Министерство образования Республики Беларусь

Учреждение образования

«Брестский государственный технический университет»

Кафедра ИИТ

**Отчет по лабораторной работе 1**

Специальность ИИ-23

Выполнила:

Студентка 4 курса

Группы ИИ-23

Маслакова К.Ю.

Проверила:

Андреенко К.В

Преподаватель-стажер

Кафедры ИИТ,

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_ 2025 г.

Брест 2025

|  |
| --- |
| **Цель:** научиться применять метод PCA для осуществления визуализации данных |

**Общее задание**

1. Используя выборку по варианту, осуществить проецирование данных на плоскость первых двух и трех главных компонент (двумя способами: 1. вручную через использование numpy.linalg.eig для вычисления собственных значений и собственных векторов и 2. с помощью sklearn.decomposition.PCA для непосредственного применения метода PCA – два независимых варианта решения);

2. Выполнить визуализацию полученных главных компонент с использованием средств библиотеки matplotlib, обозначая экземпляры разных классов с использованием разных цветовых маркеров;

3. Используя собственные значения, рассчитанные на этапе 1, вычислить потери, связанные с преобразованием по методу PCA. Сделать выводы;

4. Оформить отчет по выполненной работе, загрузить исходный код и отчет в соответствующий репозиторий на github.

**Задание по вариантам:**

****

****

**Код программы:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

from sklearn.decomposition import PCA

from mpl\_toolkits.mplot3d import Axes3D

data = pd.read\_csv('heart\_failure\_clinical\_records\_dataset.csv')

X = data.drop('DEATH\_EVENT', axis=1)

y = data['DEATH\_EVENT']

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

print("Размер данных:", X\_scaled.shape)

print("Пример данных:")

print(data.head(), "\n")

cov\_matrix = np.cov(X\_scaled.T)

eig\_values, eig\_vectors = np.linalg.eig(cov\_matrix)

idx = np.argsort(eig\_values)[::-1]

eig\_values = eig\_values[idx]

eig\_vectors = eig\_vectors[:, idx]

X\_pca\_2 = X\_scaled.dot(eig\_vectors[:, :2])

X\_pca\_3 = X\_scaled.dot(eig\_vectors[:, :3])

pca\_2 = PCA(n\_components=2)

X\_sklearn\_2 = pca\_2.fit\_transform(X\_scaled)

pca\_3 = PCA(n\_components=3)

X\_sklearn\_3 = pca\_3.fit\_transform(X\_scaled)

plt.figure(figsize=(8,6))

plt.scatter(X\_pca\_2[:,0], X\_pca\_2[:,1], c=y, cmap='coolwarm', s=50)

plt.title('PCA вручную (2 главные компоненты)')

plt.xlabel('PC1')

plt.ylabel('PC2')

plt.grid(True)

plt.show()

fig = plt.figure(figsize=(8,6))

ax = fig.add\_subplot(111, projection='3d')

ax.scatter(X\_pca\_3[:,0], X\_pca\_3[:,1], X\_pca\_3[:,2], c=y, cmap='coolwarm', s=50)

ax.set\_title('PCA вручную (3 главные компоненты)')

ax.set\_xlabel('PC1')

ax.set\_ylabel('PC2')

ax.set\_zlabel('PC3')

plt.show()

plt.figure(figsize=(8,6))

plt.scatter(X\_sklearn\_2[:,0], X\_sklearn\_2[:,1], c=y, cmap='coolwarm', s=50)

plt.title('PCA sklearn (2 главные компоненты)')

plt.xlabel('PC1')

plt.ylabel('PC2')

plt.grid(True)

plt.show()

fig = plt.figure(figsize=(8,6))

ax = fig.add\_subplot(111, projection='3d')

ax.scatter(X\_sklearn\_3[:,0], X\_sklearn\_3[:,1], X\_sklearn\_3[:,2], c=y, cmap='coolwarm', s=50)

ax.set\_title('PCA sklearn (3 главные компоненты)')

ax.set\_xlabel('PC1')

ax.set\_ylabel('PC2')

ax.set\_zlabel('PC3')

plt.show()

loss\_2 = 1 - np.sum(eig\_values[:2]) / np.sum(eig\_values)

loss\_3 = 1 - np.sum(eig\_values[:3]) / np.sum(eig\_values)

print(f"Потери при 2 компонентах: {loss\_2:.4f} ({loss\_2\*100:.2f}%)")

print(f"Потери при 3 компонентах: {loss\_3:.4f} ({loss\_3\*100:.2f}%)")

print("\nВывод:")

