

1. Dataset

Data Wine Quality Red diambil dari sumber Kaggle

<https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009> . Data ini berisi tentang informasi terkait dengan varian merah anggur Portugis "Vinho Verde" yang terdiri dari variable keasaman tetap, keasaman mudah menguap, asam sitrat, sisa gula, klorida, sulfur dioksida bebas, sulfur dioksida total, kepadatan, pH, sulfat, alkohol, dan kualitas (skor antara 0 dan 10)

2. Link Google Collab

<https://colab.research.google.com/drive/11zykDv0EQRPePFnW6TINTDOiq0uIJTC1?usp=sharing>

3. Membaca Data

```
import pandas as pd

# Memuat dataset winequality-red
red_wine_data = pd.read_csv('winequality-red.csv', sep=';')
red_wine_data = red_wine_data.fillna(method='ffill')
red_wine_data.head()
```

- Pertama, import terlebih dahulu library yang dibutuhkan untuk mengimport, menampilkan data, analisis data dan manipulasi data dalam bentuk DataFrame, yakni bernama "pandas" dan memberinya alias "pd."
- Baca dataset dari file CSV dengan **red_wine_data = pd.read_csv('winequality-red.csv', sep=';')**: Argumen **sep=';'** mengindikasikan bahwa kolom dalam dataset dipisahkan oleh tanda titik koma (;).
- Mengisi nilai-nilai yang hilang (NaN) dengan nilai dari baris sebelumnya menggunakan **red_wine_data = red_wine_data.fillna(method='ffill')**
- Menampilkan lima baris pertama dari dataset dengan **head()**

Output :

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5

Dataset terdiri 12 kolom dan 1599 baris

Visualisasi High-Dimensional dan Multivariate dengan Diagram Batang

```
[13] import matplotlib.pyplot as plt
      from mpl_toolkits.mplot3d import Axes3D
      import numpy as np

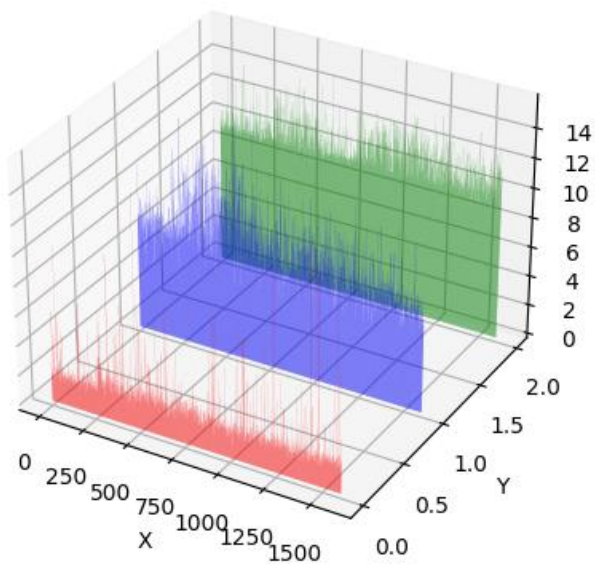
      dades01 = red_wine_data['residual sugar'].tolist()
      dades02 = red_wine_data['fixed acidity'].tolist()
      dades03 = red_wine_data['alcohol'].tolist()
      df_3d = pd.DataFrame([dades01,dades02,dades03]).transpose()
      colors = ['r','b','g']
      fig = plt.figure()
      ax = fig.add_subplot(111,projection="3d")
      z = list(df_3d)
      for n, i in enumerate(df_3d):
          print('n', n)
          xs = np.arange(len(df_3d[i]))
          ys = [i for i in df_3d[i]]
          zs = z[n]

          cs = colors[n]
          print('xs:', xs, 'ys:', ys, 'zs', zs, 'cs:', cs)
          ax.bar(xs,ys,zs,zdir="y",color=cs,alpha=0.8)
      ax.set_xlabel("X")
      ax.set_ylabel("Y")
      ax.set_zlabel("Z")
      plt.show()
```

