ApplySquare笔试反馈表

职位：数据部数据分析师

|  |  |  |  |
| --- | --- | --- | --- |
| **姓名** |  | **学校** |  |
| **专业** |  | **目前年级或工作经验（年）** |  |
| **如被录用何时可以开始工作** |  | **期望薪资** |  |
| **职位相关问题** | | | |
| 1. 你是如何理解大数据的？ | | | |
|  | | | |
| 1. 请介绍一个你做过的数据分析的项目的具体经历。(你可以把自认为做过比较有代表性的项目相关工作附件发过来。) | | | |
|  | | | |
| 1. 请评价一下你的学习能力。 | | | |
|  | | | |
| 1. 请描述一下你理解的数据分析的过程，从数据收集开始，到最后的结果展示，中间需要处理的步骤以及需要注意的问题等。 | | | |
|  | | | |
| 1. 请完成下面的测试题目，答案可以另写文档提供。(请用你熟悉的数据处理语言，推荐python) | | | |
| 1. 数据集   spam.csv  数据集介绍：  该数据集名由Hewlett-Packard实验室收集，总共4601行，58列。每行代表1封电子邮件，前57列变量分别计数电子邮件中含有某个单词或字符的频数。58列type记录该封电子邮件是否为垃圾邮件。  更多解释请参见以下网址：  <http://artax.karlin.mff.cuni.cz/r-help/library/kernlab/html/spam.html>   1. 数据描述   把数据投射在一个二维平面，使得在这个平面上数据尽可能分散，请尽可能画出便于读者阅读的二维图。  小提示：主成分分析   1. 模型搭建   我们的目的是有效地区分一封新邮件是否为垃圾邮件。你要做的是选择一个模型，使得这个模型能准确推断一封新邮件是否为垃圾邮件，请写明模型选择过程。  小提示：   1. 把数据集拆分成训练集（training dataset）和测试集（test dataset）； 2. （如果必要的话，将列数据降维）尝试各种机器学习算法； 3. 得到错误率最低的算法模型，标明每个模型的成功率。 4. 结果输出   上传spam\_or\_not.doc文件，注明模型选择的过程及用测试集测试的成功率，简述你所尝试的各类模型的原理，并附上各步骤的代码以及必要的说明。 | | | |