

V5.0 A 类赛题答疑（截止至 20201003，标黄为本周更新）

【A01】基于绿色出行场景个人碳排放积分系统开发与运营【八维通】

1、碳积分运行系统。我们是单独做一个 app 或者微信小程序，还是在八维通的 APP 里面内嵌我们的系统模块呢？

答：内嵌的，可以考虑 h5 或者 sdk 的方式

2、在所需数据接口中共有六个接口，请问支付接口中的微信和支付宝代扣需要自己实现吗？

答：不用

3、在所需数据接口中共有六个接口，请问支付接口中的微信和支付宝代扣需要自己实现吗？

答：不用

4、赛题 demo 统一 api 接口访问所需的 appid 和账户有提供么？

答：api 地址已在群文件中公布，为了方便开发，没有设置验证，http post 请求调用即可，接口返回的也都是模拟数据。在模拟环境 API 准备时，为了简化工作，去掉了接口验证要求。现在所有参数都不用带，直接调用接口即可。

5、关于碳积分的定义，题目中的碳积分的定义指二氧化碳排放配额，企业或个人可通过购买碳积分消除碳足迹，是一种惩罚措施。而本题的碳积分是鼓励绿色出行的，绿色出行的效果减少碳排放，是不是可以定义为碳积分的量化是绿色出行带来的减碳量，而不是一般的碳排放量来量化的。

答：是的，通过绿色出行带来的碳的减排量获得奖励（政府或公益组织提供）碳积分，进而获得商业价值。可以通过碳积分交易，卖给有碳排放需求的用户变现其价值。

6、是否会提供商家入驻、支付等相关接口？

答：不提供。可以自己做一些模拟数据和操作，不用实际实现。

7、在 api 接口地址中，能查到行程信息，但是没有 user_id 和 user_name 等用户信息，返回 message: "Required String parameter 'user_id' is not present", path: "/v1.0/api/app/user/getUserInfo", 请问用户信息表是本来

就没有吗？

答：userid 值取 1，在请求头中赋值。见文档更新。

8、在提供的文档，“验签过程”中提出我们应该提交 json 格式的数据，但是在给出的接口里面只能提交明文数据。如果提交明文的话，服务器端收到数据后，要按照什么格式转换为 JSON 进行签名校验，然后 null 字段要不要计算？各个字段按照什么顺序排列？是否需要换行？

答：已取消验签环节。见群文件接口文档更新。

9、如果是开发 sdk 的话，演示的时候是不是可以做个简易的 app 入口来仿八维通，在这个入口然后调用 sdk，传入相关数据进行演示，实现嵌入 app 效果？

答：可以，请自行设计，不做要求。

【A02】基于地铁出行场景的小游戏设计【八维通】

1、请问这个游戏最终是嵌入现有平台使用呢？还是单独做一个游戏？，如果是嵌入现有平台使用，游戏大小以及风格有没限制？

答：基于接口做独立的 h5 应用。

2、最后测试时，可以拿到 app 来先测试吗？

答：不用集成到 app 测，就基于接口测 demo 就可以。

3、请问一下，企业能否考虑提供 H5 开发平台？，介于题目要求做独立的 H5 应用，现阶段网上的各大 H5 开发平台均不是免费开放。

答：“独立 H5 应用”不等于要使用第三方平台。理解成是要做一个可以独立运行、前端基于 H5 的 demo 应用系统，不要求在互联网上可访问。所以不提供，也不希望使用网上的开发平台。

