Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт информационные системы и анализ данных

Центр программной инженерии

**ОТЧЁТ**

к лабораторной работе №5 по дисциплине:

|  |
| --- |
| «Методы анализа данных» |
| Построение деревьев решений для задачи классификации |
| наименование темы |

Выполнилстудент ИСТб-21-1 Е.А. Юрковский

номер группы подпись И. О. Фамилия

дата

Проверил Доцент Е.А. Осипова

Должность подпись И. О. Фамилия

дата

Иркутск 2023

**Оглавление**

[Цель работы и задание 3](#_Toc136270210)

[1 Ход работы 4](#_Toc136270211)

[1.1 Получение исходных данных 4](#_Toc136270212)

[1.2 Построение диаграммы рассеивания и дерева решений 4](#_Toc136270213)

[Заключение. 8](#_Toc136270214)

[Список использованных источников 9](#_Toc136270215)

# Цель работы и задание

**Цель работы**: изучить способы построения деревьев решений для задачи классификации.

**Задание:**

А. Выбрать среду программирования для языка Python.

Б. Получить набор данных с параметрами, согласно номеру варианта. Построить деревья решений и выполнить их анализ. При этом необходимо:

* разделить исходную выборку на две равные части: для обучения и для тестирования;
* обучить дерево решений и вывести его на экран;
* используя функцию рисования решающих поверхностей из лабораторной работы № 3, вывести на экран решающие поверхности и данные для построенного дерева решений для каждой пары признаков;
* оценить точность модели на тестирующей выборке;
* в случае, если точность модели менее 0,9, изменяя значение параметров criterion, max\_depth, min\_samples\_split, min\_samples\_leaf попытаться ее повысить.

В. Выполнить анализ полученных на каждом этапе результатов и оформить отчет по лабораторной работе.

**Варианты индивидуальных заданий**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ варианта** | **Количество  кластеров** | **Количество признаков** | **Стандартное отклонение** |
| 9 | 3 | 4 | 3 |

# 1 Ход работы

# Получение исходных данных

Получение исходных данных:

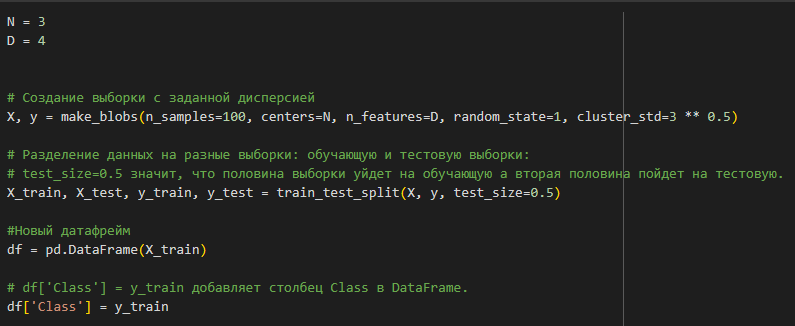


Рисунок 1 – Получение исходных данных

# Построение диаграммы рассеивания и дерева решений

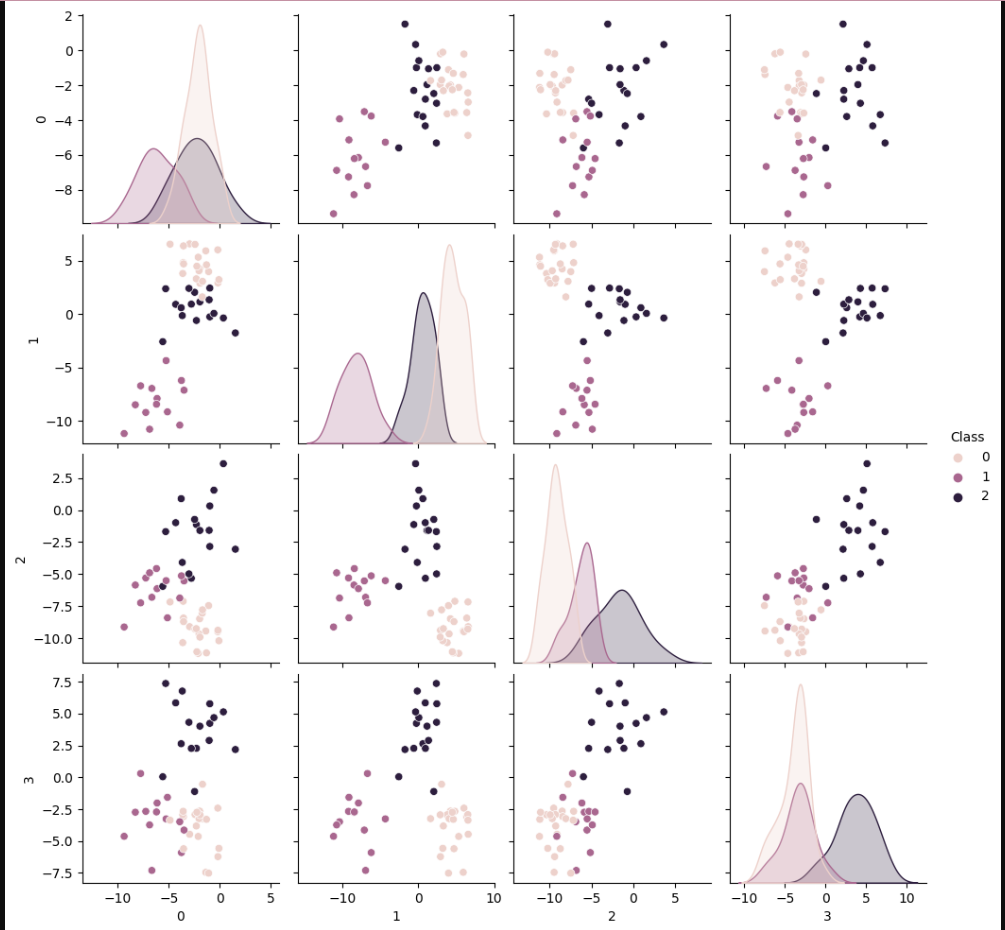


Рисунок 2 – Диаграммы рассеяния

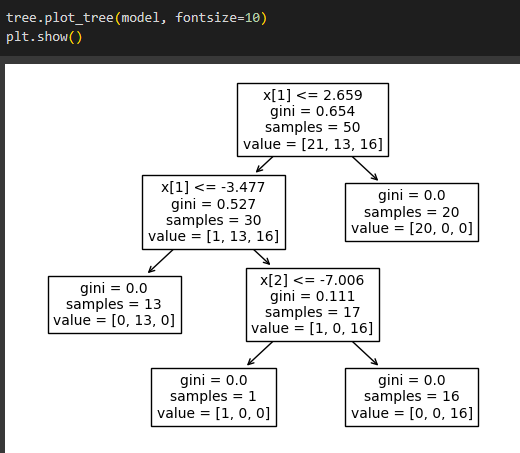


Рисунок 3 – Классификационное дерево решений

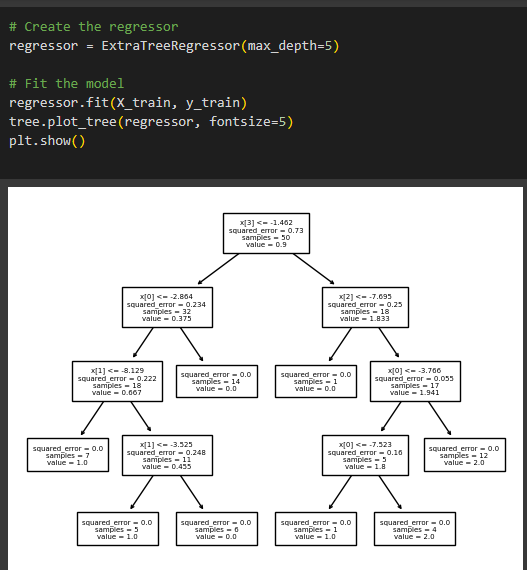


Рисунок 4 – Регрессионное дерево решений

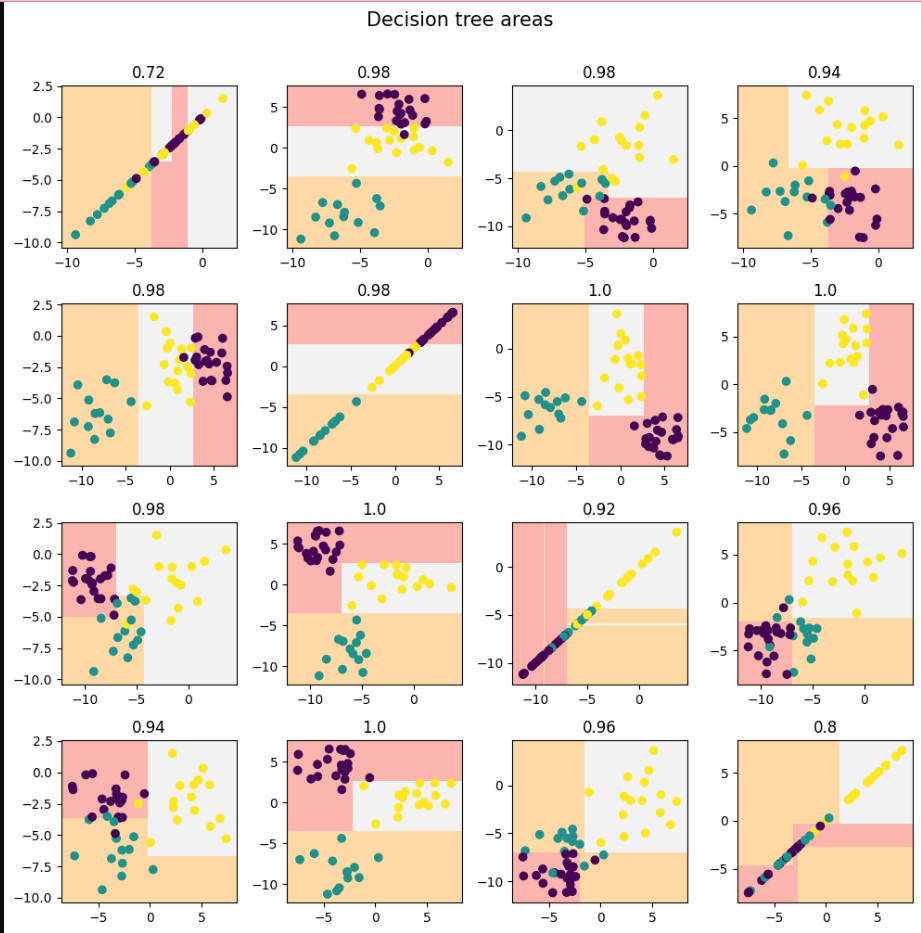


Рисунок 5 – Графики дерева решений для каждой  
 комбинации пар признаков

Теперь можно проверить, насколько точной получилась модель. На рисунке 6 представлены результаты оценки точности модели.

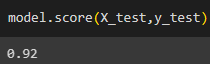


Рисунок 6 – Результат оценки точности модели

# Заключение.

В ходе лабораторной работы были получены навыки работы с деревьями решений. Были построены диаграммы рассеяния для сгенерированных данных, классификационное дерево решений, регрессионное дерево решений. Была отредактирована функция отрисовки решающих поверхностей и был выведен график дерева решений для каждой  
 комбинации пар признаков. Также была оценена точность модели, она была равна 92%

# Список использованных источников

1 Петров А.В., Бучнев О.С. Лабораторный практикум — Иркутск: ФГБОУ ВО ИРНИТУ, 2022. — 114 с.