

## Лабораторная работа № 5

### Тема: Линейная регрессия.

**Цель работы** – изучить на практике понятие корреляции, реализовать модели линейной регрессии из библиотеки *sklearn*.

1. Выберите любой доступный на просторах интернета набор данных подходящий для регрессионного анализа (как правило это прогнозирование цены, зарплаты и т.д.). Как и в прошлом задании это можно сделать, введя запрос на сайте kaggle.

**Требование к набору данных: минимум три признака, описывающих объект (можно больше).**

2. Визуализируйте матрицу корреляций в виде тепловой карты.
3. Постройте матрицу диаграмм рассеяния.
4. Проанализируйте коэффициенты корреляции и диаграммы рассеяния и выберите параметры для построения модели простой линейной регрессии.
5. Рассчитайте модель и визуализируйте график модели на диаграмме рассеяния параметра (который учтен в модели)
6. Оцените качество полученной модели.
7. Добавьте в модель линейной регрессии еще несколько параметров, в той или иной мере влияющих на целевую переменную.
8. Рассчитайте модель с несколькими параметрами и оцените качество модели.
9. Сделайте вывод: какая модель лучше описывает зависимость целевой переменной от параметров.

### Вопросы:

1. Что такое корреляция?
2. Что показывает коэффициент корреляции? Каким он может быть?
3. Что представляет из себя модель простой линейной регрессии?
4. Чем множественная линейная регрессия отличается от простой линейной регрессии?
5. Что характеризует коэффициент детерминации?
6. Как сложность модели влияет на коэффициент детерминации?