【A03】工业物联网——高性能边缘计算轻量级容器管理引擎设计【文思海辉】

1、文思海辉赛题讲解视频要求提交的作品是倾向于边缘计算的 web 容器管理工具，而赛题手册中赛题是要求设计一个应用容器引擎，请问以哪个为准提交作品？

答：最终交付以最后发的文档为准，web 的交付耗时，不存在技术上的难点，

需要学生实现的是边缘计算应用容器的设计，可以没有 web。

【A04】工夹具全寿命智能管理系统（TMS）【捷普】

1、一个工夹具就是一套还是几个工夹具就是一套？

答：一个就是一套。

2、视频中分为上线夹具和下线夹具，请问有什么区别？

答：上线是夹具从仓库到生产线上（开始使用），下线是从生产线返回仓库（结束使用）。

3、点检是专门有一个人来负责的吗？是定时的吗？

答：点检由专人定期进行。

4、工夹具的定位是简单的用表展示在某个厂区某个区域某人使用即可？请问位置跟踪的具体要求。

答：只需显示厂区，区域，线别，工站。

5、移动端的最终样式是类似独立的功能 app 吗？有无具体要求。

答：移动端 app 是主体的补充，使用同一个数据库。

6、视频中点检记录只有上下线点检，是在出库之后就进行点检，再上生产线；在下生产线时，先点检再入库？那视频中的定时点检又是如何记录到现有的这张纸质表格中？

答：上下线的点检记录是这样的，夹具上线先要点检然后分配到生产线，夹具下线也要点检然后进入到仓库库位保存。至于定时点检的记录是在另外一份定时点检表格中的记录的。

7、现有的这张表格上下线夹具不是同一套，是否可理解为这一条生产线即将下线的一套夹具和即将上线的一套夹具记录在同一张表格中。

答：是的

8、在视频中有提到“需要临时领出去”，记录的表格是产线领退记本，领出去是领到哪里去？

答：领去的地方有很多，有可能去生产线临时需要用一下，有可能去一个工作区域临时做校验验证等，地点是不固定的，临时领取的夹具数量也是单一的不是产品成套的。

9、另外夹具出入库，这个库是否就是夹具柜，临时领出去和这个出入库的区别？
这个过程在功能综述和流程概述中并没有十分明确的体现，是否需要实现？

答：库就是夹具柜，临时领出去和出入库其实性质是一样的，都是做出入库记录。不一样的地方是正常出入库一定是从仓库领去产线，产线退回仓库，数量是产品的成套夹具。而临时领出去就如上面讲的出去的地点不确定，数量不确定不是产品的成套夹具。临时出入库其实只要做一个简单的出入库记录功能就可以，便于查询这个夹具的去向。

10、 采购是否会出现一次采购多种夹具，并且夹具数量不为 1 的情况。采购之后夹具上的编号是何时印刷的（这个采购流程是否与我们系统相关，我们是否需要提供编号），另外可否提供一下采购单的照片。 采购审核主要审核哪些方面？

答：会出现一次采购多种夹具的情况，每次采购数量不同肯定不为 1. 夹具上的编号在夹具制作的时候就已经生成，系统只要负责录入记录即可。系统可以不要采购审核的内容，系统只要录入已经核实好的信息，比如：夹具名称，夹具本身的编号，夹具的接受编号，夹具的接收日期，夹具属于哪个产品，夹具的 PM 频率，夹具的 PM 类型，夹具的库位信息，夹具的负责人等。

生产线、具体夹具是否只有一个所属的工作部门，对其他工作部门不可见？

答：可编辑和不可编辑，统一由管理员授权，系统通过公司个人的 NTID 登录。

11、 Family（所属大类） 是什么含义，也是要完全独立吗？

答：Family 属于一个产品的家族大类，它下面有不同的型号，型号下面还有不同的频段号。

12、 前端框架是否可以用 vue 框架？

答：可以

13、 关于出入库：产线员工需不需要作为系统用户，如果是是，那么产线员工与仓管员是否都是作为初级用户？仓管员是否作为一个独立于部门存在的职位？

答：产线员工指的是操作这个系统的人员，他们的主要职责是将夹具从仓库中领出然后发放到生产线上，并且在系统上做登记。还有就是临时领用做登记，夹具归还做登记，夹具异常做登记，夹具保养做登记。而仓库管理员

的权限在操作员上一个等级，他要将夹具的基本信息录入到系统，夹具的库位信息的管理变更等。

- 14、 系统以 workcell 为单位隔离数据，各个 workcell 之间数据完全独立互不可见，那么请问只有本部门的产线员工能够借所属的工夹具？个人理解是产线员工都是一个生产部门的

答：Workcell 里面由工程部，生产部，质量部，测试部门，而这个夹具管理的职能部门是工程部是辅助工程部对夹具有系统化可视化的管理，所以对对这个系统有操作权限的都是工程部，其他部门只是可见基本信息即可。这里说的产线员工指的是工程部的实际操作这个系统的在线操作员工。

