

第二届“强国杯”技术技能大赛
——工业大数据管理与分析赛项

技
术
方
案

2024 年 7 月

目 录

一、大赛名称	4
二、大赛目的与意义	4
三、大赛内容、形式和成绩计算	4
(一) 竞赛内容	4
(二) 竞赛形式	5
(三) 报名条件	5
(四) 成绩计算	6
四、奖励办法	7
五、大赛命题原则	7
六、大赛范围、赛题类型和其他	7
(一) 理论知识竞赛	7
(二) 实际操作竞赛	8
七、大赛场地与设施	8
八、大赛关键环节与时间安排	9
(一) 关键环节	9
(二) 竞赛流程	9
(三) 时间安排	10
九、大赛评分标准制定原则、评分方法、评分细则及技术规范	10
(一) 评分标准制定原则	10
(二) 评分方法	10
(三) 评分要点(评分指标)	11
(四) 技术规范	12
十、大赛平台说明	13
十一、大赛安全保障	15
(一) 比赛环境	15

(二) 生活条件	15
(三) 组队责任安全	15
(四) 应急处理安全	15
(五) 处罚措施	16
十二、大赛组织与管理	16
(一) 大赛组织	16
(二) 大赛设备与设施管理	19
十三、附件（大赛样题）	20

一、大赛名称

第二届“强国杯”技术技能大赛——工业大数据管理与分析赛项。

二、大赛目的与意义

为贯彻落实国家大数据发展战略，促进工业数字化转型，激发工业数据资源要素潜力，加快工业大数据产业发展，促进工业互联网领域人才产教融合培养体系，提高工业互联网行业人才培养培训质量，扩大人才培养规模，支撑服务产业发展，工业大数据管理与分析赛项应运而生。

工业大数据竞赛对于促进技术创新、推动行业发展和解决实际问题具有重要意义。竞赛为参与者提供工业数据管理与分析平台，让选手能够运用数据科学和技术手段，探索解决工业领域中的挑战，从而促进技术创新、推动行业发展、解决实际问题，并促进跨界合作与交流，为工业领域带来新的突破和进步。

三、大赛内容、形式和成绩计算

（一）竞赛内容

竞赛内容分为理论考试和实践操作，理论考试为线上考核均为客观题，实践操作是在工业数据管理与分析平台上操作，主要考核参赛选手在工业数据管理与分析平台上的工业数据抽取与准备、数据清洗、数据挖掘、数据分析、数据增强等方面的技能。通过对工业数据的分析，让数据赋能价值创造过程，完成数据的价值化变现。此外同时考核参赛选手解决问题能力以及致力于紧跟行业发展步伐的自我学习能力。

表 1 竞赛模块内容

阶段	比赛模块	分值占比	时长	考核内容
理论阶段	理论	20%	60 分钟	参赛选手在工业大数据方面基础知识考核。
实操阶段	数据抽取与准备	15%	180 分钟	选手通过平台上传数据文件，创建待分析测点，替换或追加数据，形成一套分析数据集。
	数据清洗	15%		选手对工业数据管理与分析平台的应用和对业务问题的理解，通过过滤、公式计算等方式完成数据清洗等操作。
	数据挖掘	15%		选手对数据进行基于模式的搜索和挖掘分析。
	数据分析	15%		选手通过多图层对比、统计分析、相关性诊断分析等形成数据分析报告的视图。
	数据增强	10%		选手通过公式、窗口计算等方式，基于原始测点数据创建出虚拟测点，用于分析和数据可视化。
职业素养	职业素养	10%		参赛选手的参赛规范、文明竞赛。

（二） 竞赛形式

本次竞赛形式为线下比赛，分为职工组（含教师）和学生组，各组别均为单人赛，不得跨单位组队。每单位每组别最多报 2 个队伍。

（三） 报名条件

具有工业大数据相关职业工作经历的企业在职人员，从事相关专业工作的高等院校、职业院校（含技工院校，下同）在职人员，以及高等院校、职业院校相关专业全日制在籍学生均可报名参赛。

已获得“中华技能大奖”、“全国技术能手”称号及取得“全国技术能手”申报资格的人员，不得以选手身份参赛。具有全日制学籍的在校创业学生不得以职工身份参赛。

(四) 成绩计算

理论考试竞赛满分为 20 分，赛题均为客观题，赛题数量为 200 道，采用机考方式实现。实操部分和职业素养满分为 80 分计入竞赛总成绩。理论考试成绩、实践操作成绩和职业素养相加得出参赛选手竞赛总成绩，满分为 100 分。

