1、数据来源

训练集：train.csv

待预测集：test.csv

2、对数据的处理

1、因为数据中有部分的数据缺失，所以我进行了如下的尝试处理来解决数据缺失的问题。（1）直接把训练集中有缺失的数据删除，测试集上的缺失数据用所在列的众数来填充。（2）训练集上用所在列的众数填充，测试集上用训练集的数据的众数来填充。

2、提取出数据中与标签相关性强的属性来进行学习。提取的方式有PCA和人工寻找相关性。