



2020 数字中国创新大赛-大数据赛道

——操作手册

1 注册账号

登录 data.xm.gov.cn，点击赛事详情进入“2020 数字中国创新大赛大数据赛道”网站首页。点击右上角“登录”按钮或“报名参赛”跳转至 i 厦门登录界面，如图 1.0.0 所示：



图 1.0.0

已有 i 厦门个人实名账号的用户，输入账号密码登录系统即可进入赛事系统。

未有 i 厦门个人实名账号的用户，点击“立即注册”，跳转至 i 厦门注册页面进行账号注册，如图 1.0.1 所示：



图 1.0.1

进入 i 厦门注册页面，勾选“我是市民用户”，（“2020 数字中国创新大赛大数据赛道”仅支持 i 厦门个人实名账号登录），如图 1.0.2 所示：

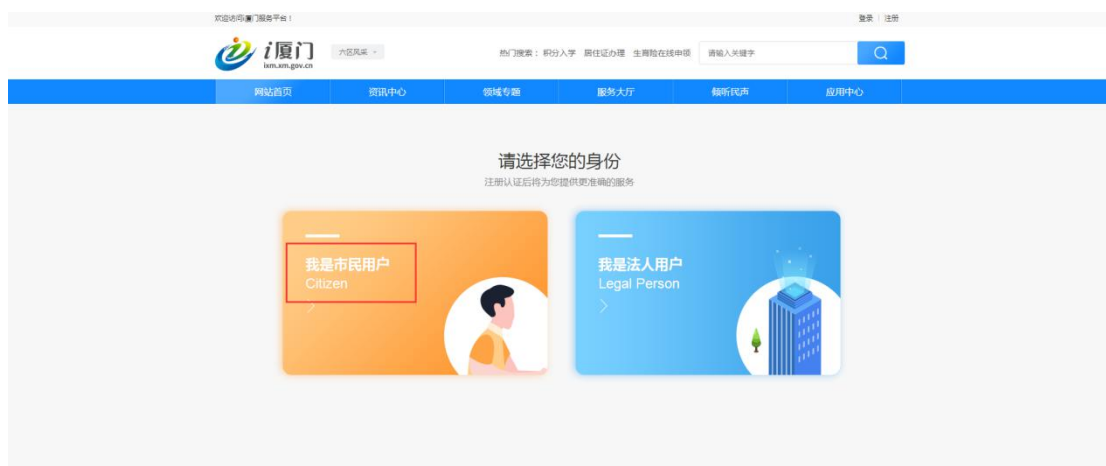


图 1.0.2

根据个人情况，选择对应的注册方式，以下以公安实名认证流程为例，若公安实名认证失败，可尝试银联卡实名认证等其他注册方式。注册完毕后，即可使用账号密码登录“2020 数字中国创新大赛大数据赛道”网站，如下图 1.0.3 所示：