表 2 各模块分值

模块		主要内容	比赛时长	分值
理论模块	理论题	理论知识主要包括信息技术、数据库、大数据接口、大数据存储与处理、大数据分析、计算机软件开发规范、大数据管理、软件维护等方面知识。	60 分钟	20
实操模块	数据抽取与准备	选手通过平台上传数据文件，创建待分析测点，替换或追加数据，形成一套分析数据集。	180 分钟	15
	数据清洗	选手对时序数据管理与分析平台的应用和对业务问题的理解，通过过滤、公式计算等方式完成数据清洗等操作。		15
	数据挖掘	选手对数据进行基于模式的搜索和挖掘分析。		15
	数据分析	选手通过多图层对比、统计分析、相关性诊断分析等形成数据分析报告的视图。		15
	数据增强	选手通过公式、窗口计算等方式，基于原始测点数据创建出虚拟测点，用于分析和数据可视化。		10
职业素养	职业素养	参赛选手的参赛规范、文明竞赛。		10

四、奖励办法

1. 赛项设置一二三等奖及优胜奖，颁发获奖证书，获奖比例分别为10%、20%、30%及10%。
2. 学生组一等奖获奖队伍的指导教师(每支参赛队伍指定1名指导教师)，颁发“优秀指导教师”证书。
3. 对做出突出贡献的承办单位和技术支持单位，颁发“突出贡献单位”证书。
4. 对大赛组织实施中表现突出的工作人员，颁发“优秀工作者”证书。
5. 对在大赛执裁工作中表现突出的个人，颁发“优秀裁判员”证书。

五、大赛命题原则

大赛命题由组委会组织专家负责，在命题方向上，聚焦工业大数据领域和产业深度融合发展领域，以行业及产业需求和痛点为导向，同时竞赛借鉴世界技能大赛命题方法，适当增加相关新知识、新技术、新设备、新技能等内容，结合企业职业岗位对人才培养的需求，参照相关国家职业标准制定。

六、大赛范围、赛题类型和其他

(一) 理论知识竞赛

理论考试为机考，分值占比20%，通过大赛指定竞赛平台完成。建立竞赛标准题库，采用专家抽取的方式；竞赛评分将由系统自动完成，理论知识主要包括信息技术、数据库、大数据接口、大数据存储与处理、大数

据分析、计算机软件开发规范、大数据管理、软件维护等方面知识。理论考试竞赛时间为 1 小时。

(二) 实际操作竞赛

实践操作考核范围是工业大数据的常用场景，分值占比 80%，在工业数据管理与分析平台完成，竞赛评分将由裁判评审。包含数据抽取与准备、数据清洗、数据挖掘、数据分析、数据增强五个阶段，题型为实操。实践操作时长为 3 小时。

七、大赛场地与设施

1. 理论场地：标准计算机房，有局域网，可断外网。

2. 实操场地：竞赛现场设置竞赛区、裁判区、服务区、技术支持区。现场保证良好的采光、照明和通风，提供稳定的电和供电应急设备，同时提供教练休息场所。

3. 竞赛设备：所有竞赛设备由赛项执委会负责提供和保障，竞赛区按照参赛队数量准备比赛所需的软硬件平台，为参赛队提供标准竞赛设备。

4. 竞赛工位：每个赛位标明编号，赛位面积不少于 4m²，竞赛现场各个工位配备单相 220V/3A 以上交流电源。每个比赛工位上标明编号。每个工位配有工作台，用于摆放 1 台计算机，同时配备 1 把工作椅（凳）。

