

ФГАОУ ВО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт космических и информационных технологий
Кафедра «Информатика»

Методы анализа данных

Практическая работа № 5
Деревья решений

Красноярск, 2024

Цель: знакомство с теоретическими основами построения деревьев решений, формирование навыков построения деревьев решений для решения задач классификации и регрессии с помощью языка программирования Python.

Задачи:

Выполнение практической работы предполагает решение следующий задач:

1. Предварительная обработка исходных данных
2. Построение базовых моделей на основе деревьев решений
3. Подбор оптимальных параметров моделей
4. Оценка качества построенных моделей на валидационной/тестовой выборке

Ссылки на данные

Для задачи классификации взять датасет из практической работы №3

Для задачи регрессии взять датасет из практической работы №4

Общая последовательность действий

1. Загрузить данные для обучения и для теста.
2. Выполнить предварительную обработку исходных данных (в случае необходимости)
3. Построить дерево решений. Визуализировать результат.
4. Исследовать зависимость качества прогнозирования от настраиваемых параметров (на тестовой выборке, на контрольной выборке, в ходе перекрестной проверки). Построить график зависимости значений метрик от значений параметров модели. Сделать выводы об оптимальных значениях параметров.
5. Исследовать, приводит ли обрезка дерева к улучшению результата на контрольной выборке.
6. Сравнить результаты, полученные с использованием дерева решений и с помощью реализованных алгоритмов в 3 и 4 работе.

Требования к выполнению практической работы:

1. Написание программного кода и формирование результатов согласно заданию и установленному варианту.
2. Составление отчета, содержащего описание решаемых задач методов решения и полученных результатов.

3. Воспроизводимость полученного результата.

Программный код и отчет должны быть выполнены в среде Jupyter notebook.