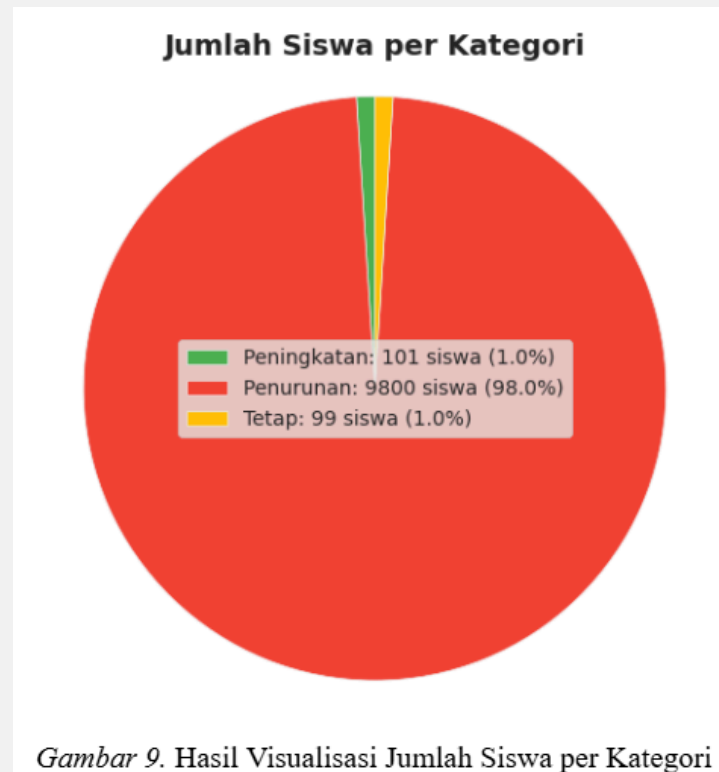
Variabel target 'Performance Index' akan dibandingkan dengan total 'Previous Score' untuk melihat tren performa siswa dari waktu ke waktu.



Ujian sebelumnya mencatat skor antara 40 hingga 99 dengan distribusi nilai yang merata. Pada ujian saat ini, skor terendah turun drastis menjadi 10, namun ada siswa yang meraih nilai sempurna (100).