5. 技术支持区为参赛选手提供竞赛相关技术支持，服务区提供医疗等服务保障。

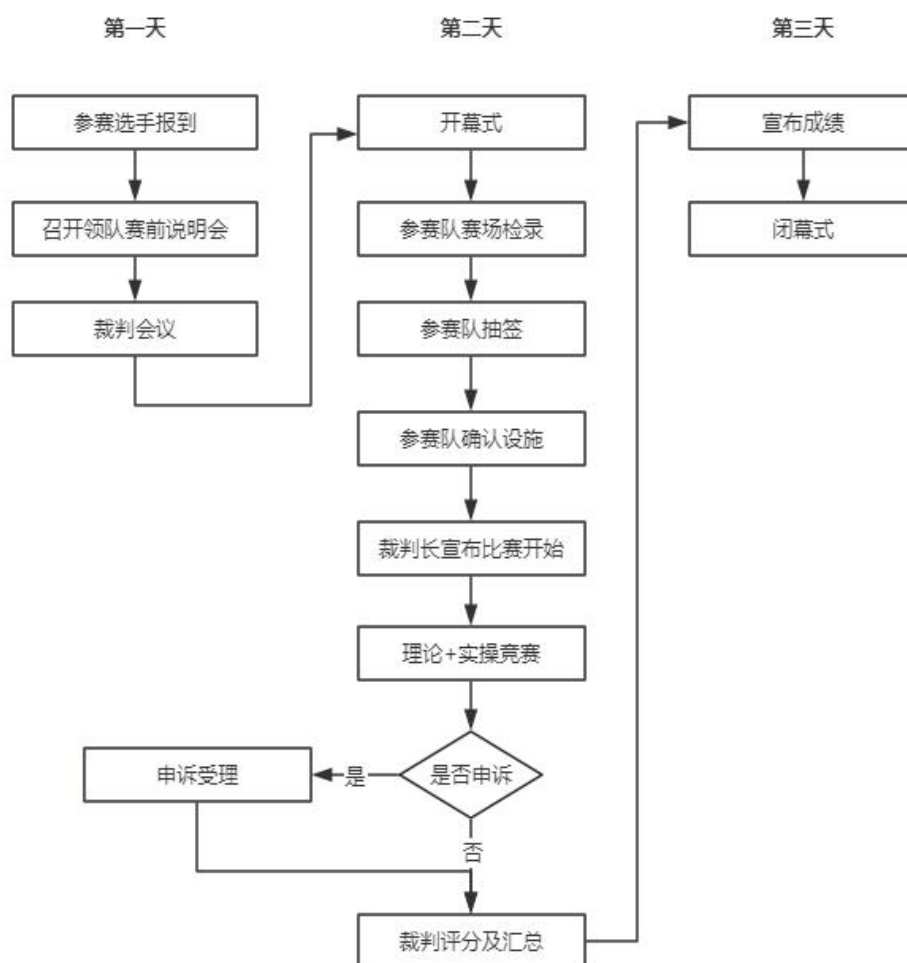
6. 竞赛环境依据竞赛需求设计，在竞赛不被干扰的前提下赛场面向媒体、行业专家开放，允许媒体、行业专家在规定的时段内沿指定路线进行现场参观。

八、大赛关键环节与时间安排

（一）关键环节

参赛选手报到—召开领队赛前说明会—参赛选手赛前熟悉场地—开幕式—理论考试—实操竞赛—比赛结束（参赛选手上交比赛成果）—成绩评定—大赛技术点评、颁奖仪式、闭幕式。

（二）竞赛流程



（三） 时间安排

比赛时间预计为 3 天，如下表所示，具体以竞赛指南日程为准。

表 3 日程表

日期	时间	内容
第一天	14:00—21:00	报到
	16:00 前	裁判报到
	17:00—18:00	裁判会议
第二天	9:00—10:00	开幕式
	10:00—11:30	领队会议、赛场参观
	12:30—13:00	检录、加密
	13:00—17:00	理论+实操考试
第三天	8:00—10:00	成绩公示
	10:00—11:00	闭幕式、技术点评、颁奖

九、大赛评分标准制定原则、评分方法、评分细则及技术规范

（一） 评分标准制定原则

基于大赛目的和相关技术规范，本着“科学严谨、公正公平、可操作性强、突出工匠精神”的原则，按照技能大赛技术专家组制定的考核标准，能够全面评价参赛选手职业能力的要求制定评分标准。

（二） 评分方法

裁判组在坚持“公平、公正、公开、科学、规范”的原则下，各负其责，按照制订的评分要点进行评分。

结果评分：比赛结束后，裁判组根据参赛选手提交的比赛结果进行评分。

成绩汇总：比赛成绩经过加密裁判组解密后进行加权计算，确定最终比赛成绩，经总裁判长审核、仲裁组长复核后签字确认。总成绩相同，以实操成绩高者优先排名，如实操成绩相同，按照实操模块五成绩高者优先排名。

(三) 评分要点(评分指标)

