## Министерство науки и высшего образования РФ Hациональный исследовательский университет ИТМО

**Факультет Программной инженерии и компьютерных технологий**

**По дисциплине:**

Системы искусственного интеллекта

Лабораторная работа №3.

*«Древо решений»*

Вариант: Нечетный

**Выполнил**: Ву Минь Хиеу

**Группа**: Р33201

Санкт–Петербург 2022 год.

# Задание

* 1. 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).
  2. Отобрать случайным образом sqrt(n) признаков.
  3. Реализовать без использования сторонних библиотек построение дерева решений (numpy и pandas использовать можно).
  4. Провести оценку реализованного алгоритма с использованием Accuracy, precision и recall.
  5. Построить AUC-ROC и AUC-PR.

# Решение

# Полный код можно посмотреть по [ссылке](https://github.com/hieuminhvuu/AI_System_ITMO)

## 1.Реализовать построения дерева решений

Я выбираю порог в 4 балла для оценки результатов обучения студента. То есть, если GRADE меньше или равен 4, он будет оцениваться как «not passed», а более 4 будет оцениваться как «passed».

Text

Description automatically generated

## Отобрать случайным образом sqrt(n) признаков.

Используйте библиотеку *train\_test\_split()*, чтобы разделить набор данных в соотношении 80:20. То есть 80% данных будет тренировочным набором, а остальные 20% — тестовым.