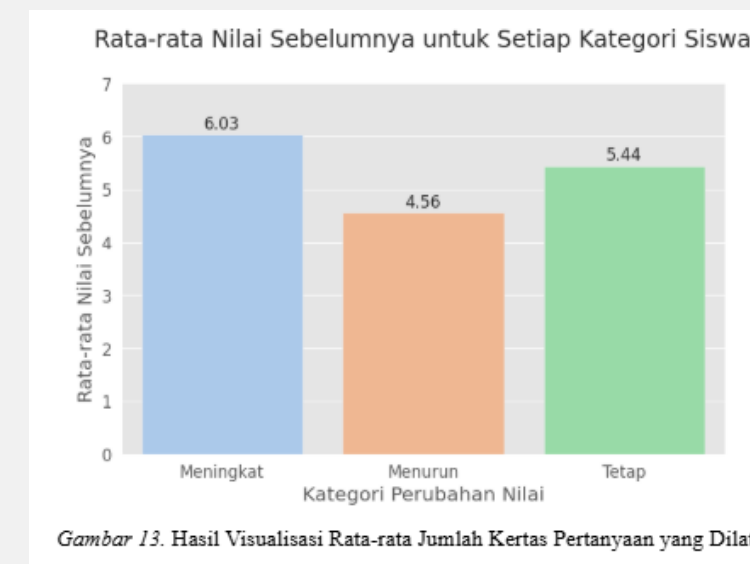
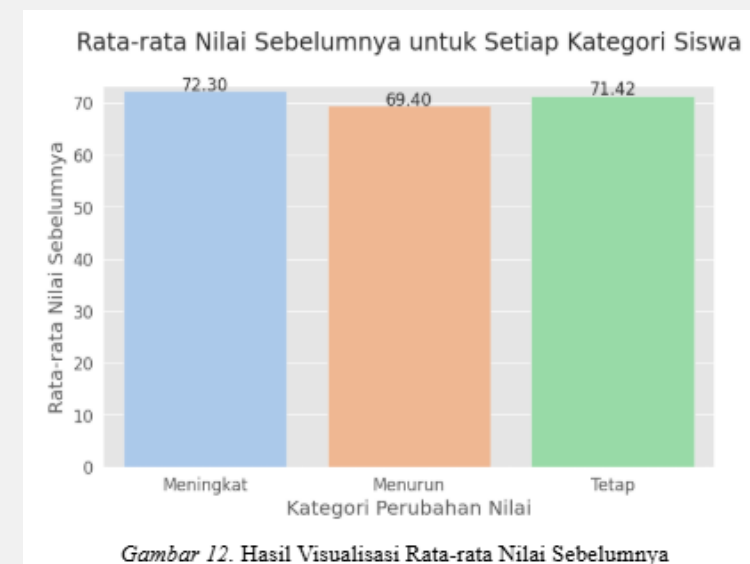
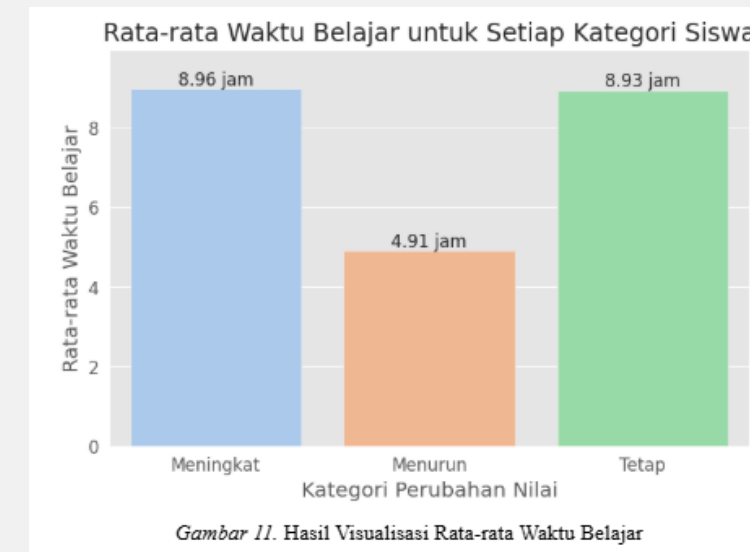
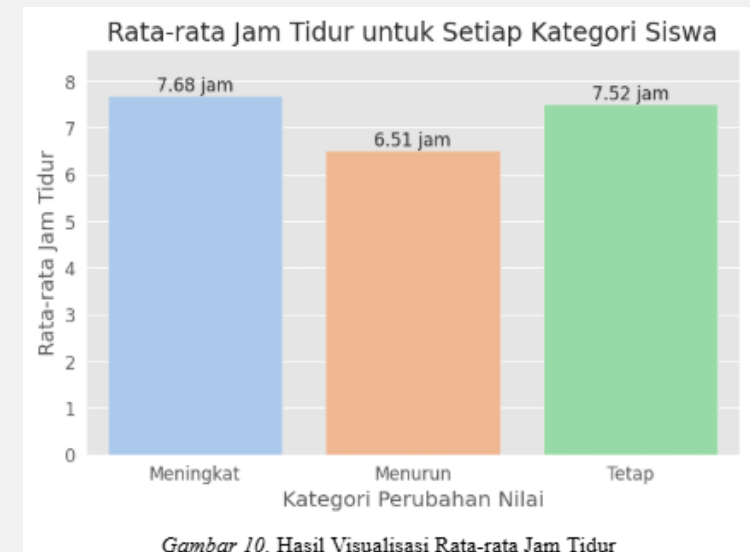
Analisis Awal Data

