



Лабораторная работа №3

по дисциплине: Системы искусственного интеллекта

Вариант: нечётный (15)

Выполнил: Неграш Андрей, Р33301

Преподаватель: Кугаевских Александр Владимирович

Санкт-Петербург, 2022

Задание

- 1) Согласно варианту использовать датасет с данными про оценки студентов инженерного и педагогического факультетов, введя для данного датасета метрику: студент успешный/неуспешный на основании грейда
- 2) Отобрать случайным образом \sqrt{n} признаков
- 3) Реализовать без использования посторонних библиотек построение дерева решений
- 4) Провести оценку реализованного алгоритма с использованием Accuracy, Precision и Recall
- 5) Построить AUC-ROC и AUC-PR

Выполнение

Общее количество признаков в датасете – $n = 32$. Отсюда требуемое количество:
 $\sqrt{n} = \sqrt{32} = 6$

Accuracy: 0.6896551724137931

Precision: 0.8461538461538461

Recall: 0.6111111111111112

Figure 1

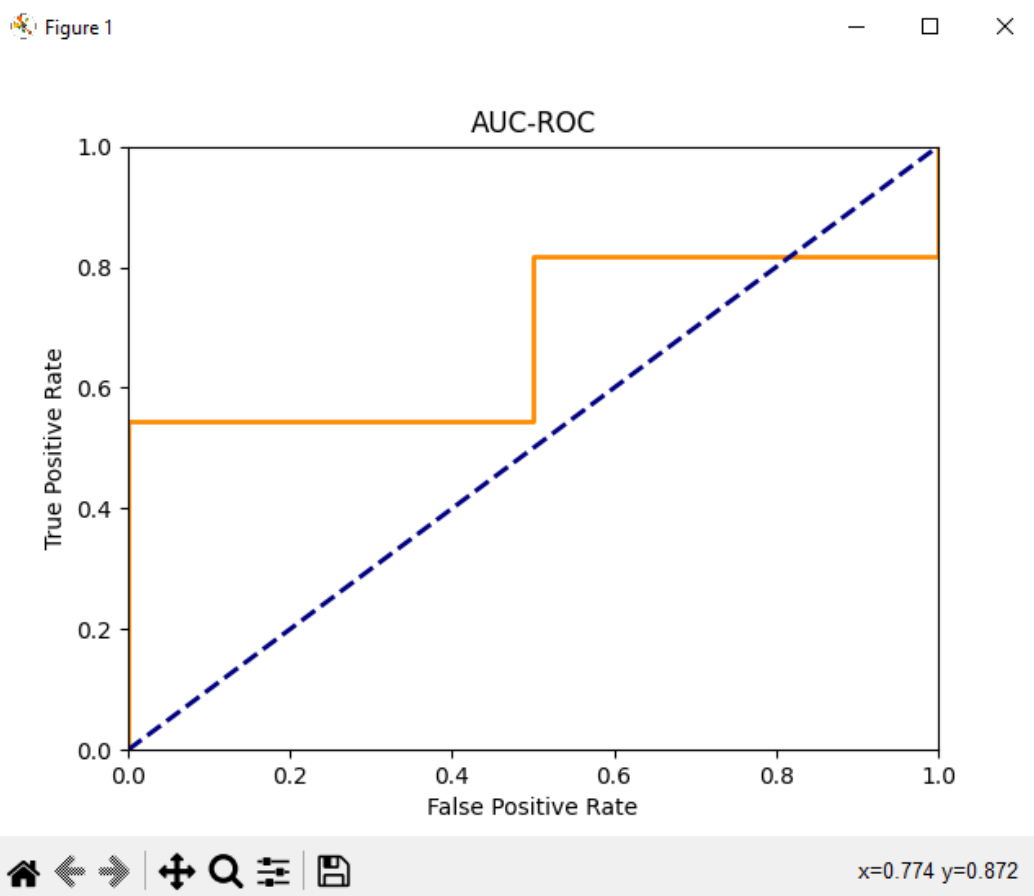
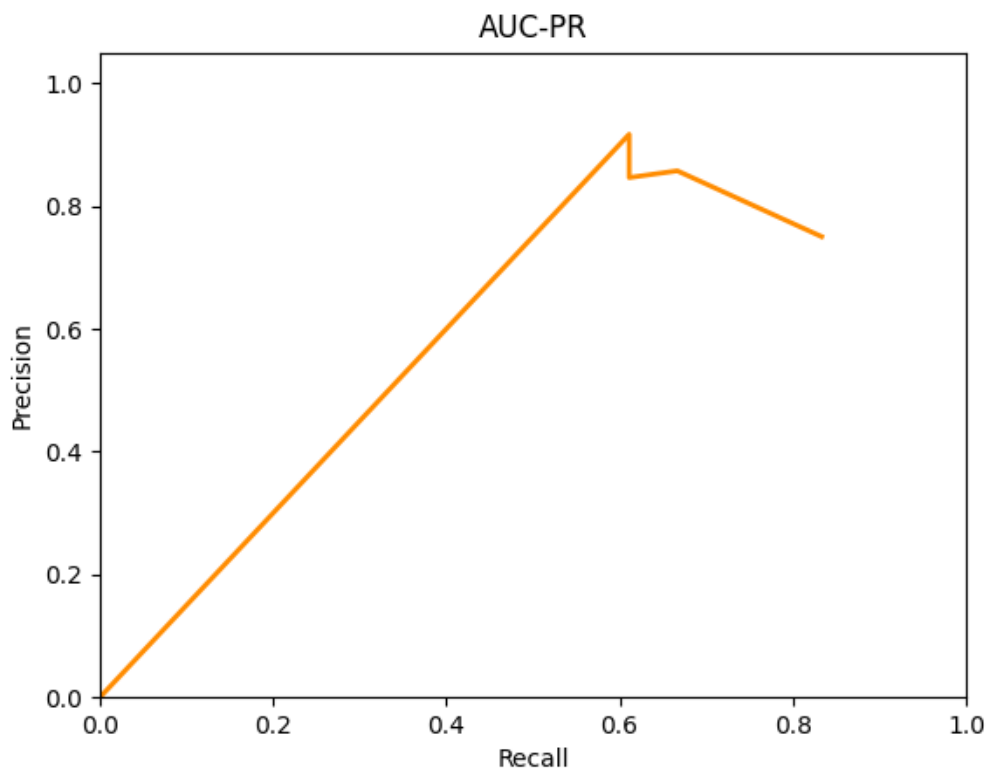


Figure 1



x=0.542 y=1.006

Вывод

В ходе выполнения данной лабораторной работы я познакомился с алгоритмом C4.5 и реализовал его на языке программирования Python. Также вспомнил, что такое Accuracy, Precision, Recall и построение графиков AUC-ROC и AUC-PR.

Данные графики дают недостаточно информации, поскольку исходные данные имеют объём в 150 строк, при тестовом разбиении 2:8. Именно поэтому количество данных, по которым учится алгоритм, а также количество тестовых данных мало, что показывает малое количество точек на графике.