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 лабораторной работе №3**

**Технологии аналитической обработки информации**

|  |  |
| --- | --- |
|  | Выполнил: студент группы КЭ-404  Кондрашин М.О.  Проверил:  Преподаватель  Гоглачев А.И. |

Челябинск, 2025 г.

**ФОРМУЛИРОВКА ЗАДАНИЯ**

1. Разработайте программу, которая выполняет классификацию заданного набора данных с помощью дерева решений. Параметрами программы являются набор данных, критерий выбора атрибута разбиения (Information gain, Gain ratio, Gini index).
2. Проведите эксперименты на наборе Census Income (данные о результатах переписи населения, в т.ч. о годовом доходе -- ниже или выше $50000: скачать обучающую выборку в формате CSV, тестовую выборку в формате CSV, скачать описание). В качестве обучающей выборки для построения дерева используйте 100% исходных данных.
3. Выполните визуализацию построенных деревьев решений.
4. Доработайте программу, добавив в список ее параметров долю, которую занимает обучающая выборка от общего размера набора данных, и обеспечив вычисление и выдачу в качестве результатов следующих показателей качества классификации: аккуратность (accuracy), точность (precision), полнота (recall), F-мера.
5. Проведите эксперименты на наборе данных, фиксируя критерий выбора атрибута разбиения и варьируя соотношение мощностей обучающей и тестовой выборок от 60%:40% до 90%:10% с шагом 10%.
6. Выполните визуализацию полученных результатов в виде следующих диаграмм:

* построенные деревья решений для заданного набора данных;
* показатели качества классификации в зависимости от соотношения мощностей обучающей и тестовой выборок для заданного набора данных.

1. Подготовьте отчет о выполнении задания и загрузите отчет в формате PDF в систему. Отчет должен представлять собой связный и структурированный документ со следующими разделами:

* формулировка задания;
* гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами;
* рисунки с результатами визуализации;
* пояснения, раскрывающие смысл полученных результатов.

**ССЫЛКА НА КАТАЛОГ РЕПОЗИТОРИЯ**

Гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами: https://github.com/Max-Wind0/Tree-classification-and-Ensemble-classification

**РИСУНКИ С РЕЗУЛЬТАТАМИ ВИЗУАЛИЗАЦИИ**

В ходе выполнения лабораторной работы по классификации с построением деревьев решений были получены следующие результаты:

Дерево решений, соответствующее модели с использованием критерия разбиения gini, приведено на рисунке 1.