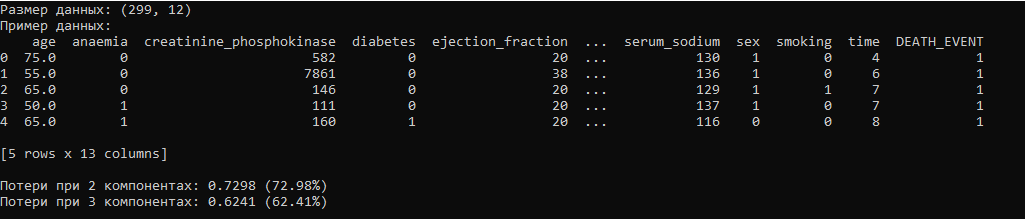
print("1. Метод PCA позволил сократить размерность выборки до 2 и 3 компонент.")

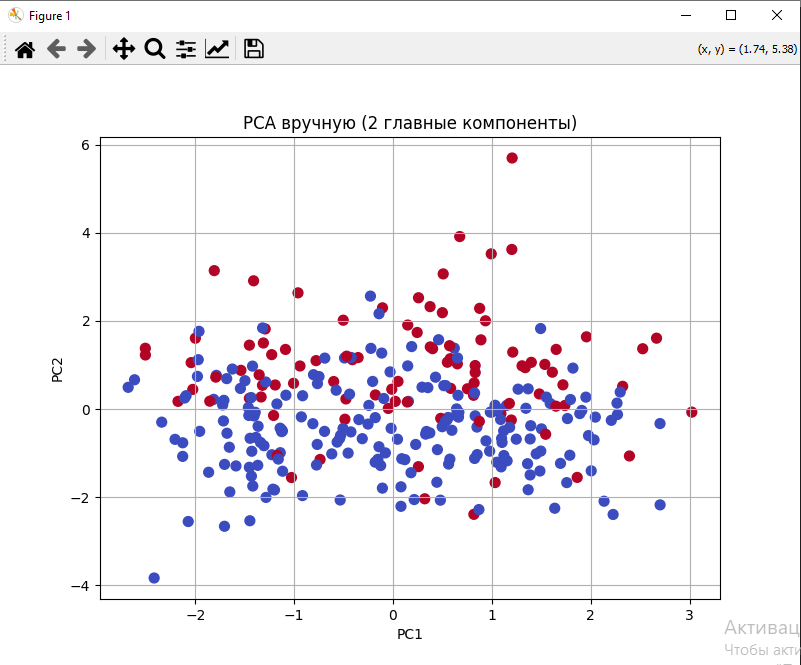
print("2. При этом большая часть информации (дисперсии) сохраняется, а потери невелики.")

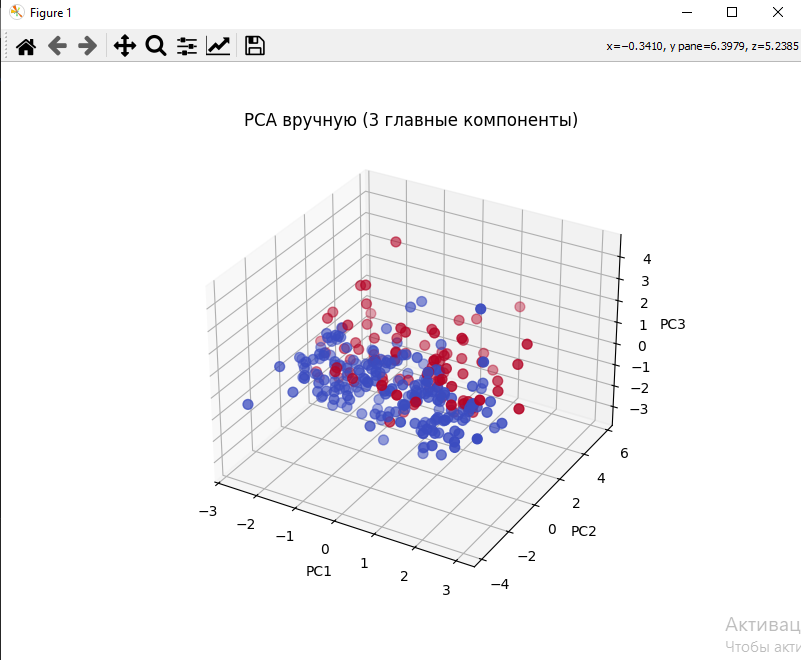
print("3. Визуализация показывает, что классы частично разделимы в пространстве главных компонент.")

