

Лабораторная работа №3

Тема: Подбор гиперпараметров модели

Цели:

- 1) Выбрать модель
- 2) Показать параметры модели
- 3) Выбрать датасет и подготовить данные
- 4) Подобрать гиперпараметры модели для модели

На самом лучшем обучении реализовать калькулятор, который показывает локальную интерпретацию при помощи LIME, а глобальную при помощи SHAP



Модель – Random Forest

Устойчив к переобучению благодаря усреднению множества деревьев.

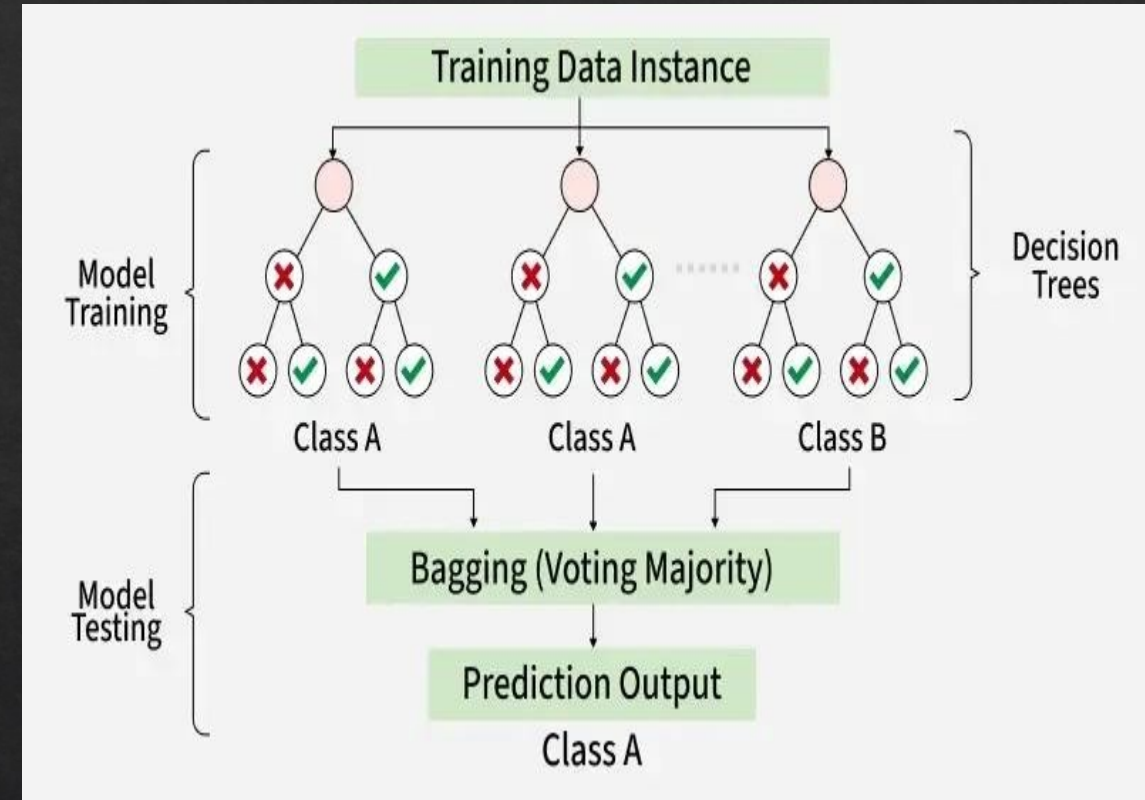
Не требует тонкой настройки — работает хорошо «из коробки».

Обработывает как числовые, так и категориальные признаки (с предобработкой).

