

Nama : Rania Yunaningrum

NRP : 5003201076

Dataset

Dataset The Breast Cancer Wisconsin (Diagnostic) terdiri atas 596 baris dan 30 kolom. Artinya dataset The Breast Cancer Wisconsin (Diagnostic) memiliki 30 variabel dengan Y atau target adalah variabel target.

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	...	worst radius	worst texture	worst perimeter	worst area	worst smoothness	worst compactness
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	0.07871	...	25.380	17.33	184.60	2019.0	0.16220	0.0661
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	0.05667	...	24.990	23.41	158.80	1956.0	0.12380	0.1871
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	0.05999	...	23.570	25.53	152.50	1709.0	0.14440	0.4201
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	0.09744	...	14.910	26.50	98.87	567.7	0.20980	0.8683
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	0.05883	...	22.540	16.67	152.20	1575.0	0.13740	0.2061
...
564	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	0.05623	...	25.450	26.40	166.10	2027.0	0.14100	0.2156
565	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	0.05533	...	23.690	38.25	155.00	1731.0	0.11660	0.1912
566	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	0.05648	...	18.980	34.12	126.70	1124.0	0.11390	0.3009
567	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	0.07016	...	25.740	39.42	184.60	1821.0	0.16500	0.8683

Modelling

Dengan pemodelan menggunakan klasifikasi sebagai berikut

```
#IMPORT PACKAGES YANG DIPERLUKAN
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.datasets import load_breast_cancer

#LOAD DATA BREAST CANCER
cancer = load_breast_cancer()
X = pd.DataFrame(cancer.data, columns=cancer.feature_names)
y = pd.Series(cancer.target, name='target')

#SPLIT DATA > TRAIN DAN TESTING
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

#STANDARISASI DATA
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

#LDA
lda = LinearDiscriminantAnalysis()
lda.fit(X_train_scaled, y_train)

#PREDICTION
```

Nama : Rania Yunaningrum

NRP : 5003201076

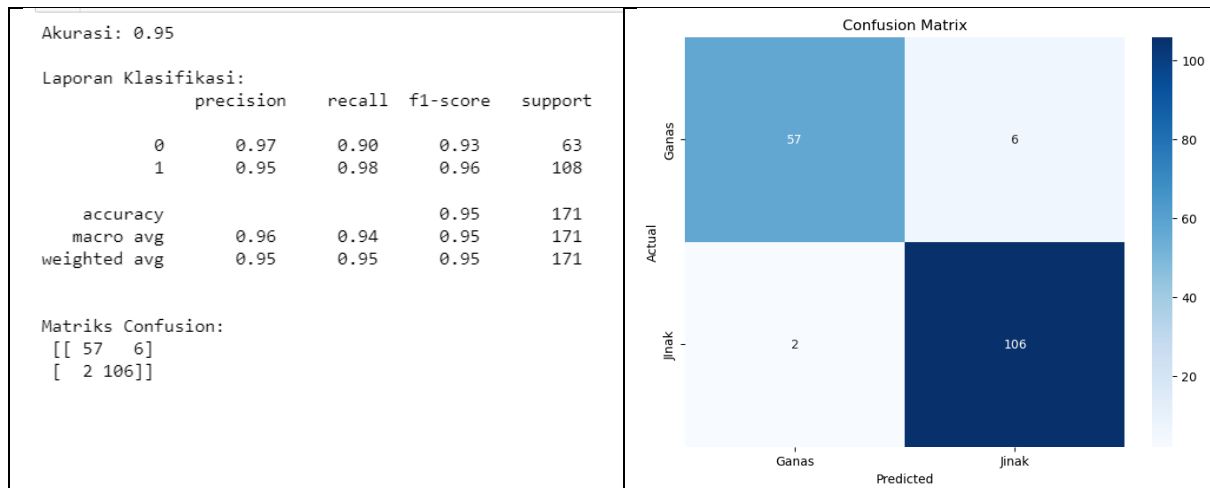
```
y_pred = lda.predict(X_test_scaled)
y_pred

# Akurasi
accuracy = accuracy_score(y_test, y_pred)
print(f'Akurasi: {accuracy:.2f}')
```

```
# Laporan Klasifikasi
print("\nLaporan Klasifikasi:\n", classification_report(y_test, y_pred))

# Matriks Confusion
cm = confusion_matrix(y_test, y_pred)
print("\nMatriks Confusion:\n", cm)
```

Diperoleh hasil sebagai berikut.



Dengan hasil tersebut, didapatkan akurasi sebesar 95% dimana model memprediksi data dengan benar sebesar 95%. Akurasi ini terbilang cukup tinggi dan model ini memiliki kinerja yang baik dalam mengklasifikasikan tumor menjadi ganas atau jinak. Selain menggunakan akurasi, bisa juga dilihat dengan laporan klasifikasi lainnya seperti :

1. **Precision** (presisi) sebesar 0.95 menunjukkan bahwa sedikit tumor jinak yang salah diklasifikasikan sebagai ganas.
2. **Recall** (sensitivitas) sebesar 0.94 artinya menunjukkan model memiliki kemampuan yang baik untuk mengidentifikasi tumor ganas. Pada kasus ini, semakin besar sensitivitas suatu model, maka akan semakin baik di kehidupan nyata.
3. **F1-score** sebesar 0.95 artinya menunjukkan keseimbangan yang baik antara precision dan recall.

Adapun komponen confusion matrix dapat dijelaskan sebagai berikut.

1. True Positive (TP): 57, jumlah tumor ganas yang benar diprediksi sebagai ganas.
2. True Negative (TN): 106, jumlah tumor jinak yang benar diprediksi sebagai jinak.
3. False Positive (FP): 2, jumlah tumor jinak yang salah diprediksi sebagai ganas.
4. False Negative (FN): 6, jumlah tumor ganas yang salah diprediksi sebagai jinak.