

Tugas Besar 1 Visualisasi Data

Adika Faris MH - 41822010137

1. Dari gambar berikut, dapat ditarik kesimpulan statistik untuk setiap variabel:

Summary Statistics:

	Min	Max	Mean	SD	Class Correlation
sepal length:	4.3	7.9	5.84	0.83	0.7826
sepal width:	2.0	4.4	3.05	0.43	-0.4194
petal length:	1.0	6.9	3.76	1.76	0.9490 (high!)
petal width:	0.1	2.5	1.20	0.76	0.9565 (high!)

- Sepal Length: Minimum: 4.3 Maximum: 7.9 Mean (Rata-rata): 5.84 Standar Deviasi (SD): 0.83 Korelasi dengan Class: 0.7826
- Sepal Width: Minimum: 2.0 Maximum: 4.4 Mean (Rata-rata): 3.05 Standar Deviasi (SD): 0.43 Korelasi dengan Class: -0.4194 (korelasi negatif)
- Petal Length: Minimum: 1.0 Maximum: 6.9 Mean (Rata-rata): 3.76 Standar Deviasi (SD): 1.76 Korelasi dengan Class: 0.9490 (korelasi sangat kuat/high!)
- Petal Width: Minimum: 0.1 Maximum: 2.5 Mean (Rata-rata): 1.20 Standar Deviasi (SD): 0.76 Korelasi dengan Class: 0.9565 (korelasi sangat kuat/high!)

Kesimpulan

Petal Length dan Petal Width memiliki korelasi yang sangat kuat (high!) dengan Class, menunjukkan bahwa kedua variabel ini sangat baik untuk membedakan kelas-kelas dalam dataset.

Sepal Width memiliki korelasi negatif dengan Class, yang berarti ketika sepal width meningkat, nilai class cenderung menurun.

Standar deviasi terbesar terdapat pada Petal Length (1.76), menunjukkan variasi data yang lebih besar pada variabel ini.

2. Code:

```
[23]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Memuat dataset Iris
df = pd.read_csv('irisdata2.csv', names=['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'species'])

# Menghitung jumlah panjang sepal untuk setiap spesies
species_sum = df.groupby('species')['sepal_length'].sum().reset_index()

# Membuat grafik batang
plt.figure(figsize=(8, 6))
sns.barplot(data=species_sum, x='species', y='sepal_length', palette='Blues')

# Menambahkan judul dan label sumbu
plt.title('Grafik Batang')
plt.xlabel('Species')
plt.ylabel('Panjang Sepal')

# Menetapkan batas sumbu y hingga 350
plt.ylim(0, 350)

# Menampilkan grafik
plt.show()
```

3. Code:

```
[24]: import matplotlib.pyplot as plt

# Membuat histogram untuk kolom sepal_length
plt.figure(figsize=(8, 6))
plt.hist(df['sepal_length'], bins=10, edgecolor='black')
plt.title('Sepal Length')
plt.xlabel('Panjang Sepal')
plt.ylabel('Frekuensi')
plt.show()
```

4. Code:

```
[21]: # Memuat dataset dari file CSV
df = pd.read_csv('irisdata2.csv')
df.columns = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class']

# Membuat Grafik Garis untuk 150 data
plt.figure(figsize=(12, 6))

# Mereset indeks agar plot menggunakan indeks dari 0 hingga 149
df_reset = df.reset_index()

# Plot setiap fitur (sepal_length, sepal_width, petal_length, petal_width)
plt.plot(df_reset.index, df_reset['sepal_length'], marker='o', label='Sepal Length', linewidth=1, markersize=4)
plt.plot(df_reset.index, df_reset['sepal_width'], marker='o', label='Sepal Width', linewidth=1, markersize=4)
plt.plot(df_reset.index, df_reset['petal_length'], marker='o', label='Petal Length', linewidth=1, markersize=4)
plt.plot(df_reset.index, df_reset['petal_width'], marker='o', label='Petal Width', linewidth=1, markersize=4)

# Kustomisasi grafik
plt.xlabel('Data Points (0-149)')
plt.ylabel('cm')
plt.title('Iris Features - All 150 Data Points')
plt.legend()
plt.grid(True)

# Mengatur batasan sumbu x agar dimulai dari 0 sampai 149
plt.xlim(-5, 154)

# Menambahkan ticks pada sumbu x setiap 10 data points
plt.xticks(np.arange(0, 150, 10))

plt.tight_layout()
plt.show()
```

5. Code:

```
[35]: fig = plt.figure(figsize=(12, 8))

# Membuat subplot pertama dari Layout 2x2
plt.subplot(2, 2, 1)
sns.scatterplot(data=df, x='sepal_length', y='sepal_width', hue='class')
plt.title('Sepal Length vs Width')

# Menyusun tata Letak agar Lebih rapi
plt.tight_layout()
plt.show()
```