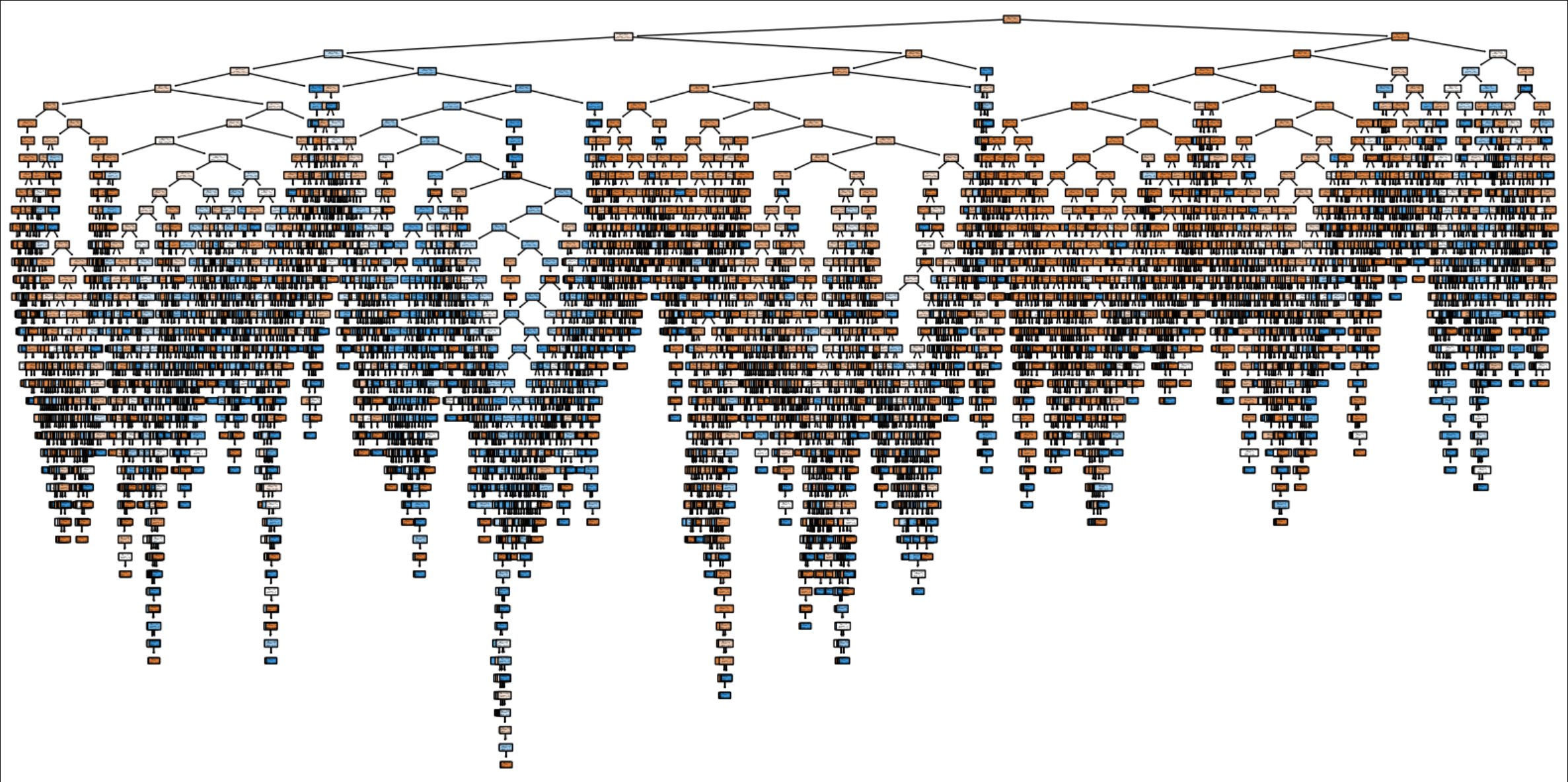


Рисунок 1 – Дерево решений gini

Дерево решений, соответствующее модели с использованием критерия разбиения entropy, приведено на рисунке 2.