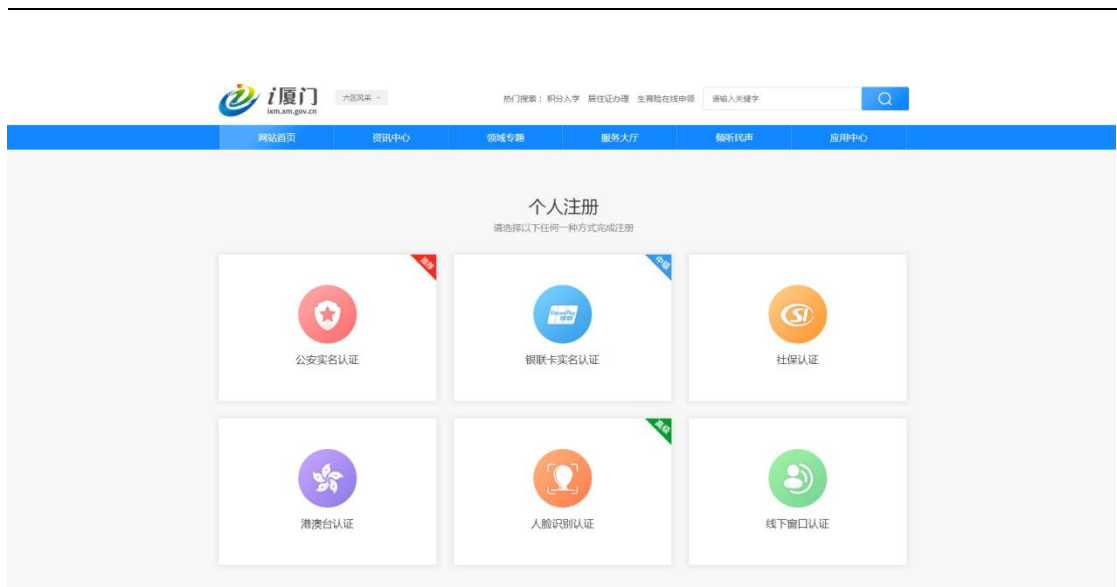


图 1.0.3

根据勾选的注册方式进入对应的注册页面，根据提示填写真实姓名和身份证号，点击下一步，如图 1.0.4 所示：

图 1.0.4

填写手机号和收到的验证码，**验证码收不到的情况下，请多次点击几次“获取验证码”**，点击下一步，如图 1.0.5 所示：

公安实名认证

1 身份认证

2 短信验证

3 设置密码

请输入您的真实信息，通过后则不能更改

手机号：

157

图片校验码：

验证成功

验证码：

请输入验证码

167秒后重新发送

下一步

图 1.0.5

填写密码，点击注册，如图 1.0.6 所示：

公安实名认证

1 身份认证

2 短信验证

3 设置密码

请输入您的真实信息，通过后则不能更改

设置密码：

.....

密码长度8-30位要包含字符和数字，可以使用特殊字符

确认密码：

.....

