

# Лабораторная работа №3

Тема: «Подбор гиперпараметров модели»

Таланкин Кирилл Александрович  
М8О-309Б-23

# Задачи

- ▶ Выбрать модель для обучения
- ▶ Построить модель, предсказывающую Sales\_Volume (объем продаж)
- ▶ Подобрать гиперпараметры с помощью трёх методов:
  - ▶ GridSearch
  - ▶ RandomizedSearch
  - ▶ Optuna
- ▶ Сделать локальную и глобальную интерпретации модели

# Описание датасета

- ▶ Датасет: BMW worldwide sales records 2010-2024. Он содержит записи о продажах автомобилей BMW по всему миру с 2010 по 2024 год, с фокусом на модели, регионы, характеристики и объемы продаж
- ▶ Источник: BMW sales data (2010-2024) (1).csv  
Размер: 50000 строк, 11 столбцов
- ▶ Ключевые поля:
  - Model
  - Year
  - Region
  - Color
  - Fuel\_Type
  - Transmission
  - Engine\_Size\_L
  - Mileage\_KM
  - Price\_USD
  - Sales\_Volume
  - Sales\_Classification

# Признаки и гиперпараметры

- ▶ Категории: Model, Region, Color, Fuel\_type, Transmission → OneHotEncoder
- ▶ Числовые: Year, Engine\_Size\_L, Mileage\_KM, Price\_USD
- ▶ Pipeline: ColumnTransformer (OHE) → RandomForestRegressor
- ▶ Гиперпараметры RandomForestRegressor (перечень). Подбор: n\_estimators – количество деревьев, max\_depth – глубина, min\_samples\_split, min\_samples\_leaf, max\_features

# Методы поиска

- ▶ GridSearchCV - перебирает все возможные комбинации указанных гиперпараметров
- ▶ RandomizedSearchCV - выбираем случайные комбинации гиперпараметров из заранее определённых диапазонов
- ▶ Optuna - строим вероятностную модель функции потерь и используем её, чтобы выбирать новые более подходящие гиперпараметры

