

第六届“华为杯”苏皖鲁豫省际交界地区职业院校计算机应用技能大赛 人工智能技术应用大赛技术文件

一、赛项名称

赛项名称：人工智能技术应用

赛项组别：中职组、高职组（含职教本科、留学生）

二、竞赛目的

进一步强化职业院校本专业学生职业技能训练和职业能力的综合运用，促进校企合作、产教融合，完善“岗课赛证”教学模式，培育工匠精神，推动职业院校“双师型”师资队伍建设，大力培养适应经济与社会发展的高素质劳动者和技术技能型人才，进一步提升中高职院校计算机类专业学生能力素质与企业用人标准的吻合度，为在新形势下全面提高信息技术类专业教学质量、为扩大就业创业、运用新技术新模式赋能传统产业转型升级、培育经济发展新动能做出新贡献。

三、竞赛内容

（一）中职组竞赛内容

竞赛旨在全面评估选手在人工智能技术应用，特别是大数据平台部署与计算机视觉任务中的实践能力、工程规范和技术素养。考核内容主要包含理论知识与两个实操核心模块：

理论知识：

理论知识部分主要考核参赛者对Linux系统基础、Hadoop架构与配置原理、Java环境部署、PyTorch及Anaconda虚拟环境使用、CUDA与GPU兼容性等工程常识。

同时考查计算机视觉的核心目标与典型任务、深度学习与传统机器学习的关键区别以及典型应用领域。通过笔试全面评估选手的技术理论基础、规范意识和问题理解能力，为其顺利完成实操任务提供支撑。

模块A：Hadoop环境搭建与启动：

参赛者需要基于主流Linux操作系统，独立完成一个完整的Hadoop分布式计算平台的部署与启动。这要求选手掌握Java环境的配置、Hadoop核心组件的安装、关键配置文件（如core-site.xml, hdfs-site.xml等）的修改，以及环境变量的正确设置。最终需要通过进程查看命令验证平台的成功启动，考察选手的系统部署与运维能力。

模块B：简单图像分类任务：

参赛者需使用竞赛调整过的Cifar10数据集，完成一个完整的图像分类项目流程。首先，需要对数据进行预处理，包括去除纯色图像和图像去重等清洗操作。随后，使用清洗后的数据训练一个图像分类模型，并对测试集进行预测，将结果按类别规整。本模块考察选手的数据预处理能力、模型训练与应用的工程实践能力。

（二）高职组竞赛内容

竞赛旨在全面评估选手在人工智能技术应用，特别是在复杂场景下的大数据平台部署与精细化计算机视觉任务中的实践能力、工程规范、优化能力和技术素养。考核内容主要包含理论知识与两个实操核心模块：

理论知识：

理论知识部分主要考核参赛者对Linux系统配置（如CentOS分区、环境变量）、Hadoop基础原理、Anaconda与PyTorch使用方法；重点考查计算机视觉的核心任务、机器学习与深度学习的关键区别、人工智能技术体系，以及主流框架特性（PyTorch、PaddlePaddle）等知识的掌握。通过单选、多选和判断题，综合评估选手的理论基础与工程认知能力。

模块A：Hadoop环境搭建与启动：

参赛者需要基于主流Linux操作系统，独立完成一个完整的Hadoop分布式计算平台的部署与启动。这要求选手熟练掌握Java环境的配置、Hadoop核心组件的安装、关键配置文件（如core-site.xml, mapred-site.xml等）的修改与优化，以及环境变量的正确设置。最终需要通过进程查看命令验证平台的成功启动，考察选手对分布式系统基础环境的部署与运维能力。

模块B：精细化图像分类任务：

参赛者需使用更为复杂的food101数据集（选取15个类别），完成一个精细化的图像分类项目。在数据预处理阶段，除了基础的纯色图像清除和去重外，还需进行图像分辨率的调整。随后，需利用清洗后的数据训练一个性能更优的分类模型，并对测试集进行精准预测与结果归类。本模块综合考察选手处理复杂数据集的预处理能力、模型的选择与优化能力，以及完成完整AI项目流程的规范性。

