**BREAST CANCER WISCONSIN**

BACK GROUND

Data mengenai kanker payudara ini didapat dengan melakukan perhitungan terhadap gambar digital atas uji Aspirasi Jarum Halus (FNA) dari massa payudara. Data ini menggambarkan karakteristik inti sel yang ada pada gambar.

Dalam pembuatan data, digunakan bidang pemisah untuk mengolah gambar yang dibuat dengan menggunakan Multisurface Method-Tree (MSM-T), sebuah metode klasifikasi yang menggunakan pemrograman linier untuk membuat pohon keputusan. Fitur yang relevan dipilih menggunakan pencarian menyeluruh di ruang dengan 1-4 fitur dan 1-3 bidang pemisah.

Tujuan dibuatnya dataset ini adalah untuk mengidentifikasi jumlah kelas kanker yang jinak (benign) atau ganas (malignant). Sample ini dikumpulkan secara periodical atas laporan klinis dari Dr. Wolberg, oleh karena ini database ini mencerminkan pengelompokan secara kronologis.

DATASET

* Nomor ID
* Diagnosis (M = malignant, B = benign)
* Sepuluh fitur yang diwakili oleh bilangan real dihitung untuk setiap inti sel:
  + a) radius (jarak rt-rata dari titik tengah ke perimeter)
  + b) texture (standar deviasi nilai skala abu-abu)
  + c) perimeter
  + d) area
  + e) smoothness (variasi lokal dalam panjang jari-jari)
  + f) compactness (perimeter^2 / area - 1.0)
  + g) concavity (keparahan bagian cekung dari kontur)
  + h) concave points (jumlah bagian cekung dari kontur)
  + i) symmetry
  + j) fractal dimension ("perkiraan coastline" - 1)

Mean, standard error dan Worst (mean dari tiga nilai terbesar) dari fitur-fitur ini dihitung untuk tiap gambar, sehingga menghasilkan 30 fitur. Sebagai contoh, kolom ke 3 adalah Mean Radius, kolom 13 adalah Radius SE, kolom 23 adalah Worst Radius. Semua nilai fitur disimpan dengan empat angka signifikan.

REQUEST

Dari data ini, silahkan buat model yang dapat mengklasifikasikan kanker dengan tipe jinak (benign) dan tipe ganas (malignant)

TARGET

* Lakukan eksplorasi data yang ada, untuk melihat perilaku dari data
* Lakukan data preprocessing
* Buatlah model untuk memprediksi jenis kanker berdasarkan fitur-fitur yang ada pada data set

REQUIREMENT

* Buatlah model klasifikasi menggunakan minimal 3 model yang sudah dipelajari sebelumnya dengan F1 score diatas 0.75
* Lakukan optimalisasi model dengan menggunakan hyperprameter search