

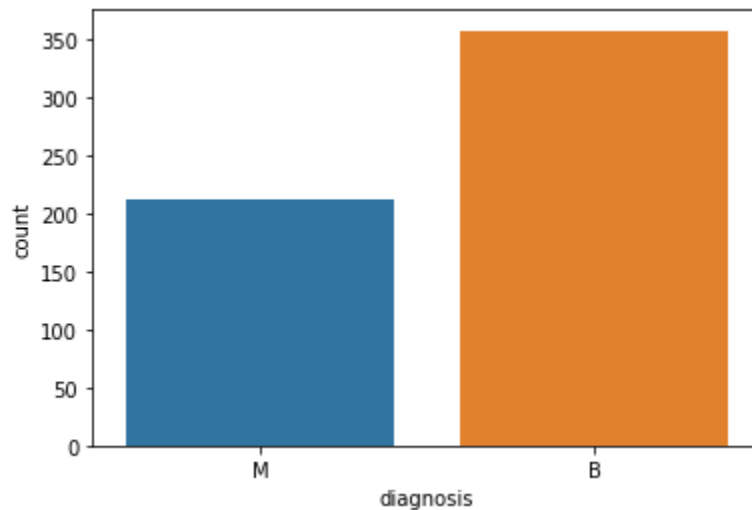
Obed Rhesa Ludwiniananda Handoko
(2101191020)

UAS Pembelajaran Secara Statistik & Optimisasi

I. Background

Klasifikasi fitur kanker dikelompokkan ke dalam dua kelas, yaitu **malignant** dan **benign**. Dataset terdiri dari **569 data** dengan **30 fitur** pada setiap data. Untuk meminimalkan jumlah fitur, dilakukan pengurangan dimensi pada dataset tersebut menjadi hanya satu fitur saja (**Dimensional Reduction**). Pada tugas ini dilakukan perbandingan metode pengurangan dimensi dengan 3 metode, yaitu **Principal Components Analysis (PCA)**, **Linear Discriminant Analysis (LDA)**, dan **Factor Analysis (FA)**. Dikarenakan LDA hanya dapat 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 **70:30**. Klasifikasi dilakukan menggunakan **Multi Layer Perceptron Classifier (MLP Classifier)** dengan jumlah hidden layer sebanyak **40** dengan **20 neuron** di setiap layer. Proses training dilakukan sebanyak **1000 iterasi**.

II. 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 **Standard Scaler** 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 **MLP Classifier** hanya menerima tipe data integer atau real

maka nama kelas dataset akan dikodekan menjadi 0 & 1 menggunakan *Label Binarizer*. Setelah itu, keseluruhan dataset akan dibagi menjadi test dan train dengan perbandingan **70:30** dimana **training sample berjumlah 398**, dan **testing sample berjumlah 171**.

III. Training Model

Training model dilakukan menggunakan *MLP Classifier* dengan langkah – Langkah sebagai berikut:

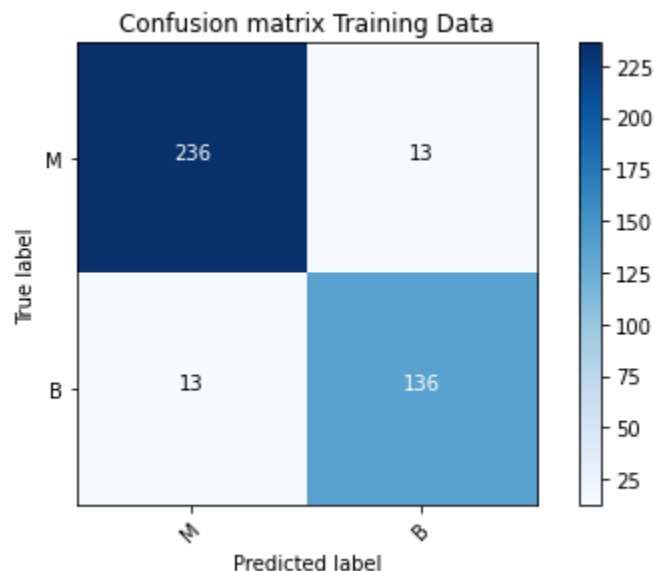
1. Import *MLP Classifier* dari library *sklearn.neural_network*.
2. Setting konfigurasi dari hyperparameter training model. Dimana pada tugas ini digunakan jumlah hidden layer sebanyak **40** dengan **20 neuron** pada setiap hidden layer. Training dilakukan dengan maksimal **iterasi sebanyak 1000**.
3. Fit dataset dengan model yang telah dikonfigurasi.

