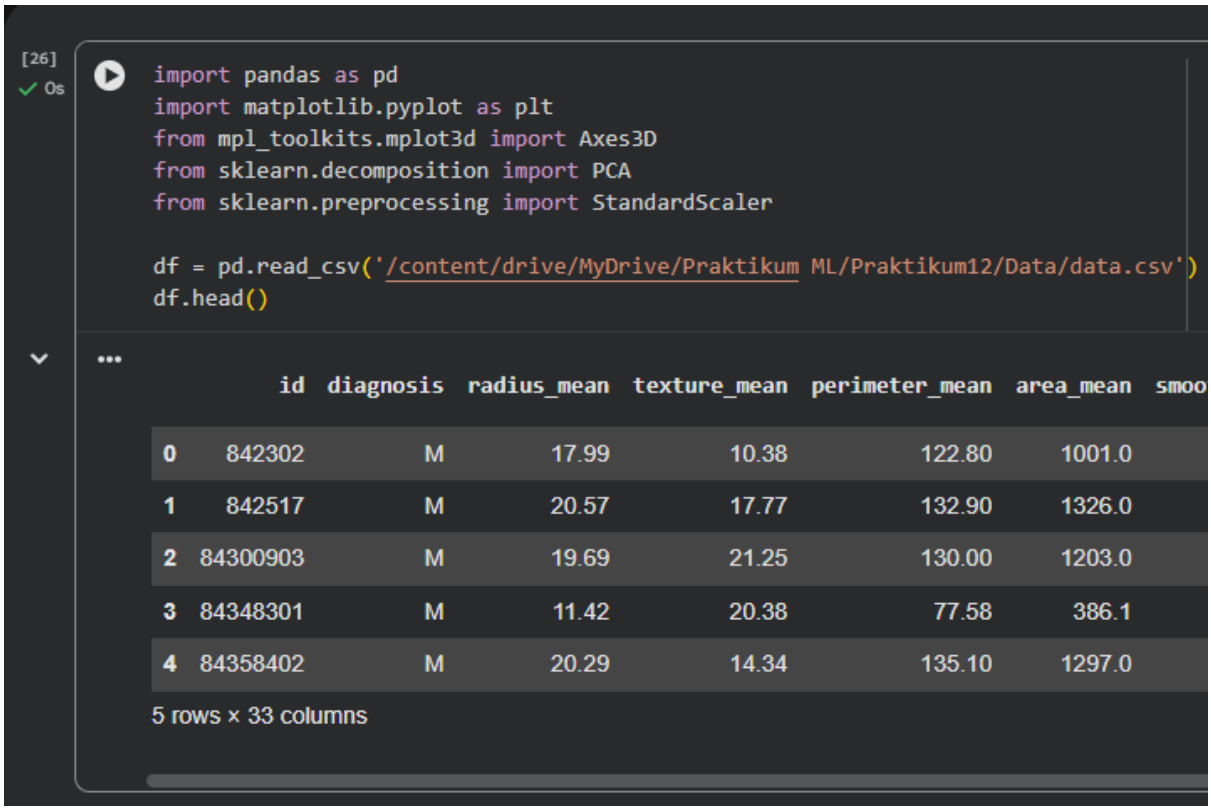
Tugas12 Praktikum Mandiri

SYAHRI GHIFARI MAULIDI 0110222217

[1] Teknik Informatika, STT Terpadu Nurul Fikri, Depok

1.1 Membaca Dataset



```
[26]
✓ 0s
import pandas as pd
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

df = pd.read_csv('/content/drive/MyDrive/Praktikum ML/Praktikum12/Data/data.csv')
df.head()
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoo
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	

5 rows × 33 columns

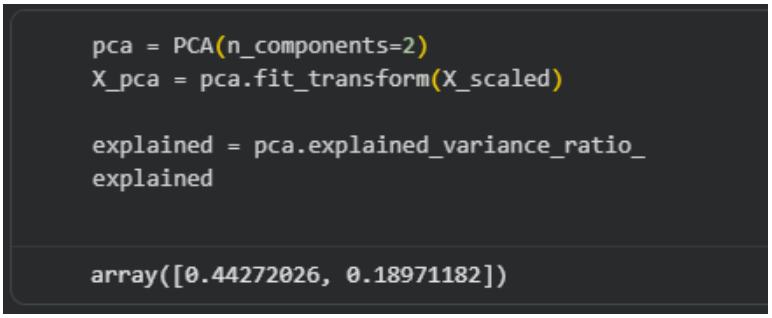
Gambar ini menunjukkan tahap awal analisis, dilakukan proses pemanggilan dan pembacaan dataset *Breast Cancer Wisconsin* menggunakan library pandas. Dataset dimuat dari direktori Google Drive dengan perintah pd.read_csv(), yang kemudian diverifikasi melalui fungsi df.head() untuk menampilkan lima baris pertama dataset. Dari hasil tampilan tersebut diketahui bahwa dataset terdiri atas 33 kolom yang mencakup kolom identitas (id), kolom diagnosis (diagnosis), serta 30 fitur numerik yang merepresentasikan karakteristik fisik sel kanker yang diperoleh melalui analisis citra digital. Nilai-nilai fitur seperti radius_mean, texture_mean, perimeter_mean, dan area_mean merupakan indikator awal yang digunakan dalam proses diagnosis kanker payudara secara komputasi.

1.2 Persiapan Data untuk PCA

Gambar ini menunjukan Proses dimulai dengan mengakses penyimpanan Google Drive menggunakan fungsi drive.mount(), yang memungkinkan sistem membaca dataset secara langsung dari direktori pengguna.

langkah berikutnya adalah melakukan pembersihan data dengan menghapus kolom yang tidak relevan, yaitu kolom id dan Unnamed: 32. Kolom id tidak memiliki nilai diagnostik dan hanya berfungsi sebagai identitas sampel, sedangkan Unnamed: 32 merupakan kolom kosong akibat format penyimpanan data.

