|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

**Факультет «Информатика и системы управления»**

**Кафедра «Системы обработки информации и управления»**

Отчет по лабораторной работе №6

«Методы обучения без учителя»

2024 г.

**Описание задания**

Цель лабораторной работы: изучение методов кластеризации и снижения размерности.

Краткое описание:

1. Выберите набор данных (датасет) для решения задачи обучения без учителя.
2. Выберите подмножество признаков датасета для анализа. Например, подмножество признаков может не включать целевой признак. Количество признаков в подмножестве должно быть более двух. Полученный датасет назовем датасетом D1.
3. С использованием метода главных компонент проведите для датасета D1 снижение размерности данных до двух. Полученный датасет назовем датасетом D2.
4. С использованием алгоритма t-SNE также проведите для датасета D1 снижение размерности данных до двух. Полученный датасет назовем датасетом D3.
5. Визуализируйте точки датасетов D2 и D3. В каком случае кластеры выделены наиболее явно?
6. Для датасетов D1, D2 и D3 проведите кластеризацию с использованием как минимум трех методов кластеризации, рассмотренных в лекции. Оцените качество кластеризации с использованием метрик, рассмотренных в лекции.
7. Сделайте выводы о том, какой метод кластеризации и почему оказался лучше для каждого из датасетов.

Текст программы и экранные формы с примерами выполнения программы приложены в файле lab2-6.ipynb