注册

图 1.0.6

给出提示：恭喜您，操作成功！点击确定可回到的登录页面，使用账号密码即可进入“2020 数字中国创新大赛大数据赛道”网站，如图 1.0.7 所示：

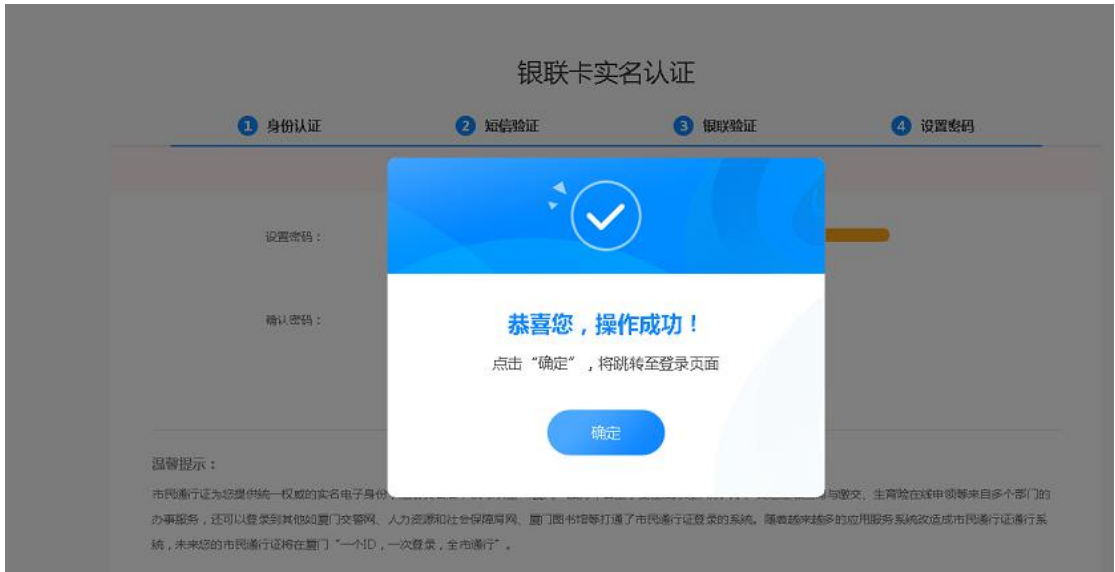


图 1.0.7

2 赛题报名

未登录的情况下，请先登录赛事系统。已登录的情况下，点击【首页】“报名参赛”或进入【赛题说明】的赛题详细根据情况对赛题进行报名，如图 2.0.0 所示：



图 2.0.0

3 队伍确认/队伍绑定

队长创建参赛队伍后，队员需用 i 厦门个人实名账号登录“2020 数字中国创新大赛大数据赛道”网站，进入【我的比赛-我的队伍】绑定参赛队伍，如图 3.0.0 所示：



图 3.0.0

队员全员实名且绑定队伍后，队长需进入【我的比赛-我的队伍】对队伍进行确认，队伍信息一旦确认，不可进行队员增减或队伍信息更改。如图 3.0.1 所示：



图 3.0.1

4 作品提交账户获取

队长对队伍信息进行确认后，可进入【我的比赛-比赛中】提交算法模型分

析结果集、算法模型源码、赛题分析报告、PPT 答辩材料等材料，**算法模型得分** 占总得分 **30%**，**分析报告及答辩得分** 占总得分 **70%**，如图 4.0.0 所示：



图 4.0.0



图 4.0.1

5 作品提交

本次大赛，作品提交是通过 SFTP 方式来完成的。SFTP 登录方式可通过命令行方式或 SFTP 工具来完成，每支队伍的 SFTP 空间大小是 5G，请控制好自己提交的文件大小。SFTP 账号密码为官方下发的账号密码，详见本文档第 4 章节-