四、竞赛规则

（一）选手报名

各职业院校按照大赛组委会规定的报名要求，进行报名参赛。

（二）熟悉场地

熟悉场地时间以组委会统一安排为准，届时将安排参赛队熟悉比赛场地，召开领队会议，宣布竞赛纪律和有关事宜。

（三）赛场规范

赛前准备：选手抽签加密入场，参赛队就位并领取比赛任务，完成比赛设备、线缆和工具检查等准备工作。

正式比赛：参赛选手需按题目要求完成选择、实操、代码填空。操作顺序和分工，由参赛队自行商定。

（四）成绩评定与结果公布

成绩评定和结果公布由裁判组、监督组和仲裁组共同负责。

裁判组实行“裁判长负责制”，设裁判长1名，全面负责赛项的裁判分工、裁判评分审核、处理比赛中出现的争议问题等工作。

裁判员根据比赛需要分为检录裁判、加密裁判、现场裁判和评分裁判。

（1）检录裁判：负责对参赛队伍（选手）进行点名登记、身份核对等工作；

（2）加密裁判：负责组织参赛队伍（选手）抽签，对参赛队信息、抽签代码等进行加密；

（3）现场裁判：按规定做好赛场记录，维护赛场纪律，评定参赛队的过程得分；

（4）评分裁判：负责按评分细则评定成绩。

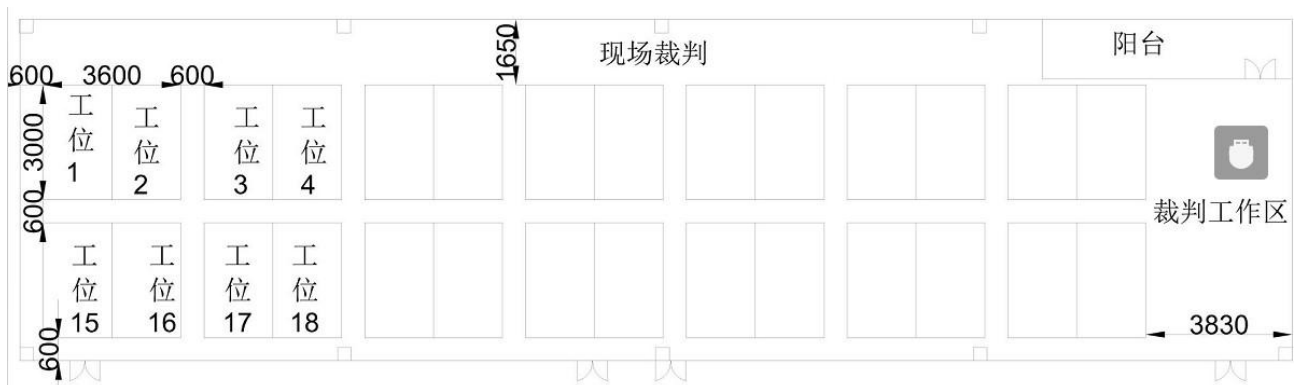
监督组对裁判组的工作进行全程监督，并对竞赛成绩抽检复核。

仲裁组负责接受由参赛队领队提出的对裁判结果的申诉，组织复议并及时反馈复议结果。

最终成绩经裁判组、监督组和仲裁组审核无误后正式公布。

五、竞赛环境

（一）竞赛场地安排



（二）理论竞赛环境要求

每个工位为选手提供 PC 机，每工位配备 220V 电源，工位内的电缆线应符合安全要求。竞赛工位标明工位号，要求保证赛场采光、照明通风良好、温度湿度适宜。

（三）技能竞赛环境要求

竞赛工位内设有操作平台，竞赛工位标明工位号和参赛设备号，并配备竞赛平台和技术工作要求的软、硬件。环境标准要求保证赛场采光、照明通风良好、温度湿度适宜；留有出入和消防通道。比赛现场内应参照相关职业岗位要求为选手提供必要的劳动保护，承办单位应提供保证应急预案实施的条件，必须明确制度和预案。

（四）医疗服务及要求

配备防疫和急救人员与设施。

（五）裁判员工作场所及要求

现场裁判工作场地与要求详见（一）竞赛场地安排；评分裁判需要2间各30平米的房间，每间需要提供两张不小于3*2的方桌，每张方桌配置6把靠椅，配备220V电源（带漏电保护装置），提供打印机。

