**Задание по машинному обучению (контрольная точка 5)**

Выполнить *любое из двух заданий*: либо по регрессии, либо по классификации.

**Задание по регрессии**:

Построить модель множественной линейной регрессии цены дома от его параметров. Для этого:

1) Импортировать файл “houses.csv”.

2) Отобрать показатели, которые, по вашему мнению, должны влиять на цену.

3) Провести попарный корреляционный анализ столбцов, для чего построить матрицу корреляции. На основании нее выбрать объясняющие переменные для регрессии. Если признак влияет на цену совсем слабо (корреляция менее 0,5), его можно отбросить. Если 2 признака имеют коэффициент корреляции между собой 0,7 и выше, они коллинеарны (сильно линейно зависят друг от друга), поэтому от одного можно избавиться (того, корреляция которого с целевой переменной ниже, т.к. он слабее на нее влияет).

В результате должно остаться несколько объясняющих переменных, слабо зависящих друг от друга и достаточно сильно влияющих на зависимую переменную (цену).

4) Построить несколько моделей множественной линейной регрессии с различными объясняющими переменными. В качестве аргумента X метода fit() будет не одномерный массив, а датафрейм из нескольких столбцов (объясняющих переменных).

5) Выбрать наилучшую модель по метрикам RMSE и R2, вычисленным на тестовой выборке.

Качество моделей оценивать при кросс-валидации. Выборку разбивать на *k* блоков (*k* вводится с клавиатуры).

6) Дать комментарии относительно полученных результатов (в коде программы).

**Задание по классификации**:

Построить модель бинарной классификации, предсказывающую доход человека (<=50 тыс. или >50 тыс.). Использовать датасет "adult.csv". Выполнить предварительную обработку данных. Использовать все рассмотренные алгоритмы классификации, сравнить их по метрикам, выбрать наилучший. Качество моделей оценивать при кросс-валидации. Дать комментарии относительно полученных результатов (в коде программы).