- Proses Pengelompokan

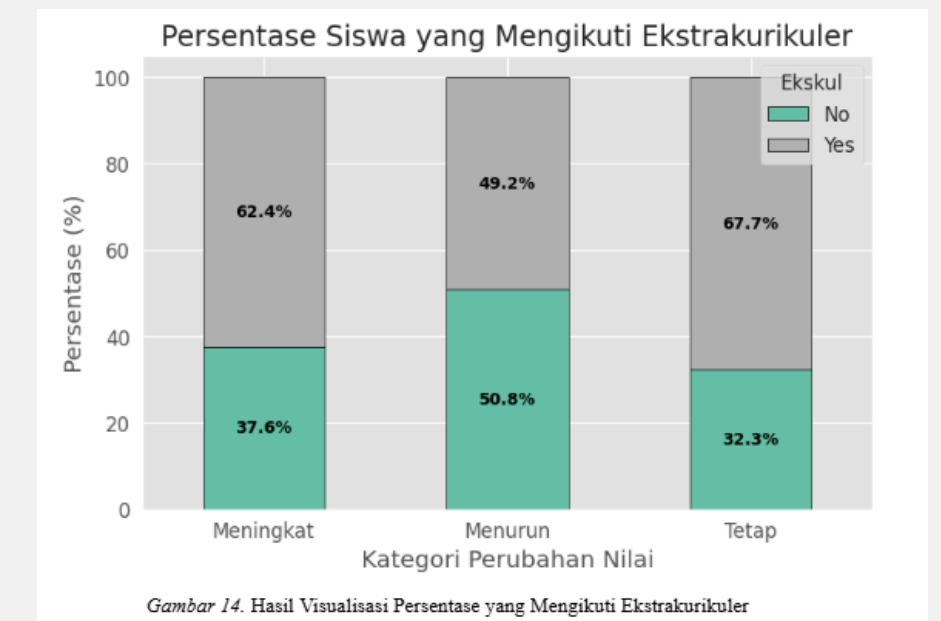


Hasil visualisasi menunjukkan 101 siswa (1%) mengalami peningkatan nilai, 9.800 siswa (98%) mengalami penurunan, dan 99 siswa (1%) nilainya tetap.