- 15、 由于提供的数据缺乏真实性，按数据中推得一个工夹具对应一个采购单号，这与实际情况相悖

答：一个夹具有唯一的自身编号和 buyoff 编号，这里说的 buyoff 编号是经过各个部门验证这个夹具符合功能的接收编号，不是采购编号。

- 16、 关于申请入库流程：请问贵公司的入库申请实在采购之前，申请之后经过审批再去购买，还是说已经购买回来了，只是需要更新仓库内工夹具信息

答：入库是在 buyoff 之后。

- 17、 工夹具的哪些基础信息是固定的，比如说同一类的工夹具的共性。

答：固定的有夹具自身编号，夹具 buyoff 编号，夹具的 jbuyoff 日期，ower，使用的产品 family，型号。

- 18、 有无用于预警预测的数据？预警模型的构建需要有大量的历史记录以及此工具夹在借用之后是否损坏，若损坏则对应的损坏程度如何。

答：目前尚无这方面的具体数据模型，有能力的团队可以充分发挥创造性思维去设计和开发。

- 19、 前端是否不限制框架？如果限制，可用范围是？

答：前端框架原则上不作限制，但希望是 bootstrap/jquery 或 vue/element-ui 等主流框架

- 20、 是否会有一个夹具属于两种不同的产品？

答：同一个产品会有多套不同的夹具，同一套夹具会用于不同的产品。

- 21、 公司有几个 WorkCell, 每个 WorkCell 都有生产部、质量部等等？ A

WorkCell 里面定义的夹具类别 在 B WorkCell 中是否能够看到？还是说完全屏蔽？

答：每个 workcell 是独立的系统，数据互不可见，用户和权限也是各自一套，互不相干。

22、 1、采购入库针对的采购还是入库？ 2、采购入库申请是在购买前进行申请还是在检查无误后入库时进行申请？ 3、我们没有在您提供的数据 excel 中找到 buyoff 编号。

答：系统中的采购入库仅指夹具第一次录入数据库，与实际采购行为（定价、合同、付款等）毫无无关系。

23、 请问工作部门是怎么划分的？你的权限里有初级用户、高级用户、监管员、经理，这都是一个部门的，还是不同部门的？这些不同的角色需要区分于不同的部门吗？

答：每个 workcell 都有自己独立的一套的用户和权限。

24、 在问题总结中，您指出临时领出去是数量不确定的，不是产品的成套夹具，请问这个非成套夹具指的是半成品吗？因为您在第一个问题中指出一套夹具就是一个，所以这个非成套夹具指的是什么意思呢？

答：成套的意思是这个产品需要的所有夹具，比如说 A 产品需要 10 个夹具，这 10 个就叫成套。非成套是临时领取 这个产品的 2 个或者 3 个夹具，这是非成套的意思。

25、 请问夹具的使用记录与领用的信息记录有哪些区别？或者说是不是等同于记录夹具的领用情况？我们想详细了解一下夹具的使用记录需要记录哪些内容。

答：夹具的使用记录和领用记录，其实都是对夹具的进入库做一个登记，可以视为同种功能。记录的内容就是，时间，班次，生产产品型号，领用人，领出时间点，夹具使用地，备注等。

26、 请问系统的部署环境是怎么样的？

答：部署环境为 windows server 2016 + IIS 9。

27、 能否不止限于局域网，比如增加连接互联网使用微信等方式使用该系统？

答：夹具数据涉及公司各个客户的商业机密，本系统不可能接入外网。

28、 工具夹报修报修何时何地地进行，工具夹报修时，应该先执行返库流程再报修，还是直接报修？

答：直接报修。

29、 请问线下工人和仓管员属于权限角色的哪一个，或者还是要单独列权限角色？

答：线下工人只是操作的角色，他的功能是执行夹具出入库登记，夹具损坏报修。仓管员的功能是夹具的信息管理，保养，报修的异常处理，等。

30、 贵公司提供的两张 excel 表格，其中一张名为夹具信息的表格中存在的信息与另一张表中的夹具定义表和夹具实体表中存在的信息均无对应关系（例如夹具信息表中有夹具编号而另外两张表格中并无此类信息），可否自行添加？

答：可以

31、 工夹具入库以后，工夹具的位置就固定在仓库的工作柜中的一个 bin 位，以后的进出库它的位置都不变吗？还是每次都进行改变？

答：bin 位需要可变。

32、 关于领走和退还功能，已知领走是夹具在生产线上被临时领走，那领走的具体操作是怎么样的，是产线上任何员工都可以进行领走产线确定吗？还有退还的时候是直接给该夹具的仓管员还是也是给任意一个产线员工呢？

答：领走不是从生产线领走的，是从仓库领走的。需要记录从哪边领走去到哪里谁领走的做登记记录。进库操作员可以登记，仓管员也可以操作。

33、 夹具在提交维护申请后，谁进行维护呢？这个是否在系统的范畴之类，如果在的话则这个维护是不是属于其他部门来进行维护而不是工程部，这个其他部门处理好后再反馈给该工程部的高级用户进行维护后的信息记录？

答：系统的夹具信息由仓管员维护，当然也可以设立一个管理员的角色来维护整个系统。

34、 一个班次有多少个仓管员，多个仓管员是联合起来一起管 50 个夹具柜不分你我，还是一个仓管员负责一部分的夹具柜？如果是一个人一部分的话需要对仓管进范围的圈限？

答：最多 2 个仓管员，仓管员负责整个 workcell 的所有夹具柜。

35、 关于点检人员级别的疑问。①点检人员的级别是 Operator I (初级人员) 还是 Operator II (高级人员)? ②夹具点检表由哪一级别的人员提交?

