

## Домашнее задание по теме «Снижение размерности. Алгоритм PCA»

### Формулировка задания

Выполните подготовку данных для решения задачи снижения размерности. Проведите снижение размерности алгоритмом KernelPCA и ещё одним алгоритмом на выбор: t-SNE или UMAP. Обратите внимание на подбор гиперпараметра `n_components`.

Для снижения размерности использовать библиотеки `scikit-learn` и `umap-learn`.

Результирующий код должен быть читаемым, с единой системой отступов и адекватными названиями переменных.

### Описание плана работы

- 1) Загрузите данные из дополнительных материалов или по ссылке:  
<https://www.kaggle.com/datasets/youssefaboelwafa/clustering-penguins-species>
- 2) Проведите EDA(Exploratory Data Analysis). Отдельно обработайте выбросы и проведите нормализацию данных.
- 3) Примените метод снижения размерности Kernel PCA (перебрав все ядерные функции - ('linear', 'poly', 'rbf', 'sigmoid', 'cosine')).
- 4) Отобразите на графиках результат работы алгоритма с разными ядерными функциями. Сделайте вывод.
- 5) Для линейного ядра выведите дисперсию и почитайте `lost_variance` по примеру с лекции. Сделайте вывод.
- 6) Примените для сравнения один из алгоритмов: t-SNE, UMAP. Отобразите на графиках результат работы алгоритма. Сделайте вывод.

### Перечень необходимых инструментов

- Python
- scikit-learn
- umap-learn
- pandas
- venv

- Jupiter Notebook
- IDE VS Code
- GigaIDE

## Форма предоставления результата

1. В поле ссылки загрузить ссылку на удаленный репозиторий с доступом для наставника.
2. В поле файла загрузить архив с папкой, в которой разместить отчет со скриншотами по заданию и решение задачи. Решение должно быть представлено в формате .ipynb или .py.

## Шкала оценивания

- 1.0 – отлично
- 0.7–0.9 – хорошо
- 0.5–0.6 – удовлетворительно
- Менее 0.5 – задание не выполнено