

## LATIHAN CLUSTERING

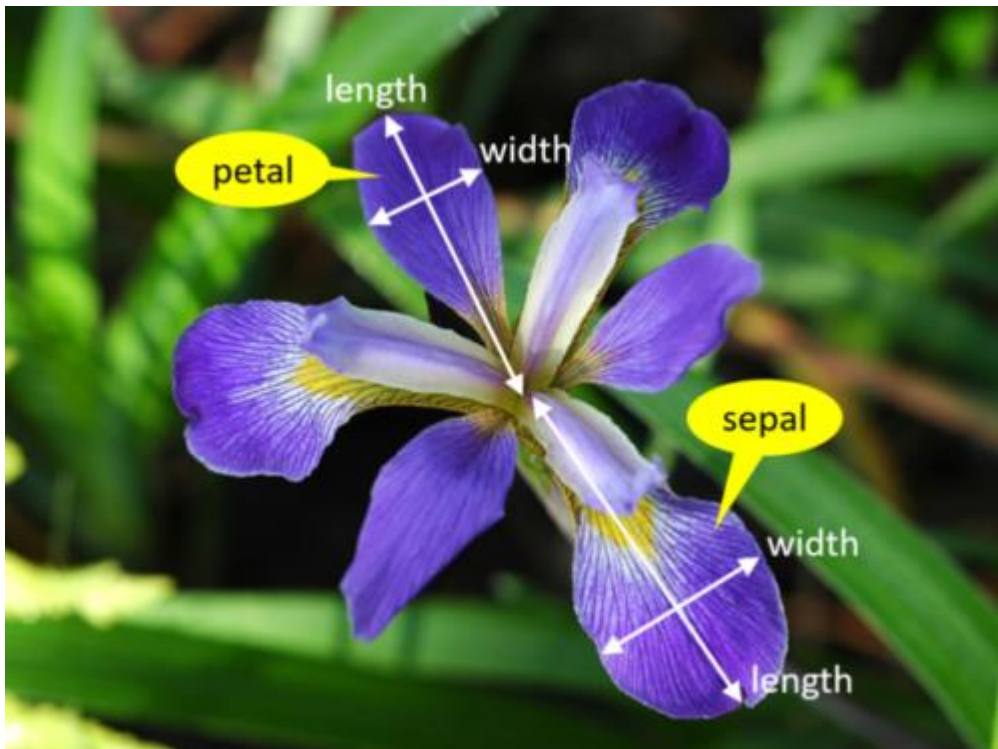
### WORKSHOP R 2022

#### “BUNGA IRIS”

##### Business Understanding

Sebuah kelompok peneliti dari suatu perusahaan pengembang spesies flora sedang melakukan penelitian terkait bunga iris. Peneliti melakukan segala macam kombinasi perkawinan untuk menciptakan kualitas bunga iris terbaik. Pada umumnya, harga bunga iris dinilai berdasarkan ukuran daun kelopak dan ukuran daun mahkotanya.

Setelah dilakukan pengumpulan data, para peneliti ingin mengelompokkan bunga-bunga iris yang dihasilkan ke dalam beberapa kelompok untuk mempermudah proses *pricing*, penjualan dan pemasaran. Pengelompokan dilakukan bukan berdasarkan spesies, tetapi berdasarkan ukuran daun kelopak (sepal) dan ukuran daun mahkota (petal).



Sumber: integratedots.com

### Data Understanding

Pada analisis *clustering* kali ini, digunakan dataset *built-in* yang terdapat di dalam R, yaitu dataset bunga iris milik Fisher atau Anderson. Dataset hasil observasi berisi informasi terkait ukuran (panjang dan lebar) dari daun kelopak (sepal) dan daun mahkota (petal) bunga iris. Selain itu, terdapat juga informasi terkait spesies dari bunga iris yang diobservasi. Namun, spesies dari bunga iris tidak akan masuk ke dalam variabel dalam rangka mengelompokkan bunga iris untuk kepentingan *pricing* nantinya. Walaupun pada hakikatnya, karakteristik suatu spesies bunga iris dapat dilihat atas ukuran daun kelopak dan daun mahkotanya.

### Your Job

Sebagai seorang *data analyst*, kamu diminta bantuan oleh peneliti-peneliti untuk membuat *cluster* bunga Iris berdasarkan panjang dan lebar dari *petal* dan *sepal* bunga Iris. Gunakanlah metode *clustering* secara khusus algoritma K-means untuk proses *clustering* yang dilakukan.