1.理论考试部分评分

理论考试部分总分 20 分，理论考试竞赛平台直接自动评分。

2.实践操作部分评分

表 4 评分要点

阶段	模块	评分要点	分值
实操阶段	模块一： 数据抽取与准备	将提供的工业数据文件上传至管理与分析平台，并创建测点，替换和追加操作。	15 分
	模块二： 数据清洗	通过查看工业数据统计指标、过滤异常值、逻辑公式计算等方式完成数据清洗、数据替换等操作。	15 分
	模块三： 数据挖掘	通过对工业数据进行基于取值的搜索、和基于相似度模式的搜索等操作，能够对典型时序模式进行探查。	15 分
	模块四： 数据分析	通过多图层对比、统计分析、相关性诊断分析等形成工业数据分析报告的视图。	15 分
	模块五： 数据增强	正确使用平台的公式计算、窗口计算等功能，基于原始测点数据创建出虚拟测点。	10 分
	职业素养	参赛选手的参赛规范、文明竞赛。	10 分

实践操作由裁判按照表 4 评分要点评分。

(四) 技术规范

GB/T 42128-2022 智能制造 工业数据 分类原则

GB/T 42203-2022 智能制造 工业数据 云端适配规范

GB/T 40693-2021 智能制造 工业云服务 数据管理通用要求

GB/T 38672-2020 信息技术 大数据接口基本要求

GB/T 38673-2020 信息技术 大数据大数据系统基本要求

GB/T 38676-2020 信息技术 大数据存储与处理系统功能测试要求

GB/T 38643-2020 信息技术 大数据分析系统功能测试要求

GB/T 38675-2020 信息技术 大数据计算系统通用要求

GB/T 38633-2020 信息技术 大数据系统运维和管理功能要求

GB/T 42127-2022 智能制造 工业数据 采集规范

GB/T 39400-2020 工业数据质量 通用技术规范

GB/T 41778-2022 信息技术 工业大数据 术语

GB/T 42130-2022 智能制造 工业大数据系统功能要求

职业素养与安全要求：

严格遵循相关职业素养要求及安全规范，安全文明参赛；操作规范；

工具摆放整齐；着装规范；资料归档完整等。

十、大赛平台说明

表 5 大赛平台说明

实操平台	工业数据管理与分析平台
功能描述	工业数据管理与分析平台是一款智能工业数据分析软件平台。通过数据接入、测点创建、公式计算、移动窗口计算、数据过滤、时序模式搜索、多层对比、相关性诊断、统计分析、数据视图和数据包络等功能，帮助企业更加便捷的分析和应用生产过程中产生的海量时间序列数据。该平台不仅能完成描述性、探查性和诊断性等多种复杂的数据分析，还解决了数据难以处理、理解和挖掘的问题，从而为优化生产过程提供有力支持，释放数据的真正价值。该平台提供了直观且易于操作的界面，使学生能够轻松接入和处理工业数据。通过多层对比和相关性诊断等功能，学生可以深入理解数据之间的关系和规律，增强分析能力。同时，平台的统计分析和数据视图功能能够帮助学生更好地掌握数据分析的基本原理和方法，培养他们的数字化操作技能，快速融入工业数字化生产中。
技术参数	1.数据接入管理 (1)支持离线导入 csv 文件。 2.数据视图 (1)支持用户根据测点名称进行搜索，添加到视图中，并选择不同时间段的数据进行展示； (2)支持包括趋势图、泳道图、散点图、箱线图以及原始数据的展示； (3)支持主视图和上下文图切换，兼顾趋势和细节的查看和筛选； 3.测点管理 (1)支持灵活的添加、删除、保存、隐藏测点； (2)支持测点在视图中以多种形式展示和统计； (3)支持查看测点加载过程中的日志记录； 4.数据搜索

(1)基于值的搜索：支持用户根据阈值大小进行搜索，支持多个条件的 AND/OR 的组合搜索；

(2)相似度搜索：支持基于特定时间段的测点数据，通过模式识别，寻找与之相似（绝对值相似、或形状相似）的时间段；

5.数据过滤

(1)支持对搜索结果进行过滤，支持手动筛选、反选、以及基于条件规则的筛选，用于排除特定的时间段落；

(2)支持过滤结果在界面上的可视化展示，将过滤内容在视图全集中也灰色显示；

(3)支持过滤结果的保存以及重新加载；

6.统计分析

(1)支持对选定数据的平均值、标准差、最小值、最大值、极差、求和、初始值、终值及阶段增量的统计；

7.虚拟测点（公式计算）

(1)支持用户根据常用的计算函数和规则，创建自定义的测点，并参与分析；

(2)支持==, >, <, <=, >=, !=, +, -, *, /, abs, trunc, exp, pow, min, max, log, sqrt, ceil, floor, sin, cos, tan, asin, atan, if, and, or, not, duration_to_numeric, numeric_to_duration, piecewise_sum 等函数。

