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

**кафедра "Системы обработки информации и управления"**

****

## **["Методы машинного обучения"](https://github.com/ugapanyuk/ml_course_2022/wiki/COURSE_MMO)**

к лабораторной работе №2

## **Обработка признаков (часть 2).**

Инструктор：Юрий Гапанюк

Email:2623859464@qq.com

Студент: Ван Чаочао

группа ИУ5И-22М

2022/04/01

# **Лабораторная работа**

## **Обработка признаков (часть 2).**

****Цель лабораторной работы:**** изучение продвинутых способов предварительной обработки данных для дальнейшего формирования моделей.

### **Требования к отчету:**

Отчет по лабораторной работе должен содержать:

1. титульный лист;
2. описание задания;
3. текст программы;
4. экранные формы с примерами выполнения программы.

### **Задание:**

1. Выбрать один или несколько наборов данных (датасетов) для решения следующих задач. Каждая задача может быть решена на отдельном датасете, или несколько задач могут быть решены на одном датасете. Просьба не использовать датасет, на котором данная задача решалась в лекции.
2. Для выбранного датасета (датасетов) на основе материалов лекций решить следующие задачи:
   1. масштабирование признаков (не менее чем тремя способами);
   2. обработку выбросов для числовых признаков (по одному способу для удаления выбросов и для замены выбросов);
   3. обработку по крайней мере одного нестандартного признака (который не является числовым или категориальным);
   4. отбор признаков:
      1. один метод из группы методов фильтрации (filter methods);
      2. один метод из группы методов обертывания (wrapper methods);
      3. один метод из группы методов вложений (embedded methods).