



**哈尔滨工业大学（深圳）**

**Harbin Institute of Technology, Shenzhen**

## 实验任务书

题 目: 机器学习实验—ID3 及剪枝算法的实现

学 院: 计算机科学与技术

## 实验目的

本实验旨在让学生通过实践掌握 ID3 算法的基本原理和实现方法。通过编写代码实现 ID3 算法，学生将能够理解决策树的构建过程、信息增益的计算以及如何根据决策树进行分类预测。同时帮助学生加深对剪枝算法的理解和应用。

## 实验内容

学生需要使用 Python 编程语言实现 ID3 算法，并基于给定的数据集构建决策树模型、剪枝（可选）及预测。实验内容包括以下几个方面：

### （1）数据集读取和使用

我们提供了两个数据集 `train.csv` 和 `predict.csv`，学生需要正确读取和使用两个数据集进行实验。

### （2）构建决策树

学生需要自行实现 ID3 算法，即利用 `train.csv` 中的数据通过对各个属性 Entropy 和 Gain 的计算，构造我们需要的多叉决策树。

### （3）剪枝（可选）

针对构建完成的决策树，由于树的结构比较复杂，且对训练样本易发生过拟合，故可以通过剪枝算法来删除某些叶结点，以达到防止过拟合的效果。

### （4）预测

使用 `predict.csv` 中的数据对 'weather' 属性进行预测，并输出针对决策树的正确预测结果。

预测结果可以提供运行界面输出截图或存储在 `result.csv` 中提交。

## 实验要求

学生需要将实现的 ID3 算法（及剪枝）和预测代码、实验报告提交给助教或老师进行评审。代码应具有良好的可读性和注释，并能正确运行。报告应陈述他们的代码如何计算信息熵和信息增益以构建决策树并进行分类预测，并提供代码的正确运行结果截图。

备注:

在实验过程中鼓励学生积极探索和尝试不同的优化方法，例如剪枝策略或其他信息增益准则的使用。此外，鼓励学生思考决策树算法的局限性和改进方向，并在实验报告中阐明。学生应独立编写自己的代码和实验报告，不得抄袭他人代码或实验报告。