

第十届中国大学生服务外包创新创业大赛企业命题 类赛题相关问题解答 F&Q (11.0)

(截止至 3 月 14 日)

一、【A01】2018 网络零售平台商品分类【浪潮】

1. 准确率是在 ppt 或者文档中提交还是提交模型后由项目方进行测试?

答：由学生团队在作品中提交。

2. 1) 企业给出的 450 万测试集的质量如何, 大致准确有多少? 关乎我们对于样本中明显错误的操作。 2) 如不方便透露数据的大致准确率, 是否回答如下情况, 如果我们评测是正确的, 但是企业是错误的情况, 是算我们对还是企业对。 3) 企业是否提供测试的机会? 是否可以通过邮箱等方式给我们的答案一个目前的准确度。

答：1) 450 万测试集商品标签均来自于采集平台网站, 准确率相对可以保证 (2) 默认采集到的标签正确 (3) 后期看具体情况。

3. 请问如果给 50W 训练集打标签, 是在原文件 type 列旁打还是去掉原来的文件有的标签重新打?

答：在原文件 type 列旁打。

4. 批量输入是指, 项目通过接收企业上传的文件 (含上万条数据), 计算后在网页上呈现所有结果嘛? 批量输出是指, 企业可以在项目的文本框中输入多条测试用例, 计算后呈现对应的结果在网页上吗?

答：对, 可视化尽量能够支持批量输入, 输出。

5. 50W 数据集中存在部分类别分类明显错误的情况, 我们是否可以重新更改训练集的标签, 我们改正这些商品条目标签, 被修改标签商品条目成绩判定时, 是否会判定错误。

答：测试集作为评判成绩的标准。

6. 有些不在训练集中的类别, 是否可以增加? 需要提交分类程序么?

答：我们给定的训练集的标签就是测试集中所有的标签; 需要提交程序, 用于测试。

7. 为什么需要对 50w 训练集打标签, 如何保证各队伍结果的真实性?

答：可以使用抽样方法, 抽取一批样本, 按照队伍分类模型测试结果和给定的结果的误差。

二、【A03】大规模资金流入流出的大数据预测【阿里创新创业中心】

1. 作品最后怎么验证结果？

答：企业要求学生通过阿里云官网系统

(<https://tianchi.aliyun.com/competition/entrance/231573/information>)来测试结果, 并将结果、阿里云账号名称等信息与其他作品材料提交至组委会官网。具体操作说明请参加各团队邮件（如未收到，请先在垃圾邮件中查询）。

三、【A04】语音人机交互技术带来的家庭娱乐体验升级【阿里人工智能实验室】

1. 请问技能上线是指 3 月 27 日（作品提交截止日）之前必须完成吗，技能上线审核这边需要 4-10 个工作日。

答：是的，需要上线，上线是最低标准。提醒团队最好提前上线，因为可能会有各种需要下线再优化的问题。

四、【A07】智能化跨境商品交易平台【东软睿道】

1. 在将测试版微信小程序转为正式版时需要选择类型，但是“电商平台”只对企业开放，供个人选择类型只有“出行与交通”、“生活服务”、“餐饮”、“旅游”、“商业服务”、“快递业与邮政”、“教育”、“工具”、“体育”，请问应该选择什么类型进行转正？

答：生活服务。

2. 微服务架构体系：指的是后端必须基于 spring cloud 架构吗？

答：不限技术。

五、【A08】智能会议室管理系统【虹软】

1. 打包 IOS 端 APP 用开发证书还是发布证书？如用开发证书，官方会提供苹果设备的 UDID 吗？

答：因每个方案可能不同，如需要 UDID 请自行解决。

六、【A10】运用数据分析和可视化技术创造完美的数据大屏交互体验【融创软通】

1. 因为在题目中要求有实时状态的展示，这似乎更贴近于流计算，这样的话数据源源不断产生即可，如果数据量限定范围是不是又可以理解为批处理？小组的作品更加贴近于流计算，到时候如何上交初始数据，因为小组的数据一直都是靠程序模拟动态生成的。

答：项目要有获取实时数据的能力，如果演示条件不允许的话，可以考虑模拟实现实时数据。

数据可视化页面要有动态数据显示效果。是使用流计算，批处理，使用爬虫或第三方数据源都可以。

七、【A11】基于微信的移动智能学习平台【文思海辉】

1. 请问任务清单中的详细设计——功能设计是要求侧重于描述产品所有功能还是侧重于编程技术上的说明呢？

答：偏编程技术。

八、【A13】工业技术类从业者与领域专家的交流平台【海尔】

1. 有关 app 展示的数据，交流平台的提问内容是由出题人提供 api 还是由参赛者自行编辑呢？

答：参赛团队可自行编辑。

九、【A14】运用文本相似度实现（证券）智能客服【恒生电子】

1. 程序应该输出对应的标准问还是输出什么？

答：标准问。

2. 赛题成果怎样验证？

答：鉴于各团队使用的开发语言和工程结构各不相同，难以统一用测试集对各团队的模型进行验证，同时也为了更好地保护各参赛团队代码的知识产权，现决定将测试集公布给各参赛团队。各团队自行对自己的模型进行验证，然后提交结果。具体安排如下：

1) 请各参赛团队于 3 月 22 日下午 18:00 前将现有模型的测试结果（P、R 和 F）提交到组委会邮箱（fwwbds@niso.edu.cn），测试结果文件中需注明：团队编号、团队名称，逾期提交测试结果的，拒不采用。

2) 组委会将于 3 月 23 日上午 10:00 将测试数据集 test.xlsx 发送到各团队邮箱；

3) 参赛队伍需在不更改模型的条件用测试数据集对模型进行验证，并将测试结果联合整个作品于 3 月 27 日下午 18:00 前发送到组委会邮箱 fwwbds@niso.edu.cn 备份，测试结果文件中需注明：团队编号、团队名称。

初审阶段将针对部分团队进行抽查。企业将数据集分成 1:1 两部分，一部分于 3 月 23 日发放给团队进行测试提交结果，另一部分数据集，用于抽查团队。抽查时，组委会将电话通知部分团队通过邮件限时提交源代码等材料，转交企业专家抽查。请各团队队长保持手机通畅。

十、【A16】AI 数据采集平台【虹软】

1. 出于大数据高可用和高并发的考虑，使用了 redis, mycat, hadoop, hive 等第三方工具。这些应该是可以使用的吧？

答：可以。