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

Лабораторная работа №1  
по дисциплине  
«Методы машинного обучения»  
на тему

# «Разведочный анализ данных. Исследование и визуализация данных»

Выполнил:  
студент группы ИУ5-22М  
ЧжаоЛян

Москва — 2022 г.

**1. Цель лабораторной работы**

Изучить различные методы визуализации данных [1].

**2. Задание**

Требуется выполнить следующие действия [1]:

• Выбрать набор данных (датасет).

• Создать ноутбук, который содержит следующие разделы:

1. Текстовое описание выбранного набора данных.

2. Основные характеристики датасета.

3. Визуальное исследование датасета.

4. Информация о корреляции признаков.

• Сформировать отчет и разместить его в своем репозитории на GitHub

**3. Ход выполнения работы  
3.1. Текстовое описание набора данных**

В этой тетради я буду использовать графики для визуализации взаимосвязи между переменными в наборе данных "Успеваемость студентов на экзаменах".

* **The dataset includes the following columns:**  
  **Gender** (object) - The gender of the participant
  + female
  + male
* **Race/Ethnicity** (object) - The race or ethnicity of the participant
  + Group A
  + Group B
  + Group C
  + Group D
  + Group E
* **Parental level of education** (object) - The parental level of education of the participant
  + bachelor's degree
  + some college
  + master's degree
  + associate's degree
  + high school
  + some high school
* **Lunch** (object) - Whether the participant is:
  + standard
  + free/ reduced
* **Test preparation course** (object) - Whether the participant took the test preparation course or not
  + none
  + completed
* **Math score** (int64) - The participants math score
* **Reading score** (int64) - The participants reading score
* **Writing score** (int64) - The participants writing score

**3.2. Основные характеристики набора данных**  
文本

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

图表, 折线图

描述已自动生成

图表, 直方图

描述已自动生成

图表

描述已自动生成

图形用户界面

描述已自动生成

图表, 散点图

描述已自动生成图表, 日程表

描述已自动生成图表, 散点图

描述已自动生成

**Список литературы**

[1] Гапанюк Ю. Е. Лабораторная работа «Разведочный анализ данных. Исследование  
и визуализация данных» [Электронный ресурс] // GitHub. –– 2019. –– Режим доступа: https://github.com/ugapanyuk/ml\_course/wiki/LAB\_EDA\_VISUALIZATION (дата  
обращения: 13.02.2019)

[2] <https://www.kaggle.com/datasets>