

Министерство образования Республики Беларусь  
Учреждение образования  
«Брестский государственный технический университет»  
Кафедра интеллектуальных информационных технологий

Отчет по лабораторной работе № 1  
Специальность ИИ-22

**Выполнила:**

С.М. Сокол

Студентка группы

ИИ-22

**Проверил:**

А. А. Крощенко,

доц. кафедры ИИТ

**Цель:** научиться применять метод PCA для осуществления визуализации данных.

**Постановка задачи:**

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 вывод

**Ход работы:**

Вариант 18

Было выполнено программа, соответствующая заданию.

Результаты:

Исходное кол-во компонент:5

Сжатие до ...компонент:2 и 3

Потери, %

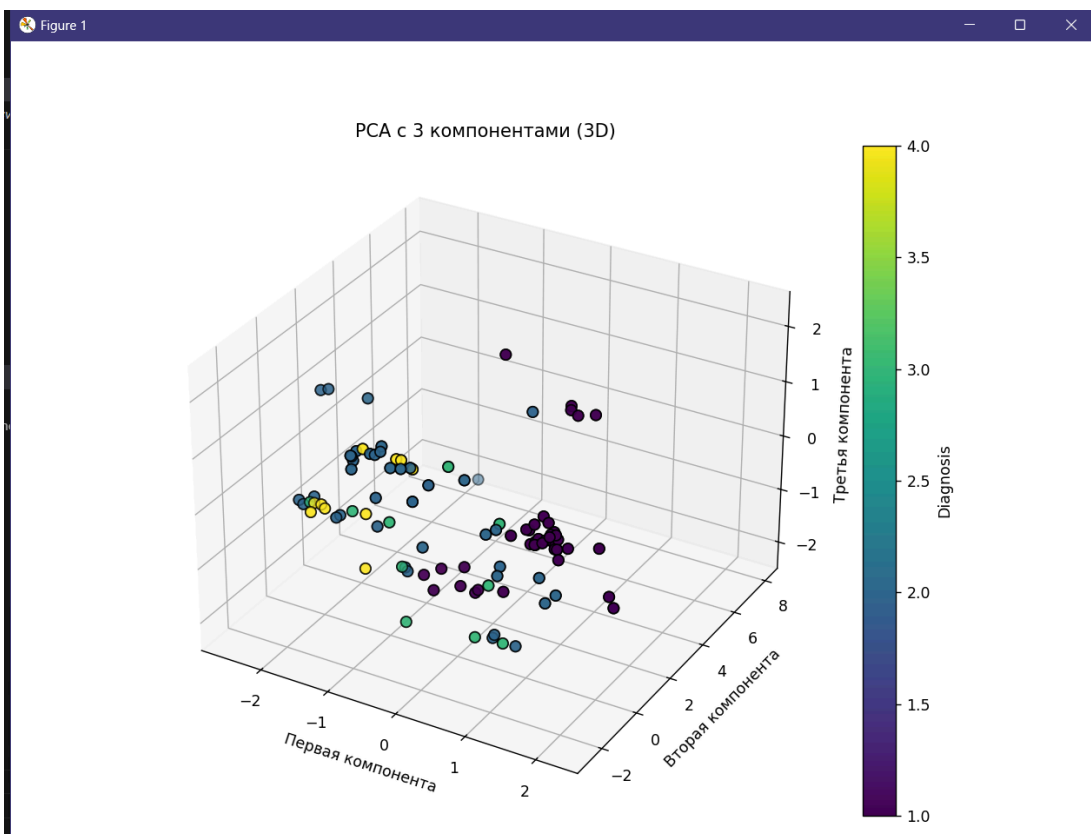
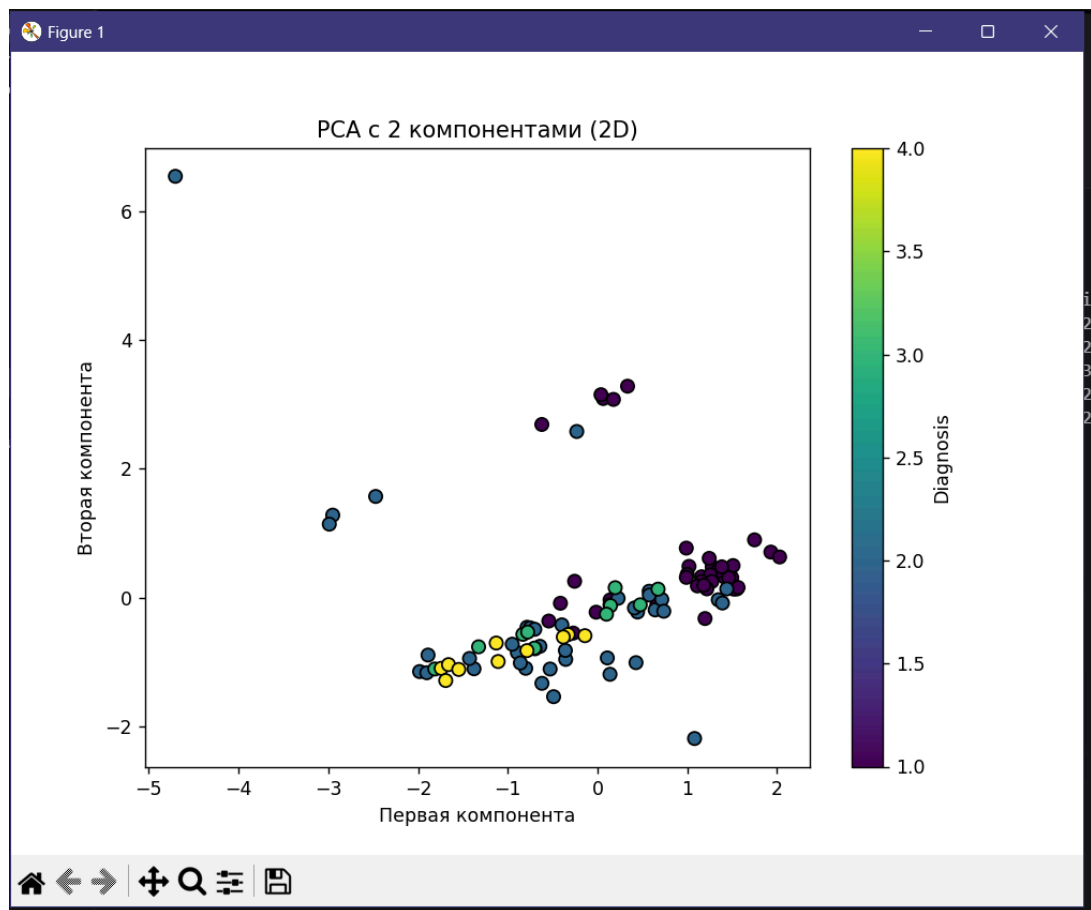
```
Первые 5 строк данных:
Diagnosis  ID  Imaginary Part  Unnamed: 3  Real Part  Unnamed: 5  Gender  Age  Smoking  Unnamed: 9  Unnamed: 10  Unnamed: 11  Unnamed: 12
