**Министерство образования и науки Российской Федерации**

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Саратовский государственный технический университет

имени Гагарина Ю.А.»

**Кафедра «Информационно-коммуникационные системы и программная инженерия» (ИКСП)**

**Практическая работа №4**

**Выполнил:**

Студент гр. М-ИВЧТ-11

Кудряшов Алексей Владимирович

Саратов – 2023

**1 задание**

Предсказать рейтинг энергопотребления здания и понять, построить модель, которая иллюстрирует, какие переменные больше всего влияют на целевую переменную.

Задачи:

1. Понять сущность имеющихся данных
   1. К какому типу данных относятся переменные в файле?
   2. Какие проблемы в данных Вы обнаружили?
   3. Какие гипотезы можно выдвинуть относительно изучаемого набора данных?
   4. Какая переменная является целевой?
   5. Какие переменные являются независимыми?
   6. К какому классу задач машинного обучения относится данная задача? (обучение с учителем / обучение без учителя)?
   7. К какому типу будет относиться будущая модель? (классификация, кластеризация, регрессия, ассоциация и т.п.)

a. Целочисленные, вещественные, строковые.

b. Отсутствуют данные в строках, разные способы заполнения данных (не по шаблону).

c. Энергопотребление зданий в основном зависит от площади здания

d. Source EUI (kBtu/ft²)

e. DOF Gross Floor Area

f. Обучение без учителя

g. Регрессия

Распределение целевой переменной: чем больше количество зданий, тем выше оценка.

**DOF Gross Floor Area, Largest Property Use Type - Gross Floor Area (ft²), 2nd Largest Property Use - Gross Floor Area (ft²), 3rd Largest Property Use Type - Gross Floor Area (ft²), Site EUI (kBtu/ft²),**