作品提交账户获取。【注】：sftp 端口是 57891。

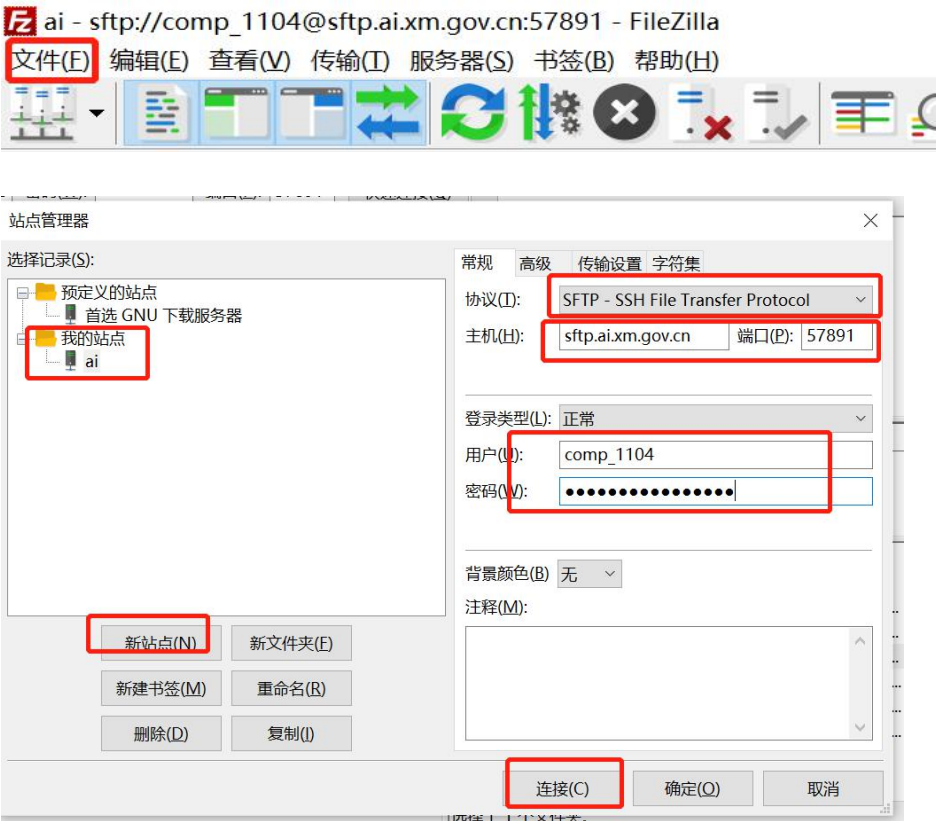
5.1 命令行方式登录 SFTP

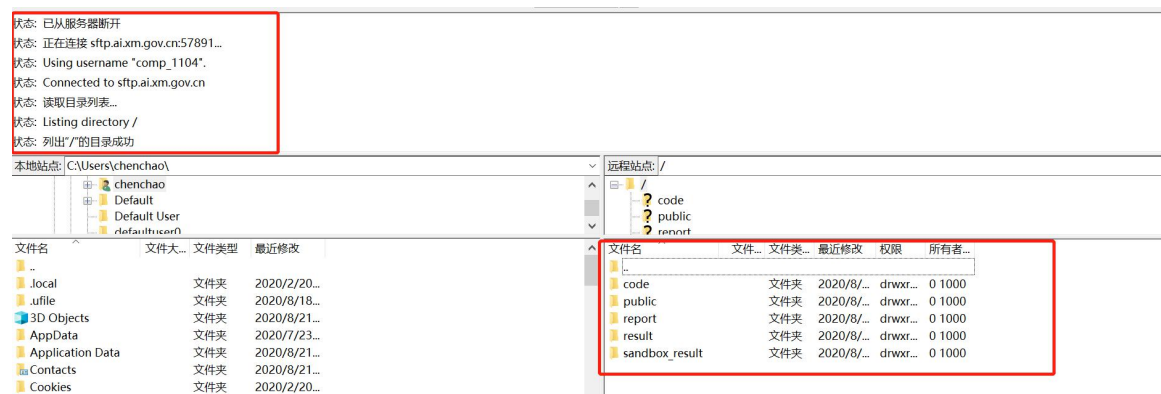
可通过 filezilla 或其它 sftp 工具，连接登录，参照如下：

```
[root@172.16.130 ~]# sftp -oPort=57891 comp_882@sftp.ai.xm.gov.cn
comp_882@sftp.ai.xm.gov.cn's password:
Connected to sftp.ai.xm.gov.cn.
sftp> ls
code                public              report              result              sandbox_result
sftp>
```

5.2 SFTP 工具方式登录 SFTP

可通过 filezilla 或其它 sftp 工具连接登录，在【文件】—【站点管理器】里设置服务端的主机、端口，参照如下：





5.3 作品所在文件路径

本次大赛，参赛队伍需提交的作品包含：①算法模型分析结果集、②算法模型源码、③赛题分析报告、PPT 答辩材料及其他自主选择提交的交付材料。

对应 SFTP 指定目录如下表：

| 序号 | 作品名称 | 存放目录 | 命名规则 |
|----|-------------------------------|---------|--|
| 1 | 算法模型源码 | /code | 参赛编号-code-提交日期 如： JT-CZJT-0000-code-20200903 |
| 2 | 赛题分析报告、PPT 答辩材料及其他自主选择提交的交付材料 | /report | 参赛编号-report-提交日期 如： JT-DWJT-0000-report-20200903 |
| 3 | 算法模型分析结果集 | /result | 分赛题： ①端午假期交通路网拥堵识别及缓堵策略研究： duanwu_result.txt ②城市巡游车与网约车运营特征对比分析： taxi_result.txt |

6 算法模型代码提交说明

参赛队伍提交的算法应为可稳定运行的算法模型，根据赛题规则要求，规范化模型输出，并整理完整代码、注释说明及补充材料，以备代码复核验证。由于算法模型源代码文件较大，整场比赛一支队伍最多提交 3 次。