Text

Description automatically generated

## 2.Реализовать построения дерева решений

### \* Код для построения дерева решений

Text

Description automatically generated

Text

Description automatically generated

Text

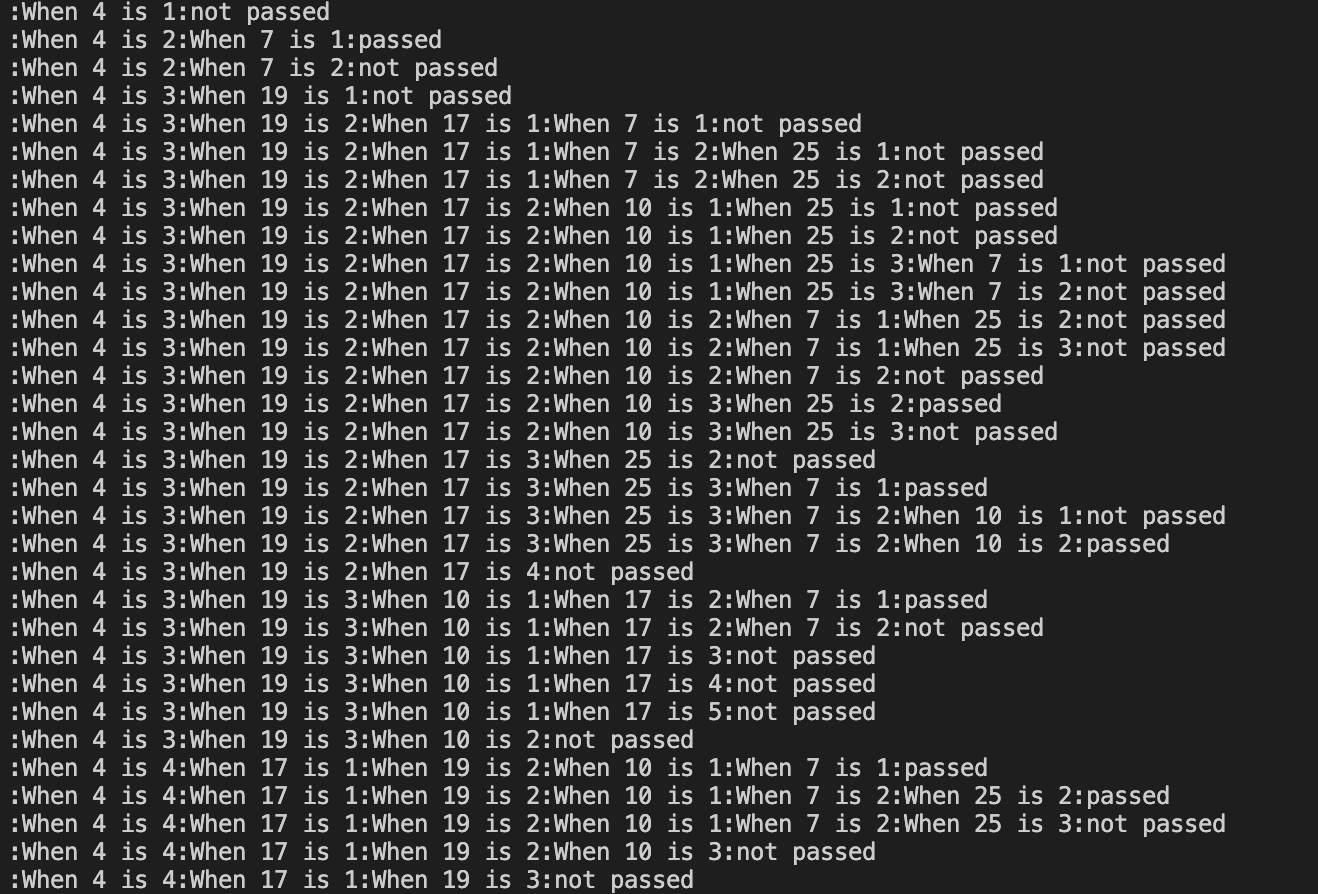
Description automatically generated

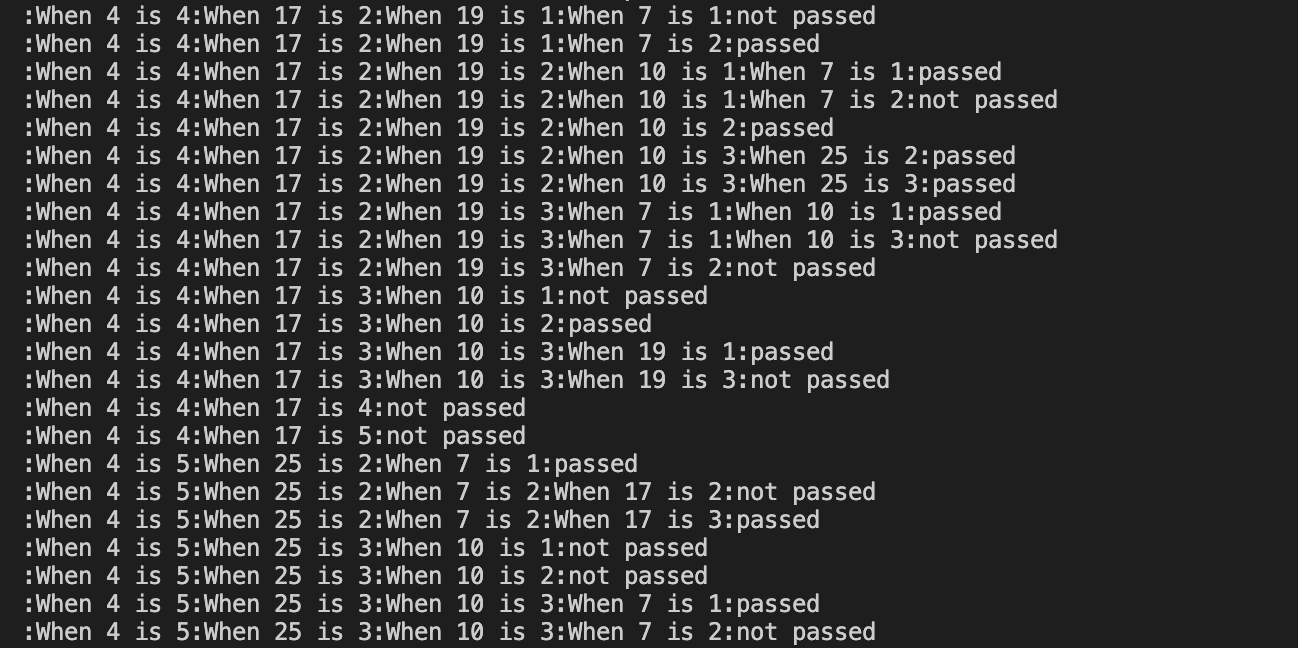


Text

Description automatically generated

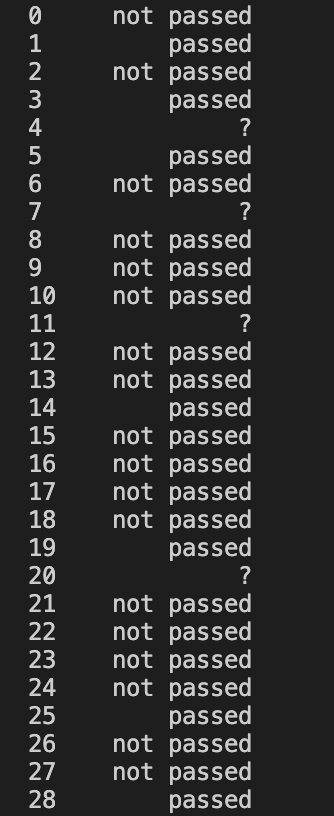
* *Пример дерева решения*

**

**

*\* Пример предсказания результата*

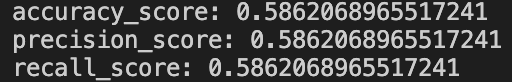
Поскольку мое дерево решений построено на дискретных переменных, будут случаи, когда оно не даст результата. В таких случаях я бы присвоил ему '?'



**3. Провести оценку реализованного алгоритма с использованием Accuracy, precision и recall.**

Text

Description automatically generated



**4.Построить AUC-ROC и AUC-PR.**

Чтобы визуализировать сравнение качества двух моделей дерева решений, я построил два графика AUC-ROC и AUC-PR.

### Код для построения AUC-ROC и AUC-PR

Text

Description automatically generated

AUC-ROC Curve

Chart, line chart

Description automatically generated

AUC-PR

Chart, histogram

Description automatically generated

## 5.Вывод

Построение модели дерева решений для классификации данных по алгоритму C4.5.

Использование характеристических параметров, таких как Accuracy, Precision, Recall для оценки и сравнения качества двух моделей дерева решений.

Построение графики AUC-ROC и AUC-PR.