2023-2024：

True positive真阳性 预测为正，实际也为正

False positive假阳性 预测为正，实际为负 (误报)

True negative真阴性 预测为负，实际也为负

False negative假阴性 预测为负，实际为正 (漏报)

Accuracy = (TP + TN) / (TP + FP + FN + TN)预测对的概率

Recall / Sensitivity召回率 = TP / (TP + FN)实际上是{1}的结果中，预测正确的比例，也称为真阳性率True Positive Rate

Specificity特异性 = TN / (TN + FP)实际上是{0}的结果中，预测正确的比例，Also known as True Negative Rate

Precision = TP / (TP + FP)被预测为{1}的结果中，预测正确的比例known as Positive Predictive Value阳性预测值

False Positive Rate假阳性率 = FP / (FP + TN)被预测为{1}的结果中，预测错误的比例，Also known as Type I Error Rate

False Negative Rate假阴性率 = FN / (FN + TP)被预测为{0}的结果中，预测错误的比例Also known as Type II Error Rate

数据根据精度分为四个等级：Nominal→Ordinal→Interval→Ratio

Nominal名目：

Unordered classes无序类，数据只能被归类categorized

可被编码为数字'dummy variables哑变量'

如：性别，种族，学位项目

Ordinal顺序：

Ordered classes有序类，可以被归类并排名

也可以被编码为numeric variables数值变量

如：年龄组，教育程度

Interval等距：

Numerical (quantitative)数值数据，有相等间隔但是没有绝对0

如：温度，年

Ratio等比：

有相等间隔还有绝对0的Numerical (quantitative)data

如：年龄，体重身高

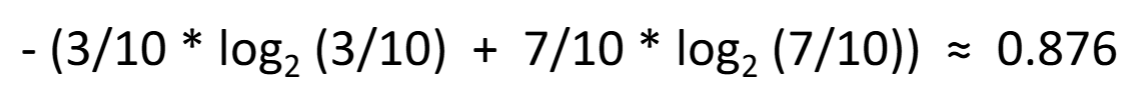
图示

AI 生成的内容可能不正确。

计算：图片包含 图示

AI 生成的内容可能不正确。pi 是类别 i 元素的比例

10 data points in our dataset with 3 belonging to positive class and 7 belonging to negative class:



Information Gain就是决策树中父节点的熵值减去子节点的熵值

信息增益越大的属性越适合作为决策树的下一个节点（因为反映了信息量的贡献）

Apriori algorithm:

先找support for each 1-itemset（所有事务中每个元素的占比）

去掉小于预定support的

得到Frequent 1-itemsets

从Frequent 1-itemsets中生成2-itemsets

再查看support for each 2-itemset

往下直到事务中没有更多组合

置信度公式：文本

AI 生成的内容可能不正确。

算出来后对比是否达到最小置信度要求