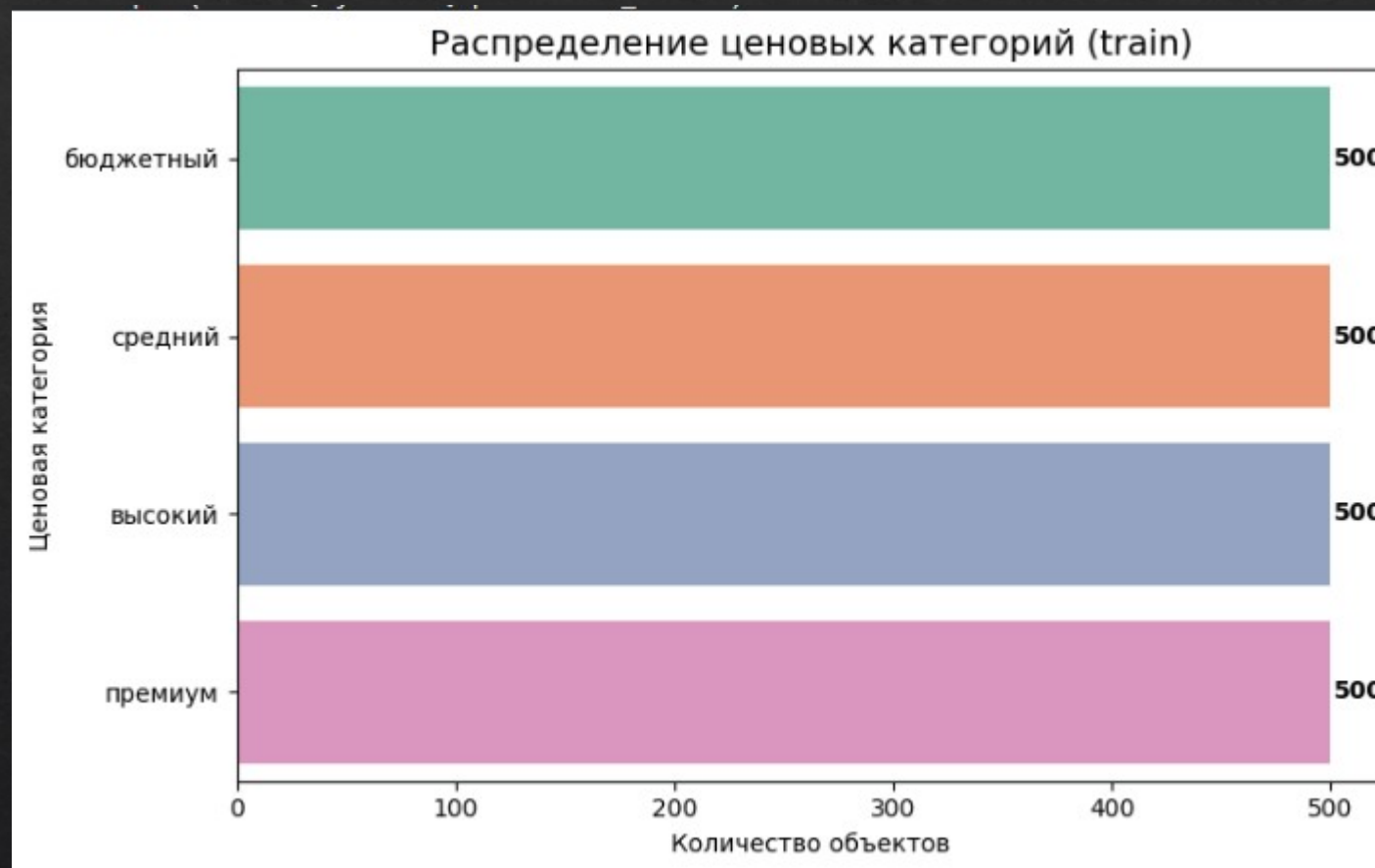
Нечувствителен к масштабу признаков и выбросам.

Даёт оценку важности признаков.

