**Shakira Angelina Ika Putri (41822010117) ||** Visualisasi Data.Iris

Diberikan folder berisi data iris dengan nama file yang saya rubah menjadi iris.csv. Data iris merupakan data sebanyak 150 yang terbagi menjadi 3 kelas dengan masing-masing 50 data disetiap kelasnya. Data tersebut akan saya sajikan kedalam bentuk tabel dengan 5 kolom yaitu:

1. Panjang Sepal dalam cm
2. Lebar Sepal dalam cm
3. Panjang Petal dalam cm
4. Lebar Petal dalam cm
5. Nama kelas yaitu:
6. Iris Setosa
7. Iris Versicolor
8. Iris Virginica

Berikut adalah gambaran hasil:

A table with numbers and symbols

Description automatically generated

Dengan keterangan:

0 (Iris Setosa) 1 (Iris Versicolor) 2 (Iris Virginica)

Alasan saya rubah menjadi angka terlebih dahulu adalah agar dapat menampilkan bagian correlation karena correlation hanya untuk kolom format numerik.

Untuk mendapatkan hasil diatas, perlu menjalankan beberapa code yaitu sebagai berikut:

1. Mengimport library yang umum digunakan

A screen shot of a computer program

Description automatically generated

Keterangan:

* Numpy : digunakan untuk melakukan analisa data numerik dan perhitungan berbasis vektor atau matriks
* Pandas : digunakan untuk melakukan pengolahan data tabular
* Matplotlib : digunakan untuk melakukan ploting atau penggambaran grafik, dapat digunakan sebagai alat bantu dalam analisa data
* Seaborn (sns) : Dibangun di atas Matplotlib, mempermudah pembuatan visualisasi statistik yang lebih menarik secara estetika.
* scikit-learn (datasets): Pustaka untuk memuat dan bekerja dengan dataset klasik (misalnya, dataset Iris) untuk keperluan pembelajaran mesin.

No.1

A screenshot of a computer

Description automatically generated

Jawaban diatas didapat dari:

A screenshot of a computer code

Description automatically generated

Keterangan**:**

Menghitung Statistik Dasar:

* + min\_values = data.min()[:4]: Menghitung nilai minimum dari empat kolom pertama dalam data.
  + max\_values = data.max()[:4]: Menghitung nilai maksimum dari empat kolom pertama.
  + mean\_values = data.mean()[:4]: Menghitung rata-rata (mean) dari empat kolom pertama.
  + std\_values = data.std()[:4]: Menghitung standar deviasi dari empat kolom pertama.
  + class\_corr = corr['class'][:4]: Mengambil nilai korelasi antara fitur data dan kelas untuk empat kolom pertama.

**Loop Melalui Setiap Kolom untuk Menampilkan Statistik**:

* + for col in data.columns[:4]: Loop melalui empat kolom pertama data.
    - corr\_value = class\_corr[col]: Ambil nilai korelasi untuk kolom saat ini.
    - high\_corr = " (high!)" if abs(corr\_value) > 0.9 else "": Jika nilai korelasi lebih dari 0.9 atau kurang dari -0.9, dianggap sebagai korelasi tinggi dan label " (high!)" ditambahkan.

No.2

A bar chart with different colored bars

Description automatically generated

Didapat dengan:

A screenshot of a computer code

Description automatically generated

Keterangan:

* **plt.figure(figsize=(5, 4))**: mengatur ukuran grafik, dengan lebar 5 inci dan tinggi 4 inci.
* **sns.barplot**: untuk membuat grafik batang.

No.3

A blue bar graph with white text

Description automatically generated

Hasil diatas didapat dari:

A computer screen shot of a code

Description automatically generated

Keterangan:

* **plt.figure(figsize=(5, 3))**: mengatur ukuran grafik, dengan lebar 5 inci dan tinggi 3 inci.
* **dataset['sepal length (cm)']**: berisi data sepal length dalam cm.
* **bins=10**: Membagi data sepal length menjadi 10 interval (bin) untuk ditampilkan sebagai batang-batang pada histogram.
* **color='blue'**: Mengatur warna batang histogram menjadi biru.
* **edgecolor='black'**: Menambahkan garis tepi hitam pada setiap batang histogram

No.4

A graph with different colored lines

Description automatically generated

Jawaban diatas didapat dari:

A screen shot of a computer code

Description automatically generated

Keterangan:

* **Mem-plot Individu:** Setiap baris plt.plot(...) membuat garis untuk masing-masing.
* **Label Legenda:** label= memberikan label untuk setiap garis di plot, sehingga dapat diidentifikasi di legenda.

No.5

A diagram of different colored dots

Description automatically generated

Hasil diatas didapat dari:

A close-up of a computer screen

Description automatically generated

Keterangan:

* Membuat **scatter plot** dari dataset yang menunjukkan hubungan antara SepalLengthCm dan SepalWidthCm pada dataset iris, dengan setiap titik berwarna berdasarkan spesiesnya.
* Membuat scatter plot menggunakan **Seaborn**