# Результаты

## ► GridSearchCV

```
{'model__max_depth': 5, 'model__min_samples_split': 5, 'model__n_estimators': 300}
```

MSE GridSearch: 8.174112e+06

## ► RandomizedSearchCV

```
{'model__n_estimators': 400, 'model__min_samples_split': 10, 'model__max_depth': 5}
```

MSE RandomSearch: 8.174370e+06

## ► Optuna

```
{'n_estimators': 146, 'max_depth': 5, 'min_samples_split': 6}
```

MSE Optuna: 8.173015e+06

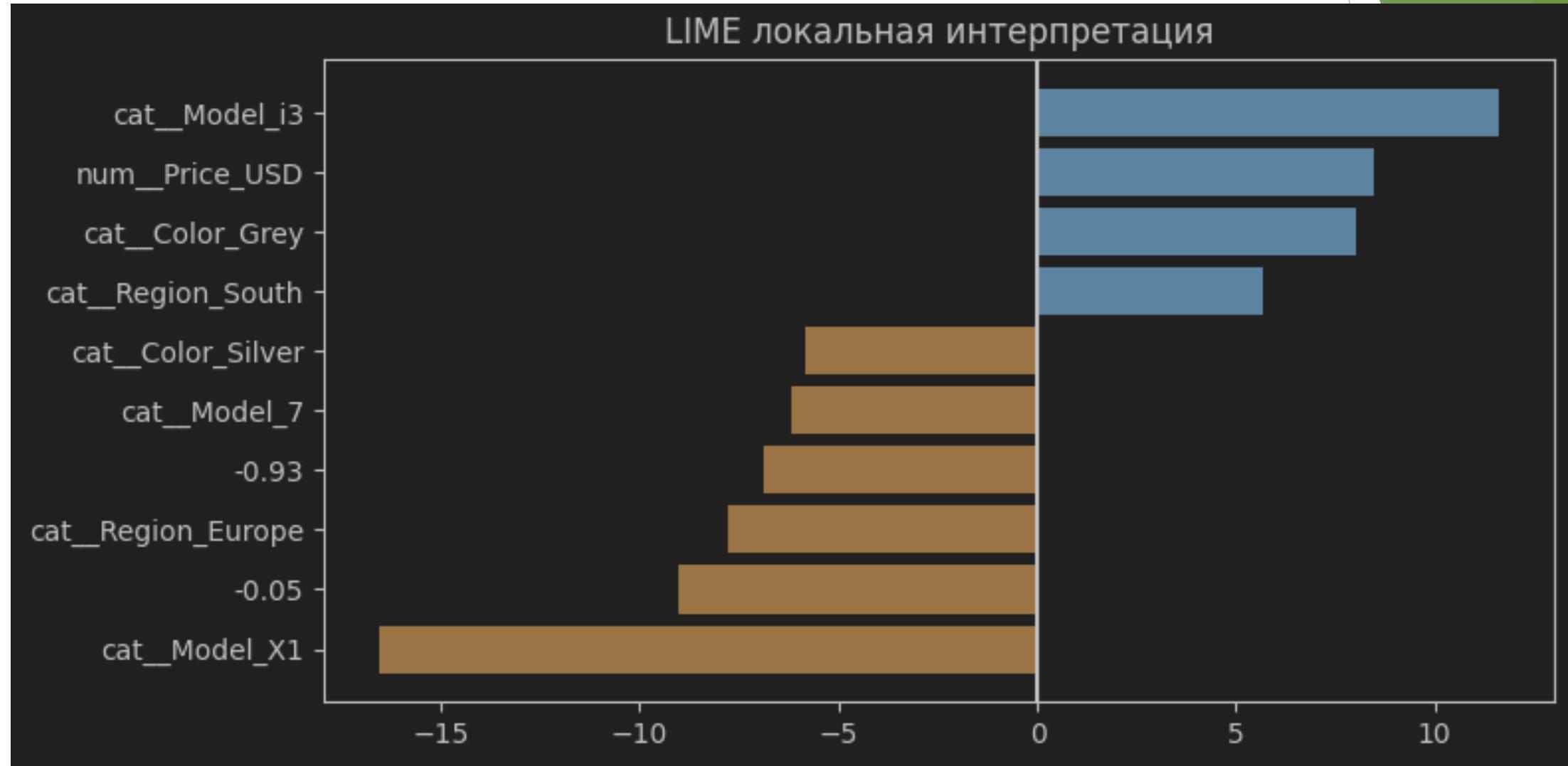
# Выбор лучшего метода

- ▶ Сравнив три метода видно, что среднеквадратичная ошибка меньше всего у метода Optuna со значением