（六）赛场保密场所及要求

保密室需要提供可容纳160*50张A4纸的保险柜，两台不低于20张/分钟的激光打印机及用于装订试卷的必要设备设施。

（七）赛场摄像头安装要求

指导教师、领队、参赛选手可在直播室观看选手操作现场。

六、技术规范

（一）教学标准

序号	标准单位	中文标准名称
1	教育部	高等职业教育专科专业简介：51 电子与信息大类-5102 计算机类-510209人工智能技术应用
2	教育部	高等职业教育专科专业简介：51 电子与信息大类-5102 计算机类-510205大数据技术
3	教育部	中等职业教育专科专业简介：71电子与信息大类-7102计算机类-710205大数据技术应用

(二) 行业标准

序号	标准号	中文标准名称
1	GB/T 38673-2020	信息技术 大数据 术语
2	GB/T 25000.24-2017	系统与软件质量要求和评价 (SQuaRE) 第24部分：数据质量评价模型
3	GB/T 38634.2-2020	系统与软件工程 软件测试 第2部分：测试过程

七、技术平台

(一) 参赛设备要求 (自备)

本赛项采用参赛队自备计算机设备方式进行，各参赛队须自行准备符合以下技术规格的笔记本电脑，并确保赛前完成所有必要软件的安装与调试：

序号	项目	技术要求
1	操作系统	Windows 10/11 专业版 (64位) 或 Ubuntu 20.04 LTS / CentOS 7 及以上 Linux 发行版 (推荐使用 Linux 环境以适配 Hadoop 部署)
2	处理器 (CPU)	≥ 四核心八线程，主频 ≥ 3.0 GHz

3	内存 (RAM)	≥ 16 GB (建议 32 GB, 以支持 Hadoop 伪分布式模式及轻量级深度学习训练)
4	存储 (硬盘)	≥ 1 TB (Hadoop 临时文件及图像数据集占用空间较大, 建议使用固态硬盘)
5	独立显卡 (GPU)	≥4G显存, 支持加速深度学习训练
6	其他	配备可用 USB 接口、电源适配器; 建议携带扩展坞 (如需多接口)

(二) 竞赛用软件清单

类别	名称	版本
Java 环境	OpenJDK 或 Oracle JDK	1.8 (Java 8)
大数据平台	Apache Hadoop	3.3.x (如 3.3.0–3.3.6)
开发工具	Visual Studio Code	Version 1.77及以上
	PyCharm 社区版	Version 2022及以上
软件环境	Anaconda	4.3.0及以上
	Python	3.9及以上
	NumPy, Pillow, scikit-learn, torch	由conda自动管理
终端工具	PowerShell / Bash / Terminal	系统自带

(三) 赛场支持条件

赛场仅提供:

- 稳定电源插座（每工位 1 个）；
- 技术咨询台（仅解答规则疑问，不提供技术支持或故障维修）。

裁判评分将在参赛队提交的成果包基础上进行，不登录选手设备操作验证，因此选手须确保提交内容完整。

（四）裁判工作需要的办公用品及设备、场所等要求及清单

赛点提供两间评分裁判工作室；竞赛区域需设置现场裁判工作区域，需配备计算机和打印机各一台。裁判用的办公用品主要包括：黑色和红色水笔、订书机、计算器、A4打印纸等。

八、成绩评定

（一）评分方法

1、裁判评分方法

模块A由裁判依据《人工智能赛项》评分标准进行评判，成绩按比例折算后计入总分；

模块B提交后，裁判依据《人工智能赛项》评分标准进行评判，成绩按比例折算后计入总分。

2、成绩审核方法

各裁判员首先审核自身对选手的原始打分成绩，并签名；裁判长对所有裁判员的打分成绩进行审核，并签名。

（二）成绩复核与解密

监督、仲裁组将对总成绩排名前30%的选手成绩进行100%复核，其余成绩抽检覆盖率≥15%。

复核错误率超过5%时，裁判组对全部成绩进行复核，成绩复核无误后进行排名。

（三）成绩公布

记分员将解密后的各参赛队竞赛成绩进行汇总制表，经裁判长、监督仲裁组签字后在指定地点，以纸质形式向全体参赛队进行公布。公布2小时无异议后，经裁判长、监督仲裁组长审核签字后，在闭赛式上宣布。

（四）评分标准

本赛项所有任务（包括理论知识、模块A Hadoop 环境搭建和模块B图像分类）均采用**一次性最终提交**方式。

仅接受一次提交，提交后不得修改或补交，逾期或重复提交视为无效。

理论知识部分于开考30分钟后收取纸质试卷；

竞赛开始**2小时**后，由裁判员在**十分钟内现场审核**所有队伍**模块A完成进度**并打分，期间**暂停作答**；

模块B提交内容必须严格符合样题要求的格式与路径规范，需包含 `./preprocessed_trainset/` 和 `./predictions/` 目录，且结构与类别命名正确；