答: 1 初级人员。2 初级人员。

36、 夹具的生产线流程是怎么样的? 具体步骤有哪些? 产线员工在去从申请到得到该夹具会是怎样的流程呢, 目前有如下理解 1. 产线工人去找管理员 (指定位置) -> 管理员在其办公处进行相关登记 -> 带他一起去拿夹具 2. 产线工人自己在手机上申请 (连接公司内网 WIFI) /附近公司提供的申请设备 (如连接至该系统的电脑) -> 管理审核完毕后 -> 通知到指定位置领用 还是说有其他方式? 对于扫描员工卡确认身份, 是已经有该保障了吗 (如有, 大致怎样扫描确认), 如果没有, 是希望该功能融入现有系统吗?

37、 答: 工夹具上线流程, 产线工 (也就是初级人员) 根据生产计划 (另外的系统) 确定该条生产线今天需要生产的产品型号, 然后工人通过手持电脑在系统中根据产品型号检索到该产品型号的所有夹具信息, 然后工人去对应的库位提取夹具, 然后在系统中登记出库记录夹具的去向及出库时间操作人员等 (每出一个夹具操作一次出库登记), 到这时一个夹具的出库流程完成了。人员的身份登记可以通过刷员工卡可以通过扫身份二维码。身份确认需要融入系统, 就是当这个工人登录系统后他的所有操作都是系统自动记录。

38、 初级用户操作进出库: 是指可以在界面选择自己想要借用和归还的工夹具; 还是指初级用户负责操作对借出的工夹具进行记录和归还的工夹具进行归还记录, 这个页面动作是记录信息还是操作借用和归还工夹具?

答: 这个页面既是操作界面也是记录界面。

39、 Model 夹具模组 (多个) 与 PartNo 夹具料号 (多个) 有什么作用, 查询的时候需要按照他们查询吗

答: 根据这个信息能快速检索定位到改夹具, 或者该型号的所有夹具。

40、 不同批次采购入库的同种工夹具的 code 是否是一样的 (比如说报废了一套, 又要买一套新的一样的夹具回来, code 会变吗)

答: 夹具的 code 在同一时间是唯一的, 但是当夹具报废后, 它所占用的 code 会被系统回收, 并赋予给新采购来的夹具。

【A05】基于人工智能的视觉识别技术【文思海辉】

1、赛题中提到，不能使用商用模块，具体是指什么？

答：只能用 Apache、mit、bsd 授权的开源库。

2、现场测试的图片集是和题目给的数据集类似的吗？

答：是。

3、有数据清洗标准吗？xm 内的 l 字段能否给出详细解释？

答：清洗没有标准，按给出的数据集实现效果就行；xml 是 voc 数据集标注的标准格式，学生团队可自行查询。

4、题目中说要实现实时检测，而且要开发可以部署到上机位的系统，对这个系统有什么要求呢？

答：没有具体要求 windows linux 都可以

5、最后的模型，赛方会验证吗，如果验证怎样的验证形式呢（视频，图片，还是...）

答：提供模型相关指标和代码 以及演示视频即可

6、最后提交的作品需要做成 APP 吗？还是基于 TensorFlow 之类的深度学习平台，只要能实现检测头盔的目的即可？

答：需要一个简单的演示界面，能够从监控视频中圈出没有合法佩戴安全帽的人员，输入可以是实时的监控视频流或者本地的视频文件。

7、这道题的一个最后有个要求是：要求开发可以部署到上位机的系统。我可以开发出一个单独的 App 吗

答：可以。

8、请问可以使用基于 LGPL 的 QT5 吗？

答：可以。

【A06】交通时空大数据分析挖掘系统【东软】

1、是否可以使用第三方数据集？是否可以使用第三方接口，或者开源工具？

答：开源工具无限制，第三方数据集及接口如果是为了实现模块任务，在现

有数据基础上进行丰富也是可以的。

2、在企业提供的数据外，学生可以自己添加数据吗？

答：数据格式为 csv，查看数据用 Notepad++ 等工具请同学熟悉赛题以及数据后再作决定。我们企业提供的数据可以支撑赛题中所涉及的所有模块。

3、(timestamp) 信息记录开始时间为什么比 (timestamp1) 信息记录结束时间要大？

答：两个时间取一个既可。

4、企业提供的数据是真实的还是模拟数据？

答：真实业务场景，真实数据。

5、(timestamp) 信息记录开始时间为什么比 (timestamp1) 信息记录结束时间要大？

答：两个时间取一个既可。

6、赛题中提到的实时热力图，一共只有 10000 多条数据 100 多台机器 如果要实现实时热力图 请问 `实时` 要求的是 小时级别的 还是分钟级别的 还是秒级的？

