

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from scipy.stats import mode
from sklearn.metrics import accuracy_score

# Step 1: Read CSV file from device
df = pd.read_csv("Breast Cancer Wisconsin.csv")

# Step 2: Preprocess the dataset
df = df.drop(columns=['id', 'Unnamed: 32'], errors='ignore')

if 'diagnosis' in df.columns:
    diagnosis = df['diagnosis'].map({'M': 1, 'B': 0}) # 1: Malignant, 0: Benign
    df = df.drop(columns=['diagnosis'])
else:
    diagnosis = None

# Step 3: Standardize features
scaler = StandardScaler()
scaled_data = scaler.fit_transform(df)

# Step 4: Apply K-Means Clustering
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(scaled_data)
labels = kmeans.labels_

# Step 5: PCA for Visualization
pca = PCA(n_components=2)
pca_data = pca.fit_transform(scaled_data)

# Prepare data for plotting
vis_df = pd.DataFrame({
    'PCA1': pca_data[:, 0],
    'PCA2': pca_data[:, 1],
    'Cluster': labels
})

if diagnosis is not None:
    vis_df['Actual'] = diagnosis

    # Map cluster numbers to label names based on majority
    cluster_map = {}
    for cluster in [0, 1]:
        majority = mode(diagnosis[labels == cluster], keepdims=True).mode[0]
        cluster_map[cluster] = 'Predicted Malignant' if majority == 1 else 'Predicted Benign'

    vis_df['Cluster_Label'] = vis_df['Cluster'].map(cluster_map)

    # Map actual labels to strings
    vis_df['Actual_Label'] = vis_df['Actual'].map({1: 'Actual Malignant', 0: 'Actual Benign'})

    # Calculate accuracy by comparing predicted labels with actual labels
    # Map the predicted clusters to actual labels (Malignant or Benign)
    predicted_labels = np.array([1 if cluster == 0 else 0 for cluster in labels]) # Map cluster 0 to Malignant, 1 to Benign
    accuracy = accuracy_score(diagnosis, predicted_labels)
    print(f"Accuracy: {accuracy * 100:.2f}%")

# Step 6: Plotting
plt.figure(figsize=(10, 6))

# Plot predicted clusters (colored)
sns.scatterplot(
    data=vis_df,
    x='PCA1',
    y='PCA2',
    hue='Cluster_Label',
    palette={'Predicted Malignant': 'red', 'Predicted Benign': 'blue'},
    s=100,
    alpha=0.6,
    legend='full'
)

# Plot actual values (black shapes)
if diagnosis is not None:
    sns.scatterplot(
        data=vis_df,
        x='PCA1',
        y='PCA2',
        style='Actual_Label',
        markers={'Actual Malignant': 'X', 'Actual Benign': 'o'},
        color='black',
        s=50,
        alpha=0.5,
        legend='brief'
    )

plt.title("K-Means Clustering on Breast Cancer Dataset (PCA Projection)")
plt.xlabel("Principal Component 1")
plt.ylabel("Principal Component 2")
plt.legend(title="Legend", bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
plt.tight_layout()
plt.show()
```

```
/home/ailba/anaconda3/lib/python3.10/site-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to
'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f163b0>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f163b0>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f163b0>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3d870>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3d870>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3d510>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3d7e0>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3d870>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f163b0>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3d870>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3c9d0>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Exception ignored on calling ctypes callback function: <function _ThreadPoolInfo._find_modules_with_dl_iterate_phdr.<locals>.match_module_callback at
0x7f6896f3d870>
Traceback (most recent call last):
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 400, in match_module_callback
    self._make_module_from_path(filepath)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 515, in _make_module_from_path
    module = module_class(filepath, prefix, user_api, internal_api)
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 606, in __init__
    self.version = self.get_version()
  File "/home/ailba/anaconda3/lib/python3.10/site-packages/threadpoolctl.py", line 646, in get_version
    config = get_config().split()
AttributeError: 'NoneType' object has no attribute 'split'
Accuracy: 91.04%
```

