决策树：

节点字段选择： 信息熵：描述信息分类的混乱程度。信息量↑，信息熵↑。因此信息熵与信息量常等价。

1.
   - 单变量信息熵：
   $$H(P_1, P_2 \cdots P_k) = -\sum_{i=1}^{k} (P_i) \log P_i$$
   - (经验)信息熵：
   $$H(D) = -\sum_{i=1}^{k} \left(\frac{|C_i|}{|D|}\right) \cdot \log \frac{|C_i|}{|D|}$$
   $\frac{|C_i|}{|D|}$ ⇒ 样本量；$|D|$ 样本总量

2. 条件信息熵：
$$H(D|A) = \sum_{i=1}^{k} \left(\frac{D_i}{D} \cdot \sum_{k=1}^{k} \left(\frac{D_{ik}}{D_i} \cdot \log \frac{D_{ik}}{D_i}\right)\right)$$

3. 信息增益指标：描述在某个条件下信息熵下降程度
$$Gain_A(D) = H(D) - H(D|A)$$

事件A对事件D影响越大，条件熵越小，信息增益 $Gain_A(D)$ 越大，在事件A影响下，事件D被划分得越"纯净"

△ 因此根/中间 结点变量选择时，即挑出各自变量(A)下因变量(D)信息增益最大的自变量。
(即选对因变量影响最大的)

此为 ID3 算法.

信息增益率：$Gain\_Ratio_A(D) = \frac{Gain_A(D)}{H_A} = \frac{H(D) - H(D|A)}{H_A}$

引入原因：克服信息增益指标缺点，即事件A取值越多，$H(D|A)$越小导致的$Gain_A(D)$越大。H_A相当于惩罚。

此为 C4.5算法.

4. 基尼指数：CART算法（分类回归树算法）字段选择指标.

$$Gini(p_1, p_2, \cdots p_k) = \sum_{k=1}^{k} p_k(1-p_k) = \sum_{k=1}^{k}(p_k - p_k^2)$$

只可能有两种情况.

$$= 1 - \sum_{k=1}^{k} p_k^2$$

$$Gini(D) = 1 - \sum_{k=1}^{k}\left(\frac{|C_k|}{|D|}\right)^2$$

引入目的：C4.5算法与ID3算法均只能对离散型因变量进行分类，为了能让决策树预测连续型因变量，引入CART算法.

基尼增益率：

$$\Delta Gini = Gini(D) - Gini_A(D)$$

表基尼指数下降的快慢，$\Delta Gini$ 越大下降的越快，

事件A对D的影响也越大。

（与信息增益类似，事件A对D的影响越大，则 $Gini_A(D)$ 也越小。 $Gini_A D = \sum_{s=1}^{k} P(A^{(s)})\left(1 - \sum_{i=1}^{k} P_{si}^2\right)$ ）

随机森林：→ k棵CART决策树构成森林，

↓ 利用 Bootstrap 抽样法，从原始数据集 $X_{N\times p}$ 中

生成k个数据集，每个数据集中也有 N个观测与 P个自变量，

对每个数据集构造 CART 决策树。

在构建过程中，没有将所有自变量作为节点字段选择，而是随机选择了个字段。

△ 在构建决策树时，① 使得树中每个节点尽可能纯净，

② 每棵决策树尽可能充分生长。

不对其限制（剪枝）

△ 随机森林中 — 对分类问题 利用投票法。

分类决策树

对回归问题 利用均值法

（回归决策树）

将该类别作为最终结果。