

Лабораторная работа 4. ОБУЧЕНИЕ С УЧИТЕЛЕМ. РЕШАЮЩИЕ ДЕРЕВЬЯ. АНСАМБЛИ МОДЕЛЕЙ

1. Изучение примеров. Изучите примеры: Lab4_Ex1
Base_DTs&ensembles.ipynb, Lab4_Ex2_DRs_classifier_Gini+Entropy.ipynb, Lab4_Ex3
DRs_classifier+PyCaret.ipynb, Lab4_Ex4_DTs_regression+PyCaret.ipynb, Lab4_Ex5
DRs_classifier DecRules.ipynb, Lab4_Ex6_SketchBoost_experiments.ipynb

2. Загрузка и подготовка данных.

- В соответствии с индивидуальным вариантом загрузите предобработанные датасеты в формате CSV для решения задачи регрессии и классификации.
- При разделении данных на обучающую и тестовую выборки используйте два метода с соответствующими параметрами: *hold-out* и *k-fold*.

3. Решение задачи регрессии и классификации.

3.1. Решите задачу регрессии и классификации на ваших данных с использованием деревьев решений: DecisionTreeRegressor и DecisionTreeClassifier из Scikit-learn.

3.2. Решите задачу регрессии и классификации на ваших данных с использованием ансамблей из Scikit-learn:

- BaggingRegressor
- BaggingClassifier
- GradientBoostingRegressor
- GradientBoostingClassifier
- StackingRegressor
- StackingClassifier

3.3. Решите задачу регрессии и классификации на ваших данных с использованием библиотек градиентного бустинга над деревьями решений:

- CatBoost
- XGBoost
- LightGBM
- *SketchBoost (**!!!Если удастся подключить в вашей среде окружения!**)

Примечание. При необходимости подобрать гиперпараметры регрессора/классификатора тремя способами.

4. Визуализация. Визуализировать дерево решений, вывести решающие правила в виде текста.

5. Оценка качества моделей. Выполнить оценку качества моделей ML. Сравнить полученные результаты обучения ваших моделей ML с результатами, полученными с помощью PyCaret¹ (см. DRs_classifier+PyCaret.ipynb, Lab4_Ex4_DTs_regression+PyCaret.ipynb). Обратите внимание на сгенерированные таблицы (и их содержание) как результаты работы алгоритмов ML посредством PyCaret.

6. Создание таблицы результатов.

- Создайте две таблицы, выведите в них наименования используемых регрессоров и классификаторов, наименования и значения вычисленных метрик оценки качества (Y1 и т.д.) как с использованием Scikit-learn, так и специализированных библиотек градиентного бустинга над деревьями решений (Образец 1 и Образец 2).

Образец 1

Алгоритм ML	Train Data				Test Data			
	Y1	Y2	Y3	Y4	Y1	Y2	Y3	Y4
...	0.XX	0.XXXX	0.XXXX	0.XXXX	0.XX	0.XXXX	0.XXXX	0.XXXX
...								

Образец 2

Алгоритм ML	Метод hold-out (отложенная выборка)				Метод k-fold (кросс-валидация)			
	Y1	Y2	Y3	Y4	Y1	Y2	Y3	Y4
...	0.XX	0.XXXX	0.XXXX	0.XXXX	0.XX	0.XXXX	0.XXXX	0.XXXX
...								

7. Вывод. Напишите вывод о выполненной **Лабораторной работе №4**, в котором выберите лучшую модель регрессии и классификации. Обоснуйте свое решение.

¹PyCaret — это библиотека ML (low-code Machine Learning library) с открытым исходным кодом и малым объемом кода на языке Python, которая автоматизирует рабочие процессы ML. – URL: <https://pycaret.org/>