

# Klasifikasi Kanker Payudara

Nama 1: Bayu Setho K.S (2101192050)

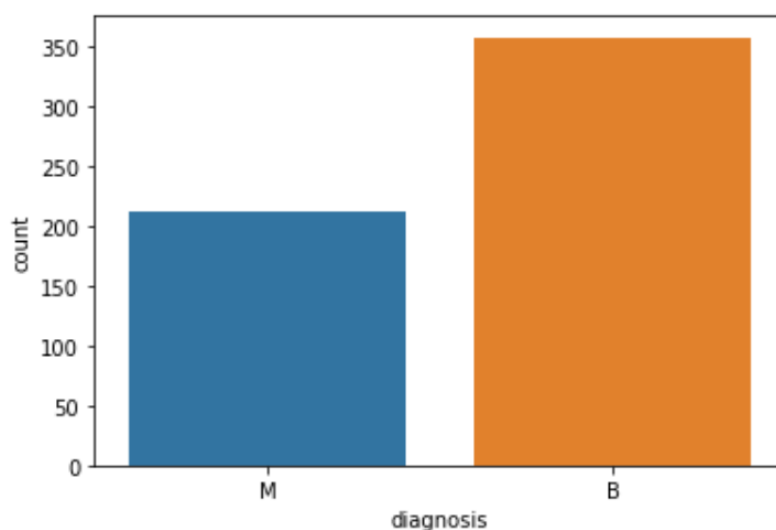
Nama 2: Akhmad Yusuf Nasirudin (2101191025)

## A. Pendahuluan

Proses klasifikasi fitur kanker payudara dikelompokkan ke dalam dua kelas, yaitu malignant dan benign. Dataset terdiri dari 569 data dengan 30 fitur disetiap datanya. Untuk meminimalkan jumlah fitur, dilakukan pengurangan dimensi pada dataset tersebut menjadi hanya satu fitur saja. Pemilihan satu fitur ini dikarenakan pada tugas ini akan dilakukan perbandingan metode pengurangan dimensi yaitu PCA, LDA, dan FA. Dikarenakan LDA hanya bisa menghasilkan jumlah komponen fitur sebanyak jumlah\_kelas-1, maka semua metode akan menggunakan jumlah komponen fitur=1. Dataset akan dibagi menjadi train dan test dengan besar perbandingan 80:20. Klasifikasi dilakukan menggunakan Multi Layer Perceptron Classifier (MLPClassifier) dengan jumlah hidden layer sebanyak 20 dimana setiap layer memiliki 10 neuron. Proses training dilakukan sebanyak 500 iterasi.

## B. Preprocessing Data

Jumlah keseluruhan dari dataset adalah 569 yang terbagi kedalam dua kelas, malignant dan benign. Perbandingan dari kelas tersebut dapat dilihat pada diagram di bawah ini.



Kolom "Unnamed: 32" dan "id" akan dihilangkan dari dataset karena kolom tersebut tidak mengandung fitur yang dibutuhkan untuk klasifikasi. Setelah mengambil fitur yang dibutuhkan, hal selanjutnya yang dilakukan adalah normalisasi seluruh fitur menggunakan StandardScaler agar nilai fitur tidak ada yang menyimpang terlalu jauh. Selanjutnya, data yang telah dinormalisasi akan menjadi input pada 3 metode pengurangan dimensi (PCA, LDA, dan FA). Dikarenakan MLPClassifier hanya menerima tipe data integer atau real, maka nama kelas dataset akan dikodekan menjadi 0 & 1 menggunakan LabelBinarizer. Setelah itu, keseluruhan dataset akan dibagi menjadi test dan train dengan perbandingan 80:20 dimana training sample berjumlah 455, dan testing sample berjumlah 114.

## C. Training Model

Training model dilakukan menggunakan MLPClassifier dengan Langkah-langkah sebagai berikut:

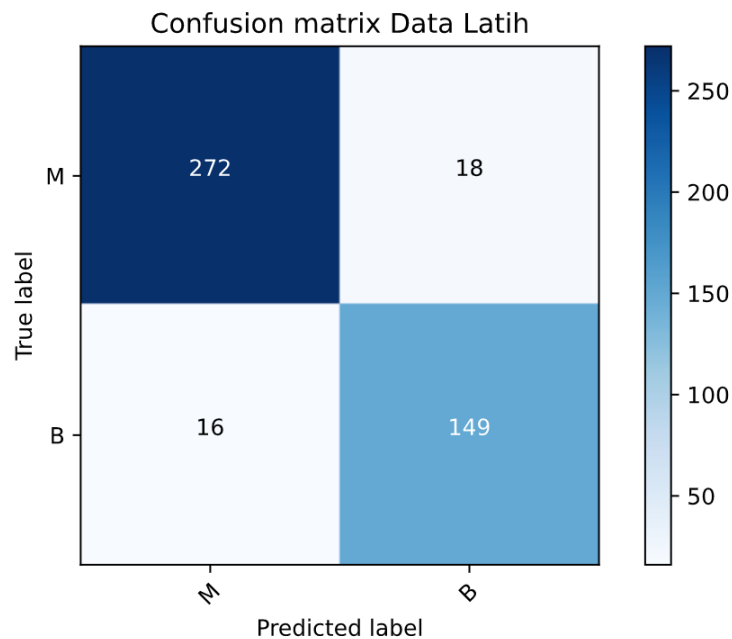
1. Import MLPClassifier dari library sklearn.neural\_network
2. Setting konfigurasi dari hyperparameter training model. Dimana pada tugas ini digunakan jumlah hidden layer sebanyak 20 dengan 10 neuron pada masing-masing hidden layer. Training dilakukan dengan maksimal iterasi sebanyak 500.
3. Fit dataset dengan model yang telah dikonfigurasi.

