

# Домашнее задание по теме «AutoML алгоритмы для работы с данными»

## Формулировка задания

Провести классификацию и задачу прогнозирования временных рядов с применением библиотеки AutoML. Обратите внимание на выбор лучшей модели и критерии качества модели.

Варианты библиотек/фреймворков autoML для выбора:

- H2O AutoML
- AutoSklearn
- Pycaret
- flaml AutoML
- LightAutoML
- FEDOT
- AutoGluon
- LAMA

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

## Описание плана работы

### Задача 1. Классификация

1) Загрузите данные из дополнительных материалов или по ссылке:

[https://www.kaggle.com/datasets/gauravduttakiit/smoker-status-prediction-using-biosignals?select=train\\_dataset.csv](https://www.kaggle.com/datasets/gauravduttakiit/smoker-status-prediction-using-biosignals?select=train_dataset.csv) из заданий “Классификация SVM” и “Классификация Decision Tree”.

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

2) Выбрать один из инструментов AutoML и применить к данным (H2O AutoML, AutoSklearn, Pycaret, flaml AutoML, LightAutoML, FEDOT, AutoGluon, LAMA).

3) Выполнить подготовку и настройку инструмента для табличных данных:

- Подготовить данные
- Настроить AutoML. Выбрать встроенные модели для задачи классификации.
- Построить сравнительную таблицу с критериями качества по моделям.
- Построить визуализацию по проведенному исследованию
- Выгрузить лучшие модели машинного обучения

Сравнить работу AutoML с результатом машинного обучения без AutoML. Сделать вывод.

## **Задача 2. Прогнозирование временных рядов**

1) Загрузите один из датасетов из дополнительных материалов:

- Продажи шампуня по месяцам sales-of-shampoo.csv
- Продажи пивной продукции monthly-beer-production-in-austr.csv
- Производство молочной продукции milk-production.csv
- Потребление электричества electric-production.csv
- Минимальная температура daily-minimum-temperatures.csv
- Количество рабочих мест в Австралии aus-employment.csv

2) Выбрать один из инструментов AutoML и применить к данным (H2O AutoML, AutoSklearn, Pycaret, flaml AutoML, LightAutoML, FEDOT, AutoGluon, LAMA). Убедитесь, что в выбранном инструменте есть модели работы с временными рядами.

3) Выполнить подготовку и настройку инструмента для табличных данных:

- Подготовить данные
- Настроить AutoML. Выбрать встроенные модели для задачи прогнозирования.
- Отобразить распечатку результата по основным статистическим показателям по данным
- Построить сравнительную таблицу с критериями качества по моделям.
- Построить визуализацию по проведенному исследованию
- Выгрузить лучшие модели машинного обучения

Сравнить работу AutoML с результатом машинного обучения без AutoML. Сделать вывод.

## **Перечень необходимых инструментов**

- Python
- scikit-learn
- pandas
- H2O AutoML
- AutoSklearn
- Pycaret
- flaml AutoML
- LightAutoML
- FEDOT
- AutoGluon
- LAMA
- 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