

## Project Eksperimen dataset menggunakan Naive Bayes dan PCA

Anggota Kelompok :

1. Adinda Aulia Sari (2117051018)
2. Shela Rana Kezyah (2157051007)
3. Nurul Fadilah (2117051098)

### Gender Classification Dataset

Gender Classification Dataset adalah kumpulan data yang digunakan untuk melatih dan menguji model machine learning atau algoritma untuk memprediksi atau mengidentifikasi jenis kelamin (gender) seseorang berdasarkan atribut-atribut tertentu. Terdapat 5001 sampel yang ada di dalam dataset. Dataset ini umumnya digunakan dalam berbagai konteks, seperti analisis sosial, pemasaran, dan penelitian medis.

Identifikasi gender pada dataset ini berdasarkan pada keterangan dari 7 fitur dan kolom label yang berisikan :

1. **long\_hair** - Kolom ini berisi angka 0 dan 1 dimana 1 adalah "rambut panjang" dan 0 adalah "rambut tidak panjang".
2. **forehead\_width\_cm** - Kolom ini dalam CM. Ini adalah lebar dahi.
3. **forehead\_height\_cm** - Ini adalah tinggi dahi dan dalam satuan Cm.
4. **nose\_wide** - Kolom ini berisi angka 0 dan 1 dimana 1 adalah "hidung lebar" dan 0 adalah "hidung tidak lebar".
5. **nose\_long** - Kolom ini berisi angka 0 dan 1 dimana 1 adalah "Hidung panjang" dan 0 adalah "hidung tidak panjang".
6. **lips\_thin** - Kolom ini berisi angka 0 dan 1 dimana 1 melambangkan "bibir tipis" sedangkan 0 berarti "Bibir tidak tipis".
7. **distance\_nose\_to\_lip\_long** - Kolom ini berisi angka 0 dan 1 dimana 1 mewakili "jarak jauh antara hidung dan bibir" sedangkan 0 adalah "jarak pendek antara hidung dan bibir".
8. **gender** - Ini bisa berupa "Pria" atau "Wanita".

Link Dataset : [https://github.com/shelakzy/Dataset-PrakPola/blob/main/gender\\_classification\\_v7.csv](https://github.com/shelakzy/Dataset-PrakPola/blob/main/gender_classification_v7.csv)

Sumber Dataset : Kaggle