答：时间的数据格式是精确到秒的 建议最好能达到秒级，分钟也可

7、请问给出的是静态数据，如何展示实时的热力图？

答：给出的数据是可实现模块内容的，这个问题涉及到模块的实现，请学生回去好好思考，不要再咨询如何设计模块等问题

8、出行方式静态数据中，已经明确了出行方式。所谓标签化如何理解？

答：将每个人标签化，定位每个人的出行方式

9、数据说明中提到 imsi 是用户唯一 id，而原始数据中，学生观察到大部分都是同一个用户，也就是说原始数据中都是相同的一两个人的数据，这个是否有问题？

答：请学生自行查找 `信令数据` 方向的相关文章，这块不难理解，原始数据中是几百人的数据

10、1. 原始数据集中“nid -信息类别 1 ， npid -信息类别 2”，分别代表什么？

答：参考数据说明，其他清洗后的字段不影响各模块设计。

11、出行方式判断结果只能从已给出行方式（公交和地铁）中选择吗？还是说类似于驾车步行等出行方式需要自己进行判断？另外，请问基本清洗后数据量已不足万条，是否可以提供更多数据？

答：出行方式可拓展，可自行判断，不能提供更多数据

12、是否需要考虑设计相关 APP（或小程序），为用户推荐出行方式？比如根据“非实时的”原始数据进行分析，下午五点三台子地铁站人口密集，相关 APP 倾向于不推荐用户前往此地铁站乘坐地铁；又比如根据目前的“实时的”原始数据分析，蒲田路地铁站人数异常突增，是否需要在 APP 上提醒用户或者在要求的前端页面上标注异常？

答：有能力的团队可以考虑更多的合理设计。

13、本系统面向的用户是居民百姓，还是政府或商业机构？

答：系统面向政府。

14、企业 只提供了一天的数据，能否给出一个时间段的数据

答：现有数据可支撑赛题要求，企业不提供其他数据。

【A07】酒店视觉 AI 解决方案【虹软】

1、是否只能使用虹软提供的人脸识别接口

答：推荐使用虹软的人脸识别

2、是否可以引入外部的一些开源项目，比如 OpenCV 等，对项目的部分功能起到辅助作用，但人脸识别、活体检验等部分仍由虹软官方的 SDK 进行识别。

答：对开源项目的引用无限制

3、酒店类型众多，主要需要考虑哪种类型的酒店？

答：酒店类型没有限制，关键是给酒店带来价值，给客人带来便利

4、请问关于赛题设计范围，设计是要贯穿酒店多个环节还是可以只针对某个或者几个特定场景来深度挖掘？

答：不限制场景，以真正能够给酒店客人提供便利和给酒店带来价值为核心，当然如果能有整套的解决方案更佳

5、只做网页还是要做网页，app，小程序都要做

答：只做网页可能无法满足需求，建议加上 APP 等

【A08】智·会学习——今目标“智慧校园”场景挖掘与实践应用【今目标】

1、是否允许自己开发应用？

答：关于赛题的延伸，我们可以让有能力的团队在完成基础赛题后进行基于“教师管理场景”的深入思考，开发一套今目标外的小程序用于教师工资的自动核算与发布。

2、我们需要对题目中所列举的几个场景都找到管理难点给出解决方案，还是通过自己调研所总结的难点给出解决方案？

答：关于场景选择，可以根据团队的时间和精力，选择多个场景，可以是题目中罗列的，也可以是其他的场景。题目中的场景只是抛砖引玉，同学们可以先通过调研和走访找到几个具体痛点，然后再通过今目标，搭建解决方案。

3、可以给今目标移动端设计 UI 界面么？

答：设计 UI 界面跟本次赛题关联性不强，不作为评分参考项。

4、该怎样展现我们的成品，因为业务搭建是在 app 中进行，那我们该怎么说明我们是如何搭建的，是需要将每一步文字化然后配合截图放入在提交的详细方案里吗？详细方案里具体我们需要写些什么？演示视频是指整个模拟场景的展示还是指操作今目标的整个流程？