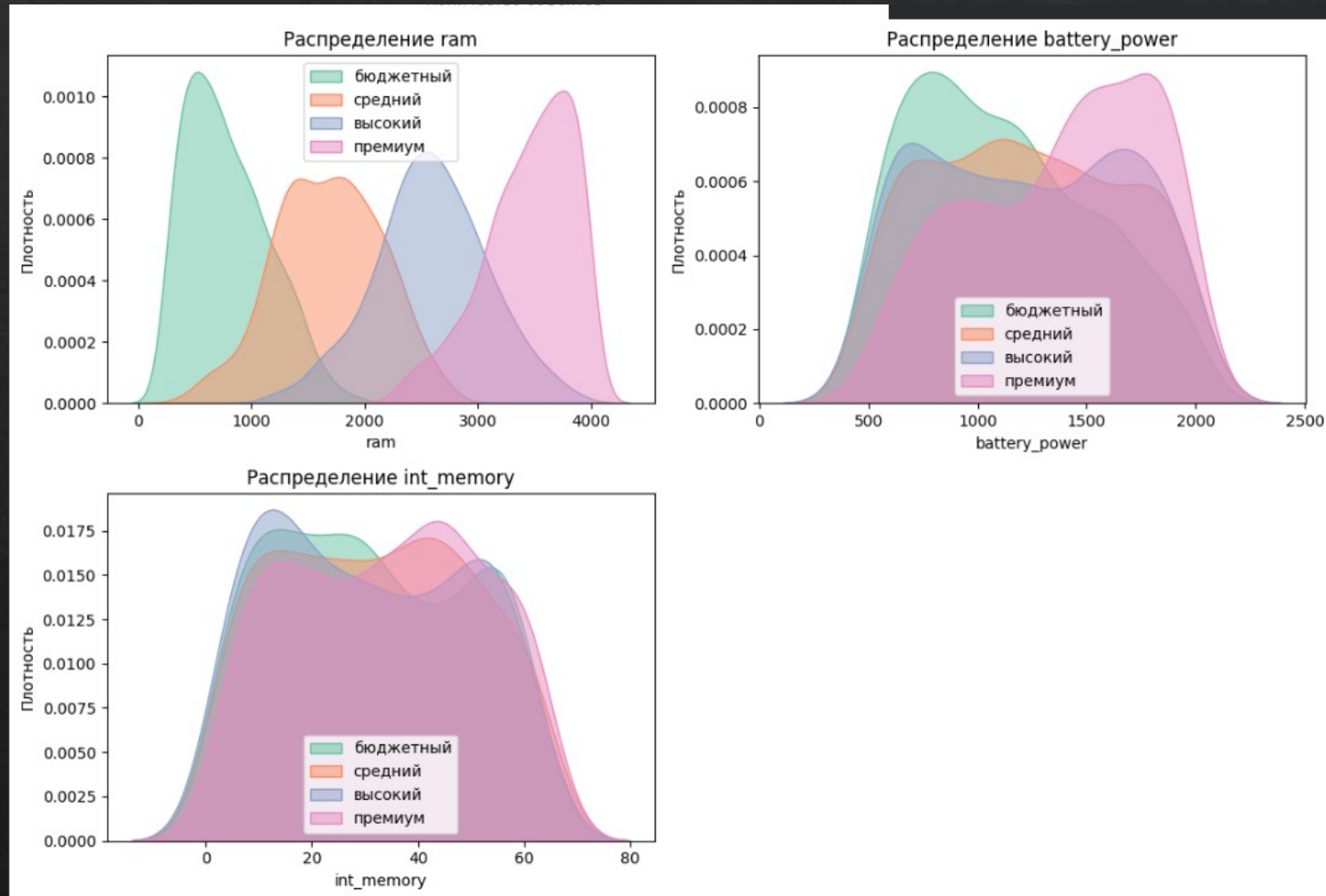
Работает с пропущенными данными (в некоторых реализациях).



Распределение ценовых категорий



Распределение параметров по ценовой категории



Разбиение данных для валидации

```
X_train_split, X_val, y_train_split, y_val = train_test_split(  
    X_train, y_train, test_size=0.2, random_state=RANDOM_STATE, stratify=y_train  
)  
  
print(f"Разбиение для тренировки: {X_train_split.shape}, Валидация: {X_val.shape}")
```

```
Train split: (1600, 20), Validation: (400, 20)
```