任何格式错误、路径不符或缺失关键文件，将按评分细则相应扣分或判为零分。

团队协作规则：

竞赛过程中，队伍成员可充分讨论与分工协作；

最终提交操作须由**一名选手**完成，仅记录一次提交为有效成绩。

成绩评定以提交成果的完整性、正确性及任务完成度为准，不设过程加分或时间优先排名机制。

中职组与高职组分别依据各自样题中的评分细则独立评分（如数据集不同、去重要求一致但准确率基准不同等）。

九、竞赛须知

（一）参赛队须知

1、参赛队名称统一使用规定的代表队名称；

2、参赛队员在报名获得审核确认后，原则上不再更换，如筹备过程中，选手因故不能参赛，所在学校需出具书面说明、按相关规定补充人员、经组委会议同意后方可参赛；

3、参赛队按照大赛赛程安排，凭大赛组委会颁发的参赛证、有效身份证件参加比赛及相关活动；

4、各参赛队统一参加比赛前熟悉场地环境的活动；

5、各参赛队准时参加赛前领队会，领队会上举行抽签仪式、抽取场次号；

6、各参赛队要注意饮食卫生，防止食物中毒；

7、各参赛队要发扬良好道德风尚，听从指挥、服从裁判、不弄虚作假。

（二）指导老师须知

1、各指导老师要发扬良好道德风尚，听从指挥，服从裁判，不弄虚作假。指导老师经报名、审核后确定，一经确定不得更换。

2、对申诉的仲裁结果，领队和指导老师应带头服从和执行，还应说服选手服从和执行。

3、指导老师应认真研究和掌握本赛项比赛的技术规则和赛场要求，指导选手做好赛前的一切准备工作。

4、领队和指导老师应在赛后做好技术总结和工作总结。

（三）参赛选手须知

1、参赛选手应遵守比赛规则，尊重裁判和赛场工作人员，自觉遵守赛场秩序，服从裁判的管理。

2、参赛选手应佩戴参赛证，带齐身份证、注册的学生证。在赛场的着装，应符合职业要求。在赛场的表现，应体现自己良好的职业习惯和职业素养。

3、进入赛场前须将手机等通讯工具交赛场相关人员保管，不能带入赛场。未经检验的工具、电子储存器件和其他不允许带入赛场物品，一律不能进入赛场。

4、比赛过程中不准互相交谈，不得大声喧哗；不得有影响其他选手比赛的

行为，不准有旁窥、夹带等作弊行为。

5、参赛选手在比赛的过程中，应遵守安全操作规程，文明的操作。通电调试设备时，应经现场裁判许可，在技术人员监护下进行。

6、比赛过程中需要去洗手间，应报告现场裁判，由裁判或赛场工作人员陪同离开赛场。

7、完成比赛任务后，需要在比赛结束前离开赛场，需向现场裁判示意，在赛场记录上填写离场时间并签工位号确认后，方可离开赛场，离开赛场后不可再次进入。未完成比赛任务，因病或其他原因需要终止比赛离开赛场，需经裁判长同意，在赛场记录表的相应栏目填写离场原因、离场时间并签工位号确认后，方可离开；离开后，不能再次进入赛场。

