**《Python与数据科学实验》期末大作业报告**

**1. 期末大作业报告任务**

请同学们自主选择实验6、实验7、实验8和实验9中的任意两个实验，撰写一份期末实验报告。四个实验（实验6-9）的主题可以归纳如下：

实验6 重大疾病分类实验

实验7 蛋白质功能分类实验

实验8 新闻文本分类实验

实验9 表情识别多分类实验

**2. 报告章节架构**

期末大作业(实验报告)撰写的组织结构与章节安排参考如下：

**第1章 数据挖掘与分类预测**

简要综述数据科学与数据挖掘等领域的发展现状与趋势，概述数据挖掘中的分类预测任务与目标，及其应用价值。

**第2章 分类模型与数据挖掘库sklearn**

阐述实验所涉及分类模型的理论基础，并简要介绍数据挖掘的Python库sklearn。注：与所选实验不相关的算法无需介绍。

**第3章 文本分类实验**

该章标题可根据选择的具体实验而更改；本章内容可包括实验数据、实验方法的介绍，并较为详细地阐述实验设计与步骤（附关键步骤的Python源代码截图），以及实验结果与分析。

**第4章 蛋白质功能分类实验**

该章标题可根据选择的具体实验而更改；本章内容可包括实验数据、实验方法的介绍，并详细阐述实验设计与步骤（附关键步骤的Python源代码截图），以及实验结果与分析。

**第5章 分类预测实验总结**

该章内容可包括实验工作及其创新点总结、数据挖掘经验总结、未来学习计划与数据科学实践展望。

**3. 报告撰写与提交**

(1) 请依据以上章节安排以及模板文件“Python与数据科学实验\_期末大作业报告.docx”撰写期末大作业报告。正文撰写之前，在报告的封面(首页)处，除“成绩”一栏留白之外，请完整填写各自有关信息。

(2) 页数要求：小四字，1.5倍行距，一般情况下20页以上。

(5) 一般要求在《Python与数据科学》理论课程结束之前完成报告撰写，具体截止时间由任课老师自行指定。最后，将撰写完成的报告“Python与数据科学实验\_期末大作业报告.docx”以附件形式提交至学习通。注：报告的文件名无需修改。

**4. 期末考核说明**

此次大作业报告的成绩等同于《Python与数据科学实验》课程的期末测试成绩。本课程的期末总成绩构成如下：

期末总成绩=期末测试成绩(大作业报告成绩)\*60% +平时成绩\*40%

请同学们务必认真对待本次期末大作业报告，千万不可出现文字雷同、互相抄袭现象。一旦发现，将认定两份报告均为抄袭者，后果自负。

最后，针对大作业报告的评分考核，我们认为，**一份优秀的报告**需具备的几大要素如下：

**(1)** 清晰介绍所用算法的基本原理。针对采用算法的sklearn相关函数要进行一定的功能说明。

**(2)** 给出解决问题的基本思路或技术路线，并配以简洁的流程图。

**(3)** 清晰给出分类器的评价方法及实施方案。即使没有测试集的正确答案，也要根据训练集给出分类器模型的评价方案与实施过程。如果采用的模型存在超参数，还需合理说明参数调优策略。

**(4)** 需针对实验结果进行一定的分析。每个实验中，均需包含至少1个基准模型，以体现自己所用模型的优越性。

如你分别尝试过GNB和KNN两种分类器，发现GNB效果更好而采用GNB为最终的算法，那么也请同时报道KNN的验证结果，并进行分析讨论或经验总结。

与最终采用算法进行基准比较的方法模型越多越好，由此可进行结果分析的内容也会越多。

**(5)** 需在报告正文中适当位置，截图粘贴采用算法的Python代码，并展示运行结果。