**深 圳 大 学 实 验 报 告**

**课程名称：­ 概率论与数理统计**

**实验项目名称：Bayes Classifier in Natural Language Processing**

**学院： 电子与信息工程学院**

**专业： 电子信息工程**

**指导教师： 陈昌盛**

**报告人： 陈闻天 学号： 2023280259**

**班级： 04**

**实验时间： 2024年11月4日**

**实验报告提交时间： 2024年11月25日**

**教务处制**

|  |
| --- |
| Aim of Experiment:   1. Familiar with the Bayes theorem. 2. Understand the implementation of the Bayes theorem in python. 3. Know how to use Naive Bayes for a practical task, e.g., text classification. |
| Experiment Content:   1. Read the background information about Bays Theorem and Navie Bayes. 2. Naïve Bayes for Text Classification in python |
| Experiment Process：   1. Download the Enron email dataset and import it into the environment of experiment 2. Verify whether the dataset is loaded successfully without loss of messages 3. Define a SpamDetector\_1 class for the process of messages 4. Compute the log prior probability of spam/ham, using 和 5. Verify whether the log prior probability of spam/ham is computed correctly 6. Compute the conditional probability of spam/ham with Laplace smoothing 7. Calculate the sum of the conditional probabilities of spam/ham 8. Calculate the score of the prediction of spam/ham 9. Compare the score and determine whether the message is spam 10. Calculate the accuracy of prediction |
| Data Logging and Processing:   1. the result which indicates the dataset is loaded successfully      1. the result of the log prior probability of spam\ham      1. the result of log conditional probability of words given spam/ham |
| Experimental Results and Analysis:  The final outcome is below:    No error and correct results indicate that classification for spam and ham messages is successful.  The analysis of this experiment is Bayes and Navie Bayes are useful and powerful for prediction, which actually are used at the task of classification for text or image. |
| 指导教师批阅意见：  成绩评定：  指导教师签字：  年 月 日 |
| 备注： |

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后10日内。