IV. Prediction

Berikut merupakan prediksi dari ketiga model dengan metode *dimensional reduction* yang berbeda:

1. *Principal Components Analysis (PCA)*

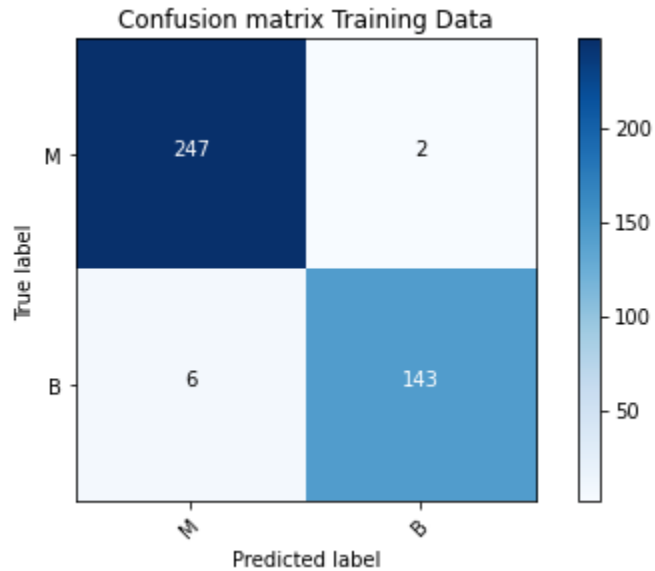
	precision	recall	f1-score	support
0	0.95	0.95	0.95	249
1	0.91	0.91	0.91	149
accuracy			0.93	398
macro avg	0.93	0.93	0.93	398
weighted avg	0.93	0.93	0.93	398



Pada **PCA**, didapatkan akurasi sebesar **93%** dengan data *malignant* memiliki prediksi benar sebanyak **236**, dan salah sebanyak **13**. Sementara data *benign* memiliki prediksi benar sebanyak **149** dan salah sebanyak **18**.

2. *Linear Discriminant Analysis (LDA)*

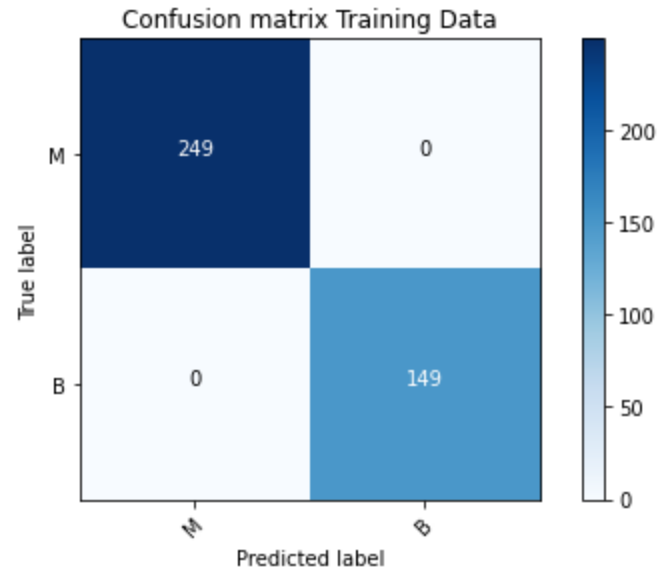
	precision	recall	f1-score	support
0	0.98	0.99	0.98	249
1	0.99	0.96	0.97	149
accuracy			0.98	398
macro avg	0.98	0.98	0.98	398
weighted avg	0.98	0.98	0.98	398



Pada **LDA**, didapatkan akurasi sebesar **98%** dengan data *malignant* memiliki prediksi benar sebanyak **247**, dan salah sebanyak **6**. Sementara data *benign* memiliki prediksi benar sebanyak **143** dan salah sebanyak **2**.

3. *Factor Analysis (FA)*

	precision	recall	f1-score	support
0	1.00	1.00	1.00	249
1	1.00	1.00	1.00	149
accuracy			1.00	398
macro avg	1.00	1.00	1.00	398
weighted avg	1.00	1.00	1.00	398



Pada **FA**, didapatkan akurasi sebesar **100%** dengan data *malignant* memiliki prediksi benar sebanyak **249**, dan salah sebanyak **0** dan data *benign* juga memiliki prediksi benar sebanyak **143** dan salah sebanyak **0**.

V. Conclusion

Setelah dilakukan klasifikasi data cancer dengan menggunakan 3 metode *dimensional reduction* yaitu **PCA**, **LDA**, dan **FA**. Dari percobaan tersebut, hasil terbaik didapatkan pada metode *Factor Analysis* (**FA**) dengan nilai akurasi mencapai **100%**.