(3)支持使用窗口计算创建虚拟测点。利用各种算子在移动窗口中对原有测点值进行聚合计算。窗口计算能够平滑带有噪声的信号，或突出显示时序趋势的特定部分，例如令特定时间段内的最小值或最大值持续整个窗口。

8.图层处理

(1)支持用户手动设置时间段，或根据搜索的结果，将不同时间段的结果叠加到同一个视图中进行比对和分析；

(2)支持对图层中各种指标的统计值及图层的相似度进行计算；

(3)支持最多 10 个图层的叠加；

9.容器化基础架构

	(1)平台底层应兼容支持容器基础架构。支持基于容器架构提供大数据分布式计算框架，数据计算支持流处理和批处理引擎，最大化硬件资源和云计算资源的使用。
--	---

十一、大赛安全保障

(一) 比赛环境

须在赛前组织专人对比赛现场、住宿场所和交通保障进行考察，并对安全工作提出明确要求。赛场布置，赛场内器材、设备，应符合国家有关安全规定，承办院校赛前须按照要求排除安全隐患。

赛场周围设立警戒线，要求所有参赛人员必须凭有效证件进入场地，防止无关人员进入发生意外事件。

(二) 生活条件

比赛期间，原则上由赛项承办院校统一安排参赛选手食宿，并确保大赛期间人员的住宿、卫生、饮食安全等。

赛项执委会和承办院校须保证比赛期间选手、领队和裁判员、工作人员的交通安全。

(三) 组队责任安全

各学校组织代表队时，须安排为参赛选手购买大赛期间的人身意外伤害保险，并进行安全教育。

(四) 应急处理安全

比赛期间发生意外事故时，发现者应在第一时间报告赛项执委会，同

时采取措施，避免事态扩大。赛项执委会应立即启动预案予以解决并向赛区执委会报告。出现重大安全问题的赛项可以停赛，是否停赛由赛区组委会决定。事后，赛区执委会应向大赛执委会报告详细情况。

(五) 处罚措施

因参赛队伍原因造成重大安全事故的，取消其获奖资格。

赛事工作人员违规的，按照相应的制度追究责任。

十二、大赛组织与管理

(一) 大赛组织

1. 大赛裁判

(1) 裁判人员具有良好的职业道德和心理素质，严守竞赛纪律，服从组织安排，责任心强；

(2) 评分裁判须从事计算机专业（职业）相关工作 2 年以上（含 2 年），具备深厚的专业理论知识和较高的实践技能水平，具有省级或行业职业技能竞赛执裁经验；

(3) 裁判人员有较强的组织协调能力和临场应变能力；

(4) 裁判人员年龄原则上不超过 60 周岁，身体健康，无任何违法违纪记录，且获得工作单位支持，能在规定时间内到岗，并按要求完成指定裁判工作。

(5) 加密裁判、检录裁判由大赛办公室指派责任心强的专业人员担任。

(6) 竞赛中所有裁判人员不得影响选手正常竞赛。

(7) 评分裁判严格执行竞赛项目评分标准，做到公平、公正、真实、准确，杜绝随意打分；严禁利用工作之便，弄虚作假、徇私舞弊。

(8) 现场裁判要提醒选手注意操作安全，对于选手的违规操作或有可能引发人身伤害、设备损坏等事故的行为，应立即制止并按规定扣分，对于参赛选手作弊的情况取消其比赛资格。

(9) 竞赛期间，因裁判人员工作不负责任，造成竞赛程序无法继续进行或评判结果不真实的情况，由赛项组委会视情节轻重，给予通报批评或停止裁判资格，并通知其所在单位做出相应处理。

2. 大赛监督

(1) 监督组在大赛办公室领导下，负责对大赛筹备与组织工作，实施全程现场监督。

(2) 监督组的监督内容包括大赛场地和设施的部署、选手抽签、裁判培训、大赛组织、成绩评判及汇总、成绩发布、申诉仲裁、成绩复核等。

(3) 监督组对比赛过程中明显违规现象，应及时向大赛办公室提出改正建议，同时采取必要技术手段，留取监督的过程资料。比赛结束后，向大赛组委会提报监督工作报告。

(4) 监督组不参与具体的赛事组织活动。

3. 申诉与仲裁

(1) 各参赛选手对不符合大赛规程规定的仪器、设备、工装、材料、物件、计算机软硬件、大赛使用工具、用品，大赛执裁、赛场管理、比赛成绩，以及工作人员的不规范行为等，可向大赛仲裁组提出申诉。