- Analisis Deskriptif berdasarkan Kategori Murid

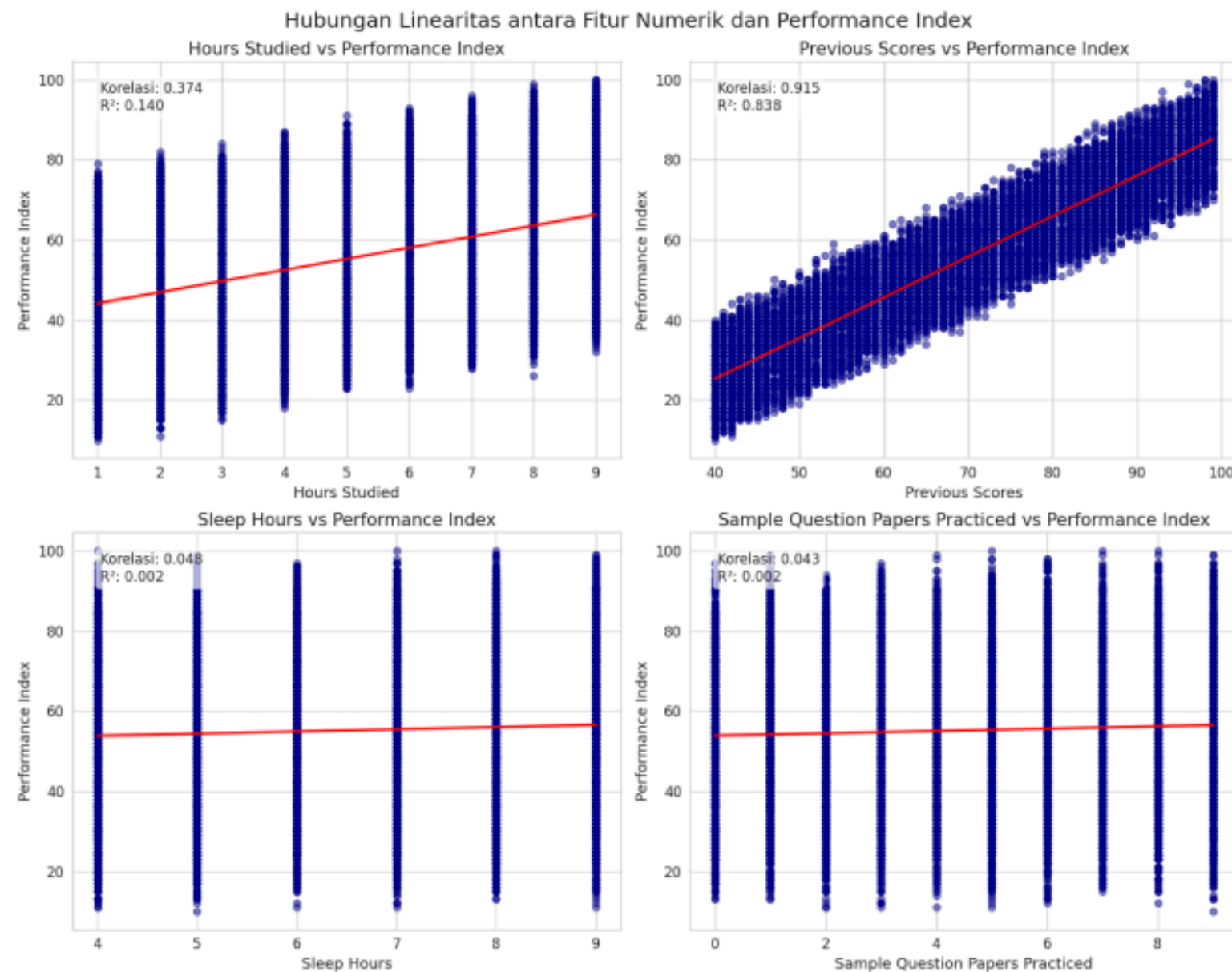


- Analisis 'Extracurricular Activities' dari Ketiga Kelompok



Siswa dengan nilai stabil atau naik lebih banyak ikut