Результаты подбора гиперпараметров

Fitting 5 folds for each of 324 candidates, totalling 1620 fits

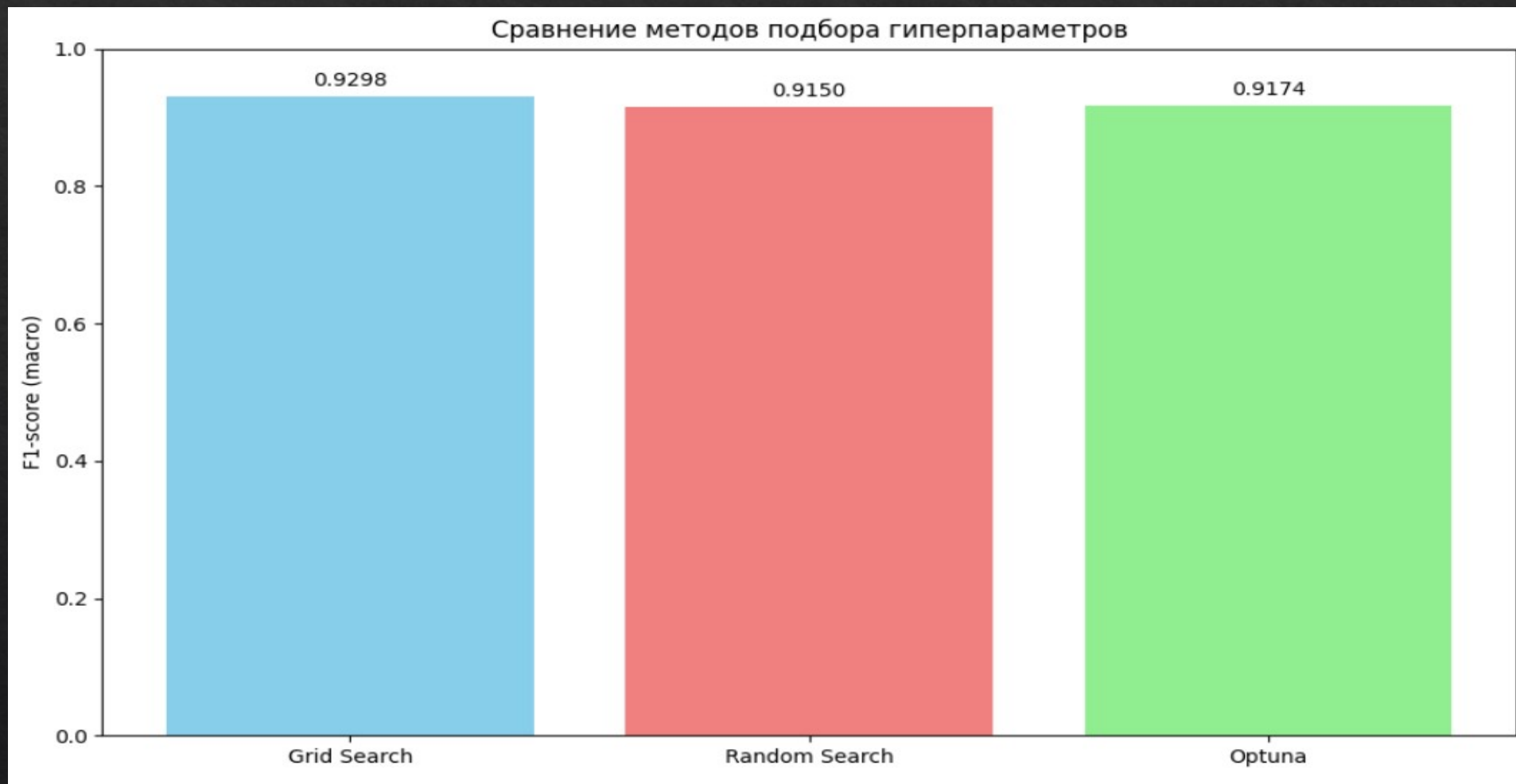
Лучшие параметры GridSearch: {'max_depth': 15, 'max_features': 0.5, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}
Лучший средний F1 (CV): 0.8844
F1 на validation: 0.9298

Fitting 5 folds for each of 50 candidates, totalling 250 fits

Лучшие параметры Random Search: {'max_depth': 20, 'max_features': 0.8, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 225}
Лучший средний F1 (CV): 0.8843
F1 на validation: 0.9150

Лучшие параметры (Optuna):
{'n_estimators': 383, 'max_depth': 13, 'min_samples_split': 3, 'min_samples_leaf': 3, 'max_features': 0.5, 'bootstrap': False}
Лучший средний F1 (CV): 0.8776
F1 на validation: 0.9146

Выбор и обучение лучшей модели



Пример работы реализованного калькулятора

Предсказанная категория: премиум

Вероятности категорий:

