**2021-2022-2数据分析应用实践期末试题**

要求：

给定一份原始数据customdata.xls，（数据的最后一列是预测值，其余为输入值）用Python编程语言编写程序进行数据分析，其中前1988条数据为训练数据，最后5条数据作为预测数据，调用合适的学习算法进行数据预测，并输出训练模型的混淆矩阵及分类模型的各评价得分。实验报告需要对数据做简单说明，解释整个数据挖掘过程，所使用的方法以及关键的步骤。用文字**解释关键代码**，但不需要把代码拷贝至文档中，报告中可以有少量运行结果的截图。Python源代码中需要至少50%的关键语句注释，并具有可读性。文件需按规定命名：“张三\_学号”，word报告需添加给定封面，请勿修改封面格式！！！纸质报告除封面外双面打印并装订，由班委收齐交给我，python源程序以及电子版的报告提交到网络教学平台的期末作业中，提交时间截至时间为6月26日24点整。

评分标准：

1、报告格式符合要求，无错别字，排版合理，5分

2、数据理解正确，10分

3、数据预处理方法使用正确，15分

4、数据基本统计分析正确、完整，10分

5、模型使用正确，20分

6、使用多种模型、合理调参，15分

7、程序输出正确，美观，10分

8、代码简洁，正确，15分