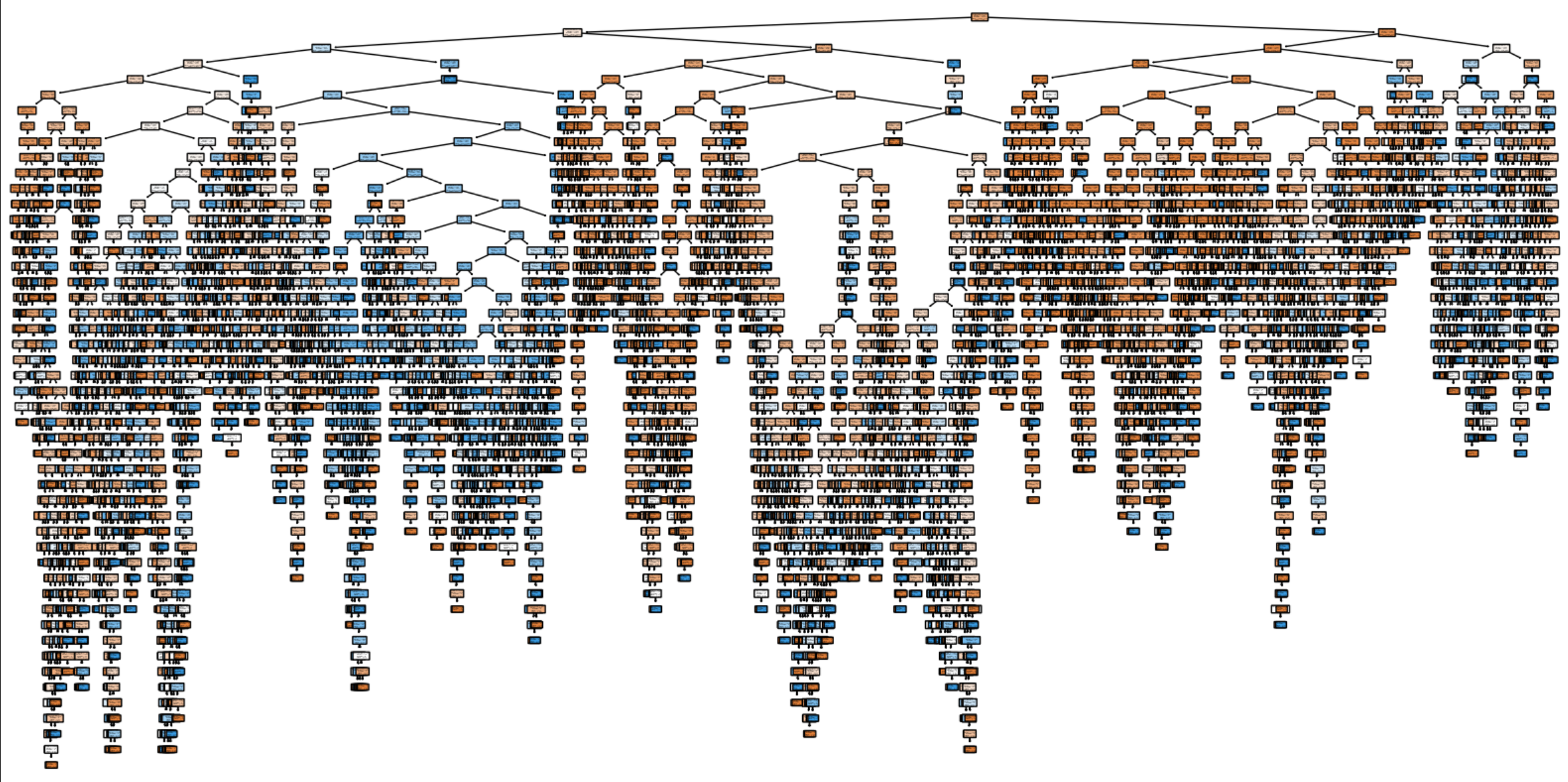


Рисунок 2 – Дерево решений entropy

Показатели качества классификации в зависимости от соотношения мощностей обучающей и тестовой выборок для заданного набора данных для критерия gini представлены на рисунке 3.

