Kode program yang akan diajarkan di sini bisa diunduh di tautan [berikut ini](https://colab.research.google.com/drive/1DESTMdcE_8M4kjN8GrMNPzE4z9uNBc5x?usp=sharing). Silahkan buat salinan dari kode ini dan jalankan di Google Colab Anda.

**Tujuan**

Setelah mengenal *support vector regression*, kita akan belajar menerapkan teknik tersebut ke dalam model ML dengan *library SKLearn*.

**Tahapan Latihan**

Tahapan pada latihan kali ini adalah sebagai berikut:

1. Ubah data menjadi Dataframe.
2. Pisahkan atribut dan label.
3. Latih model SVR.
4. Buat plot dari model.

**Codelab**

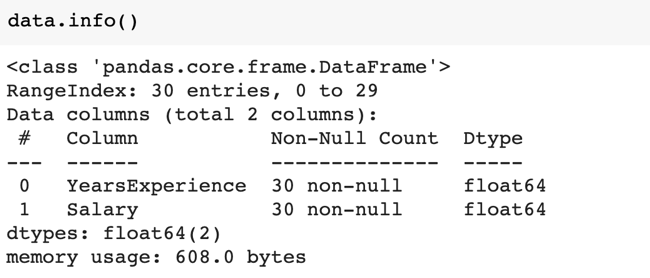
Dataset yang akan kita gunakan adalah data tentang lama kerja seseorang dan gajinya. Dataset ini dapat diunduh pada [tautan](https://www.kaggle.com/karthickveerakumar/salary-data-simple-linear-regression) berikut.

Setelah mengunggah dataset yang sudah di-*unzip* pada Google Colaboratory, pada cell pertama notebook kita impor library dasar yang dibutuhkan. Jangan lupa untuk mengubah berkas csv dari dataset menjadi dataframe Pandas.

1. import pandas as pd
3. # membaca dataset dan mengubahnya menjadi dataframe
4. data = pd.read\_csv('Salary\_Data.csv')

Selanjutnya kita bisa melihat apakah terdapat missing value pada dataset dengan fungsi *.info()*. Keluaran dari cell di bawah menunjukkan bahwa tidak ada missing value pada dataset.

1. data.info()



Selanjutnya kita tampilkan 5 baris pertama dari dataframe.

1. data.head()



Kemudian kita pisahkan antara atribut dan label yang ingin diprediksi. Ketika hanya terdapat satu atribut pada dataframe, maka atribut tersebut perlu diubah bentuknya agar bisa diterima oleh model dari library SKLearn. Untuk mengubah bentuk atribut kita membutuhkan library numpy.

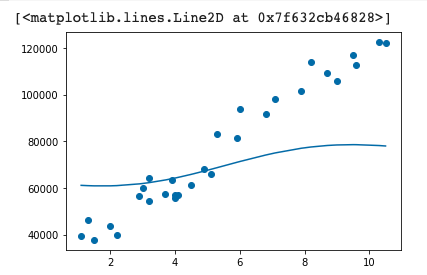
1. import numpy as np
3. # memisahkan atribut dan label
4. X = data['YearsExperience']
5. y = data['Salary']
7. # mengubah bentuk atribut
8. X = np.array(X)
9. X = X[:,np.newaxis]

Berikutnya kita buat objek support vector regression dan di sini kita akan mencoba menggunakan parameter C = 1000, gamma = 0.05, dan kernel ‘rbf’. Setelah model dibuat kita akan melatih model dengan fungsi fit pada data.

1. from sklearn.svm import SVR
3. # membangun model dengan parameter C, gamma, dan kernel
4. model  = SVR(C=1000, gamma=0.05, kernel='rbf')
6. # melatih model dengan fungsi fit
7. model.fit(X,y)

Terakhir kita bisa memvisualisasikan bagaimana model SVR kita menyesuaikan terhadap pola yang terdapat pada data menggunakan library matplotlib.

1. import matplotlib.pyplot as plt
3. # memvisualisasikan model
4. plt.scatter(X, y)
5. plt.plot(X, model.predict(X))



Hasil visualisasi menunjukkan bahwa model yang kita kembangkan, belum mampu menyesuaikan terhadap pola pada data dengan baik. Nah, pada modul berikutnya kita akan mencoba meningkatkan performa model kita dengan menggunakan *grid search*.