ekstrakurikuler, menunjukkan dampak positifnya. Sebaliknya, nilai turun cenderung dipengaruhi faktor lain.

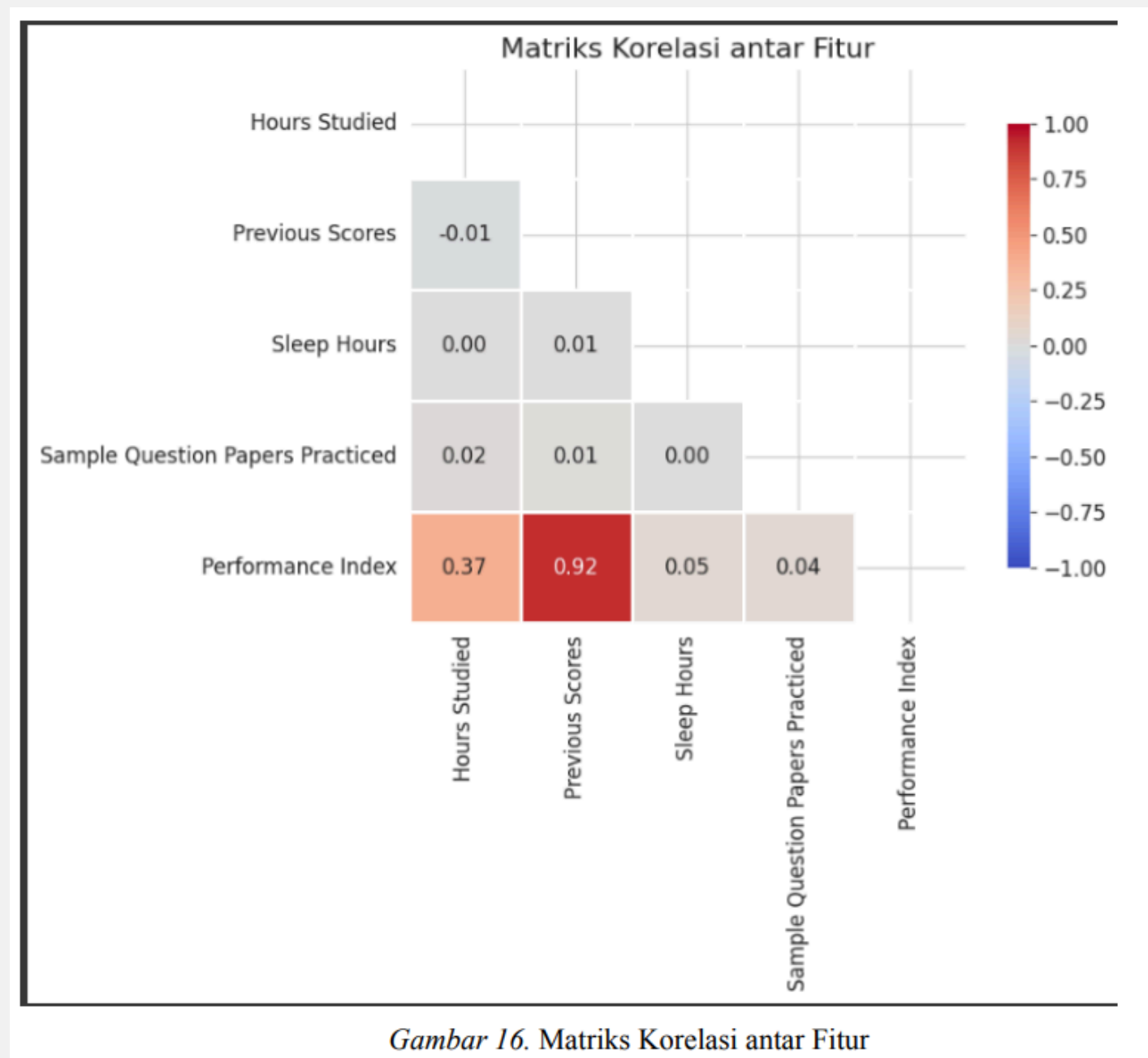
Analisis Hubungan Linear Fitur Numerik dengan Performance Index



