**Mengatur Warna Scatter Plot Berdasarkan Kategori di Python**

Scatter plot adalah salah satu jenis visualisasi data yang digunakan untuk melihat hubungan antara dua variabel numerik yang diproyeksikan pada sumbu x dan y.

Data yang kita miliki adakalanya tidak hanya berisi variabel numerik, tetapi juga sering mengandung variabel kategori. Nah, bagaimana caranya jika kita ingin melihat hubungan variabel numerik dan kategori dalam satu scatter plot?

Sebagian dari kita mungkin akan menjawab dengan memberinya warna yang berbeda untuk setiap kategori. Jadi, bagaimana caranya kita mengatur warna scatter plot berdasarkan kategori di Python?

Setidaknya kita dapat menggunakan 3 cara di Python untuk memberikan warna yang berbeda berdasarkan katagori.

**Menggunakan Matplotlib**

[Matplotlib](https://matplotlib.org/3.3.3/api/_as_gen/matplotlib.pyplot.scatter.html) memiliki parameter **c** untuk mengatur warna, dimana kita dapat memberinya nilai berupa warna tetap seperti ‘red’, ‘blue’, dan ‘green’, ataupun berupa list.

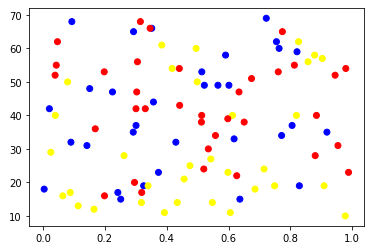
Sebagai contoh saya akan membuat dataframe terlebih dahulu yang terdiri dari tiga variabel, dimana dua di antaranya berupa nilai acak yang dihasilkan dari **numpy.random** serta satu variabel kategori seperti di bawah ini.

import numpy as np  
import pandas as pd  
  
np.random.seed(10)  
x = np.random.rand(100)  
y = np.random.randint(10, 70, 100)  
cat = ['Bahasa', 'IPA', 'IPS']  
jurusan = np.random.choice(cat, 100)  
  
df = pd.DataFrame(dict(X=x, Y=y, Jurusan=jurusan))  
df.head(10)

Lalu kita gunakan Matplotlib untuk membuat scatter plot.

import matplotlib.pyplot as plt  
  
colors = {'Bahasa':'red', 'IPA':'blue', 'IPS':'yellow'}  
  
fig, ax = plt.subplots()  
ax.scatter(df['X'], df['Y'], c=df['Jurusan'].map(colors))  
plt.show()

**Output:**



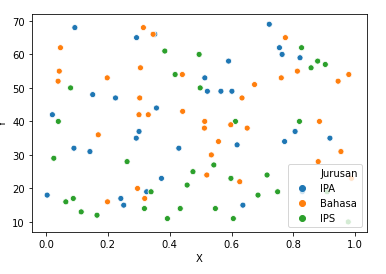
**Menggunakan Seaborn**

Jika ingin menggunakan [Seaborn](https://seaborn.pydata.org/generated/seaborn.scatterplot.html), kita dapat mengatur warna pada parameter **hue**. Parameter **hue** berfungsi mengelompokkan variabel yang akan menghasilkan data point dengan warna berbeda sesuai kategorinya.

Contoh penggunaannya seperti di bawah ini.

import seaborn as sns  
  
sns.scatterplot('X', 'Y', data=df, hue='Jurusan')

**Output:**

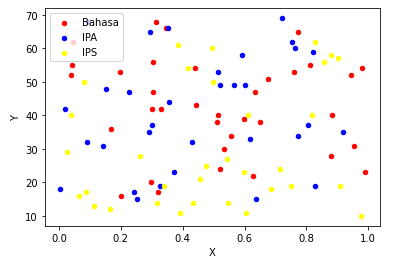


**Menggunakan Groupby Pandas + Matplotlib**

Dengan menggunakan perulangan, kita akan membuat scatter plot untuk tiap data yang telah dikelompokkan dengan **groupby** kemudian menentukan warnanya. Perhatikan kode di bawah ini.

fig, ax = plt.subplots()  
  
df\_group = df.groupby('Jurusan')  
colors = {'Bahasa':'red', 'IPA':'blue', 'IPS':'yellow'}  
  
for name, group in df\_group:  
    group.plot(ax=ax, kind='scatter', x='X', y='Y', label=name, color=colors[name])  
  
plt.show()

**Output:**



Contoh di atas merupakan contoh sederhana untuk mengatur warna pada scatter plot berdasarkan kategori. Tentunya Anda dapat melakukan improvisasi agar visualisasinya terlihat lebih menarik.