****Московский государственный технический университет им. Н.Э. Баумана  
Кафедра «Системы обработки информации и управления»

Лабораторная работа №3  
по дисциплине  
«Методы машинного обучения»  
на тему

# «Обработка признаков, часть 2.»

Выполнил:  
студент группы ИУ5-22М  
Лун Сыхань

Москва — 2024г.

**1. Цель лабораторной работы**

Изучение продвинутых способов предварительной обработки данных для дальнейшего формирования моделей.

**2. 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:

один метод из группы методов фильтрации (filter methods);

один метод из группы методов обертывания (wrapper methods);

один метод из группы методов вложений (embedded methods).

**3. Ход выполнения работы** В этой тетради я буду использовать графики для визуализации взаимосвязи между переменными в наборе данных "Титаник".

**The dataset includes the following columns:**  
Survived

Pclass

Name

Sex

Age

Siblings/Spouses Aboard

Parents/Children Aboard

Fare

`文本

描述已自动生成

**масштабирование признаков (не менее чем тремя способами)**

表格

低可信度描述已自动生成

**обработку выбросов для числовых признаков (по одному способу для удаления выбросов и для замены выбросов)**表格

中度可信度描述已自动生成

**обработку по крайней мере одного нестандартного признака (который не является числовым или категориальным)**

表格

中度可信度描述已自动生成

**отбор признаков:**

**один метод из группы методов фильтрации (filter methods);**

**один метод из группы методов обертывания (wrapper methods);**

**один метод из группы методов вложений (embedded methods).**

文本

描述已自动生成