# 树相关代码需求

开发目的：

1. 获得最佳树模型及最佳树的参数
2. 对于目标变量为分类型的数据集，通过树做数据的补缺。

开发内容：

1. 模块（文件）名：

ym\_tree.py

1. 对外接口：
   1. 易名树模型类
      1. 类名：
         1. **YMTree(data: pd.DataFrame, method = “both”, n\_jobs = 4)**

**Parameter:**

**Data**: pd.DataFrame, 训练树的数据

**method**：str类型，默认值为both 包含下面5个值。

‘none’: sklearn 默认设置  
‘cal’：通过样本量换算相关参数  
‘tune’：暴力搜索获取树的最佳深度  
‘prune’：通过后剪枝算法获取最优参数  
‘both’：先通过暴力搜索获取最佳深度后再后剪枝

**n\_jobs**: int 默认为4

**Attributes:**

**record：**记录对象。在每个叶子节点的数据集上，每个字段的平均值。如果该字段全部为空，则计算该叶子节点父节点上该字段的平均值，以此类推。结构为字典key为节点序号，从左至右，从1开始。每个key对应的value为一个pd.Series, 2行n列。第一行为data中的head=0，即标题行， 第二行为每列的平均值。第二行不能有空值.

**parameters:** **dic**

类方法：

**impute(data: pd.DataFrame)**

**Parameter:**

**data**: 需要补缺的数据

**return：**

pd.DataFrame返回补缺好的数据集

**get\_features()**

**return**: 列表类型list， 特征提取后的重要变量