一 Project内容

1：用MapReduce算法实现贝叶斯分类器的训练过程，并输出训练模型；

2：用输出的模型对测试集文档进行分类测试。测试过程可基于单机Java程序，也可以是MapReduce程序。输出每个测试文档的分类结果；

3：利用测试文档的真实类别，计算分类模型的Precision，Recall和F1值。

二 贝叶斯分类器理论介绍

三 贝叶斯分类器训练的MapReduce算法设计

解释说明所设计的算法的执行过程，要求给出MapReduce的Data Flow示意图，并说明作业输入文件格式是是什么，Map的Key是什么含义，Value是什么含义，Reduce的输出Key，Value是什么含义

四：源代码清单

要求对Map，Reduce方法的关键代码进行解释；源代码关键地方要有注释

五：数据集说明

用了多少个文档，多少个类别，每个类别的文档个数；

六：程序运行说明

要求说明运行时有多少个Map任务，多少个Reduce任务，并说明为什么会有这么多Map任务；要求给出程序运行时基于Web页面的作业监控截图和程序运行截图

七：实验结果分析

计算分类结果的Precision，Recall和F1值。