

Nama : Taufik Ramdan
Nim : 20210040154
Kelas : TI21F

Tugas Dasar Pemrograman Sesi 16 UAS

Soal

Soal Tugas Proyek: Analisis dan Visualisasi Data

Deskripsi Tugas:

Anda diminta untuk melakukan analisis data menggunakan dataset dari Kaggle.com dengan fokus pada statistik deskriptif dan regresi linier sederhana, serta mendokumentasikan seluruh proses kerja dalam bentuk laporan dan video presentasi.

Instruksi Pengerjaan:

1. Pilih Dataset:

- Kunjungi situs Kaggle.com dan pilih satu dataset yang menurut Anda menarik dan relevan untuk dilakukan analisis regresi linier.
- Pastikan dataset tersebut memiliki minimal 2 variabel numerik yang bisa dianalisis secara statistik dan digunakan untuk regresi linier.

2. Lakukan Analisis Data:

- Lakukan statistik dasar (rata-rata, median, standar deviasi, dll).
- Lakukan regresi linier sederhana (1 variabel independen dan 1 variabel dependen).
- Sertakan visualisasi data (gunakan matplotlib/seaborn/plotly).
- Sajikan hasil secara visual menarik dan tidak menggunakan format laporan baku (gunakan gaya infografis, poster digital, atau laporan interaktif).

3. Tulis Laporan Singkat yang terdiri dari:

- Latar Belakang: Jelaskan alasan pemilihan dataset dan permasalahan yang ingin dikaji.
- Metode Analisis: Langkah-langkah analisis yang dilakukan.
- Hasil dan Interpretasi: Hasil statistik dan regresi serta visualisasinya.
- Kesimpulan dan Rekomendasi: Apa yang bisa disimpulkan dari analisis, serta rekomendasi lebih lanjut.

Ini link Github nya : <https://github.com/taufik-01/UTS.git>

Ini link youtube nya : <https://www.youtube.com/watch?v=Dm-TUIZPyf0>

Jawaban

Link Dataset Kaggle yang Digunakan : <https://www.kaggle.com/datasets/joebeachcapital/students-performance>

Laporan Singkat Analisis dan Visualisasi Data

1. Latar Belakang

Pada analisis ini, saya menggunakan dataset nilai Test_1 dan Test_2 dari Kaggle karena datanya sederhana, relevan, dan cocok untuk pemahaman awal regresi linier sederhana. Permasalahan yang ingin dikaji adalah apakah terdapat hubungan linear antara nilai Test_1 dan Test_2 serta seberapa besar pengaruh nilai Test_1 terhadap Test_2.

2. Metode Analisis

Langkah-langkah yang dilakukan :

1. Mengimport dataset menggunakan pandas.
2. Melakukan statistik deskriptif (mean, median, standar deviasi).
3. Membuat model regresi linier sederhana dengan LinearRegression dari scikit-learn.
4. Menghitung slope, intercept, dan R^2 .
5. Membuat visualisasi scatter plot dengan garis regresi menggunakan seaborn.

3. Hasil dan Interpretasi

Statistik Deskriptif

- Test_1: mean = 50.72, median = 50.5, std = 29.56
- Test_2: mean = 51.36, median = 51.0, std = 29.97

Hasil Regresi Linier

- Slope: 1.036
- Intercept: -1.89
- R^2 : 0.994

Interpretasi: Terdapat hubungan linear yang sangat kuat antara Test_1 dan Test_2, dengan R^2 mendekati 1 yang menunjukkan hampir seluruh variasi nilai Test_2 dapat dijelaskan oleh nilai Test_1.

4. Kesimpulan dan Rekomendasi

Kesimpulan: Nilai Test_1 dan Test_2 memiliki hubungan linear yang sangat kuat. Setiap kenaikan 1 poin pada Test_1 akan meningkatkan Test_2 sekitar 1.036 poin.

Rekomendasi: Analisis selanjutnya dapat menambahkan variabel lain untuk regresi berganda agar hasil prediksi lebih akurat dan bervariasi.