

# Кросс-валидация

## Практическое задание для самостоятельного выполнения

### Задание 1.

1. Сформировать текстовый файл, содержащий список группы с указанием для каждого студента имеется/не имеется задолженность за последнюю сессию (например, в формате 0/1). В списке должно быть не менее 15-20 записей. Импортировать эти данные в объект *Pandas DataFrame*.
2. С полученным набором данных реализовать различные стратегии кросс-валидации и проанализировать полученные результаты.
  - 2.1. Применить стратегию **KFold** с разбиением на 5 блоков без перемешивания. Вывести списки индексов объектов, попавших в обучающую и тестовую выборку на каждой итерации, и сформировать собственно обучающие и тестовые выборки (данные и метки классов).
  - 2.2. Применить стратегию **KFold** с разбиением на 5 блоков с перемешиванием. Вывести списки индексов объектов, попавших в обучающую и тестовую выборку на каждой итерации, и сформировать собственно обучающие и тестовые выборки (данные и метки классов). Сравнить с результатами п. 2.1, дать объяснения.
  - 2.3. Применить стратегию **StratifiedKFold** с разбиением на 5 блоков. Вывести списки индексов объектов, попавших в обучающую и тестовую выборку на каждой итерации, и сформировать собственно обучающие и тестовые выборки (данные и метки классов). Сравнить с результатами п. 2.1 и 2.2, дать объяснения.
  - 2.4. Применить стратегию **ShuffleSplit** с разбиением на 10 блоков. Вывести списки индексов объектов, попавших в обучающую и тестовую выборку на каждой итерации, и сформировать собственно обучающие и тестовые выборки (данные и метки классов). Сравнить с предыдущими результатами, дать объяснения.
  - 2.5. Применить стратегию **StratifiedShuffleSplit** с разбиением на 10 блоков. Вывести списки индексов объектов, попавших в обучающую и тестовую выборку на каждой итерации, и сформировать собственно обучающие и тестовые выборки (данные и метки классов). Сравнить с результатами п. 2.4, дать объяснения.
  - 2.6. Применить стратегию **LeaveOneOut**. Вывести списки индексов объектов, попавших в обучающую и тестовую выборку на каждой итерации, и сформировать собственно обучающие и тестовые выборки (данные и метки классов).

### Задание 2.

1. Выполнить оценку качества моделей линейной классификации, созданных при выполнении задания 2 по бинарной линейной классификации (в файле 5\_2\_Бинарная классификация), с помощью кросс-валидации (использовать метрику **accuracy**, количество блоков, равное 10). Для каждой модели вывести

массив показателей качества, итоговую оценку (среднее значение) и среднеквадратичное отклонение. Выполнить анализ полученных результатов.

2. Создать отчет: оценка качества каждой модели, выводы.