代码 (含源代码、算法模型等，不含算法输出结果、分析报告等)提交要固

定输出到 SFTP 的/code 目录下，如果选手参与多个赛题，则每个赛题对应一份算法模型代码。

注意：各文件命名要清晰罗列，提交时建议文件压缩后上传，文件命名为“**参赛编号-code-提交日期**”。**代码及算法模型可在比赛结束后 2 天内提交。**

7 报告及答辩材料提交说明

参赛队伍提交的作品(含分析报告、答辩 PPT 材料等，不含算法结果文件、源代码、算法模型)提交要固定输出到 SFTP 的/report 目录下，如果选手参与多个赛题，则每个赛题对应一份分析报告和 PPT 答辩材料。

注意：各文件命名要清晰罗列，作品如包含多份材料统一打包成一份文件夹，文件命名格式为“**参赛编号-report-提交日期**”。**分析报告需在 2020 年 9 月 21 日 24 时前提交，逾期提交则视为无效提交。**

8 算法分析结果输出及提交说明

8.1 端午假期交通路网拥堵识别及缓堵策略研究

8.1.1 模型任务说明

为更好地掌握道路交通运行状态的变化规律与趋势，参赛方需利用赛事方提供的 A 城市近两年端午假期出租车（含网约车）的轨迹、订单等多种数据，一是完成 GPS 点位地图匹配，提供 A 城市交通运行指数计算方法说明、指数计算模型和指数计算结果，主要包括 TOP10 的拥堵路段名称、方向描述、平均运行速度（主要围绕每日早高峰时间 7-9 时，晚高峰 17-19 时）、路段拥堵延迟指数（主要围绕每日早高峰时间 7-9 时，晚高峰 17-19 时）。二是（如有）参赛选手

可以结合地理信息图层或其他算法软件，进一步挖掘 A 城市交通路网其他运行特征，提供相关的分析结论、计算结果和图层输出。三是针对算法输出结果，为 A 城市市民游客提出端午假期旅游出行建议，并结合拥堵特征提出相应地解决拥堵的策略（具体可以结合某一具体拥堵点），为解决城市拥堵治理提供辅助决策。

8.1.2 赛题数据说明

端午假期交通路网拥堵识别及缓堵策略研究赛题，提供 2019.06.06-2019.06.09 和 2020.06.24-2020.06.27 两年共计 8 天的 A 城市网约车和巡游车的 GPS 数据、订单数据，以及 A 城市路网矢量数据，上亿条数据，已个人实名参赛报名的选手可于大赛官网进行下载使用。

8.1.3 运行结果提交格式

算法运行结果要固定输出到 sftp 的 **/result** 目录下，结果文件为 **duanwu_result.txt**，格式为 **utf-8**。

比赛平台最终会去取这个结果文件算出结果，**如果提交结果为非 duanwu_result.txt 文件命名，或没有提交到/result 目录下，则不会有分数产生。**

算法运行结果文件内容中，**每个字段间以英文半角竖线 “|” 符分隔。**

8.1.4 输出结果字段说明

➤ 算法输出结果格式：

year_id,day_id,peak_type,road_rank,road_id,road_tti,road_vec

2019|0|1|1|R56|3.8|7.95

2019|1|1|5|R57,R301|3.25|19.57

2019 | 2 | 1 | 6 | R58,R302,R303 | 3.45 | 23.44

2019 | 3 | 1 | 4 | R58,R302,R303 | 3.45 | 23.44

2020 | 0 | 2 | 3 | R56 | 3.8 | 7.95

2020 | 1 | 2 | 7 | R59,R304,R305,R306 | 4.2 | 25.69

2020 | 2 | 2 | 9 | R60 | 2.8765 | 28.98

2020 | 3 | 2 | 2 | R61 | 3.28 | 18.43

注：输出结果文件里需要有表头，结果文件为 Utf-8 格式的 duanwu_result.txt 文件，文件内每字段以**英文半角竖线**“|”间隔。其中，road_id 字段，如果有多个值，则每个以**英文半角逗号**“,”间隔。如上。

