

Университет ИТМО

Факультет программной инженерии и компьютерной техники

Системы искусственного интеллекта

Лабораторная работа №3

Вариант 15

Выполнил:

Ким Даниил Кванхенович

Группа:

P33302

Преподаватель:

Королёва Юлия Александровна

2022

5 семестр

Задание:

0. Для студентов с четным порядковым номером в группе – датасет с классификацией грибов, а нечетным – датасет с данными про оценки студентов инженерного и педагогического факультетов (для данного датасета нужно ввести метрику: студент успешный/неуспешный на основании грейда).
1. Отобрать случайным образом \sqrt{n} признаков
2. Реализовать без использования сторонних библиотек построение дерева решений (numpy и pandas использовать можно)
3. Провести оценку реализованного алгоритма с использованием Accuracy, precision и recall
4. Построить AUC-ROC и AUC-PR

Выполнение:

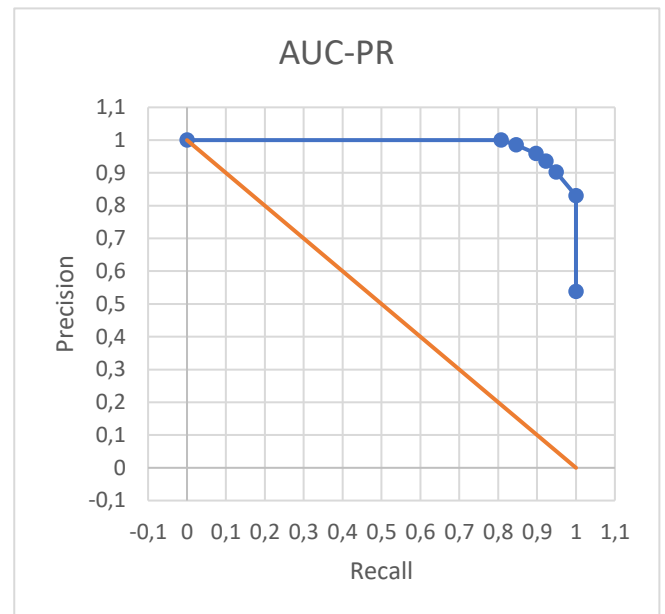
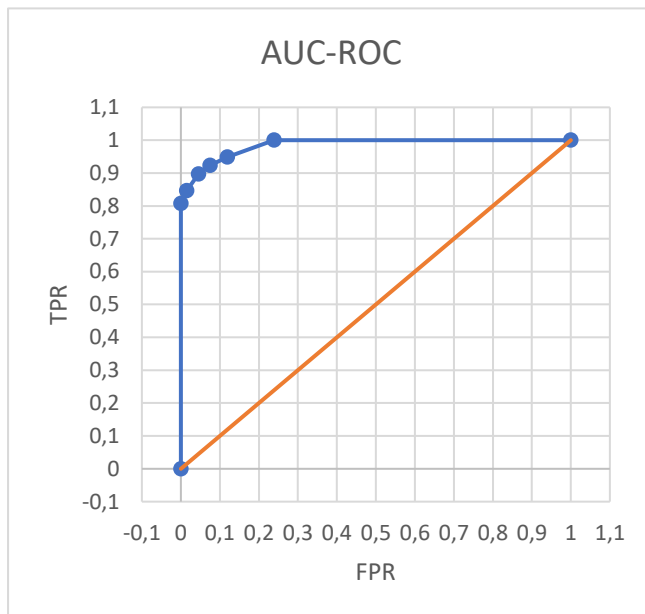
Этап 1. Выборка содержит 145 элементов и 33 атрибута, где атрибут 1 – идентификатор студента, а атрибут 33 – показатель его успеваемости. Принято, что при показателе успеваемости от 3 и выше, студент считается успешным. Количество отбираемых признаков = $\sqrt{33} \sim 6$.

Этап 2. Алгоритм построения дерева на языке Java.
Репозиторий с кодом: github.com/KIMdaniell/AISystems-Lab-3

Этап 3. При выборе атрибутов с индексами 13, 19, 31, 2, 24, 12 метрики Accuracy, Precision, Recall принимают следующие значения.

Accuracy	0,924138
Precision	0,935065
Recall	0,923077

Этап 4. Построение графиков зависимости AUC-ROC и AUC-PR при выборе атрибутов с индексами 13, 19, 31, 2, 24, 12.



Вывод:

В данной лабораторной работе были закреплены знания о методе построения дерева решений, о способах оценки неопределенности таких как энтропия и неопределенность Джинни. Кроме того, были изучены способы оценки качества работы обученной модели с помощью метрик Accuracy, Precision и Recall. На основе данных метрик были построены графики зависимости AUC-ROC и AUC-PR для различных пороговых значений.