答：1、是的，在 app 中搭建后，可以通过截图+文字或者示意图的形式放到提交的方案中。2、详细方案中可以包括：痛点描述、痛点分析、解决思路、具体解决方案、实践效果等。3、演示视频可以是两者的结合，一方面是场景的模拟，同时加入操作界面的演示，另外可以配合制作人的介绍、目标客群的采访（原有工作模式与使用系统后的对比）等，内容题材不做硬性要求。

5、目前已加入今目标所给的企业号相关组织，但因为管理权限我们很多操作无法开展，所以想问作品提交的时候是用此企业号搭建相关业务，还是说此企业号只是为了让我们更快熟悉和方便交流，而提交的作品是需在另创的一个新的模拟组织中来完成业务搭建？

答：请参赛团队及时注册企业账号（邀请企业号：86110797，注册链接 <https://web.jingoal.com/user/account.html#/join>，邀请码定期将在官方群

里更新)，组委会将以邮件的形式收集团队的模拟组织的企业号、组员人数，交于企业进行账号升级。请各团队队长关注组委会邮件（fwwbds@niso.edu.cn）。

【A09】僵尸企业画像及分类【科创信息】

1、最后放出的用于评估模型的测试集的特征属性？

答：与已发布数据特征一致，只是不同的数据集。可能在建模的时候不同的竞赛单位有特征组合，但是最后的评判还是输入原来的这些特征。

2、题目中判断僵尸企业的条件是.csv 文件中的 flag 列吗？若是 flag，题目叙述的 2000 家僵尸企业为什么在测试集中会有近 9000 个僵尸企业判定，若不是，如何通过训练判断其是僵尸企业？

答：训练集和测试集是分开的，僵尸企业数量不一样没什么问题，主要看学生怎么去用数据。另，每个企业对应多条数据。

3、项目运行中，数据模型提前载入内存会极大提升运行速度，但相对成本较高，如果每次执行任务的时候载入则成本相对较低，但运行速度相对较慢，想问问企业更偏向于成本控制还是运行速度？

答：可以考虑运行速度为主。

4、测试集和验证集必须按赛方给定的比例使用吗？请问在训练模型时必须按照赛方给定的测试集/验证集比例来训练模型吗？是否允许选手自行调整测试集验证集比例，如：将一部分赛方给定的验证集数据用于模型训练。

答：可以。

5、为什么训练集中会有一万多 flag 为 NA 的数据，是给错了吗？

答：提供给参赛队伍的数据集包括：训练数据集（5000 例僵尸企业，10000 例标签不确定的企业，包含僵尸企业数据和非僵尸企业数据）；验证数据集（30000 例企业数据，僵尸企业数据和非僵尸企业数据比例 3：7）。出题方保留的标准测试数据集：10000 例有标签的企业数据。

6、僵尸画像的用户是哪个政府部门？

答：市场监督管理局。

7、题中给出数据集只有六个省份，企业保留数据集亦是如此？

答：是。

8、视频说训练集共 15000 个样本，我下载数据发现有 15050 个样本？验证集说是 30000 多出一部分 884 个没有标签样本，是否可以将其补充到训练集或直接删除？

答：数目稍有差异，不会影响后续研究

9、可视化 web 端开发的开发语言有限制或推荐么

答：没有

10、这个系统在具体实施的时候面向的用户群体有多大？需要考虑很高的并发量吗？

答：暂时不用考虑目标群体大小，开放式的命题主要是考察分类规则的可解释性以及结果的准确率。

11、baseline 在验证集上结果已经达到 99%以上的准确率，这样的话这道赛题的数据集会不会太简单了（光是使用线性回归模型就能在验证集上达到 91%的准确率），还是说除了模型的精确性和高效性有什么别的需求是这道题的重点？

答：简单与否不是参赛方应该考虑的问题，作为一个目前没有确定标准的问题，可以作为经济领域一种现象的有益的探索。

12、赛题给的训练集中没有控制人 ID 这一属性，而验证集存在，请问再最后结果的测试集上是否会有这一特征？

答：测试时不会加入这个特征做评判。

13、关于净利润问题，本人非金融专业，但也觉得这净利润计算有问题，比如 id28 号数据，其利润总额为 7216，纳税总额为 0，那么根据公式，净利润则为利润总额 $7216 - \text{纳税总额}$ $0 = 7216$ ，而数据给出的结果却是 -7216

答：由于工商登记的时候没有认真审核，也可能有一些未知的原因导致误差，因此，利润总额这个特征的处理可以更加灵活一些。这也是目前大数据的一种困境，数据收集难，收集到的数据不一定都有价值。

