

Домашнее задание по теме «Регрессия Decision Tree»

Формулировка задания

Выполните подготовку данных для решения задачи регрессии. Проведите регрессию методом решающих деревьев. Качество оценить минимум по 3 критериям качества для регрессии: MAE, MSE, RMSE, MAPE, RMSLE, R^2 и др.

Для регрессии 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 как в задаче “Регрессия SVM” из дополнительных материалов или по ссылке: <https://www.kaggle.com/competitions/regression-with-an-insurance-dataset-clone/data>

Если на наборе данных задача решается долго, то провести исследование на части данных. Использовать случайное сэмплирование или другие методы.

2) EDA (Exploratory Data Analysis) используем из домашнего задания “Полиномиальная регрессия”.

3) Подготовка датасета к построению моделей ML использовать из задания “Полиномиальная регрессия”. Обратить внимание на то, что для регрессии деревом решений данные используем без нормализации.

4) Обучить модель регрессии дерева решений.

5) Оценить качество алгоритма. Сравнить с прошлыми решениями. Получилась ли модель лучшего качества?

6) Подобрать гиперпараметры DecisionTreeRegressor через GridSearchCV или другой метод подбора гиперпараметров. Какие гиперпараметры будут наиболее подходящими? Как изменилось качество модели?

Перечень необходимых инструментов

- Python

- scikit-learn
- pandas
- matplotlib
- seaborn
- 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