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）

查看前几项 查看是否存在null值

#### 2使用统计学与绘图对数据初步分析

1. 乘客等级更高 获救的几率越大
2. 女生获救概率更高
3. 有旁系亲友存活率高
4. 有一个兄弟姐妹的一般有更高的生存率
5. 婴儿的生存率奇高
6. 有cabin存活率高
7. 途径港口上船的人生存几率较大

#### 3数据清洗

1. 查看缺失值
2. 对cabin和ticket进行处理。
3. 对embarked进行处理，选取最多的值进行填充。
4. 对title进行处理，数字化
5. Age 选择平均值作为缺失项的填充值
6. sex embarked数字化
7. Fare 均值填充 映射fare值 丢弃Null

#### 4建模和优化

比较几个模型的预测性能，选取性能最好的进行预测。

网格搜索进行调参。

#### 5进行预测

使用调参好的模型进行预测，将预测值放入submission.csv文件中。