14、请问 Web 端的用户输入，是要求用户输入所有的维度特征，而后进行实时的模型验证，还是只需要用户输入企业的 ID 编号，在预先分类完成的数据中进行查询？

答：验证的时候需要输入维度特征。

15、完成这一算法需要要用到什么软件？

答：web 客户端不做要求，模型也仅仅建议 Python 和 R 语言实现。

【A10】金融科技服务平台企业数据的无监督分类系统【浪潮】

1、最终模型建完后，一个企业可以有多个标签还是一个标签？

答：多个标签

2、测试集数据的形式是怎么样的？或者说题目中要求：最终以完整系统的形式接受企业的信息输入。这个企业信息输入的格式是怎么样的？企业的信息输入，如果是 excel 的格式，那么数据是不是和给的训练集一样，分成这么多不同的表格，还是在一个表格里面。会不会有些表格因为没有数据，企业就不输入？

答：最终完整系统的实现接收的企业输入数据只是企业名称或者统一信用代码或者序号之类的简单信息，和给的训练集是不一样的。

3、这道题是构建无监督系统，既然是无监督打标，那在评价指标里面有准确率、召回率等，那是不是就说明在给的测试集里面的数据是有标签的？如果有标签，那是不是对于无监督分类又是一个导向？还是在“技术要求与指标”的（3）模型评估指标只是一个参考？

答：训练数据是无标签的，模型评估指标只是一个参考。

【A11】多方数据安全计算平台的设计与实现【浪潮】

1、多方安全计算的政府管理员和企业管理员分别负责什么？图中说的计算节点管理具体是什么？

答：政府管理员负责监管数据的使用情况，企业管理员负责查看计算结果，节点管理是管理个计算节点的运行情况。

2、平台可否使用第三方开源工具，评判标准除了加密解密功能还有哪些是相对重要的？

答：可以用开源工具。没有其他附加的评判标准。

3、题中给出风控模型公式 $\text{Math. max}(\text{Math. min}(\text{Math. max}(A * 0.8, B * 15) *$

10, 150000), 50000) 没有涉及到加法, 是否需要实现全同态加密算法? 也就是加法同态和乘法同态都必须实现?

答: 加减乘除根据团队能力做

4、想问下贵公司是否对输入风控模型的数据集来源有提供? 如果不提供是否有格式要求?

答: 不提供; 没要求。

【A12】基于算法的配送路线优化系统【华云+中诺思】

1、是否应该考虑配送人员的费用问题: 题目假设中提到了工作人员每日时长是 8h, 那么是否可以认为一辆货车无法配送二次, 但是人员可以多次参与配送。2、是否派送点只能经过一次: 每个派送点只能由一辆车服务一次, 是否意味着默认派送点只被经过一次, 还是要考虑经过多次, 但只派送一次的情况。3、要求中满载率的定义是否为 单辆车实际所载货物重量/该辆车最大载重量: 题目要求满载率最大化, 是指所有车辆满载率之和最大化, 还是指单辆车满载率最大化?

答: 1、可以不用考虑; 2、派送点只经过一次, 经过多次这个属于设计问题, 我们只考察一个点由一辆车派送一次, 不重复派送; 3、单个车辆的满载率。学生团队在能力允许的范围内, 可以做更优化的设计。

2、请问最终, 我们最终测试的数据是按照发出来 excel 表中的数据还是题目中给出的数据呢? 还有就是题目中的连线代表是只能这样走吗? 能不能 A 走到 C、D、E 等等这些点呢?

答: 题目中的数据中根据线的连接走, 没有连接的没有给出两点的距离, 不能直接走, 但是表格中的数据因为是给出的坐标或距离, 任何两点都可以走, 可以以题目中的做为最终的, 表格中的是验证你题目中的设计是否是最优, 可以通过不同的数据来验证。

3、题中所提到限定时间是所有货物运输完的时间, 还是每辆车的限定工作时间, 例如: A 车要在 1 小时完成货物配送, 本来的路径需要 1 个半小时, 加上限定条件后这个环境就不满足, 就要重新规划路线?