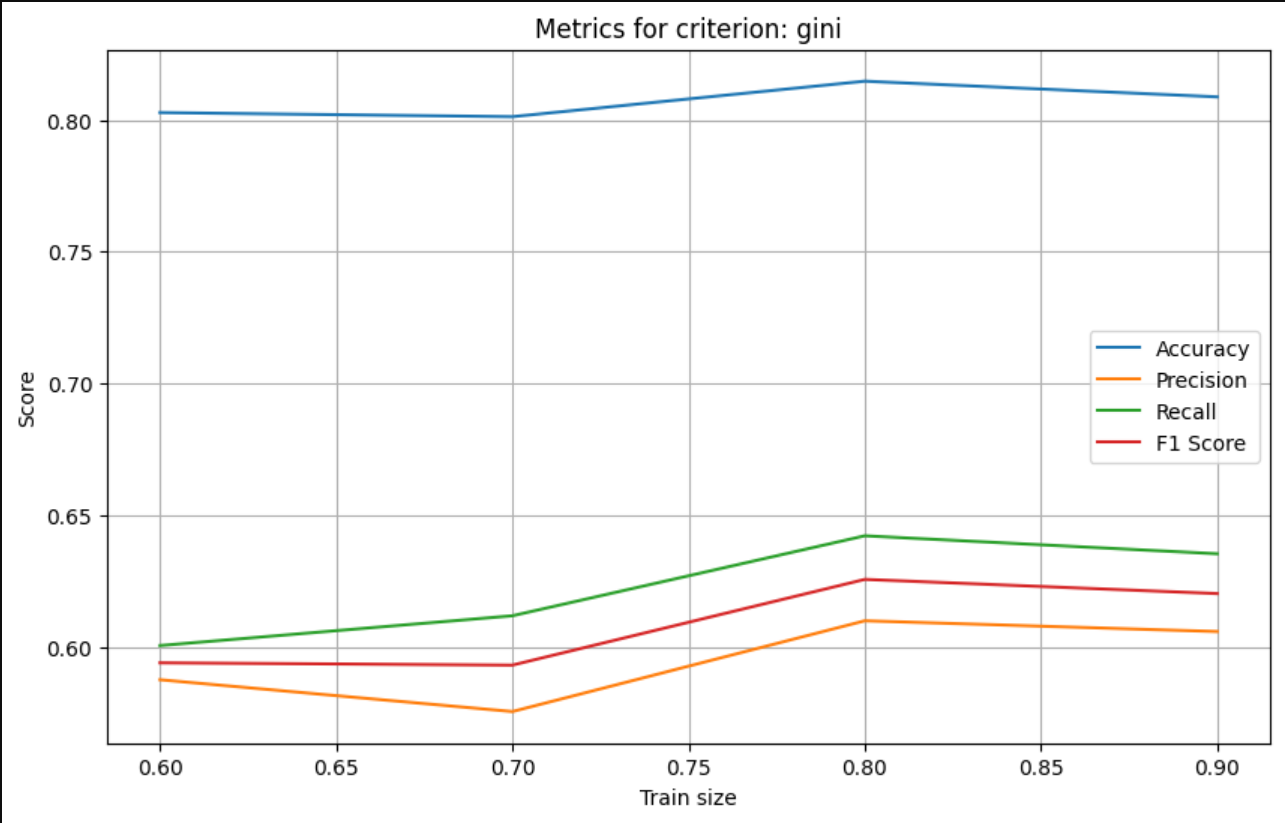


Рисунок 3 – Показатели качества gini

Показатели качества классификации в зависимости от соотношения мощностей обучающей и тестовой выборок для заданного набора данных для критерия entropy представлены на рисунке 4.

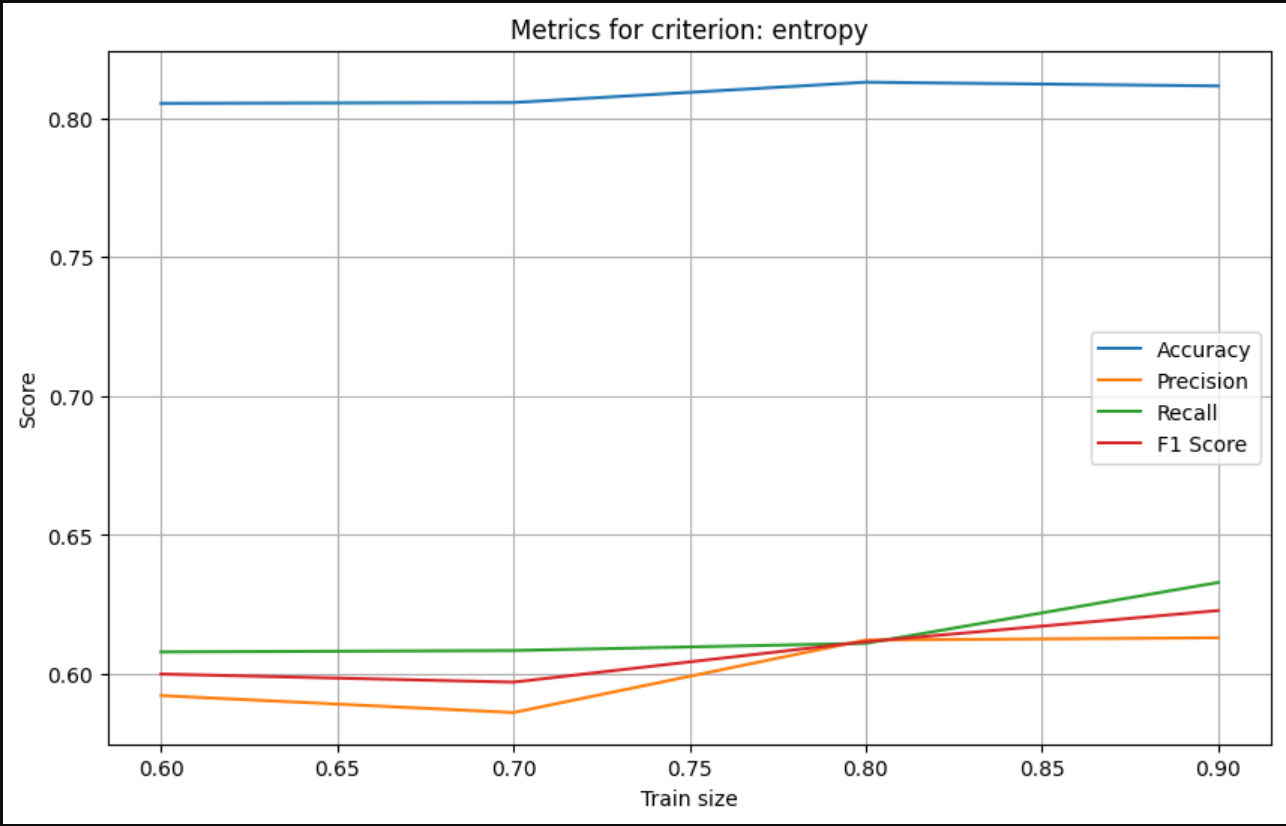


Рисунок 4 – Показатели качества entropy

**ВЫВОДЫ ПО РЕЗУЛЬТАТАМ**

Были получены визуализации построенных деревьев решений для различных критериев разбиения и объёмов обучающей выборки. Все деревья представлены в репозитории проекта.

При увеличении объёма обучающей выборки растут все показатели качества.