

Лабораторная работа 3 - подбор гиперпараметров модели

- Датасет: European Soccer Database

Выбор модели – Random Forest

Модель: Random Forest

Тип задачи: Классификация (предсказание
предпочитаемой ноги футболиста)

Причины выбора:

Устойчивость к переобучению

Хорошая интерпретируемость

Возможность работы с категориальными и числовыми
признаками

Датасет и подготовка данных

Датасет: European Soccer Database (Kaggle)

Целевая переменная: preferred_foot (правая/левая нога)

Числовые признаки: 37 (рейтинги игрока, навыки и т.д.)

Категориальные признаки: отсутствуют после обработки

Размер выборки: 183 978 строк

Баланс классов:

Правая нога: 138 409

Левая нога: 44 733

Гиперпараметры Random Forest

Гиперпараметр	Описание
criterion	Критерий, используемый для оценки качества разбиения ветки (например, 'squared_error')
max_depth	Максимальная глубина дерева, ограничивает сложность, предотвращает переобучение
min_samples_split	Минимальное количество объектов для разделения узла (увеличение снижает переобучение)
min_samples_leaf	Минимальное количество объектов в листе (увеличение сглаживает дерево)
max_features	Максимальное количество признаков, используемых для поиска лучшего разбиения
random_state	Фиксирует генератор случайных чисел для воспроизводимости результатов
splitter	Стратегия выбора разбиения: 'best' (лучшее) или 'random'
max_leaf_nodes	Ограничивает максимальное число листьев в дереве
min_weight_fraction_leaf	Минимальная доля веса всех объектов в листе (актуально для взвешенных объектов)