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 «Системное программирование»

**ОТЧЁТ**

по лабораторной работе №6

на тему «Ансамблевая классификация с помощью случайного леса»

|  |  |
| --- | --- |
|  | Выполнил  Студент группы КЭ-120  \_\_\_\_\_\_\_\_\_\_\_\_\_ Д.А. Снегирева  «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_2020 г. |
|  | Email: dashasneg@mail.ru |

Челябинск 2020

**ЗАДАНИЕ**

Выполните классификацию набора данных из задания 3 с помощью случайного леса, варьируя количество участников ансамбля (от 50 до 100 с шагом 10). Вычислите показатели качества классификации: аккуратность (accuracy), точность (precision), полнота (recall), F-мера. Выполните визуализацию полученных результатов в виде диаграмм. Нанесите на диаграммы соответствующие значения, полученные в заданиях 3, 4, 5.

СОДЕРЖАНИЕ

[ЗАДАНИЕ 2](#_Toc68258457)

[СОДЕРЖАНИЕ 3](#_Toc68258458)

[1 КРАТКИЕ СВЕДЕНИЯ О НАБОРАХ ДАННЫХ И СРЕДСТВАХ РЕАЛИЗАЦИИ 4](#_Toc68258459)

[2 ВИЗУАЛИЗАЦИЯ РЕЗУЛЬТАТОВ 5](#_Toc68258460)

# КРАТКИЕ СВЕДЕНИЯ О НАБОРАХ ДАННЫХ И СРЕДСТВАХ РЕАЛИЗАЦИИ

В данной работе был использован набор данных Wine Data Set (https://archive.ics.uci.edu/ml/datasets/wine), содержащий результаты химического анализа вин, выращенных в одном и том же регионе Италии, и состоящий из 13 атрибутов.

В качестве одного из средств реализации была использована библиотека scikit-learn, простое средство для анализа данных.

Репозиторий задания: https://github.com/DasHaSneg/BigDataMiningCourse

Каталог задания: 6 random\_forest

ВИЗУАЛИЗАЦИЯ РЕЗУЛЬТАТОВ

Для визуализации были вычислены показатели качества классификации: аккуратность (accuracy), точность (precision), полнота (recall) и F-мера для разного количество участников ансамбля (от 50 до 100 с шагом 10). Для каждого показателя были построены диаграммы, которые представлены на рис. 1. Кроме того, были нанесены соответствующие значения, полученные в заданиях 3, 4 и 5.

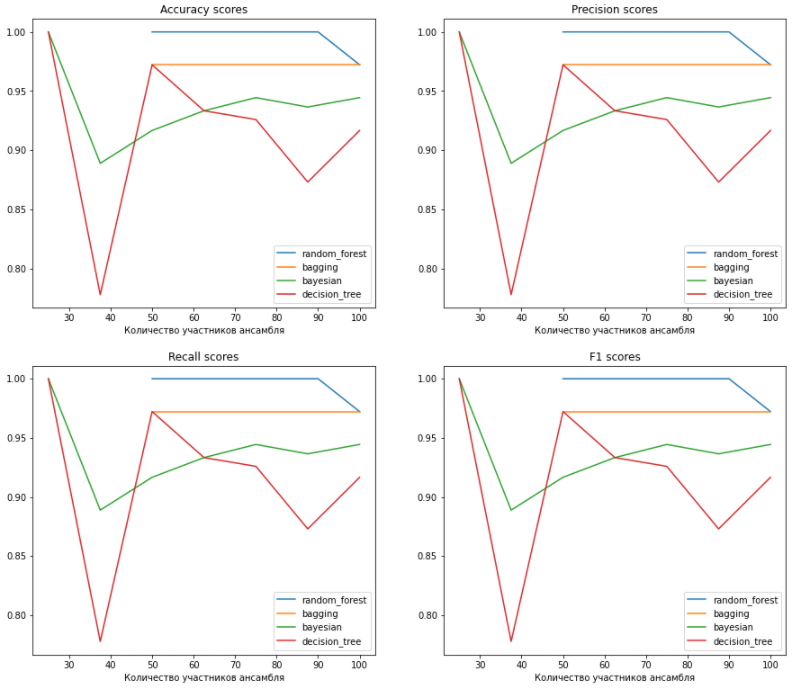


Рис. 1. Результаты

Исходя из рисунка, можно сказать, что классификация с помощью бэггинга показала результат хуже, чем Байесовская классификация, деревья решений и случайный лес, но при классификации с помощью случайного леса при достижении количества участников равное 90 результат становится равным результату классификации при помощи бэггинга.