

# ML. ЛАБОРАТОРНАЯ РАБОТА №3.

## ПОДБОР ГИПЕРПАРАМЕТРОВ МОДЕЛИ.

СТУДЕНТ: ЛЕУ ИВАН АЛЕКСАНДРОВИЧ

ГРУППА: М80-309Б-23

# ЦЕЛЬ И ЗАДАЧИ

ЦЕЛЬ И ЗАДАЧИ РАБОТЫ:

1. ВЫБРАТЬ МОДЕЛЬ ДЛЯ ОБУЧЕНИЯ
2. ПОСТРОИТЬ БИНАРНУЮ МОДЕЛЬ, ПРЕДСКАЗЫВАЮЩУЮ REFUSED\_FLAG (ЕСТЬ ЛИ ОТКАЗ ЗА НЕДЕЛЮ В БОЛЬНИЦЕ);
3. ПОДОБРАТЬ ГИПЕРПАРАМЕТРЫ С ПОМОЩЬЮ ТРЁХ МЕТОДОВ: GRIDSEARCH, RANDOMIZEDSEARCH, OPTUNA;
4. ВЫПОЛНИТЬ ЛОКАЛЬНУЮ И ГЛОБАЛЬНУЮ ИНТЕРПРЕТАЦИЮ МОДЕЛИ (LIME, SHAP).

# ДАТАСЕТ И ДАННЫЕ

Датасет: Медицинские коляски. Датасет предназначен для моделирования реальной деятельности больницы среднего размера с акцентом на укомплектование персоналом, прием пациентов и распределение коек между отделениями.

Источник: services\_weekly.csv (недельная агрегация по отделениям)  
Размер: 208 строк, 10 столбцов

КЛЮЧЕВЫЕ ПОЛЯ: WEEK, MONTH, SERVICE, AVAILABLE\_BEDS,  
PATIENTS\_REQUEST, PATIENTS ADMITTED, PATIENTS\_REFUSED,  
PATIENT\_SATISFACTION, STAFF\_MORALE, EVENT

# ЦЕЛЕВАЯ ПЕРЕМЕННАЯ И БАЛАНС КЛАССОВ

ЦЕЛЕВАЯ ПЕРЕМЕННАЯ: REFUSED\_FLAG = (PATIENTS\_REFUSED > 0)

ОПРЕДЕЛИМ ТАК: ОНА ОТВЕЧАЕТ НА ВОПРОС “БЫЛ ЛИ ХОТЯ БЫ ОДИН ОТКАЗ В ГОСПИТАЛИЗАЦИИ ЗА ЭТУ НЕДЕЛЮ ДЛЯ ДАННОЙ БОЛЬНИЦЫ?”

БОЛЬНИЦА ЗА НЕДЕЛЮ НЕ ОТКАЗАЛА НИ ОДНОМУ ПАЦИЕНТУ (ТО ЕСТЬ ВСЕХ ПРИНЯЛА) - 0,  
БОЛЬНИЦА ЗА НЕДЕЛЮ ОТКАЗАЛА ХОТЯ БЫ ОДНОМУ ПАЦИЕНТУ - 1

БАЛАНС:

148 — 1 (ЕСТЬ ОТКАЗ),

60 — 0 (НЕТ ОТКАЗА)

# ПРИЗНАКИ. ГИПЕРПАРАМЕТРЫ RANDOMFOREST

Категории: service, event → OneHotEncoder

Числовые: week, month, available\_beds, patients\_request, patients\_admitted, patient\_satisfaction, staff\_morale

Pipeline: ColumnTransformer (ОНЕ) → RandomForest

Гиперпараметры RandomForest (перечень).