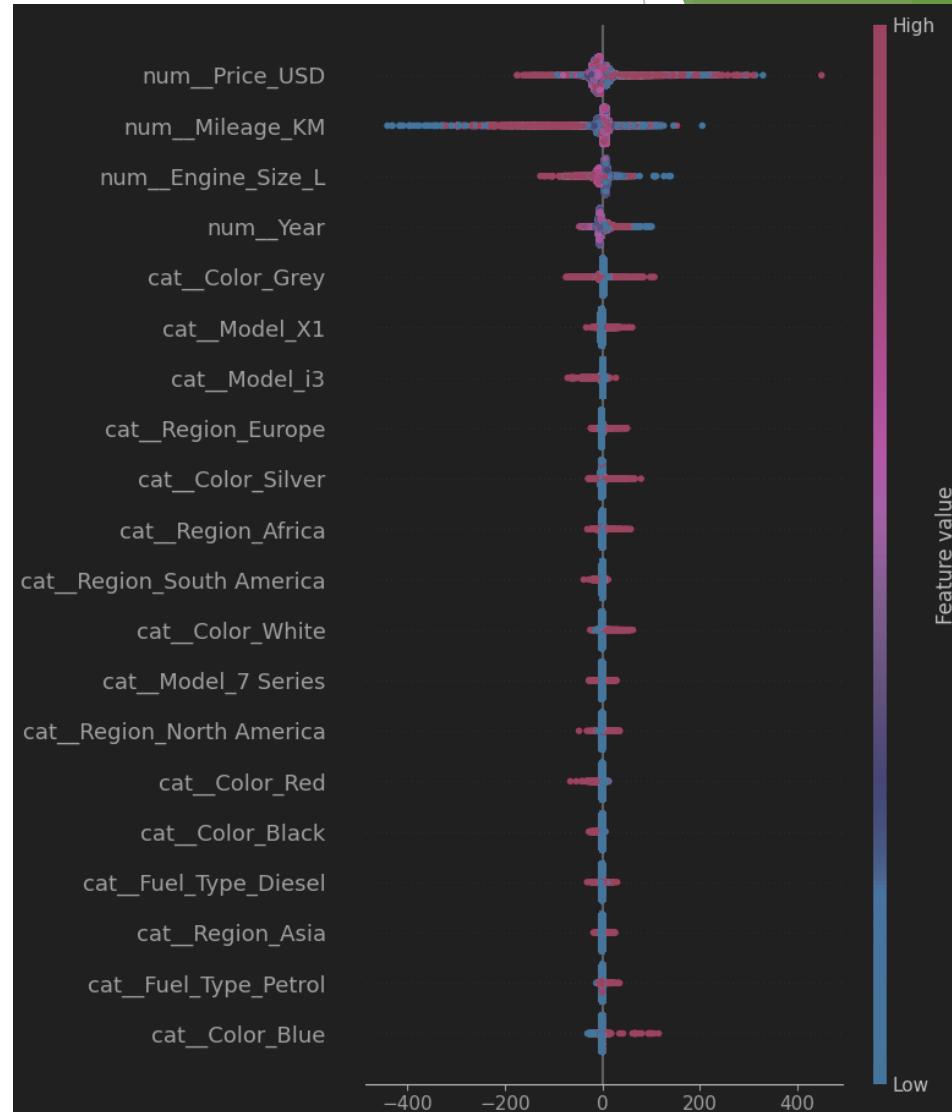
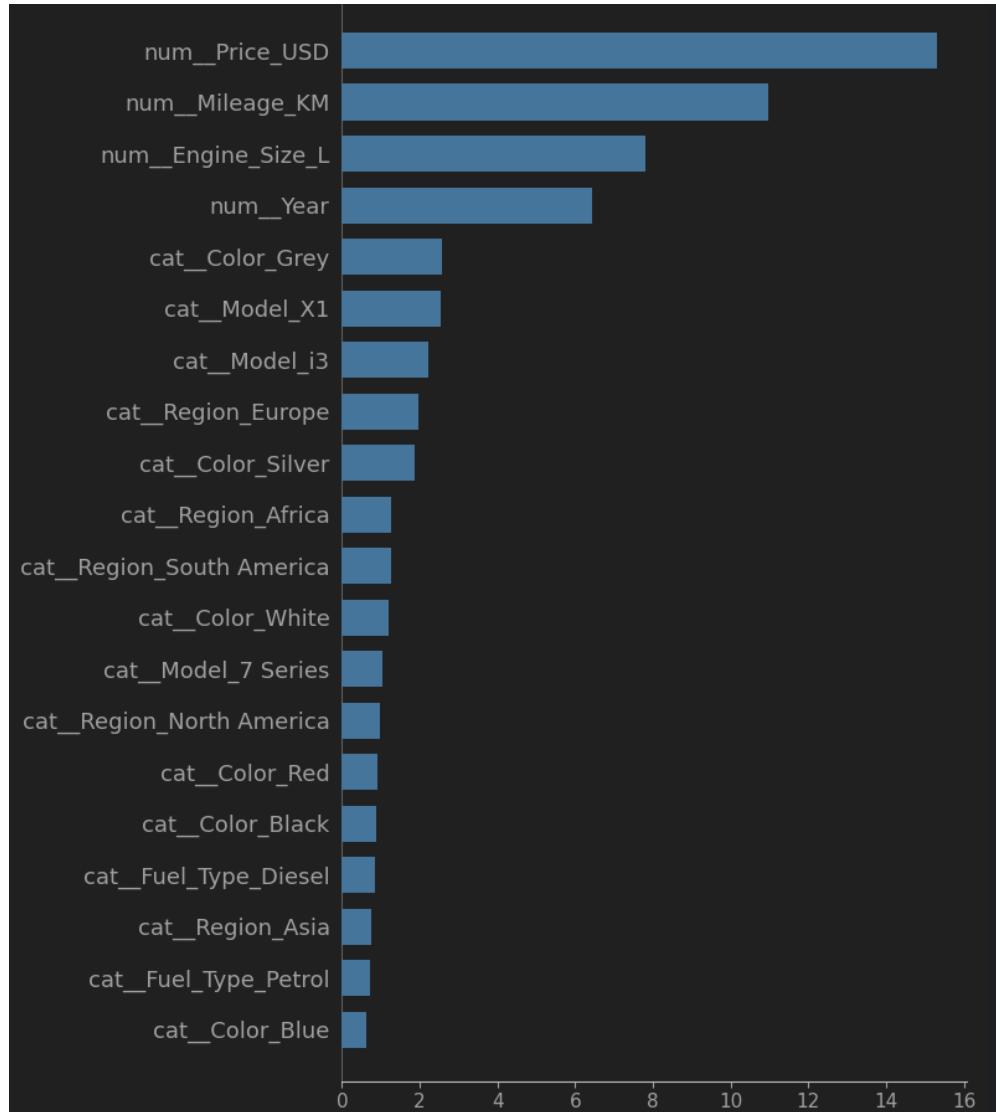
```
{'n_estimators': 146, 'max_depth': 5, 'min_samples_split': 6}
```

MSE Optuna: 8.173015e+06

# Локальная интерпретация



# Глобальная интерпретация



# Глобальная интерпретация

Топ признаков

- ▶ Price\_USD наибольший вклад
- ▶ Mileage\_KM - сильный вклад
- ▶ Engine\_Size\_L - сильный вклад
- ▶ Year - сильный вклад