这次竞赛进行了多次尝试。

1. 首先，我观察数据发现训练集和测试集数据中有问号，但是其中有四列是没有有问号的，所以我第一次尝试就是，只提取其中没有问号的四列进行决策树训练、测试、验证，得到正确率是30%。
2. 第二次，我通过查阅资料将问号先通过data.replace(‘?’,np.nan,inplace=True),将问号改为NAN，然后在对NAN进行缺失值填充。
3. 我第一次对缺失值填充是该列的均值
4. 第二次对缺失值的填充是该列的众数
5. 然后我划分训练集为训练集和测试集，分别进行决策树，KNN，神经网络各种试验，生成CSV格式，提交。
6. 因为第一次是直接对原始数据进行训练，没有对数据进行更改，所以提交之后成功，但是后几次是先修改了数据，在提交，所以数据格式不正确，所以每次提交正确率都是0。然后我把训练集的预测值都改为了字符串类型。
7. 通过在kaggle网站观察数据，发现其中第6,9,11,12,13列数据摸一个数值出现频率极大，和其他数据有明显差异，数据不适合使用。
8. model = SelectFromModel(clf, prefit=True)

X\_new = model.transform(X)

通过代码测试，发现各列数据的重要程度是：[0.03951554 0.06141498 0.13340667 0.11424999 0.13348698 0.08063093 0.0264212 0.11025933 0.06537234 0.05404779 0.06530974 0.09048582 0.02539869]

其中第1,2,6,7,9,10,11,12,13列重要程度低。