**Лабораторная работа № 2**

**Тема**: Обработка данных.

Для выполнения данной лабораторной работы можно использовать любой найденный датасет с пропусками, а если пропусков нет, то необходимо их сгенерировать.

Примеры датасетов с пропусками:

<https://www.kaggle.com/datasets/uciml/german-credit>

<https://www.kaggle.com/datasets/parulpandey/2020-it-salary-survey-for-eu-region>

1. Выявите пропуски данных несколькими способами (визуальный, расчетный…)
2. Удалите или замените пропуски. *При удалении (замене) пропусков необходимо рассуждать: можно ли удалить данный параметр и чем целесообразно заменять пропуски данных в конкретных параметрах, руководствуясь описанием параметров датасета и предметной областью.*
3. Постройте гистограмму распределения исходного датасета до и после обработки пропусков. Сделайте выводы как обработка данных повлияла на их распределение.
4. Проверьте датасет на наличие выбросов, удалите найденные аномальные записи.
5. Приведите все параметры к числовому виду.
6. Проверьте сбалансированность данных.
7. Уменьшите размерность датасета применив методы низкой дисперсии и высокой корреляции.
8. Постройте любую модель классификации для датасета со всеми параметрами и для датасета с уменьшенным числом параметров. Сравните точности моделей.