бюджетный: 0.000
средний: 0.010
высокий: 0.267
премиум: 0.723

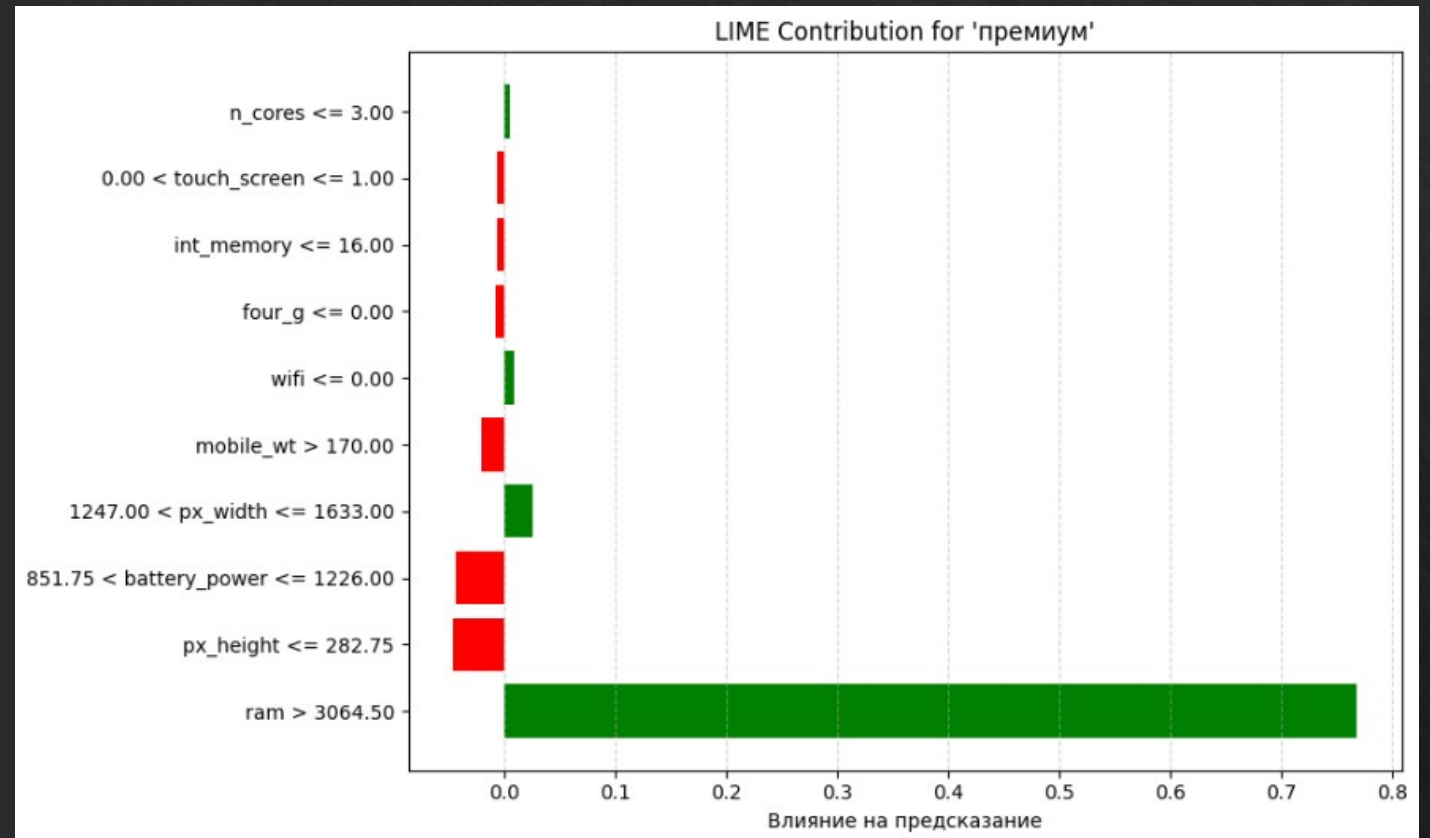
Основные характеристики:

ram: 3476
battery_power: 1043
px_height: 226
px_width: 1412
int_memory: 5
pc: 16

LIME Объяснение премиум?

LIME ANALYSIS FOR PREDICTION: премиум

Feature	Contribution	Sign
ram > 3064.50	0.768651	positive
px_height <= 282.75	-0.045465	negative
851.75 < battery_power <= 1226.00	-0.042809	negative
1247.00 < px_width <= 1633.00	0.026138	positive
mobile_wt > 170.00	-0.020906	negative
wifi <= 0.00	0.009130	positive
four_g <= 0.00	-0.007067	negative
int_memory <= 16.00	-0.006682	negative
0.00 < touch_screen <= 1.00	-0.006006	negative
n_cores <= 3.00	0.005728	positive



Sharp интерпретация результата