➤ 字段注释

| 字段名 | 字段释义 | 备注 |
|-----------|-----------|--|
| year_id | 年份 | [2019,2020] |
| day_id | 端午节假日日期编号 | {0:节前一天; 1:节中第一天; 2:节中第二天; 3:节中第三天} |
| peak_type | 时段标识 | {1:早高峰; 2:晚高峰} |
| road_rank | 拥堵路段排名 | 1~10 |
| road_id | 拥堵路段编号 | 例如: R23,路段编号来源于提供给参赛者的路网数据中,需要取出排名前 10 的路段 |
| road_tti | 路段拥堵延时指数 | 参赛者计算所得数据 |

| | | |
|---|----------|-----------|
| road_vec | 路段平均行车速度 | 单位要求 km/h |
| 注：输出结果为一张有以上 7 个字段、 160 条记录（2 年*4 天*2 时段*10 拥堵路段）的表格 | | |

8.1.5 评测得分规则

- (1) 本赛题算法模型得分占总得分 30%，即满分 30 分。
- (2) 计算分值的三个字段：road_id、road_tti、road_vec；
- (3) 其中 road_tti、road_vec 答案给出的将是区间值，在区间内则得分，不在区间内则不得分，每个答案值 0.05625 分，输出 320 个答案值，共 18 分。
- (4) road_id，每个路段编号答案值 0.075 分，输出 160 个答案值，不对应则不得分，共 12 分。

【注】：同一天路段拥堵排序均不属于同一路名，例如，某一天，某一拥堵排行结果路段编号属于**成功大道（南向北）**，那么其他排序路段则不能再属于**成功大道（南向北）**。

8.1.6 算法平台成绩结果说明

参赛队伍提交的算法分析结果文件，平台**每隔 2 小时**评分一次，如果参赛队伍提交的算法结果文件在同一个周期内多次提交，则以最后一次提交的文件参与评分。最终算法分取参赛队伍**历史分数最高**的一次。算法模型分析结果集**每日最多提交 3 次**。

参赛选手可登录 [https://data.xm.gov.cn/opendata-competition/index.html/#/](https://data.xm.gov.cn/opendata-competition/index.html#/) 具体查看每次算法得分结果及历史提交明细。

8.2 城市巡游车与网约车运营特征对比分析

8.2.1 模型任务说明

参赛方需依据赛事方提供的出租车（包括巡游车和网约车）订单数据，一是综合应用统计分析方法分别对巡游车和网约车运营的时间、空间分布特征进行量化计算，包括平均每天 24 小时的分布变化，按网格划分的空间分布（网络划分颗粒度选手自选），并分别对比分析网约车、巡游出租车的日均空驶率、订单平均运距、订单平均运行时长、上下客点分布密度等时变特性；二是根据巡游车和网约车的时空运营特征，并尝试对巡游车与网约车的融合发展提出相关建议。

8.2.2 赛题数据说明

城市巡游车与网约车运营特征对比分析赛题，提供 2019.05.31-2019.06.09 和 2020.06.18-2020.06.27 两年共计 20 天的 A 城市网约车和巡游车的 GPS 数据、订单数据，以及 A 城市路网矢量数据，上亿条数据，已个人实名参赛报名的选手可于大赛官网进行下载使用。

8.2.3 运行结果提交格式

算法运行结果要固定输出到 sftp 的 **/result** 目录下，结果文件为 **taxi_result.txt**，格式为 **utf-8**。比赛平台最终会去取这个结果文件算出结果，**如果提交结果为非 taxi_result.txt 文件命名，或没有提交到/result 目录下，则不会有分数产生**。算法运行结果文件内容中，**每个字段间以英文半角竖线 “|” 符分隔**。

8.2.4 输出结果字段说明

➤ 选手算法输出结果格式：

year_id,day_type,load_type,daily_unloaded_rate,order_load_distance,order_load_time

