

Домашнее задание
по теме «Введение в машинное обучение.
KNN. Метрики качества. Матрица ошибок.
Виды данных (числовые / категориальные)»

Формулировка задания

Выполните подготовку данных для решения задачи классификации. Проведите классификацию методом kNN. Обратите внимание на подбор параметра k и на оценку качества алгоритма. Качество оценить минимум по 3 критериям качества для классификации: `confusion_matrix`, `accuracy`, `precision`, `recall`, `f1_score`, `roc_auc`.

Для классификации и оценки качества использовать библиотеку `scikit-learn`.

Результирующий код должен быть читаемым, с единой системой отступов и адекватными названиями переменных.

Описание плана работы

- 1) Загрузите данные из дополнительных материалов или по ссылке: <https://www.kaggle.com/competitions/playground-series-s3e12/data>
- 2) Проведите EDA(Exploratory Data Analysis), то есть для каждой переменной посчитайте:
 - a) Долю пропусков
 - b) Максимальное и минимальное значение
 - c) Среднее значение
 - d) Медиану
 - e) Дисперсию
 - f) Квантиль 0.1 и 0.9
 - g) Квартиль 1 и 3
- 3) Столбец “target” принимает значения 0 (нет камней в почках) или 1(есть камни в почках) и является целевой переменной. Для целевой переменной target посчитайте долю 1.
- 4) Постройте минимум 2 информативных графика, которые будут полезны для анализа
- 5) Обучите алгоритм KNN (k-ближайших соседей), посчитайте качество классификации и напишите ответы на следующие вопросы:
 - a) Чему равен гиперпараметр k? Почему вы выбрали именно такой k?

b) Насколько ваш алгоритм верно предсказывает целевую переменную?

Перечень необходимых инструментов

- Python
- scikit-learn
- pandas
- venv
- Jupiter Notebook
- IDE VS Code
- GigaIDE

Форма предоставления результата

1. В поле ссылки загрузить ссылку на удаленный репозиторий с доступом для наставника.
2. В поле файла загрузить архив с папкой, в которой разместить отчет со скриншотами по заданию и решение задачи. Решение должно быть представлено в формате .ipynb или .py.

Шкала оценивания

- 1.0 – отлично
- 0.7–0.9 – хорошо
- 0.5–0.6 – удовлетворительно
- Менее 0.5 – задание не выполнено