Gambar 15. Hubungan Linearitas antara Fitur Numerik dan Performance Index

Nilai akademik sebelumnya dan jam belajar adalah faktor yang paling berkontribusi terhadap performa siswa, sedangkan jam tidur dan jumlah latihan soal tidak menunjukkan pengaruh yang berarti dalam konteks dataset ini.

Matriks Korelasi antar Fitur



Matriks korelasi menunjukkan bahwa Previous Scores memiliki hubungan paling kuat dengan Performance Index (nilai korelasi 0.92), memperkuat temuan sebelumnya bahwa nilai akademik terdahulu merupakan indikator terbaik untuk memprediksi performa siswa. Sementara itu, Hours Studied juga menunjukkan korelasi positif yang cukup berarti (0.37), menandakan bahwa durasi belajar masih berpengaruh terhadap performa.

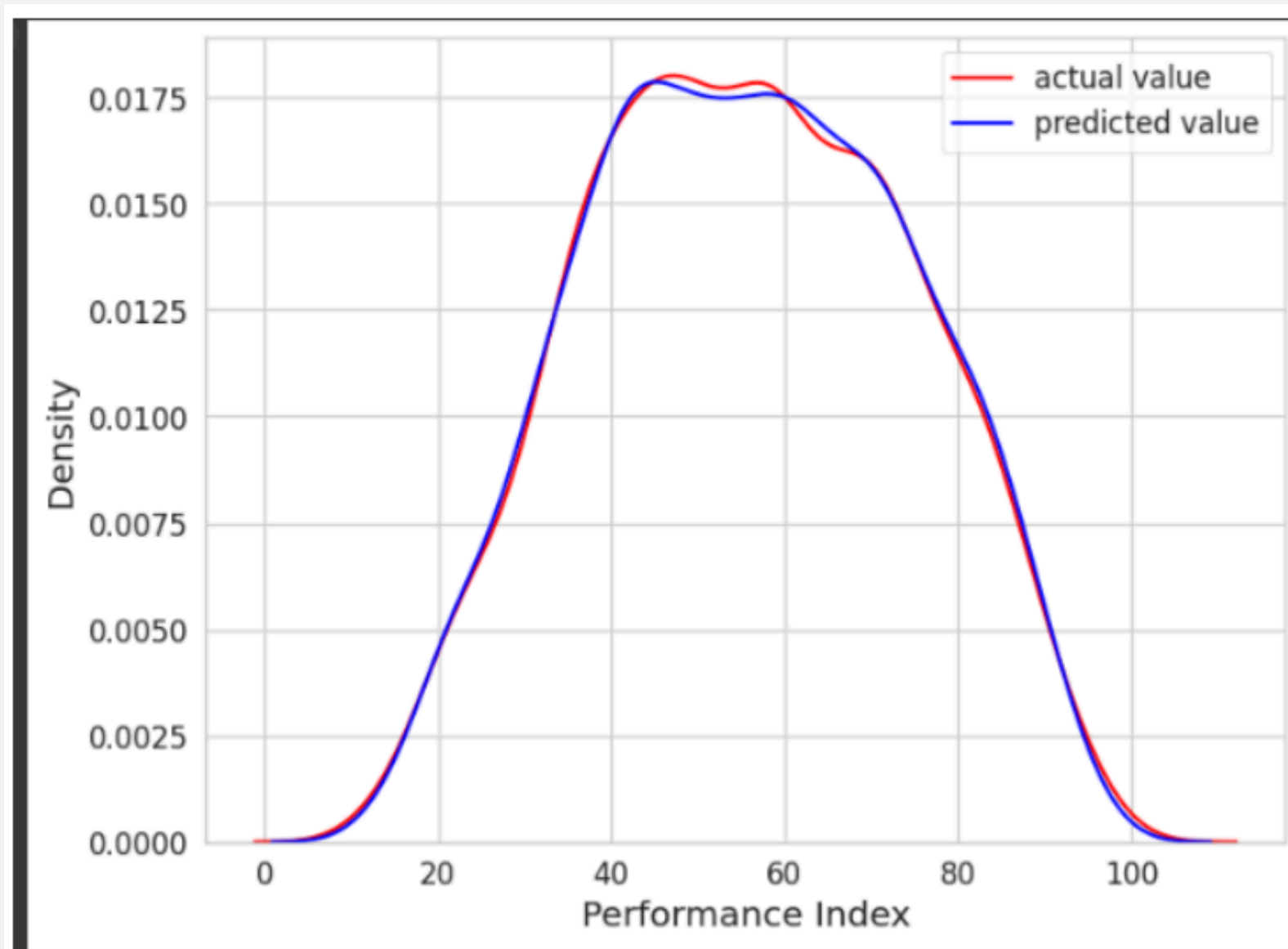
Pemilihan Fitur

Berdasarkan hasil analisis korelasi dan uji signifikansi p-value, dapat disimpulkan bahwa dua fitur yang paling berpengaruh terhadap Performance Index adalah Previous Scores dan Hours Studied.

Kedua fitur ini menunjukkan hubungan yang cukup kuat dan signifikan secara statistik dengan target variabel, menjadikannya kandidat yang relevan untuk dimasukkan ke dalam model prediksi.



