

Лабораторная работа 3.

Тема: Классификация методом дерева решений.

Освоить практическое применение метода решающего дерева для задач классификации. Исследовать влияние гиперпараметров на качество модели и научиться проводить базовый анализ важности признаков.

Теоретическая часть

Решающее дерево — алгоритм машинного обучения, который строит иерархическую структуру правил «если-то» для классификации или регрессии. Основные гиперпараметры:

`max_depth` — максимальная глубина дерева

`max_leaf_nodes` — максимальное количество листовых узлов

`min_samples_split` — минимальное количество samples для разделения узла

Задание

1. Подготовка данных

1.1. Выберите любой датасет для бинарной или многоклассовой классификации с сайта [Kaggle.com](https://www.kaggle.com/).

Рекомендуемые датасеты: Titanic, Iris, Wine Quality, Heart Disease Prediction и др.

1.2. Загрузите данные и выполните предобработку:

Обработайте пропущенные значения.

Закодируйте категориальные признаки (One-Hot Encoding или Label Encoding).

Разделите данные на обучающую и тестовую выборки (70/30 или 80/20).

2. Базовое дерево

2.1. Обучите решающее дерево с параметрами по умолчанию.

2.2. Вычислите ассигуру на тестовой выборке.

2.3. Определите 3 наиболее важных признака для классификации.

3. Подбор гиперпараметров

3.1. Проведите поиск оптимальных параметров методом перебора `max_depth` и `max_leaf_nodes` в интервалах, подходящих вашему набору данных.

3.2. Для каждой комбинации параметров:

Обучите модель на тренировочных данных;

Вычислите ассигуру на тестовых данных;

Зафиксируйте результаты

4. Анализ результатов

- 4.1. Определите комбинацию параметров, дающую наивысшую точность.
- 4.2. Визуализируйте зависимость ассигасы от глубины дерева и количества листьев
- 4.3. Сравните ассигасу лучшей модели с базовой (параметры по умолчанию)

5. Визуализация

- 5.1. Обучите финальную модель с оптимальными параметрами.
- 5.2. Визуализируйте дерево с помощью `plot_tree` или `export_graphviz`.
- 5.3. Сохраните изображение дерева с оптимальными параметрами для отчета.

6. Отчёт

Оформите отчет в электронном виде, приложив ссылку на Jupyter Notebook/ Google Colab, где код сопровождается краткими выводами по каждому шагу, электронный вид отчёта загрузите в формате pdf в ЭИОС.