

## Rangkuman Hasil Klasifikasi Dataset

Dataset : breast-cancer-winsconsin.csv (data tentang kanker payudara)

Sumber dataset : <https://www.kaggle.com/salihacur/breastcancerwisconsin>

Jumlah kolom / variabel: 11 (Sebelas)

Rincian variabel :

- Id : id sampel
- Clump Thicknes : Skala (1-10)
- Uniformity of Cell Size : Skala (1-10)
- Uniformity of Cell Shape : Skala (1-10)
- Marginal Adhesion : Skala (1-10)
- Single Epithelial Cell Size : Skala (1-10)
- Bare Nuclei : Skala (1-10)
- Bland Chromatin : Skala (1-10)
- Normal Nucleoli : Skala (1-10)
- Mitoses : Skala (1-10)
- Benormal : 2 dan 4 (2 untuk kanker jinak, 4 untuk kanker ganas)

Variabel benormal digunakan sebagai atribut utama untuk diklasifikasikan berdasarkan skala yang muncul pada atribut lainnya apakah sampel tersebut merupakan kategori kanker jinak atau kanker ganas.

Total data : 699 sampel (sebelum dilakukan preprosesing data, masih terdapat data yang tidak lengkap)

Total data : 683 sampel (sesudah dilakukan preprosesing data, data sudah dibersihkan dan disamakan tipe datanya)

Tingkat error : 1.5%

Atribut utama : uniform cell size (persentase tertinggi sebanyak 100 persen dalam mengklasifikasikan kanker jinak / ganas)

Klasifikasi : 438 sampel diklasifikasikan sebagai kanker jinak, sedangkan 245 sampel diklasifikasikan sebagai kanker ganas. 64,12% sampel menderita kanker dan 35,87% sampel kanker ganas.