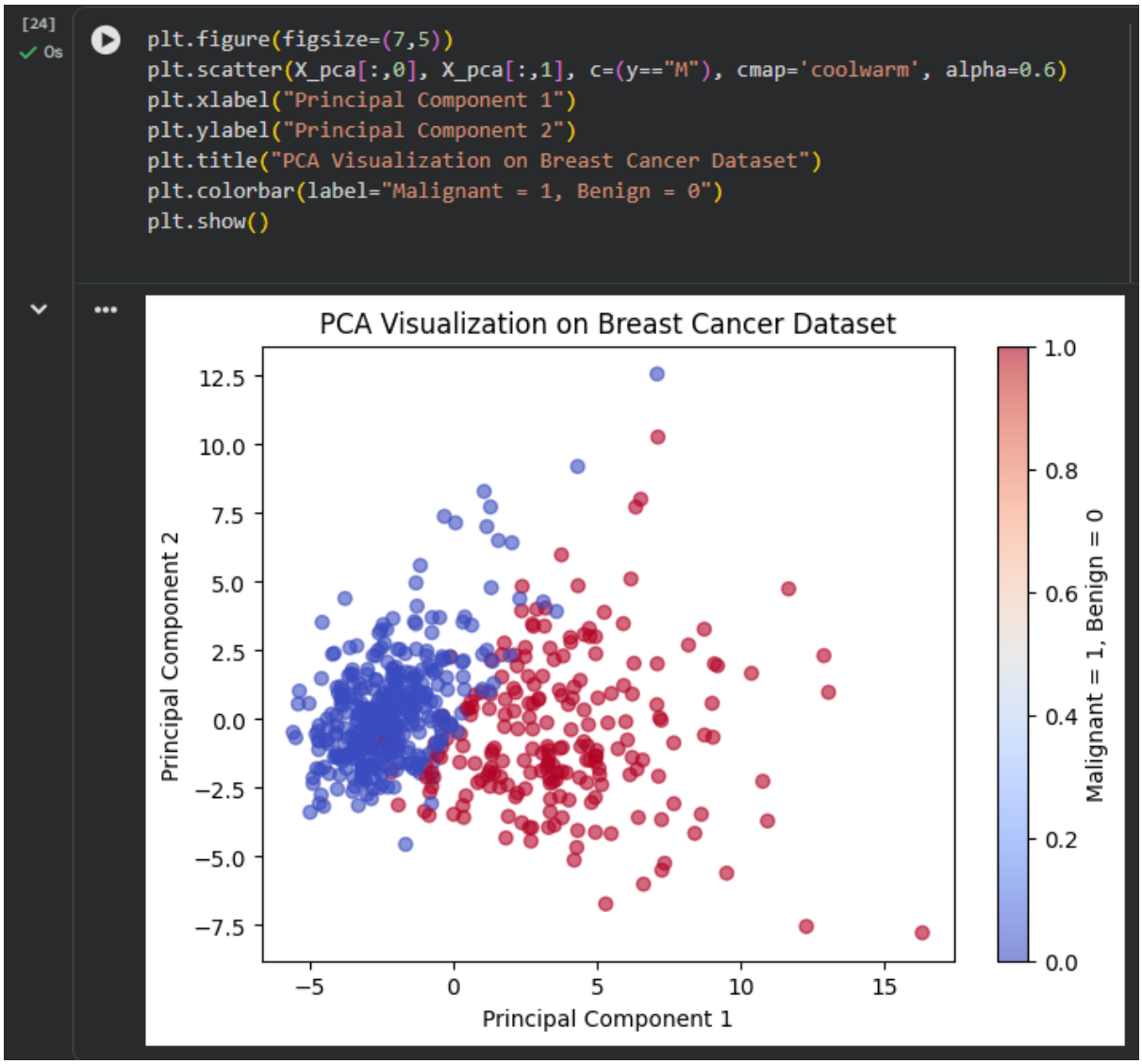
1.3 Penerapan PCA



Gambar ini menampilkan proses reduksi dimensi menggunakan metode Principal Component Analysis (PCA). PCA diterapkan dengan menetapkan dua komponen utama (n_components=2)

dengan tujuan merangkum informasi terpenting dari 30 fitur numerik ke dalam dua komponen baru yang lebih sederhana namun tetap informatif.

2,1 Visualisasi PCA



Gambar ini merupakan Tahap visualisasi PCA dua komponen dilakukan untuk memberikan gambaran yang lebih intuitif mengenai distribusi data setelah dilakukan reduksi dimensi dari 30 fitur numerik menjadi dua komponen utama. Visualisasi ini dibuat menggunakan scatter plot yang memetakan setiap sampel ke dalam ruang dua dimensi berdasarkan nilai *Principal Component 1 (PC1)* dan *Principal Component 2 (PC2)* yang dihasilkan oleh PCA. Warna titik pada scatter plot ditentukan berdasarkan nilai diagnosis,mampu memisahkan area dengan kepadatan gempa tinggi dari area yang memiliki titik gempa terpencar.

Dari tampilan grafik terlihat bahwa terdapat kecenderungan pemisahan klaster antara tumor jinak dan ganas. Sampel benign cenderung terkonsentrasi di wilayah dengan nilai PC1 yang lebih rendah, membentuk kluster padat pada sisi kiri grafik. Sebaliknya, sampel malignant tersebar lebih luas dengan nilai PC1 dan PC2 yang lebih tinggi, menunjukkan karakteristik morfologis yang lebih beragam. Pola penyebaran ini mengonfirmasi bahwa PCA mampu mengekstraksi struktur laten dalam data yang berhubungan erat dengan status diagnosis tumor.

2.2 Visualisasi PCA 3D

[23]
✓ 0s