ПОДБОР:

n\_estimators — количество деревьев, max\_depth — глубина, min\_samples\_split,  
min\_samples\_leaf, max\_features, criterion

# МЕТОДЫ ПОИСКА ГИПЕРПАРАМЕТРОВ

1. GRIDSEARCHCV — ПОЛНЫЙ ПЕРЕБОР (ДЕТЕРМИНИРОВАН). ПЕРЕБИРАЕТ ВСЕ ВОЗМОЖНЫЕ КОМБИНАЦИИ УКАЗАННЫХ ГИПЕРПАРАМЕТРОВ.
2. RANDOMIZEDSEARCHCV — СЛУЧАЙНАЯ ВЫБОРКА КОМБИНАЦИЙ (БЫСТРЕЕ). ВЫБИРАЕМ СЛУЧАЙНЫЕ КОМБИНАЦИИ ГИПЕРПАРАМЕТРОВ ИЗ ЗАРАНЕЕ ОПРЕДЕЛЁННЫХ ДИАПАЗОНОВ.
3. OPTUNA — ПОСЛЕДОВАТЕЛЬНАЯ ОПТИМИЗАЦИЯ (ИНТЕЛЛЕКТУАЛЬНЫЙ ПОИСК). СТРОИМ ВЕРОЯТНОСТНУЮ МОДЕЛЬ ФУНКЦИИ ПОТЕРЬ И ИСПОЛЬЗУЕМ ЕЁ, ЧТОБЫ ВЫБИРАТЬ НОВЫЕ ГИПЕРПАРАМЕТРЫ «УМНЕЕ», ЧЕМ СЛУЧАЙНО.

GRID — ГАРАНТИРОВАННО ПЕРЕБЕРЕТ СЕТКУ, RANDOMIZED — ЭКОНОМИТ ВРЕМЯ, OPTUNA — ЧАСТО ДАЁТ ХОРОШИЕ РЕЗУЛЬТАТЫ ПРИ ОГРАНИЧЕННОМ БЮДЖЕТЕ.

# РЕЗУЛЬТАТЫ

- GRIDSEARCHCV – THE BEST! -> ВЫБИРАЕМ КАК ФИНАЛЬНУЮ МОДЕЛЬ

TEST ACCURACY: 0.9337970478998298, {'CRITERION': 'ENTROPY', 'MAX\_DEPTH': 5, 'MAX\_FEATURES': NONE, 'MIN\_SAMPLES\_LEAF': 2, 'MIN\_SAMPLES\_SPLIT': 2, 'N\_ESTIMATORS': 150}

