Ps：Drop掉c5和c7就不说在ppt里了吧，感觉就是没考虑到这不合适orz

1. **Replacement invalid value**

在完成特征工程后获得的数据存在大量的缺失值（invalid value），这是在python中计算时遇到的缺失变量导致的，所以我们针对每个指标的含义对数据进行了填充。对于本身有含义的数据，我们填充0；对于无含义的数据，我们选择-1进行填充。

1. **初试模型**

将处理好的数据放入经典决策树模型，能发现training and validation accuracy都很高，首先推测模型出现了过拟合的现象。再通过混淆矩阵发现target=-1和0的数据在模型中表现较好，而target=1的表现很差，计算几个measure的指标也发现target=-1的 f-measure很低，只有0.05，此时意识到数据不平衡。

1. **使用不平衡算法Adasyn(Adaptive Synthetic Sampling Approach for Imbalanced Learning)**

采取四个不平衡算法，根据表现结果选择ADASYN不平衡算法：

计算accuracy发现，training accuracy高的validation accuracy低，意味着模型在这样的数据集会出现过拟合现象。为了避免后续模型出现这样的情况，我们选择adasyn平衡后的数据集。