

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-403
О.С. Мазжухин

Проверил: _____
преподаватель
А.И. Гоглачев
Дата: _____
Оценка: _____

Формулировка задания

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.

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

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/LN4rkot1k/informationProcessing>

Было принято решение использовать ансамблевую технику случайный лес для выполнения задания. Визуализация полученных результатов представлена на рисунке 1.

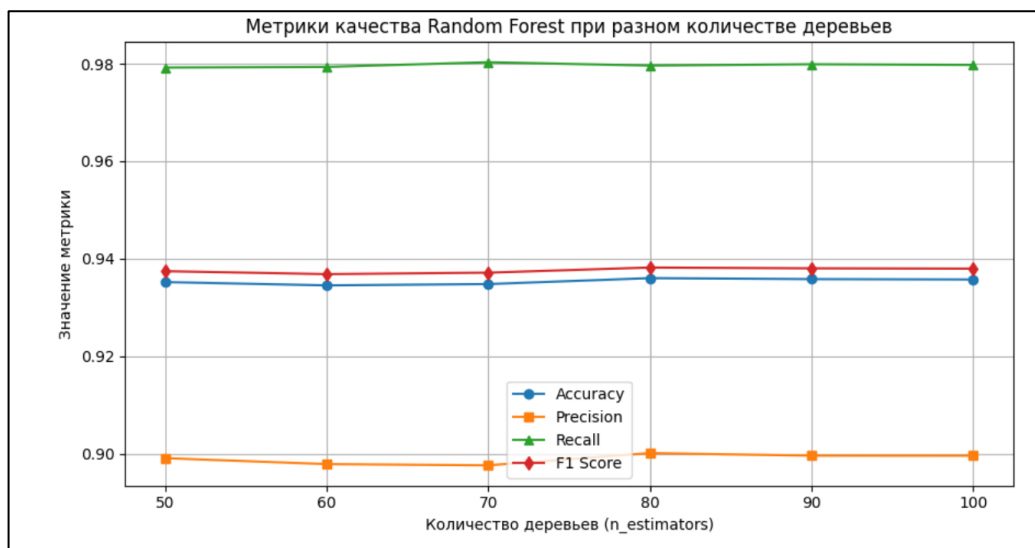


Рисунок 1 – Метрики качества случайного леса при разном количестве деревьев

Глядя на график, можно сделать вывод о том, что случайный лес хорошо справился с задачей классификации. Однако, модель уже на 50 деревьях обучилась достаточно хорошо, поэтому увеличение деревьев не приносит большой пользы. Переобучение не наблюдается, потому что наши метрики не снижаются.