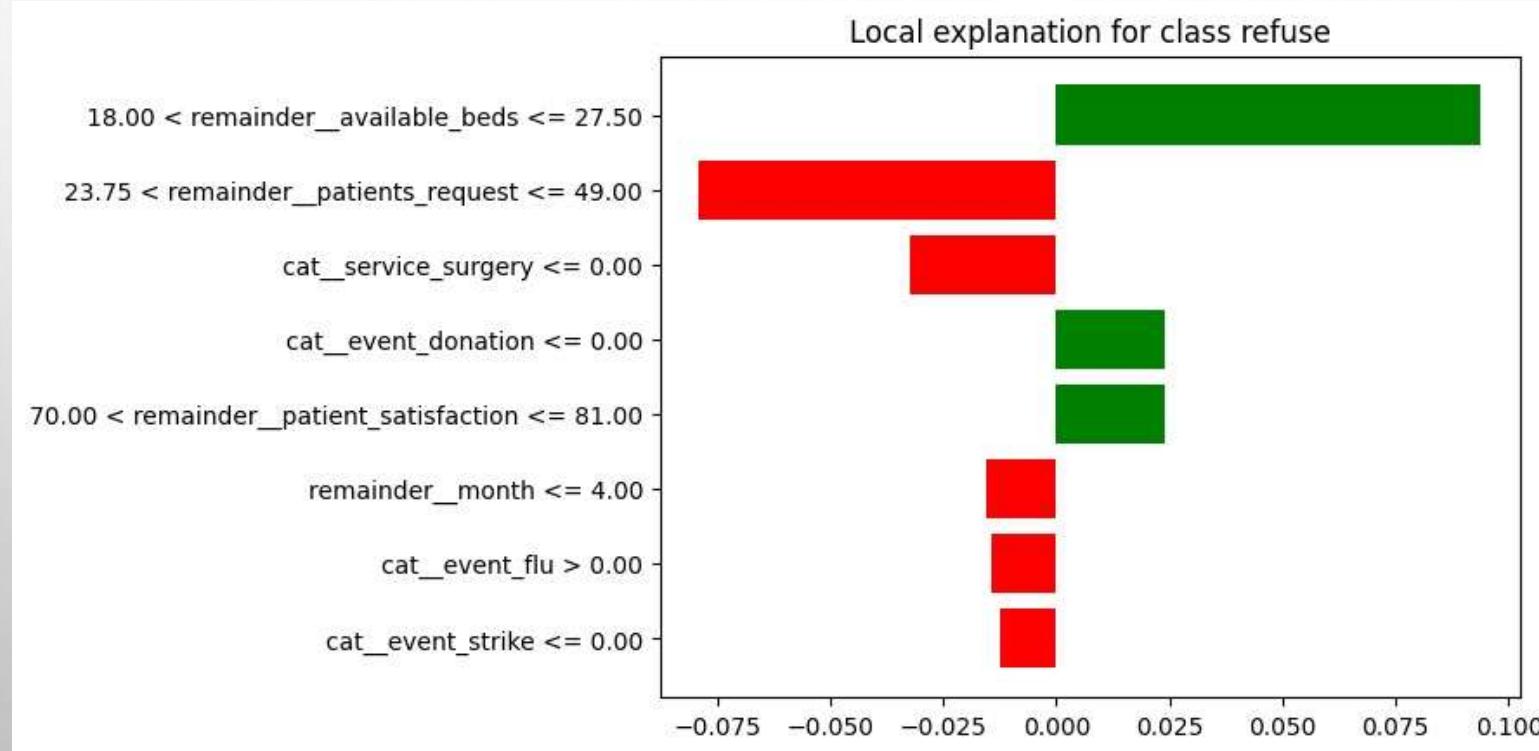
- RANDOMIZEDSEARCHCV

TEST ACCURACY: 0.9337970478998298, {'CRITERION': 'ENTROPY', 'MAX\_DEPTH': 20, 'MAX\_FEATURES': NONE, 'MIN\_SAMPLES\_LEAF': 3, 'MIN\_SAMPLES\_SPLIT': 2, 'N\_ESTIMATORS': 185}

- OPTUNA

TEST ACCURACY: 0.8959156785243741, {'N\_ESTIMATORS': 61, 'MAX\_DEPTH': 20, 'MIN\_SAMPLES\_SPLIT': 2, 'MIN\_SAMPLES\_LEAF': 4, 'MAX\_FEATURES': 'SQRT', 'CRITERION': 'GINI'}

# ЛОКАЛЬНАЯ ИНТЕРПРЕТАЦИЯ: LIME



ЧТО ДЕЛАЕТ: ОБЪЯСНЯЕТ ПРЕДСКАЗАНИЕ ДЛЯ ОДНОЙ СТРОКИ (ЛОКАЛЬНО).

ПРИМЕР: ПОКАЗАН ВЫВОД LIME для 7-й недели: PATIENTS\_REQUEST и AVAILABLE\_BEDS дали наибольший вклад в сторону отказа.

# ГЛОБАЛЬНАЯ ИНТЕРПРЕТАЦИЯ: SHAP

(SHAP - SHapley Additive exPlanations). Что делает: показывает средний, “глобальный” вклад признаков в предсказания модели.

Результат (топ-признаки):

patients\_request 0.3207 – **самый сильный**  
available\_beds 0.1262 – **самый сильный**  
patients\_admitted 0.0039  
week 0.0031  
patient\_satisfaction 0.003  
staff\_morale 0.0012  
month 0.0011  
event\_flu 0.0006  
service\_emergency 0.0006

# ГЛОБАЛЬНАЯ ИНТЕРПРЕТАЦИЯ: SHAP. ВИЗУАЛИЗАЦИЯ