8、裁判长发出停止比赛的指令，选手（补时选手除外，等延时结束）应立即停止操作进入通道，在现场裁判的指挥下离开赛场。

9、遇突发事件，立即报告裁判和赛场工作人员，按赛场裁判和工作人员的指令行动。

（四）工作人员须知

1、工作人员必须服从赛项组委会统一指挥，佩戴工作人员标识，认真履行职责，做好服务赛场、服务选手的工作。

2、工作人员按照分工准时上岗，不得擅自离岗，应认真履行各自的工作职责，保证竞赛工作的顺利进行。

3、工作人员应在规定的区域内工作，未经许可，不得擅自进入竞赛场地。如需进场，需经过裁判长同意，核准证件，有裁判跟随入场。

4、如遇突发事件，须及时向裁判长报告，同时做好疏导工作，避免重大事故发生，确保竞赛圆满成功。

5、竞赛期间，工作人员不得干涉及个人工作职责之外的事宜，不得利用工作之便，弄虚作假、徇私舞弊。如有上述现象或因工作不负责任的情况，造成竞

赛程序无法继续进行，由赛项组委会视情节轻重，给予通报批评或停止工作，并通知其所在单位做出相应处理。

（五）裁判员须知

1、裁判员执裁前应参加培训，了解比赛任务及其要求、考核的知识与技能，认真学习评分标准，理解评分表各评价内容和标准。不参加培训的裁判员，取消执裁资格。

2、裁判员执裁期间，统一佩戴裁判员标识，举止文明礼貌，接受参赛人员的监督。

3、遵守执裁纪律，履行裁判职责，执行竞赛规则，信守裁判承诺书的各项承诺。服从赛项专家组和裁判长的领导。按照分工开展工作，始终坚守工作岗位，不得擅自离岗。

4、裁判员有维护赛场秩序、执行赛场纪律的责任，也有保证参赛选手安全的安全事故的责任。时刻注意参赛选手操作安全的问题，制止违反安全操作的行为，防止安全事故的出现。

5、裁判员不得有任何影响参赛选手比赛的行为，不得向参赛选手暗示或解答与竞赛有关的问题，不得指导、帮助选手完成比赛任务。

6、公平公正的对待每一位参赛选手，不能有亲近与疏远、热情与冷淡差别。

7、赛场中选手出现的所有问题如：违反赛场纪律、违反安全操作规程、提前离开赛场等，都应在赛场记录表上记录，并要求学生签工位号确认。

8、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；对评分表的理解和宽严尺度把握有分歧时，请示裁判长解决。严禁利用工作之便，弄虚作假、徇私舞弊。

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。

十、申诉与仲裁

1、各参赛队对不符合赛项规程规定的设备、工具、材料、计算机软硬件、竞赛执裁、赛场管理及工作人员的不规范行为等，可向赛项仲裁组提出申诉。

2、申诉主体为参赛队领队，申诉启动时，参赛队以该队领队签字同意的书面报告的形式递交赛项仲裁组。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。

3、提出申诉应在赛项比赛结束后2小时内提出，超过2小时不予受理。

4、赛项仲裁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，可由领队向大赛仲裁工作组提出申诉。大赛仲裁工作组的仲裁结果为最终结果。

5、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。

6、申诉方可随时提出放弃申诉。

十一、其他

本技术文件的最终解释权归大赛组织委员会。

附件：

中职组样题

理论知识（10%）

比赛时间：30分钟（30分钟后收取纸质试卷）

题型：单选、多选、判断，总共约5-10题左右

单选题：安装CentOS系统时，swap分区的大小建议设置为内存的多少倍？

- A. 0.5-1倍
- B. 1-2倍
- C. 2-3倍
- D. 3-4倍

单选题：计算机视觉的主要目标是什么？

- A. 模拟人类视觉能力
- B. 实现语言翻译
- C. 进行数据存储
- D. 开发操作系统

多选题：机器学习与深度学习的区别在于？

- A. 随着数据规模增大，深度学习性能提升更明显
- B. 机器学习需要手动提取特征
- C. 深度学习可自动提取特征
- D. 机器学习是人工智能的子领域，深度学习不是

多选题：安装CentOS系统时需要进行哪些设置？

- A. 时间设置
- B. 磁盘分配
- C. 网络设置
- D. 数据库配置

判断题：PyTorch只能在GPU环境下运行。

判断题：在AnacondaPrompt中输入“conda list”可以查看已安装的包。

模块A：Hadoop环境搭建与启动（40%）

比赛时间：60分钟

简介

选手需基于主流Linux操作系统，独立完成Hadoop环境的搭建与启动，掌握分布式系统基础环境的部署能力，具备系统配置、服务启动与验证的综合技术素养。

任务内容

技术规格要求：

- 操作系统：主流Linux发行版（自备设备）
- 软件环境：
 - Java 1.8 或兼容版本
 - Hadoop 3.3 或兼容版本
- 配置内容：
 - 环境变量配置（JAVA_HOME，HADOOP_HOME，PATH等）
 - Hadoop核心配置文件：
 - core-site.xml
 - hdfs-site.xml
 - mapred-site.xml
 - yarn-site.xml
- 启动验证：
 - 使用 jps 命令查看并确认5个核心进程已启动

