

## ***Principal Component Analysis (PCA)***

PCA atau *Principal Component Analysis* merupakan sebuah teknik yang digunakan untuk mengurangi ukuran suatu dimensi menjadi ukuran dimensi yang lebih kecil (*reducing high dimensionality*).

Adapun langkah-langkahnya adalah sebagai berikut:

1. Melakukan normalisasi data. Hal ini ditujukan untuk memastikan rata-rata dari dataset bernilai nol.
2. Setelah itu, hasil normalisasi data dibagi dengan standar deviasinya. Hal ini ditujukan untuk membuat kovarians dari setiap pasangan bernilai satu. Hasil tersebut disebut juga standarisasi.
3. Selanjutnya, cari nilai eigen dan vektor eigen dari dataset tersebut. Vektor eigen dengan norm tertinggi merupakan komponen utamanya (*principal component*). Adapun subruang utama (*principal subspaces*) merupakan ruang sepanjang vektor eigen tersebut.
4. Terakhir, proyeksikan setiap titik data ke subruang utama.

Perlu diingat bahwa PCA merupakan salah satu bentuk metode clustering dalam *unsupervised learning*.

### **Notebook:**

#### **Basic PCA Explanation**

<https://colab.research.google.com/drive/15gx7NYQaP1-ppRbSabFTavGvKH8wvvh32?usp=sharing>

#### **Laptop KMeans Clustering and PCA**

<https://www.kaggle.com/code/fernandasubekti/laptop-clustering-kmeans-pca/notebook>

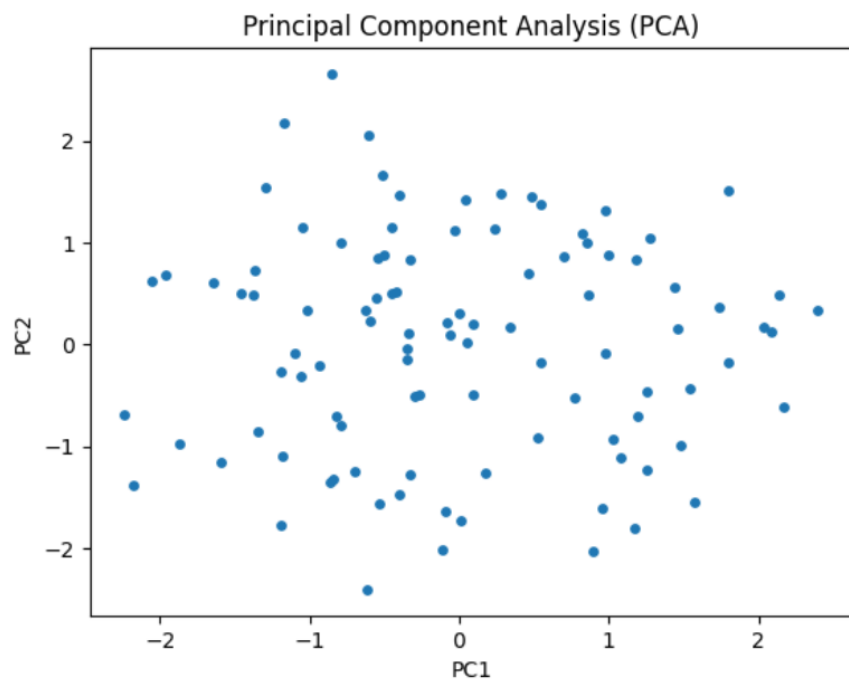
### **Visualisasi Singkat:**

Andaikan saya memiliki 100 titik data dengan 5 titik data pertama adalah sebagai berikut:

```
# return the first five rows
data[:5]

array([[16.06367555, 17.77246269, 13.867177 , 13.58605659, 15.75835698],
       [11.24415911, 18.05377665, 18.77966263, 17.88642799, 18.1911257 ],
       [10.37459984, 18.30735858, 18.01014279, 22.76574671, 20.35185393],
       [16.24982655, 15.27288836, 14.23919903, 19.84971392, 21.06583338],
       [12.76348674, 11.73679675, 21.59475482, 15.66564382, 16.52981865]])
```

Jika saya lakukan reduksi dimensi menjadi dua dimensi, maka hasilnya adalah sebagai berikut:



Sementara itu, hasil visualisasi dari KMeans clustering dengan bantuan PCA pada dataset laptop adalah sebagai berikut:

