数据分析：

1总览

ID:  
•Survived(存活与否)  
•Pclass（客舱等级，较为重要）  
•Name（姓名，可提取出更多信息）  
•Sex（性别，较为重要）  
•Age（年龄，较为重要）  
•Parch（直系亲友）  
•SibSp（旁系）  
•Ticket（票编号）  
•Fare（票价）  
•Cabin（客舱编号）  
•Embarked（上船的港口编号）

#### 2使用统计学与绘图对数据初步分析

主要分析各类乘客存活率情况

（大致分为10类左右）

#### 3数据清洗

填充缺失值（构建随机森林模型）

按姓氏等进行划分，之后同组之间进行识别，查看存活率情况

之后按存活率情况进行分组，挑出反常的一组

选取特征，进行特征转换，划分训练集与测试集

#### 4建模和优化

网络搜索选最优参数，之后训练模型，最后交叉验证

#### 5进行预测