(2) 申诉主体为参赛选手。

(3) 申诉启动时，参赛选手以亲笔签字的书面报告的形式递交大赛仲裁组。报告应对申诉事件的现象、发生时间、涉及人员、申诉依据等进行充分、实事求是的叙述。非书面申诉不予受理。

(4) 提出申诉应在比赛结束后不超过 2 小时内提出。超过时效不予受理。

(5) 大赛仲裁工作组在接到申诉报告后的 2 小时内组织复议，并及时将复议结果以书面形式告知申诉方。申诉方对复议结果仍有异议，可由省（市）大赛组委会机构向大赛办公室提出申诉。大赛办公室的仲裁结果为最终结果。

(6) 申诉方不得以任何理由拒绝接收仲裁结果；不得以任何理由采取过激行为扰乱赛场秩序；仲裁结果由申诉人签收，不能代收；如在约定时间和地点申诉人离开，视为自行放弃申诉。

(7) 申诉方可随时提出放弃申诉。

(二) 大赛设备与设施管理

表 6 各工位配套设施

序号	设备	数量	介绍
1	工业数据管理与分析平台	1 套	工业数据管理与分析平台以数据智能为核心理念，通过数据接入、测点创建、公式计算、移动窗口计算、数据过滤、时序模式搜索、多图层对比、相关性诊断、统计分析、数据视图、数据包络等功能，帮助企业分析和应用生产过程产生的海量时间序列数据，完成描述性、探查性、诊断性等多种不同难度的数据分析，解决数据看不到、读不懂、难利用、难分析、无法为优化赋能等问题，实现盘活数据资产、辅助业务决策、释放数据价值。
2	竞赛电脑	1 台/人	CPU i5 及以上，内存 $\geq 8\text{GB}$ 内存，硬盘 $\geq 500\text{GB}$ ，显示器分辨率要求 1024*768 以上。
3	服务器	5 台	系统：CentOS Linux release 7.9.2009 (Core)，处理器 ≥ 32 核，内存 $\geq 64\text{GB}$ ，硬盘（SSD）1TB 以上，系统盘 $\geq 150\text{G}$ ，千兆网口 2 个及以上
4	U 盘	1 个/人	存储空间 $\geq 8\text{G}$

十三、附件（大赛样题）

附件 1：实操样题

第一部分 竞赛须知

一、竞赛要求

- 1.正确使用工具，操作安全规范；
- 2.竞赛过程中如有异议，可向裁判员反映，不得扰乱赛场秩序
- 3.遵守赛场纪律，尊重裁判，服从安排。

二、职业素养与安全意识

- 1.完成竞赛任务所有操作符合安全操作规范，注意用电安全。
- 2.遵守赛场纪律，尊重赛场工作人员；爱惜赛场设备、器材。

三、扣分项

- 1.在竞赛过程中，因操作不当导致设备破坏性损坏或造成安全事故的，视情节扣 10~20 分，情况严重者取消比赛资格。
- 2.衣着不整、污染赛场环境、扰乱赛场秩序、干扰裁判工作等不符合职业规范的行为，在职业素养模块中扣除相应职业素养分，情节严重者取消竞赛资格。

四、选手须知

- 1.出现任务书缺页、字迹不清等问题，请及时向裁判申请更换任务书。
- 2.竞赛任务完成过程配有一台编程计算机，参考资料放置在U盘“参考资料”文件夹下。
- 3.参赛人员应在3小时内完成任务书规定内容；选手在竞赛过程中创建的程序文件和提交的答案必须存储到U盘“竞赛编号”文件夹下，未存储到指定位置的运行记录或程序文件均不予给分。
- 4.选手提交的试卷不得出现学校、姓名等与身份有关的信息，否则成绩无效。
- 5.在完成任务过程中，请及时保存程序及数据。

第二部分 竞赛任务

一、任务要求

随着工业互联网产业发展，以工业互联网时序数据分析技术在生产系统管理中的应用为主要场景，结合相应的统计学、机器学习与人工智能算法、数据可视化等技术应用基础，进行工业互联网时序分析与挖掘过程技术设计；在促进工业互联网数据智能技术在工业生产、社会生活服务等多场景中的推广应用具有积极意义。

