**Московский государственный технический   
университет им. Н. Э. Баумана**

Курс «Технологии машинного обучения»

Отчёт по лабораторной работе №9

Выполнил: Проверил:  
Мажитов В. Гапанюк Ю.Е.  
группа ИУ5-62Б

Дата: 24.05.25 Дата:

Подпись: Подпись:

Москва, 2025 г.

**Цель лабораторной работы:** изучение методов кластеризации и снижения размерности.

**Задание:**

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](https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html) проведите для датасета D1 снижение размерности данных до двух. Полученный датасет назовем датасетом D2.
4. С использованием [алгоритма t-SNE](https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html) 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.](https://nbviewer.jupyter.org/github/ugapanyuk/courses_current/blob/main/notebooks/cluster/cluster.ipynb) Оцените качество кластеризации [с использованием метрик, рассмотренных в лекции.](https://nbviewer.jupyter.org/github/ugapanyuk/courses_current/blob/main/notebooks/cluster/cluster.ipynb)
7. Сделайте выводы о том, какой метод кластеризации и почему оказался лучше для каждого из датасетов.

**Ход выполнения:**