Evaluasi Model Regresi Linear

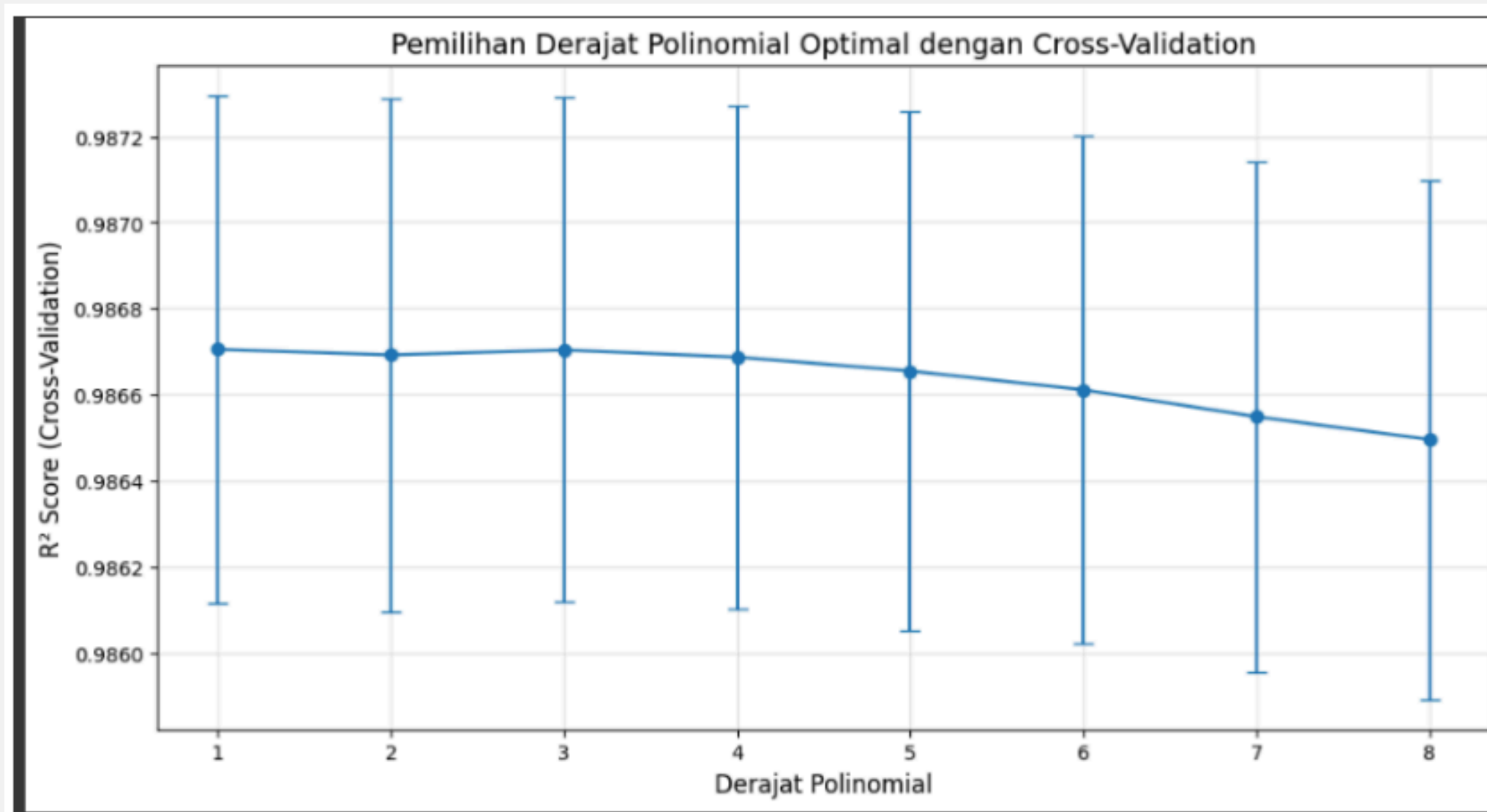


Gambar 17. Evaluasi Model Linear Regression

- Mean Absolute Error (MAE): 1.7512
- Mean Squared Error (MSE): 4.8326
- Root Mean Squared Error (RMSE): 2.1983
- R-squared (R^2): 0.9860

Nilai R^2 sebesar 0.9860 menunjukkan bahwa model mampu menjelaskan sekitar 98.60% variabilitas data. Artinya, model sangat baik dalam memprediksi Performance Index berdasarkan fitur yang digunakan.

Evaluasi Model Polynomial Regression



- Derajat 1 (model linear) memiliki skor R^2 yang tinggi dan stabil.
- Seiring peningkatan derajat, skor R^2 cenderung menurun sedikit dan standar deviasi tetap tinggi.