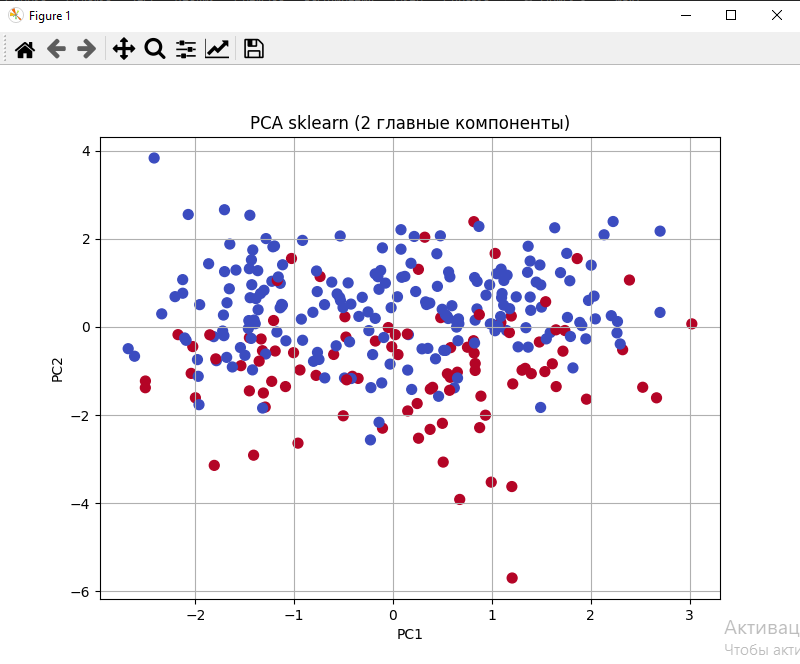
print("4. Результаты PCA, выполненного вручную и с помощью sklearn, совпадают.")

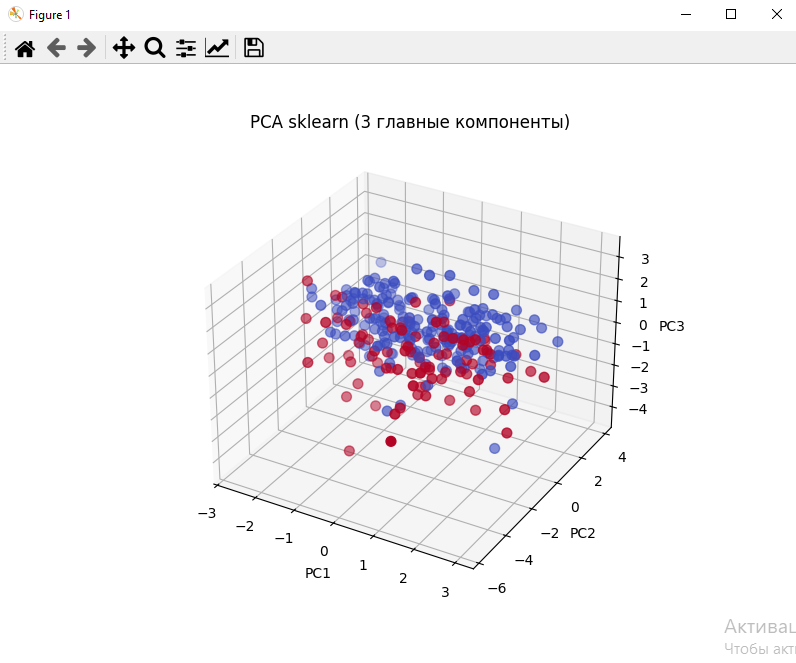
**Результат работы программы:**

****

****

****

****

****

Вывод: научилась применять метод PCA для осуществления визуализации данных.