2  COPD  301-4  -320.61  -300.5635307  -495.26  -464.1719907  1.0  77.0  2.0  NaN  NaN  Male=1  Non-smoker=1
3  COPD  302-3  -325.39  -314.7503595  -473.73  -469.2631404  0.0  72.0  2.0  NaN  NaN  Female=0  Ex-smoker=2
4  COPD  303-3  -323.00  -317.4360556  -476.12  -471.8976667  1.0  73.0  3.0  NaN  NaN  NaN  Active-smoker=3
5  COPD  304-4  -327.78  -317.3996698  -473.73  -468.856388  1.0  76.0  2.0  NaN  NaN  NaN  NaN
6  COPD  305-4  -325.39  -316.1557853  -478.52  -472.8697828  0.0  65.0  2.0  NaN  NaN  NaN  NaN

Запуск PCA с 2 компонентами...
Размеры X_pca: (100, 2)

Результаты для 2 компонент:
Точность: 0.70
Процент ошибки: 30.00%
Запуск PCA с 3 компонентами...
Размеры X_pca: (100, 3)

Результаты для 3 компонент:
Точность: 0.75
Процент ошибки: 25.00%
```

Визуализация:



**Вывод:** научились применять метод PCA для осуществления визуализации данных, потери данных при существенном сжатии являются незначительными.