Лабораторная работа №1

Данные и проблемы в них

Постановка задачи:

Необходимо сформировать два набора данных для приложений машинного обучения. Первый датасет должен представлять из себя табличный набор данных для задачи классификации. Второй датасет должен быть отличен от первого, и может представлять из себя набор изображений, корпус документов, другой табличный датасет или датасет из соревнования Kaggle, предназначенный для решения интересующей вас задачи машинного обучения.

Необходимо провести анализ обоих наборов данных,

1) поставить решаемую вами задачу,

2) определить признаки необходимые для решения задачи,

3) в случае необходимости заняться генерацией новых признаков,

4) устранением проблем в данных,

5) визуализировать распределение и зависимость целевого признака от выбранных признаков.

В отчете описать все проблемы, с которыми вы столкнулись и выбранные подходы к их решению