#### D. Prediksi

Berikut merupakan prediksi dari ketiga model dengan metode pengurangan dimensi yang berbeda.

##### 1. PCA

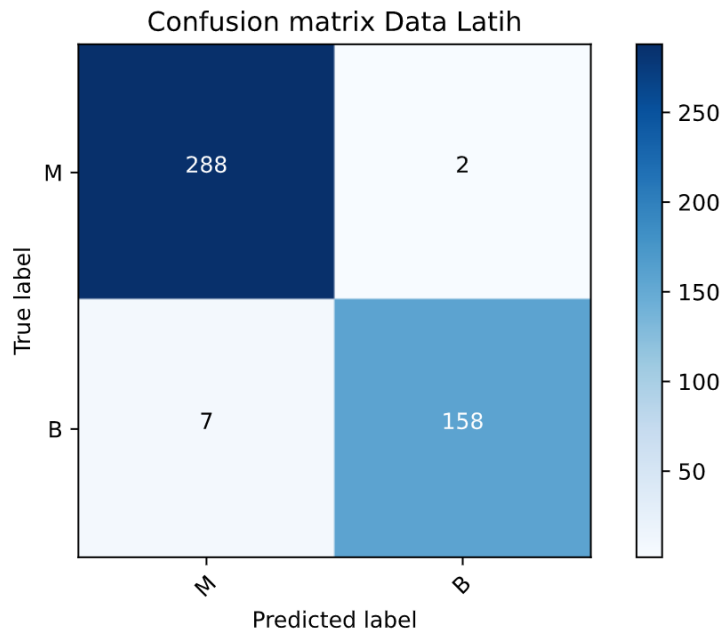
	precision	recall	f1-score	support
0	0.94	0.94	0.94	290
1	0.89	0.90	0.90	165
accuracy			0.93	455
macro avg	0.92	0.92	0.92	455
weighted avg	0.93	0.93	0.93	455



Pada PCA, didapatkan akurasi sebesar 93% dengan data Malignant yang diprediksi benar sebanyak 272, dan yang salah sebanyak 18. Sedangkan benign yang diprediksi benar sebanyak 149 dan yang salah sebanyak 18.

##### 2. LDA

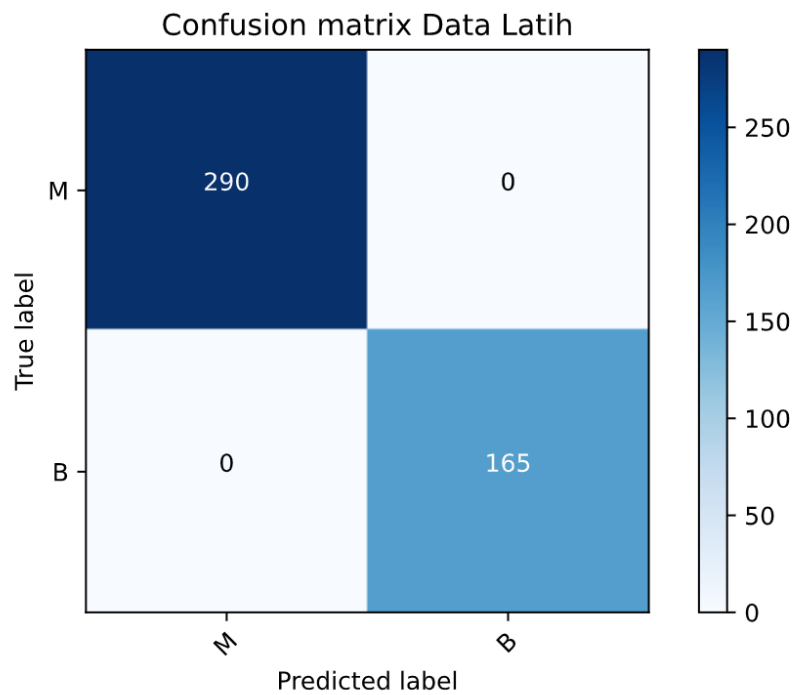
	precision	recall	f1-score	support
0	0.98	0.99	0.98	290
1	0.99	0.96	0.97	165
accuracy			0.98	455
macro avg	0.98	0.98	0.98	455
weighted avg	0.98	0.98	0.98	455



Pada metode LDA, didapatkan akurasi sebesar 98% dimana Malignant yang diklasifikasi benar sebanyak 288, dan yang salah sebanyak 7. Sedangkan Benign yang diklasifikasi benar sebanyak 158 dan yang salah sebanyak 2.

### 3. FA

	precision	recall	f1-score	support
0	1.00	1.00	1.00	290
1	1.00	1.00	1.00	165
accuracy			1.00	455
macro avg	1.00	1.00	1.00	455
weighted avg	1.00	1.00	1.00	455



Pada metode FA didapatkan akurasi sebesar 100% dimana semua data diklasifikasikan dengan benar.

#### E. Kesimpulan

Telah dilakukan klasifikasi data kanker payudara dengan menggunakan tiga metode pengurangan dimensi yaitu PCA, LDA, dan FA. Dari percobaan yang telah dilakukan, hasil terbaik didapatkan saat percobaan menggunakan metode pengurangan dimensi FA dengan nilai akurasi 100%.