**PENJELASAN DARI CODE YANG TELAH DIBUAT**

**JUPYTER NOTEBOOK/GOOGLE COLLAB**

**PENJELASAN PROGRAM ANALISIS DAN VISUALISASI DATA**

**1. Pendahuluan**

Program ini dirancang untuk melakukan analisis dan visualisasi data menggunakan bahasa pemrograman Python. Dengan bantuan beberapa pustaka seperti Pandas, NumPy, Matplotlib, dan Seaborn, kita dapat memahami pola dalam data serta membangun model prediksi menggunakan metode regresi linier.

**2. Pustaka yang Digunakan**

Kode dalam notebook ini menggunakan pustaka berikut:

* **Pandas**: Untuk membaca, mengolah, dan menganalisis data dalam bentuk tabel.
* **NumPy**: Untuk operasi matematis dan manipulasi array.
* **Matplotlib & Seaborn**: Untuk membuat berbagai jenis visualisasi data.
* **Scikit-Learn**: Untuk membangun model regresi dan evaluasi kinerja model.
* **Tkinter**: Untuk membuka file secara manual jika tidak ditemukan.

**3. Pengolahan Data**

**a. Memuat Dataset**

Kode memuat dua dataset utama:

* day.csv → Berisi data dalam skala harian.
* hour.csv → Berisi data dalam skala per jam.

Jika file tidak ditemukan dalam direktori, pengguna akan diminta untuk memilih file secara manual.

**b. Menampilkan Data**

Setelah data berhasil dimuat, beberapa baris pertama dari dataset akan ditampilkan untuk memberikan gambaran awal tentang struktur data yang digunakan.

**4. Analisis Data**

**a. Eksplorasi Data**

Dataset dianalisis menggunakan statistik deskriptif seperti:

* **Nilai rata-rata (mean)**, **nilai minimum dan maksimum**, serta **standar deviasi**.
* **Distribusi data** untuk melihat penyebaran nilai-nilai dalam dataset.

**b. Visualisasi Data**

Program membuat berbagai jenis grafik untuk memahami pola dalam data, termasuk:

1. **Histogram** → Untuk melihat distribusi data.
2. **Scatter Plot** → Untuk melihat hubungan antar variabel.
3. **Heatmap (Peta Korelasi)** → Untuk melihat hubungan antar fitur dalam dataset.

**5. Model Prediksi**

**a. Pembagian Data**

Dataset dibagi menjadi **data latih (training data)** dan **data uji (testing data)** menggunakan fungsi train\_test\_split.

**b. Regresi Linier**

Model **regresi linier** digunakan untuk membuat prediksi berdasarkan data yang tersedia. Model ini mencari hubungan linear antara variabel input dan output.

**c. Evaluasi Model**

Kinerja model diukur menggunakan tiga metrik utama:

* **Mean Absolute Error (MAE)** → Rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual.
* **Mean Squared Error (MSE)** → Rata-rata kuadrat dari kesalahan prediksi.
* **R-squared (R²)** → Mengukur seberapa baik model dapat menjelaskan variasi dalam data.

**6. Kesimpulan**

Dengan menggunakan metode eksplorasi data, visualisasi, serta model prediksi, program ini memberikan wawasan yang lebih dalam mengenai pola dalam data. Model regresi linier yang diterapkan dapat digunakan untuk memprediksi tren berdasarkan data historis. Dengan evaluasi menggunakan metrik tertentu, kita dapat menilai seberapa baik model yang dibuat dalam memprediksi nilai target.

**STREAMLIT**

**PENJELASAN PROGRAM VISUALISASI DATA DENGAN STREAMLIT**

**1. Pendahuluan**

Program ini menggunakan **Streamlit**, sebuah framework berbasis Python, untuk membangun antarmuka web interaktif yang memungkinkan pengguna untuk melakukan analisis data, visualisasi, dan prediksi model dengan lebih mudah.

**2. Pustaka yang Digunakan**

Kode ini menggunakan pustaka berikut:

* **Streamlit**: Untuk membangun antarmuka web interaktif.
* **Pandas**: Untuk membaca dan mengolah data dalam format tabel.
* **NumPy**: Untuk operasi numerik.
* **Matplotlib & Seaborn**: Untuk visualisasi data.
* **Scikit-Learn**: Untuk membangun model regresi linier dan evaluasi performa model.

**3. Fitur Aplikasi**

**a. Menu Navigasi**

Aplikasi memiliki sidebar yang berisi beberapa menu utama:

1. **Upload Data** → Untuk mengunggah dataset yang akan dianalisis.
2. **Visualisasi Data** → Untuk menampilkan grafik dan statistik deskriptif.
3. **Analisis Korelasi** → Untuk melihat hubungan antar variabel menggunakan heatmap.
4. **Distribusi Data** → Untuk menganalisis distribusi nilai dalam dataset.
5. **Prediksi Model** → Untuk membangun model regresi linier dan mengevaluasi kinerjanya.

**b. Upload Data**

* Pengguna dapat mengunggah **dua file CSV**.
* Data yang diunggah akan ditampilkan dalam bentuk tabel interaktif.
* Dataset akan disimpan dalam **session state** untuk diproses lebih lanjut.

**c. Visualisasi Data**

* Menampilkan **statistik deskriptif** dari dataset.
* Pengguna dapat memilih kolom numerik untuk dibuat **histogram**.
* Grafik dihasilkan menggunakan **Matplotlib** dan **Seaborn**.

**d. Analisis Korelasi**

* Menampilkan **heatmap** untuk melihat korelasi antar fitur numerik.
* Pengguna dapat memilih dataset yang akan dianalisis.
* Korelasi dihitung menggunakan df.corr() dan divisualisasikan dengan **Seaborn**.

**e. Prediksi dengan Model Regresi Linier**

1. Pengguna memilih **kolom target** dan **kolom fitur**.
2. Dataset dibagi menjadi **data latih (training data)** dan **data uji (testing data)** menggunakan train\_test\_split.
3. Model **Linear Regression** dilatih dengan data latih.
4. Hasil prediksi dibandingkan dengan data aktual menggunakan **scatter plot**.
5. Evaluasi model menggunakan metrik berikut:
   * **Mean Absolute Error (MAE)**
   * **Mean Squared Error (MSE)**
   * **R-squared Score (R²)**

**4. Kesimpulan**

Aplikasi ini mempermudah pengguna dalam melakukan **analisis data dan visualisasi secara interaktif** tanpa harus menjalankan kode secara manual di Jupyter Notebook.

Dengan fitur unggah data, eksplorasi statistik, analisis korelasi, serta prediksi model, pengguna dapat dengan mudah memahami pola data dan mendapatkan wawasan yang lebih dalam.