竞赛平台设备主要由工业互联网时序数据分析平台、应用电脑等组成。实现数据的上传、可视化、对比、筛选、诊断、保存分析结果等全过程；平台具有测点创建、特定方式搜索时序段落、图层堆叠、数据过滤、包络分析、相关性诊断、保存分析视图等功能。

二、任务环境

工位设备：计算机、时序数据分析平台等。

参考资料：见U盘中“参考资料”文件夹。

三、任务实施

任务说明

任务中需要提交的截图，放在U盘“答案提交”文档中。任务中配置信息、数据、帮助文档信息详见U盘“参考资料”。

阶段一 数据抽取与准备（15分）

（1）任务描述

参赛选手按工作任务书要求，在竞赛平台上通过提供的文件完成测点上传、测点追加并添加测点到活动列表等数据抽取与准备的工作任务。提供的文件中主要包含在间歇生产过程中，不同批次的工艺参数，包括time、batch、step_num、pressure和temp等数据。

（2）任务要求：

任务1 测点上传

1)完成指定文件 1 下载、上传指定测点 batch、pressure 和 temp

任务 2 测点追加

2)完成指定文件 2 下载、上传并追加 batch、pressure 和 temp

任务 3 保存基本视图

1)添加 batch、pressure 和 temp 测点到活动测点列表，按指定时间设置上下文图范围 2022/03/07 15:00:45–2022/03/09 18:54:21 和主图范围 2022/03/07 22:39:17–2022/03/08 02:15:00，隐藏 batch 测点，完成泳道图、表格和箱线图的可视化方式切换

2)保存视图到指定路径“/任务 1”。

(3) 提交内容：

1.截图测点管理列表中第一次上传测点（追加前）页面、详情表，保存为“01_测点新增页面”、“01_测点详情_新增_测点名”（若干）。

2.截图第二次上传测点（追加后）的页面，保存为“01_测点追加页面”、“01_测点详情_追加_测点名”（若干）。

3.截图泳道图、箱线图以及表格的可视化页面，分别保存为“01_泳道图页面”、“01_箱线图页面”和“01_表格页面”

4.截图工作空间中指定路径下保存的视图，保存为“01_视图保存页面”

阶段二 数据清洗（15 分）

(1) 任务描述

根据任务 1 中上传的测点数据，通过观察分析，在竞赛平台上将生产过程中发生异常的情况过滤，并通过创建公式找到数据突变的过程，完成数据清洗的工作任务。

(2) 任务要求：

任务 1 数据过滤

1)使用任务 1 中保存的泳道图视图，在主图中查看全部数据，观察所有批次中 temp 和 pressure 变化趋势。根据给定的起止时间 2022/03/07 19:29:22–2022/03/08 02:00:06 ， 2022/03/08 13:56:25–2022/03/08 20:25:07 手动创建 2 个过滤时间段，完成过滤的保存和视图的保存

任务 2 创建公式

2)创建公式测点，要求新测点可以反映数据中突然下降的变点

(3) 提交内容：

- 1.截图视图保存页面，保存为“02_视图加未保存过滤的页面”
- 2.截图其他视图加载已保存的过滤，保存为“02_过滤和视图保存后的页面”
- 3.截图公式创建成功页面，保存为“02_公式创建页面”

阶段三 数据挖掘（15 分）

(1) 任务描述

在任务 2 清洗后的数据基础上，在竞赛平台上通过基于值的搜索的功

能找到符合工艺参数要求的生产批次，并利用相似度搜索找到最佳工艺参数过程，从而完成数据挖掘工作任务。

(2) 任务要求：

任务1 基于值的搜索

- 1)通过基于值的搜索，找到在同一 batch 中过程参数 pressure 大于等于 23 并且持续时间超过 15 分钟的过程，按持续时间降序排序，平铺结果
- 2)针对上一步产生的搜索结果，添加计算 temp 最大值以及 pressure 平均值
- 3)将 temp 最大值计算的最大结果值中持续时间最短的设置为基层
- 4)保存搜索和视图到指定路径“/任务2”

任务2 基于相似度搜索

某个连续生产过程涉及到对温度敏感的物质，如果反应过程中温度出现较大波动，对反应产物的品质会有较大影响。请根据以下步骤，对反应器温度的波动模式，进行相似度搜索，找到曾经发生过的类似事件。