2019 | 1 | 1 | 0.18 | 6.78 | 34

2019 | 1 | 2 | 0.23 | 5.34 | 45

2019 | 2 | 1 | 0.23 | 9.65 | 32

2020 | 2 | 2 | 0.16 | 4.2 | 12

2020 | 3 | 1 | 0.08 | 5.79 | 56

2020 | 3 | 2 | 0.34 | 6.8 | 23

注：输出结果文件里需要有表头，结果文件为 Utf-8 格式的 taxi_result.txt 文件，文件内每字段以**英文半角竖线**“|”间隔。

➤ 字段注释

| 字段名 | 字段释义 | 备注 |
|----------|------|------------------------------------|
| year_id | 年份 | [2019,2020] |
| day_type | 时期类型 | {1:正常工作日; 2:正常周六周末; 3:端午节假日} |

| | | |
|---|----------|----------------|
| load_type | 运营车辆类型 | {1:巡游车; 2:网约车} |
| daily_unloaded_rate | 日均空驶率 | |
| order_load_distance | 订单平均运距 | 单位要求 km |
| order_load_time | 订单平均运行时长 | 单位要求 min |
| 注：输出结果为一张有以上 6 个字段、 12 条记录（2 年*3 时期*2 运营方式）的表格 | | |

8.2.5 评测得分规则

- (1) 本赛题算法模型得分占总得分 30%，即满分 30 分。
- (2) 计算分值的三个字段：daily_unloaded_rate、order_load_distance、order_load_time；
- (3) 答案给出的将是区间值，在区间内则得分，不在区间内则不得分，选手提交的答案分别各给出 2 年（2019、2020）*3 类时期（正常工作日，正常周六日、端午节假日）*2 种运营方式（网约车、巡游出租车）*三个指标（daily_unloaded_rate、order_load_distance、order_load_time），共 36 个结果值；
- (4) daily_unloaded_rate 日均空驶率，每个答案得 1.5 分，输出 12 个结果值，共 18 分；
- (5) order_load_distance 订单平均运距、order_load_time 订单平均运行时长，每个答案的 0.5 分，输出 24 个结果值，共 12 分。
- (6) daily_unloaded_rate、order_load_distance、order_load_time 的答案是一个区间匹配，选手这三个字段的输出值分别在对应答案字段的区间里，则分别得分，否则不得分。

8.2.6 算法平台成绩结果说明

参赛队伍提交的算法分析结果文件，平台**每隔 2 小时**评分一次，如果参赛队伍提交的算法结果文件在同一个周期内多次提交，则以最后一次提交的文件参与评分。最终算法分取参赛队伍**历史分数最高**的一次。算法模型分析结果集**每日最多提交 3 次**。

选手可登录 <https://data.xm.gov.cn/opendata-competition/index.html#/> 具体查看每次算法得分结果及历史提交明细。

附录 1：作品提交常见问题 Q&A

1. Sftp 登录时报 Permission Denied?

可能的原因有 2 种，一种是密码输出错误，请确保密码输入正确，可使用复制粘贴密码的方式，注意密码前后不要有空格。

另一种是当前 sftp 服务器正忙，可稍等一会重新提交。为保证选手能顺利通过 sftp 提交，避免尽量在比赛结束前的两三天内集中提交。

2. Sftp 上传文件后空间不足怎么办？

每个选手是 5G 的空间，选手尽量把作品等文件压缩打包后上传，如果压缩后空间还是满足不了，可联系官方赛事小助手帮忙后台人工协助处理。

3. 为什么上传了算法结果集，却没有展示算法得分？

有以下几种原因：

①算法结果集提交目录、格式或命名不正确。算法运行结果要固定输出到 sftp 的 **/result** 目录下，结果文件为 **taxi_result.txt 或 duanwu_result.txt（按赛题要求）**，格式为 **utf-8**。比赛平台最终会去取这个结果文件算出结果，**如果提交结果为非**

指定文件命名，或没有提交到/result 目录下，则不会有分数产生。

②算法跑分程序正常运行时长为 1-2 小时，提交后不会立即得到分数，请耐心等待。

③创意赛题所提交算法结果不参与算法评测，将不展示算法得分。