- Import library yang dibutuhkan:
 - **matplotlib.pyplot as plt**: Untuk membuat visualisasi grafik.
 - **from mpl_toolkits.mplot3d import Axes3D**: Untuk mengaktifkan visualisasi 3D.
 - **numpy as np**: Untuk melakukan operasi numerik.
- Mengambil data dari dataset:
 - Data dari kolom 'residual sugar', 'fixed acidity', dan 'alcohol' diambil dan diubah menjadi daftar Python.
- Membuat DataFrame:
 - Data yang diambil dari langkah sebelumnya digabungkan menjadi DataFrame Pandas yang disebut **df_3d** untuk membuat grafik.
- Menyiapkan warna:
 - Daftar warna ('r' untuk merah, 'b' untuk biru, 'g' untuk hijau) disiapkan untuk digunakan dalam grafik.
- Membuat plot 3D:
 - Objek **fig** dan **ax** dibuat untuk membuat plot 3D.
 - Setiap atribut yang akan divisualisasikan dipisahkan dan diplot sebagai batang 3D. Iterasi dilakukan untuk setiap atribut.
 - Untuk setiap atribut, **xs** adalah daftar nomor baris data, **ys** adalah data aktual dari atribut tersebut, **zs** adalah nilai z yang akan menentukan ketinggian batang, dan **cs** adalah warna yang sesuai dengan atribut tersebut.
 - Batang 3D diplot pada koordinat (**xs, ys, zs**) dengan warna **cs**.
- Menambahkan label sumbu x, y, dan z ke plot.
- Menampilkan grafik3D

Output :

```
[5]
n 0
xs: [ 0 1 2 ... 1596 1597 1598] ys: [1.9, 2.6, 2.3, 1.9, 1.9, 1.8, 1.6, 1.2, 2.0, 6.1, 1.8, 6.1, 1.6, 1.6, 3.8, 3.9, 1.8, 1.7, 4.4, 1.8, 1.8, 2.3, 1.6, 2.3, 2.4, 1.4, 1.8, 1.6, 1.9, 2.0, 2.
n 1
xs: [ 0 1 2 ... 1596 1597 1598] ys: [7.4, 7.8, 7.8, 11.2, 7.4, 7.4, 7.9, 7.3, 7.8, 7.5, 6.7, 7.5, 5.6, 7.8, 8.9, 8.9, 8.5, 8.1, 7.4, 7.9, 8.9, 7.6, 7.9, 8.5, 6.9, 6.3, 7.6, 7.9, 7.1, 7.8, 6.
n 2
xs: [ 0 1 2 ... 1596 1597 1598] ys: [9.4, 9.8, 9.8, 9.8, 9.4, 9.4, 9.4, 10.0, 9.5, 10.5, 9.2, 10.5, 9.9, 9.1, 9.2, 9.2, 10.5, 9.3, 9.0, 9.2, 9.4, 9.7, 9.5, 9.4, 9.7, 9.3, 9.5, 9.5, 9.4, 9.1
```



Visualisasi grafik batang 3D yang menunjukkan hubungan antara tiga atribut ('residual sugar', 'fixed acidity', dan 'alcohol') dalam dataset "winequality-red." Setiap atribut diwakili oleh warna yang berbeda pada grafik