- 1)新建视图并将启用的过滤移除
- 2)添加测点 temp 到活动测点列表
- 3)设置主图范围 2022/03/08 13:48:17–2022/03/08 16:59:35，将给定的起止时间、测点名称，作为模板，进行相似度搜索。

4)通过精选结果，筛选持续时间在 11405s-12000s 之间且得分在 90-95 分之间的结果。

(3) 提交内容：

1.截图搜索结果列表页面，保存为“03_基于值搜索的排序结果页面”

2.截图添加计算后的搜索结果列表页面，保存为“03_基于值搜索的添加计算结果页面”

3.截图相似度搜索结果列表页面，保存为“03_基于相似度搜索的结果页面”

4.截图精选结果设置完成的页面以及精选设置应用后的结果页面，分别保存为“03_精选结果设置页面”和“03_精选结果列表页面”

阶段四 数据分析（15 分）

(1) 任务描述

在产品生产过程中，异常动态行为的发生往往可能是有其他过程参数变化引起的。上传指定文件 3，文件中是采集到的所有过程参数，在平稳运行一段时间后，发现参数 y 明显异常。期望在竞赛平台上通过图层对比、统计表导出、相关性诊断等数据分析工作任务，找到过程参数 y 发生变化的可能原因。

(2) 任务要求：

任务 1：图层对比

- 1)上传指定文件 3，选择文件中的所有测点，完成测点上传
- 2)新建视图并将测点 y 添加到活动测点列表，设置主图范围为 2024/04/07 05:39:15 – 2024/04/07 13:52:29 并设置为基层，添加图层 2024/04/07 21:57:43 – 2024/04/08 06:10:57

任务 2：统计表导出

- 1)图层对比平稳运行和发生故障过程中，展示参数 y 在各个图层的“标准差”结果
- 2)导出统计表中数据

任务 3：数据根因诊断

- 1)对添加的测点 y 进行相关性诊断，获取与 y 相关性绝对值大于 0.7、并超前其至少 5 分钟发生故障征兆的测点列表。
- (3) 提交内容：
- 1.截图添加图层页面，保存为“04_多图层页面”
 - 2.截图各个图层的“标准差”结果比较页面，保存为“04_多图层比较页面”
 - 3.导出统计表数据，保存为“04_统计结果.xlsx”
 - 4.截图相关性诊断页面，保存为“04_诊断结果页面”

阶段五 数据增强（10 分）

(1) 任务描述:

光伏发电过程中，每个小时会采集一次光伏的功率以及地面的辐照数据，数据详见指定文件 4，在竞赛平台上需计算完成每小时的累计发电量、每天的发电量，通过包络功能形成光伏周期发电的总体趋势，并利用包络诊断发现光伏发电过程中偏离总体趋势的周期。

(2) 任务要求:

任务 1：窗口计算滤波

1)上传指定文件 4 中的测点，添加测点到活动列表，将上下文图范围映射为主图范围，保存视图到指定路径“/任务 5”。

2)利用窗口计算得到每小时的累计发电量

3)利用窗口计算和公式计算每天发电量

任务 2：包络及诊断

1)在主图中观察分析，将 2024/04/03 23:39:48 – 2024/04/05 00:57:23 作为模板，进行相似度搜索，相似度阈值设置为 90

2)根据搜索结果将标记 Q 的设置为基层，相似度分数最高的两个添加图层

3)历史的几个黄金批次创建包络并保存

4)加载任务 5 中保存的视图并重新设置主图范围 2024/04/12 00:05:17 – 2024/04/12 19:55:33

5)将保存的包络加载到视图，确保包络完整显示

6)诊断超限的测点

(3) 提交内容：

1.截图窗口计算测点完成后的页面，保存为“05_每小时累计发电量页面”

2.截图公式计算测点完成后的页面，保存为“05_每天发电量”

3.截图相似度搜索后添加图层完成的页面，保存为“05_形成包络的图层页面”

4.截图创建的包络保存完成页面，保存为“05_保存的包络页面”

5.截图加载保存的视图页面，保存为“05_加载视图页面”

6.截图指定视图中加载包络的页面，保存为“05_加载包络页面”

7.截图包络诊断结果页面，保存为“05_包络诊断结果页面”

职业素养（10分）

现场操作安全保护：符合安全操作规程。

操作岗位：设施摆放、工位整洁等的处理符合职业岗位标准。

选手参赛纪律：遵守赛场纪律，尊重赛场工作人员，爱惜赛场的设施和器材。

