

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное  
учреждение высшего образования  
**"Южно-Уральский государственный университет  
(национальный исследовательский университет)"**  
Высшая школа электроники и компьютерных наук  
Кафедра системного программирования

**ОТЧЕТ**  
**по практической работе №4**  
**«Ансамблевая классификация»**  
по дисциплине  
**«Технологии аналитической обработки информации»**

Выполнил: \_\_\_\_\_

студент группы КЭ-404

А.Ю. Емельянова

Проверил: \_\_\_\_\_

преподаватель

А.И. Гоглачев

Дата: \_\_\_\_\_

Оценка: \_\_\_\_\_

## Формулировка задания

1. Разработайте программу, которая выполняет классификацию заданного набора данных с помощью одной из техник ансамблевой классификации. Параметрами программы являются набор данных, ансамблевая техника (бэггинг, случайный лес или бустинг), количество участников ансамбля, а также параметры в соответствии с выбранной техникой ансамблевой классификации.

2. Проведите эксперименты на наборе данных из задания [Классификация с помощью дерева решений](#), варьируя количество участников ансамбля (от 50 до 100 с шагом 10).

3. Выполните визуализацию полученных результатов в виде следующих диаграмм: показатели качества классификации в зависимости от количества участников ансамбля для заданного набора данных; нанесите на диаграмму соответствующие значения, полученные в задании [Классификация с помощью дерева решений](#).

Гиперссылка на каталог репозитория с исходными текстами, наборами данных и другими материалами: <https://github.com/Sunless-sea/Technologies-of-analytical-information-processing>

## Визуализация

В данной практической работе было принято решение использовать метод случайного леса для выполнения задания.

Визуализация полученных результатов представлена на рисунке 1. На графике представлено сравнение числового значения метрик при использовании случайного леса, дерева решений на основе критерия индекса Джини и другого дерева на основе критерия энтропии. Значения метрик решений деревьев были взяты из предыдущего задания «Классификация с помощью дерева решений».

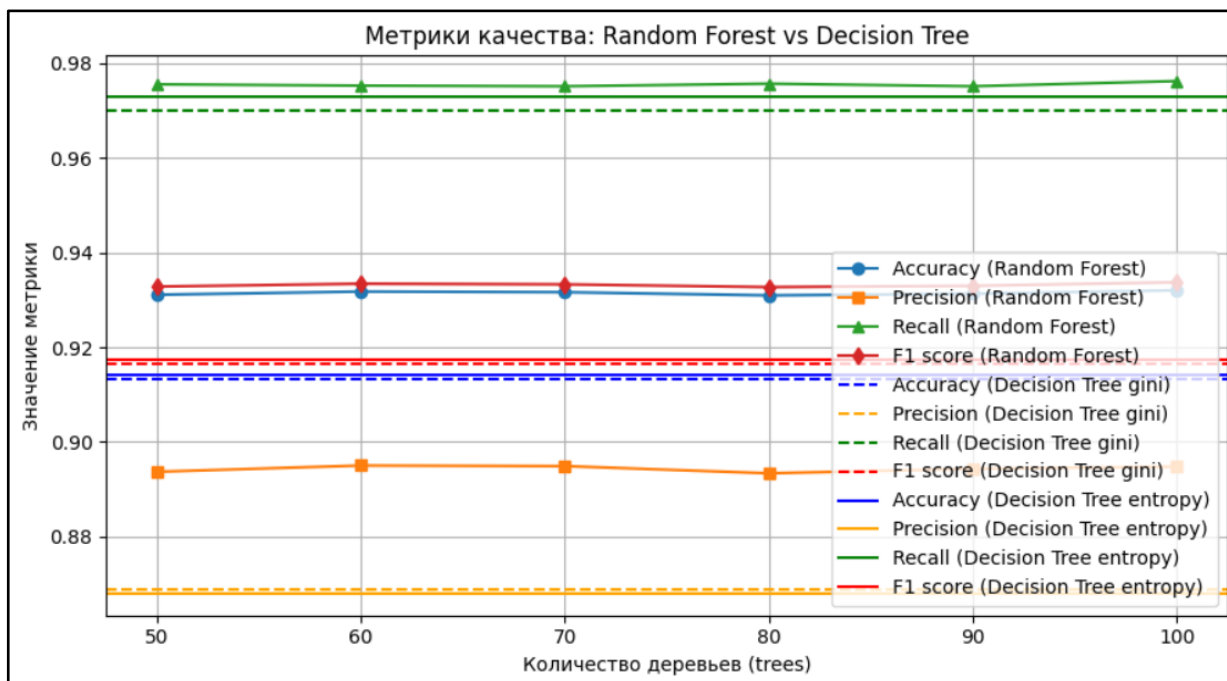


Рисунок 1 – Метрики качества в сравнении случайного леса и деревьев решений

На графике видно, что случайный лес отлично справился с задачей классификации. Однако, задача была выполнена хорошо уже на количестве в 50 деревьев. Увеличение количества деревьев не приносит пользы, но переобучения не случается, так как метрики остаются в своих пределах.

Также можно сравнить работу случайного леса с деревом решений из предыдущей работы. Он лучше справляется, чем одно дерево решений. Его значения метрик выше и стабильнее, а значит модель обучена лучше, поэтому эффективность леса больше, чем у дерева решений.