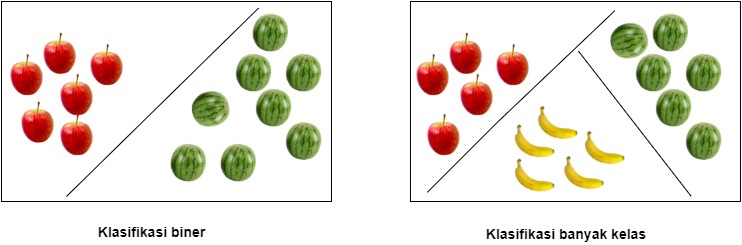
**Binary Classification**

Seperti yang telah Anda ketahui, klasifikasi adalah teknik untuk menentukan kelas atau kategori berdasarkan atribut yang diberikan. Klasifikasi yang menghasilkan dua kategori disebut klasifikasi biner, sedangkan klasifikasi yang menghasilkan 3 kategori atau lebih disebut multiclass classification atau klasifikasi banyak kelas.

[](https://www.dicoding.com/academies/184/tutorials/8387?from=16345)

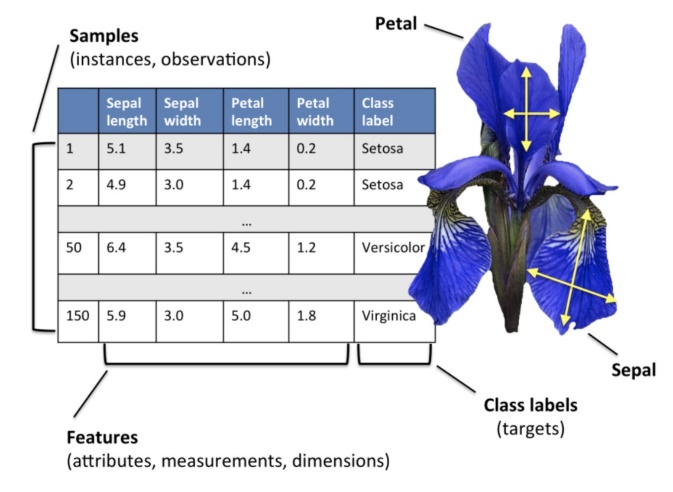
Bayangkan sebuah pertanyaan yang jawabannya adalah iya atau tidak, atau pertanyaan yang membuat Anda memilih antara satu pilihan atau pilihan yang lain. Klasifikasi biner bertujuan untuk membedakan dua kelas yang berbeda. Contohnya, klasifikasi buah semangka atau apel, klasifikasi laki-laki atau perempuan, klasifikasi email spam, dan lain-lain. Pada kasus klasifikasi email sebagai spam atau bukan, pertanyaan yang diajukan adalah: “Apakah email ini adalah spam?”

Sampai sini sudah paham ya. Pada modul selanjutnya, kita akan belajar tentang klasifikasi banyak kelas.