答: 可以先不用考虑时间。

4、题目中提到每个员工工作时间为八小时, 请问这个八小时是指类似于上下班

的固定时间段的八小时，还是其他情况？

答：暂时不考虑其它情况，这个工作时间也是个不是必要考虑的因素，固定的几个因素配送中心的客户量，每个客户的需求量，每辆车的满载率，车辆一次最大运行的最大距离，这些因素是必须考虑的，学生可以先不考虑工作人员的时间。

5、在内容说明给的样例中，题目让用算法求出车辆数量，而在下面的开发需求中需要用户设置车辆数目，能否给出更加详细的要求？

答：在没有给定车辆数的情况下，根据车辆的承载量计算需要车辆，还有一种情况就是给出了车辆数和承载，怎么去规划。

6、是只需要考虑车辆的满载率还是需要考虑车辆的耗油量的服务成本？请问关于车辆的服务成本有没有一些参考数据？

答：只考虑车辆的承载量。学生给出的成果中只要考虑了客户数，客户与配送中心的距离，客户需求量，车辆配载这些就可以，工作人员时间和车辆时间做为非必选项。

7、题目的具体数据以 excel 表格的数据集为准还是命题文档中的为准？

答：以命题文档中为准，表格中的可作为测试数据

8、最优路径的判断标准是否只有总路程，在官方数据集中只提到总路程最短？但是题目中需要输出配送完成时间 配送完成时间是否指的是配送路线中的需要完成时间最多的那一辆车的完成时间(所有配送车在同一时间从配送点出发) 那么就有可能出现配送路线中某一条路线的总路程小于另一条。而配送完成时间却大于另一条 是否需要我们确定一个权重。

答：需要路程最优，不一定是最短，还有车辆的使用情况需要考虑进去，不是车辆越多越好，完成时间只作为参考指标，根据配送量和车辆的满载率去考虑路径最优

9、在官方数据集中，要求我们需要随机生成各种数据，只明确了配送量和位置信息需要随机生成，请问其他数据是需要提供数据接口去设置吗？例如，车辆有哪几种类型，载重量等等。

答：车辆的载重量，数量需要能够配置

10、对于车辆速度，员工的工作时间等等，它是一个定量吗 能规范出哪些是定量吗

答：对题目的考核暂时可以不用考虑时间问题，这个后续可以作为拓展内容学生去研究。如果学生能在这方面有所创新，想法当然更好

11、系统部署指什么？是要封装出可供调用的 API？还是要求做出一个网站一样的东西吗？有必要根据作品，结合生活场景，如快递公司做出一个应用系统吗？如果是对界面有要求吗？

答：开发作品需要能够安装到计算机上，通过简单的界面，设置一些参数后可以看到结果

12、把算法应用到实际场景中，可以用其他公司如百度, 阿里的 API 吗？

答：这个没要求

13、题目中求解是按照给定的工作人员工作时间，在工作时间之内将选定的各个位置的派送路线计算出来，并且耗时最短，花费最少。是按照这样的要求求解吗？

答：耗时，花费是整体的，总共那么多客户的量，完成所有客户的配送整体的路线，耗时。

14、时间复杂度会是算法的一个评判标准吗？

答：时间仅作为一个参考的评判。

【A13】运用文本相似度实现主观题自动评阅【华云+中诺思】

1、赛题中所说的测试样例在哪获取？

答：物流专业的简答题基本都是名词解释和概念释义类的题目，语句结构都比较简单，网络上有很多物流相关专业的简答题文档，学生团队可以案例题的样式自己准备数据集。

2、赛题要求将标准答案与学生答案进行对比分析，请问我们所获得的学生答案是否已经经过了 OCR 文本识别？

答：无须经过 OCR 文本识别，目前的场景是在线答题，而非扫描的纸质答卷。

3、题目没有告知评价模型的标准，那我们做完之后，该如何判断自己的模型好不好？

答：题目里给出的标准是对比人工评分，意即模型评分与人工评分的结果一致度越高，模型越好；如学生答案人工评分是 7 分，系统评分为 3 分，就偏差比较大。这里的人工评分标准在于语义是否正确，以及是否正确命中得分点验证的话，

可以在网络上找几道物流仓储作业相关的简答题，模拟不同的答题情景让系统评分，然后自己对比参考答案给分，来判断模型是否基本满足需求。基本上物流相关的简答题都是名词解释，释义类的，答案都是比较好理解的。

4、是否需要按照类似人工的给分方式，将答案拆分为各小点分点给分？（例：有 ABC 三个点，每点 3 分） 还是只需要最后模型给出的总分和人工阅卷的给分一致，不需要按点给分，这样就可以了？

答：如果可以不按点给分也能实现总分与人工评分一致，也是可以的

5、企业后续是否还会提供数据？

答：企业将于 1 月 9 日前提供新的数据集，请关注大赛官网和官方 QQ 群文件更新。

【A14】金融领域公司实体消歧系统构建【恒生电子】

1、可以实时提交测试集的实体消歧结果，还是只能自己利用训练集判断模型好坏呢？假设我的模型可以得出测试集的实体消歧结果，我是否可以提交结果得到实时反馈，以判断我的模型在测试集上的效果呢？

答：验证集上可以充分验证结果。