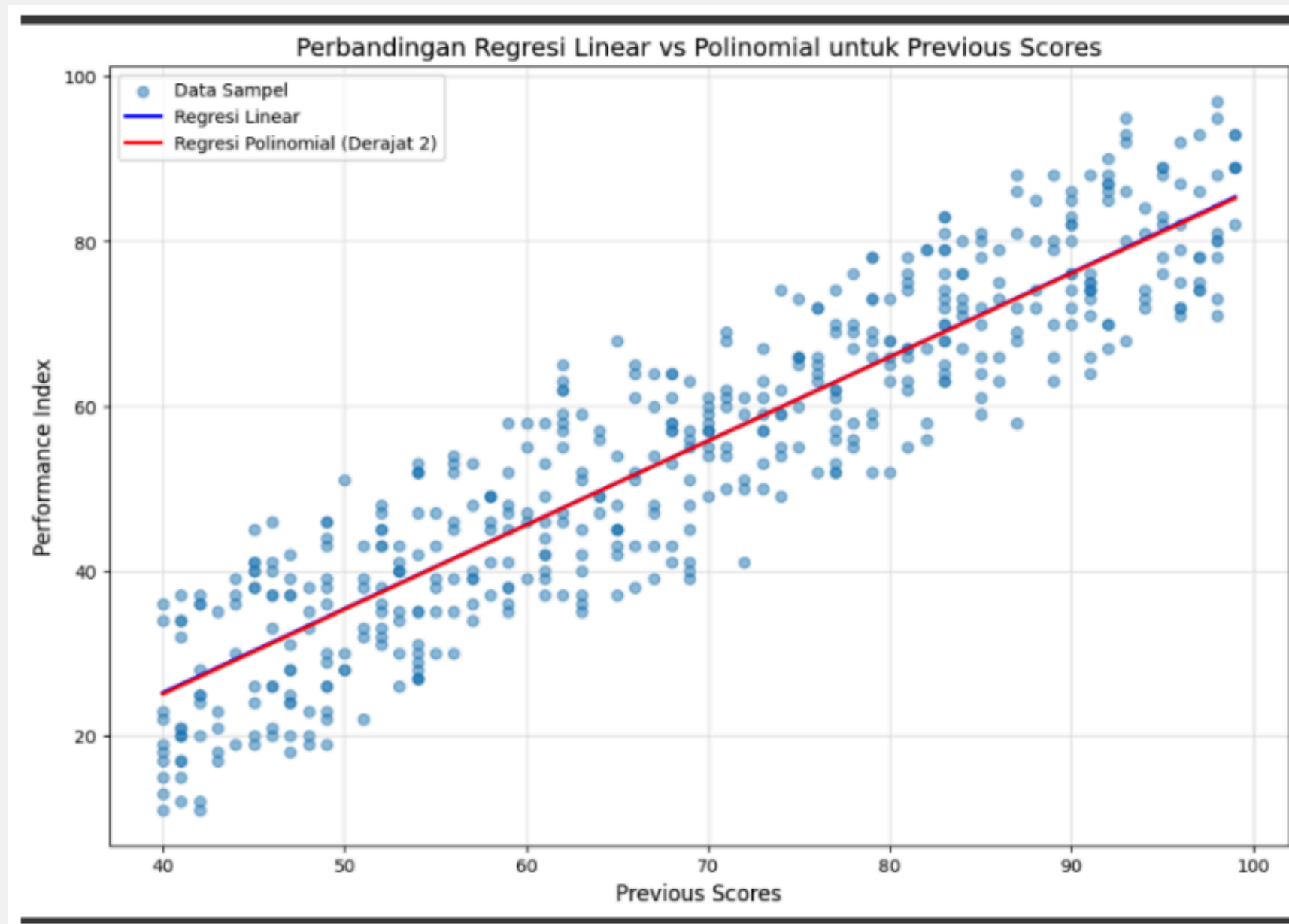
Derajat polinomial optimal adalah derajat 1, karena memberikan hasil terbaik dengan kompleksitas paling rendah.

Evaluasi Hasil

Fitur	Koefisien
Previous Scores	17.7924
Hours Studied	7.3928
Sample Question Papers Practiced	0.5431

Rekomendasi untuk Meningkatkan Performance Index Siswa:
Tingkatkan Previous Scores (koefisien: 17.7924)
Tingkatkan Hours Studied (koefisien: 7.3928)
Tingkatkan Sample Question Papers Practiced (koefisien: 0.5431)

Evaluasi Hasil



Gambar tersebut menunjukkan perbandingan antara model Regresi Linear (garis biru) dan Regresi Polinomial derajat 2 (garis merah) dalam memprediksi Performance Index berdasarkan Previous Scores. Titik-titik biru merepresentasikan data sampel asli. Artinya, model regresi linear saja sudah cukup akurat dan efisien untuk digunakan dalam konteks ini tanpa perlu menambah kompleksitas dengan model polinomial.

Kesimpulan

Berdasarkan hasil analisis, model yang dibangun dapat disimpulkan berkinerja dengan baik dalam memprediksi Performance Index siswa. Hal ini ditunjukkan oleh nilai koefisien determinasi (R^2) sebesar 0.9860, yang berarti model mampu menjelaskan sekitar 98,6% variasi data target. Dari sisi interpretasi, model juga menunjukkan hasil yang logis, di mana nilai akademik sebelumnya (Previous Scores) menjadi faktor paling dominan dalam memengaruhi performa siswa, diikuti oleh Hours Studied. Sementara itu, variabel lain seperti Sleep Hours, Sample Question Papers, dan Extracurricular Activities tidak memberikan pengaruh signifikan. Dapat disimpulkan bahwa model regresi linear yang dikembangkan cukup andal dan layak digunakan untuk memprediksi kinerja akademik siswa.

The background is a light gray color, decorated with various hand-drawn blue doodles. These include several loops and swirls at the top, a series of horizontal lines on the right side, a wavy line at the bottom center, and several checkmarks at the bottom right. The central text is in a bold, black, sans-serif font with a white outline.

**TERIMA
KASIH**