

ЛАБОРАТОРНАЯ РАБОТА #3

Подбор гиперпараметров модели

Ивченко Матвей Сергеевич
М80-307Б-23

Задание

- Выбрать модель для обучения (Decision tree, Random forest, SVM, KNN, Boosting)
- Показать какие гиперпараметры есть у выбранной модели (В виде таблички).
- Выбрать датасет для обучения и в зависимости от модели подготовить данные
- Подобрать гиперпараметры для модели и сравнить лучшие подборы, для (Grid Search, RandomSearch, Optuna)
- На самом лучшем обучении (Grid Search, RandomSearch, Optuna) сделать калькулятор, который показывает *локальную интерпретацию* с помощью LIME и *глобальную интерпретацию* с помощью SHAP

Ссылки:

1. <https://habr.com/ru/companies/otus/articles/779430/>
2. <https://habr.com/ru/companies/ods/articles/599573/>
3. <https://habr.com/ru/companies/wunderfund/articles/739744/>

Датасет

- Датасет Student Performance Factors
- Предсказываем оценку студента за экзамен
- Содержит категориальные и вещественные признаки
- Все вещественные признаки распределены нормально или равномерно

Модель

■ CatBoostRegressor:

- Умеет работать с категориальными признаками «из коробки»
- Имеет высокое качество среди конкурентов при умелой настройке

Таблица всех гиперпараметров CatBoost:

Гиперпараметр	Описание	Возможные значения
<code>iterations</code>	Количество деревьев в модели (число итераций бустинга)	Положительное целое число (например, 100, 500, 1000)
<code>learning_rate</code>	Скорость обучения; влияет на вклад каждого дерева	Положительное число, обычно от 0.01 до 0.3
<code>depth</code>	Глубина деревьев решений	Целое число от 1 до 16 (обычно 6–10)
<code>l2_leaf_reg</code>	Коэффициент L2-регуляризации для листьев	Положительное число (например, 1, 3, 5)
<code>random_strength</code>	Сила случайности при выборе разбиений	Положительное число (например, 0.1, 1.0)
<code>bagging_temperature</code>	Параметр, управляющий стратегией бэггинга (для стохастического бустинга)	Неотрицательное число; 0 — стандартный бэггинг, выше — более случайный
<code>loss_function</code>	Функция потерь для оптимизации	'RMSE', 'MAE', 'Quantile', 'LogLinQuantile' и др.
<code>eval_metric</code>	Метрика для оценки качества	Например, 'RMSE', 'MAE', 'R2'
<code>grow_policy</code>	Стратегия роста деревьев	'SymmetricTree', 'Depthwise', 'Lossguide'
<code>min_data_in_leaf</code>	Минимальное количество объектов в листе (для <code>Lossguide</code>)	Положительное целое число

