

Практическая работа №3

Задание

1. Загрузить файл 'telecom_churn.csv'.
2. Проверить данные на наличие пропусков и выбросов. Обработать пропуски и выбросы.
3. Выполнить кодирование категориальных переменных.
4. Выполнить масштабирование признаков.
5. Построить модели k-ближайших соседей, логистической регрессии, SVM
 - a. Обучить модель по кросс-валидации и выведите на экран значения метрик.
 - b. Подобрать значение параметров модели по кросс-валидации, используя GridSearchCV.
 - c. Добавить в модели все квадратичные признаки, используя функцию PolynomialFeatures. Сравнить качество моделей. Построить матрицу корреляций признаков. Если есть признаки, которые практически не коррелируют с целевой переменной, можно попробовать их удалить. Если есть пары сильно коррелирующих между собой признаков, удалить по одному признаку из каждой пары. Оценить качество моделей.
6. Сравнить полученные результаты. Выделить наилучшую модель. Построить ROC-кривую и вычислить ROC-AUC на тренировочной и тестовой выборках.

Дополнительное задание*

1. Провести отбор информативных признаков с помощью разных подходов (встроенные методы, методы фильтрации, методы-обертки). Обучить модель на информативных признаках. Сравнить полученные результаты.
2. Применить технику сэмплирования к исходным данным (over-sampling, under-sampling, ансамблевые методы). Обучить модель на данных, обработанных разными методами.
3. Сравнить полученные результаты с результатами из основного задания.