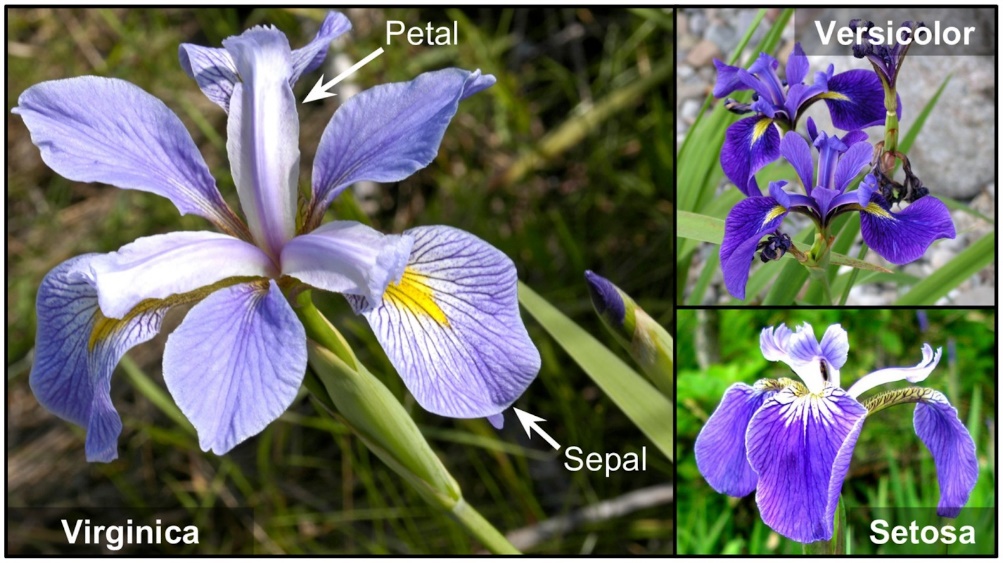
**Multiclass Classification**

Untuk lebih mudah dalam memahami klasifikasi banyak kelas, kita akan menggunakan contoh dataset Iris. Dataset iris merupakan salah satu dataset populer untuk belajar bagaimana ML dipakai dalam klasifikasi. Dataset ini berisi 150 sampel dari 3 spesies bunga iris.

Pada dataset Iris terdapat 4 kolom atribut yaitu panjang sepal, lebar sepal, panjang petal, dan lebar petal. Untuk label terdapat 3 kelas yaitu Setosa, Versicolor dan Virginica. Kelas adalah kategori atau jenis yang terdapat pada dataset. Dalam hal ini pada dataset terdapat 3 kelas yaitu Setosa, Versicolor, dan Virginica.

[](https://www.dicoding.com/academies/184/tutorials/8387?from=16345)

Sebuah model classification bertujuan untuk menentukan kelas berdasarkan atribut tertentu. Pada kasus klasifikasi Iris sebuah model bertugas untuk memprediksi spesies sebuah bunga iris berdasarkan atributnya yaitu panjang sepal, lebar sepal, panjang petal, dan lebar petalnya.

[](https://www.dicoding.com/academies/184/tutorials/8387?from=16345)

Contohnya panjang petal dari Iris Setosa lebih pendek dari spesies versicolor dan virginica. Jika panjang petal pendek maka kemungkinan spesies Iris tersebut adalah Setosa.

Seperti yang sudah dijelaskan pada modul sebelumya, klasifikasi biner terdapat hanya 2 kelas pada dataset. Sedangkan pada klasifikasi banyak kelas terdapat lebih dari 2 kelas. Pada contoh dataset Iris seperti di atas terdapat 3 kelas maka dataset tersebut adalah kasus klasifikasi banyak kelas.

**Decision Tree**

*Decision tree* atau pohon keputusan adalah salah satu algoritma supervised learning yang dapat dipakai untuk masalah klasifikasi dan regresi. Decision tree merupakan algoritma yang powerful alias mampu dipakai dalam masalah yang kompleks. Decision tree juga merupakan komponen pembangun utama algoritma Random Forest, yang merupakan salah satu algoritma paling powerful saat ini.

Decision tree memprediksi sebuah kelas (klasifikasi) atau nilai (regresi) berdasarkan aturan-aturan yang dibentuk setelah mempelajari data.

Misalnya kita memiliki data seperti di bawah. Data berisi informasi mengenai kondisi cuaca pada hari tertentu dan apakah cocok untuk bermain golf di kondisi cuaca tersebut.

| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play Golf** |
| --- | --- | --- | --- | --- |
| Rainy | Hot | High | False | No |
| Rainy | Hot | High | True | No |
| Overcast | Hot | High | False | Yes |
| Sunny | Mild | High | False | Yes |
| Sunny | Cold | Normal | False | Yes |
| Sunny | Cold | Normal | True | No |
| Overcast | Cold | Normal | True | Yes |
| Rainy | Mild | High | False | No |
| Rainy | Cold | Normal | False | No |
| Rainy | Mild | Normal | False | No |

Sebuah pohon keputusan dapat dibuat dari data sebelumnya. Perhatikan contoh pohon keputusan di bawah. Pohon ini menggunakan hanya 2 atribut yaitu kondisi langit dan kecepatan angin untuk menentukan bermain golf atau tidak.

