

Домашнее задание по теме «Логистическая регрессия»

Формулировка задания

Выполните подготовку данных для решения задачи классификации. Проведите классификацию методом логистической регрессии. Обратите внимание на подбор параметров solver, penalty, C. 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/datasets/gauravduttakiit/smoker-status-prediction-using-biosignals?select=train_dataset.csv из прошлых заданий классификации. Например, “Классификация SVM”.
- 2) EDA(Exploratory Data Analysis) и подготовку данных использовать из задания “Классификация SVM”. Рекомендуется нормализовать данные.
- 3) Обучите алгоритм LogisticRegression из библиотеки scikit-learn. Посчитайте качество классификации и напишите ответы на следующие вопросы:
 - a) Какие значения гиперпараметров алгоритма подойдут для задачи?
 - b) Насколько ваш алгоритм верно предсказывает целевую переменную?
 - c) Какие критерии качества классификации получились для задачи?

Перечень необходимых инструментов

- Python
- scikit-learn
- pyspark
- 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