Memvisualisasi Multiple Data Frame dengan Diagram Batang

```
[16] # Multiple DataFrames
list_1 = [red_wine_data['residual sugar'].head(3).tolist(), red_wine_data['fixed acidity'].head(3).tolist(), red_wine_data['alcohol'].head(3).tolist()]
df_first = pd.DataFrame(list_1, columns=list("ABC"))

list_2 = [red_wine_data['residual sugar'][3:6].tolist(), red_wine_data['fixed acidity'][3:6].tolist(), red_wine_data['alcohol'][3:6].tolist()]
df_second = pd.DataFrame(list_2, columns=list("ABC"))

list_3 = [red_wine_data['residual sugar'][7:10].tolist(), red_wine_data['fixed acidity'][7:10].tolist(), red_wine_data['alcohol'][7:10].tolist()]
df_third = pd.DataFrame(list_3, columns=list("ABC"))

columns = ["A", "B", "C"]
df_names = ['residual sugar', 'fixed acidity', 'alcohol']
df = [df_first, df_second, df_third]

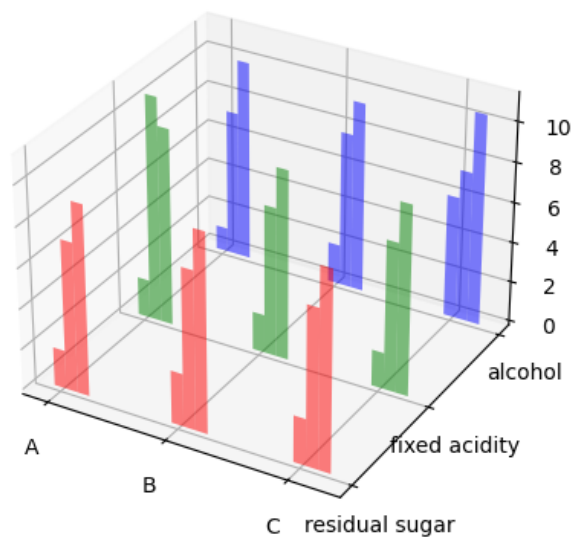
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

# make sure x and y axis get the right tick labels
plt.xticks([i for i in range(len(columns))], columns)
plt.yticks([i for i in range(len(df_names))], df_names)

# define a list for x position
xs = list()
for i in range(len(df)):
    for j in range(len(columns)):
        xs.append(i+j*0.1)
for cl, c in enumerate(['r', 'g', 'b']):
    ys = list()
    for i in range((len(columns))):
        ys.extend(df[cl].iloc[:, i:i+1].unstack().tolist())
    cs = [c] * len(xs)
    ax.bar(xs, ys, zs=cl, zdir='y', color=cs, alpha=0.5, width=0.1)

plt.show()
```

- Membuat beberapa DataFrame:
 - Tiga kelompok data yang berbeda dibentuk dengan mengambil sejumlah baris pertama dari setiap atribut ('residual sugar', 'fixed acidity', dan 'alcohol') dalam dataset "winequality-red." Setiap kelompok data ini disimpan dalam DataFrame terpisah: **df_first**, **df_second**, dan **df_third**.
- Mendefinisikan variabel:
 - Variabel **columns** berisi label kolom yang akan digunakan pada sumbu x pada grafik.
 - Variabel **df_names** adalah nama atribut yang akan digunakan pada sumbu y pada grafik.
 - Variabel **df** adalah daftar dari tiga DataFrame yang telah dibuat sebelumnya.
- Membuat plot 3D:
 - Objek **fig** dan **ax** digunakan untuk membuat plot 3D.
 - Posisi label pada sumbu x dan y diatur sesuai dengan **columns** dan **df_names** menggunakan **plt.xticks** dan **plt.yticks**.
- Membuat bar chart:
 - Dalam loop **for**, data dari setiap atribut diambil untuk setiap kelompok data dan disimpan dalam daftar **xs** dan **ys**.
 - Atribut yang berbeda pada setiap kelompok data akan memiliki posisi yang sedikit berbeda pada sumbu x agar tidak tumpang tindih.
 - Setiap bar pada grafik diberi warna yang berbeda sesuai dengan atribut menggunakan variabel **cs**.
- Menampilkan grafik:
 - Grafik batang 3D ditampilkan.



Visualisasi grafik batang 3D yang menunjukkan data dari tiga atribut ('residual sugar', 'fixed acidity', dan 'alcohol') dalam tiga kelompok data yang berbeda. Grafik ini untuk membandingkan nilai atribut yang berbeda antara kelompok data dengan warna yang berbeda pada tiga sumbu yang berbeda.