```
pca3 = PCA(n_components=3)
X_train_pca = pca3.fit_transform(X_scaled)

y_numerical = y.map({'M': 1, 'B': 0})

fig = plt.figure(figsize=(8, 6))
ax = fig.add_subplot(111, projection='3d')

scatter = ax.scatter(
    X_train_pca[:, 0],
    X_train_pca[:, 1],
    X_train_pca[:, 2],
    c=y_numerical,
    cmap='coolwarm',
    s=60
)

ax.set_title('Visualisasi PCA (3 Komponen) - Breast Cancer Dataset')
ax.set_xlabel('PC1')
ax.set_ylabel('PC2')
ax.set_zlabel('PC3')

legend1 = ax.legend(
    *scatter.legend_elements(),
    title="Kelas"
)
ax.add_artist(legend1)

plt.show()
```

Pada tahap ini dilakukan penerapan Principal Component Analysis (PCA) dengan tiga komponen utama sebagai upaya mereduksi dimensi dari 30 fitur numerik menjadi representasi yang lebih sederhana namun tetap informatif. Data fitur yang telah distandardisasi kemudian

ditransformasikan menggunakan PCA, menghasilkan tiga komponen utama yang menangkap proporsi variansi terbesar dalam dataset. Variabel target dikonversi menjadi format numerik agar dapat divisualisasikan secara konsisten, di mana kelas *Malignant* direpresentasikan sebagai 1 dan *Benign* sebagai 0. Hasil transformasi divisualisasikan dalam bentuk scatter plot tiga dimensi dengan masing-masing sumbu mewakili komponen utama PC1, PC2, dan PC3. Visualisasi menunjukkan bahwa sampel tumor ganas dan jinak membentuk pola distribusi yang relatif terpisah, mengindikasikan bahwa PCA berhasil mengungkap struktur laten dalam data medis ini. Penggunaan *colormap* "coolwarm" membantu membedakan kedua kelas secara jelas, sementara penambahan legenda memperkuat interpretasi visual.