Подбор гиперпараметров

RMSE: 1.9016
MAE: 0.5948
 R^2 : 0.7442

До подбора

RMSE: 1.9241
MAE: 0.6121
 R^2 : 0.7533

После подбора Optuna

Тест RMSE: 1.9557
Тест MAE: 0.7011
Тест R^2 : 0.7452

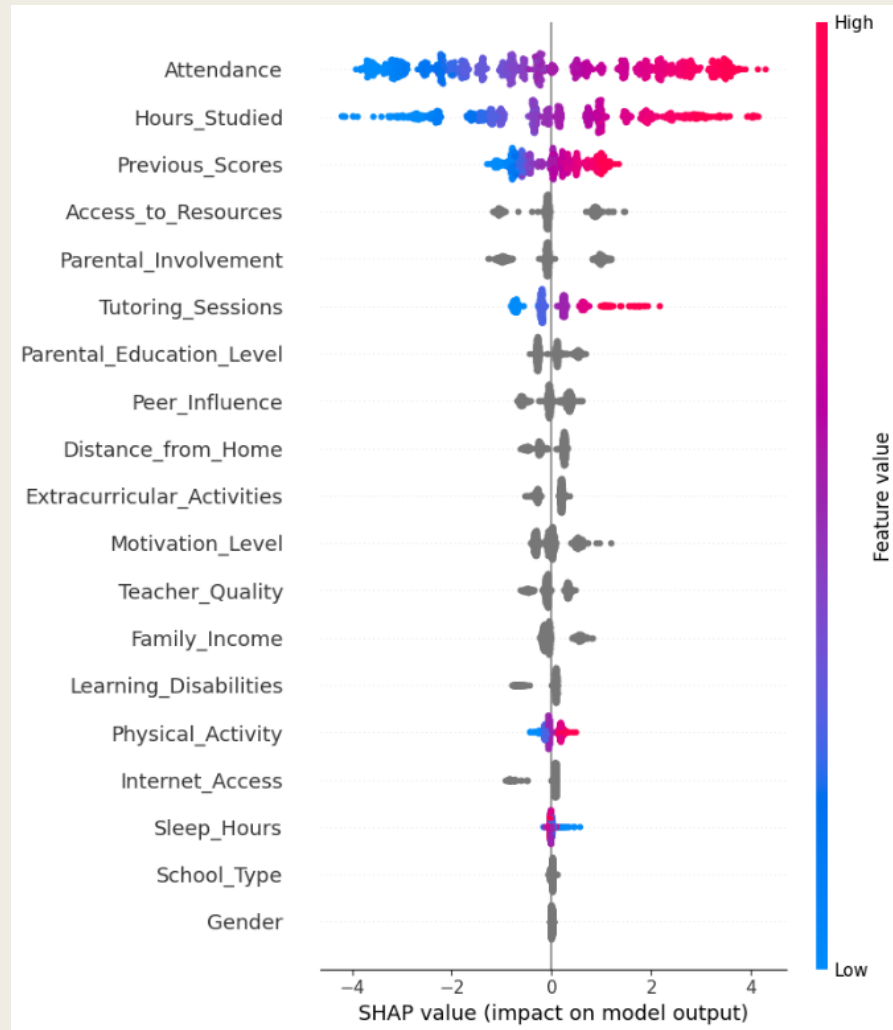
После подбора GridSearchCV

Тест RMSE: 1.9378
Тест MAE: 0.6168
Тест R^2 : 0.7498

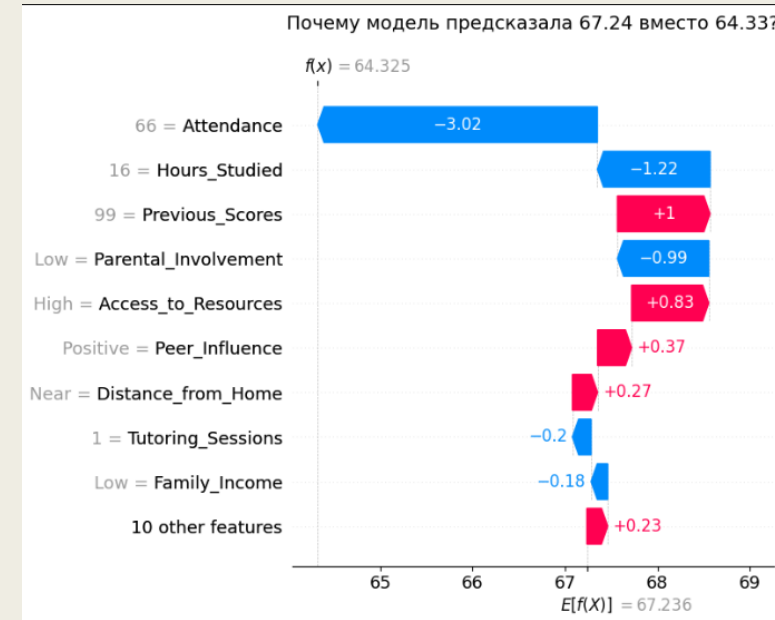
После подбора RandomSearch

Интерпретация

Глобальная (SHAP)



Локальная (SHAP)



Локальная (LIME)