Memvisualisasi Data dengan Scatter Plot

```
[17] from sklearn.preprocessing import MinMaxScaler
list_1 = red_wine_data['residual sugar'].head(60).tolist()
list_2 = red_wine_data['fixed acidity'].head(60).tolist()
list_3 = red_wine_data['alcohol'].head(60).tolist()

healthcare = (list_1, list_2, list_3)
print(healthcare)
colors = ("purple", "pink", "blue")
groups = ('residual sugar', 'fixed acidity', 'alcohol')

# Create the plot
fig = plt.figure()
ax = fig.add_subplot(111, projection="3d")

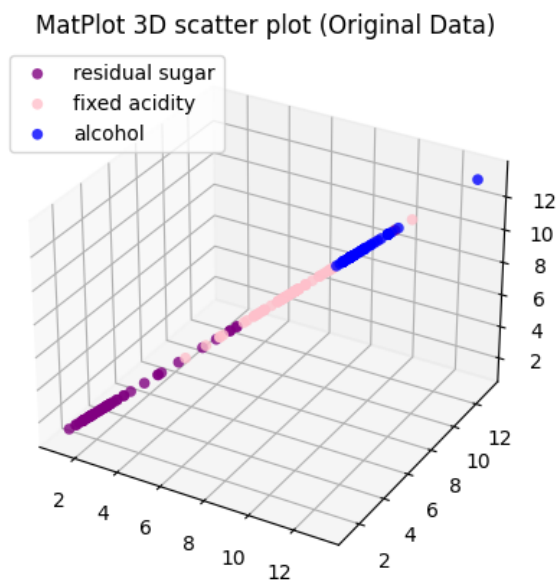
for healthcare_val, color, group in zip(healthcare, colors, groups):
    x = healthcare_val
    y = [j + 0.05 for j in x] # Adjust y-coordinate to spread points
    z = [k + 0.1 for k in y] # Adjust z-coordinate to spread points
    ax.scatter(x, y, z, alpha=0.8, c=color, edgecolors="none", s=30, label=group)

ax.set_facecolor("1.0") # Set background color of the 3D plot
plt.title('MatPlot 3D scatter plot (Original Data)')
plt.legend(loc=2)
plt.show()
```

- Mengimpor library **MinMaxScaler** dari **sklearn.preprocessing** untuk pra-pemrosesan data untuk melakukan penskalaan fitur ke dalam rentang tertentu, biasanya 0 hingga 1
- Membuat tiga daftar data (**list_1**, **list_2**, **list_3**) yang berisi 60 data pertama dari atribut 'residual sugar', 'fixed acidity', dan 'alcohol' dalam dataset "winequality-red".
- Mengelompokkan data: Menggabungkan ketiga daftar data tersebut menjadi satu tuple bernama **healthcare**.

- Menentukan warna untuk setiap atribut ('residual sugar', 'fixed acidity', 'alcohol') dalam variabel **colors**.
- Membuat tuple **groups** yang berisi nama atribut.
- Membuat objek plot 3D dengan **plt.figure()** dan **fig.add_subplot(111, projection='3d')**.
- Menggunakan loop **for** untuk memplot setiap atribut dalam warna yang berbeda.
- Posisi data pada sumbu y dan z disesuaikan agar titik-titik data terlihat tersebar dengan lebih baik.
- Mengatur warna latar belakang plot 3D dengan **ax.set_facecolor("1.0")**.
- Memberikan judul plot dengan **plt.title**.
- Menampilkan legenda dengan **plt.legend** untuk mengidentifikasi setiap atribut yang diplot.

Output :



Hasil scatter plot 3D yang menunjukkan distribusi data untuk atribut 'residual sugar', 'fixed acidity', dan 'alcohol' dari dataset anggur merah. Setiap grup memiliki warna yang berbeda dan data dikelompokkan berdasarkan atributnya. Ini memvisualisasikan bagaimana atribut-atribut ini berkorelasi dalam bentuk ruang tiga dimensi.

Memvisualisasi Data dengan Scatter Plot

```
[18] N = 60
g1 = (0.2 + 0.1 * np.random.rand(N), 0.3 + 0.2 * np.random.rand(N), 0.2 * np.random.rand(N))
g2 = (0.4 + 0.3 * np.random.rand(N), 0.5 * np.random.rand(N), 0.1 * np.random.rand(N))
g3 = (0.3 * np.random.rand(N), 0.3 * np.random.rand(N), 0.3 * np.random.rand(N))

healthcare = (g1, g2, g3)
colors = ("purple", "pink", "blue")
groups = ('residual sugar', 'fixed acidity', 'alcohol')

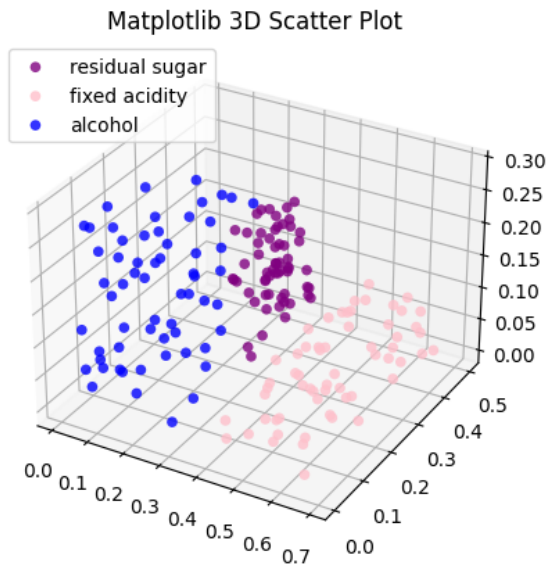
fig = plt.figure()
ax = fig.add_subplot(111, projection="3d")

# Iterasi melalui data dan plot scatter plot
for data, color, group in zip(healthcare, colors, groups):
    x, y, z = data
    ax.scatter(x, y, z, alpha=0.8, c=color, edgecolors="none", s=30, label=group)

ax.set_facecolor("white")
plt.title("Matplotlib 3D Scatter Plot")
plt.legend(loc=2)
plt.show()
```

- Mendefinisikan variabel **N** dengan nilai 60 yang berarti akan membuat 60 titik data untuk setiap kelompok (**g1, g2, g3**).
- Membuat tiga set data (**g1, g2, g3**) yang masing-masing berisi tiga array numerik (**x, y, z**). Setiap set data ini akan mewakili satu kelompok dalam scatter plot 3D.
- Setiap array **x, y**, dan **z** dalam set data tersebut diisi dengan bilangan acak. Anda menggunakan **np.random.rand(N)** untuk menghasilkan nilai acak dalam rentang [0, 1] dan mengalikannya dengan beberapa konstanta untuk mengatur distribusi data yang berbeda untuk setiap kelompok.
- Menggabungkan ketiga set data (**g1, g2, g3**) ke dalam satu tuple bernama **healthcare**.
- Mendefinisikan tiga warna (**colors**) dan tiga grup (**groups**) yang akan digunakan dalam plot.
- Membuat sebuah figure (**fig**) dan sebuah subplot 3D (**ax**) menggunakan **fig.add_subplot(111, projection="3d")** sebagai basis untuk membuat scatter plot 3D.
- Melakukan iterasi melalui data dalam **healthcare**, warna dalam **colors**, dan grup dalam **groups**. Untuk setiap kelompok, di plot titik-titik data dalam bentuk scatter plot 3D dengan koordinat **x, y**, dan **z** yang sesuai serta memberikan warna, ukuran marker (**s**), dan label untuk setiap grup.
- Mengatur warna latar belakang subplot dengan **ax.set_facecolor("white")**, memberikan judul untuk plot, dan menambahkan legenda.
- **plt.show()** untuk menampilkan plot secara visual.

Output :



Hasil scatter plot 3D memvisualisasikan tiga kelompok data (**g1**, **g2**, **g3**) dalam tiga dimensi (x, y, dan z). Setiap kelompok memiliki warna yang berbeda dan titik-titik data tersebar di seluruh ruang tiga dimensi.