评分细则：

- 1 Java版本验证（10分）
- 2 Hadoop版本验证（10分）
- 3 核心配置文件内容（30分）
- 4 jps进程（50分）

模块B：简单图像分类任务（50%）

比赛时间：90分钟

简介

选手需掌握基本的图像分类流程，使用调整后的Cifar10数据集完成数据清洗、模型训练与预测任务。

示例代码和数据发布在网盘：

链接：<https://pan.baidu.com/s/1sfPW-7adeXsLnSM4YgUbeQ?pwd=2rgk>

提取码：2rgk

文件存储要求

1. 在 D 盘创建Task02文件夹。
2. 文件夹中包含：
 - 2.1 "Original" 文件夹：存放数据集文件。
 - 2.2 "Final" 文件夹：存放最终提交的文件。
3. 所有文件命名使用英文，条理清晰。

任务内容

技术规格要求：

- 数据集：Cifar10（经过调整）
- 数据路径：
 - 训练集：./train/
 - 测试集：./test/
- 预处理要求：
 - 去除纯色图像
 - 图像去重
- 输出路径：
 - 清洗后训练集：./preprocessed_trainset/
 - 预测结果：./predictions/
- 模型要求：
 - 配置正确的深度学习环境
 - 使用清洗后数据训练分类模型
 - 对测试集进行预测并归类至对应类别文件夹

你必须提交：

- 最终提交至 "Final" 文件夹：
 - preprocessed_trainset/ 文件夹：包含所有清洗后的训练图像，保持原始目录结构

- predictions/ 文件夹：包含测试集所有图像的分类结果，按类别文件夹组织
- 将文件夹压缩为ZIP格式提交
- 命名规范：中职组-队伍编号.zip（包含Final文件夹与两个子文件夹）

评分细则：

- 1 数据预处理（50分）：
 - 1.1 纯色图像清除率（30分）
 - 1.2 去重正确性（20分）
- 2 分类准确率（50分）：测试集宏观准确率

高职组样题

理论知识（10%）

比赛时间：30分钟（30分钟后收取纸质试卷）

题型：单选、多选、判断，总共约5-10题左右

单选题：以下哪个是中国自主研发的深度学习框架？

- A. TensorFlow
- B. PyTorch
- C. PaddlePaddle
- D. Caffe

单选题：以下哪个命令可以查看Anaconda中已安装的包列表？

- A. `conda info`
- B. `conda list`
- C. `pip list`
- D. `python --version`

多选题：Anaconda安装完成后，可通过哪些方法验证安装成功？

- A. 启动Anaconda Navigator
- B. 执行`conda list`命令
- C. 检查环境变量设置
- D. 执行`python --version`命令

多选题：以下属于人工智能核心技术的有？

- A. 深度学习
- B. 机器学习
- C. 计算机视觉
- D. 数据结构

判断题：机器学习算法需要专家手动提取特征，而深度学习可以自动提取特征。

判断题：在CentOS系统安装时，boot分区建议设置为512MB-1024MB。

模块A：Hadoop环境搭建与启动（40%）

比赛时间：60分钟

简介

选手需基于主流Linux操作系统，独立完成Hadoop环境的搭建与启动，掌握分布式系统基础环境的部署能力，具备系统配置、服务启动与验证的综合技术素养。

任务内容

技术规格要求：

- 操作系统：主流Linux发行版（自备设备）
- 软件环境：
 - Java 1.8 或兼容版本
 - Hadoop 3.3 或兼容版本
- 配置内容：
 - 环境变量配置（JAVA_HOME，HADOOP_HOME，PATH等）
 - Hadoop核心配置文件：
 - core-site.xml
 - hdfs-site.xml
 - mapred-site.xml
 - yarn-site.xml
- 启动验证：
 - 使用 jps 命令查看并确认5个核心进程已启动

评分细则：

- 1 Java版本验证（10分）
- 2 Hadoop版本验证（10分）
- 3 核心配置文件内容（30分）
- 4 jps进程（50分）

模块B：精细化图像分类任务（50%）

比赛时间：90分钟

简介

选手需完成一个完整的图像分类任务，包括数据预处理、模型训练与测试预测。考察数据处理能力、模型构建与优化能力，以及任务流程的规范性。

示例代码和数据发布在网盘：

链接：<https://pan.baidu.com/s/1sfPW-7adeXsLnSM4YgUbeQ?pwd=2rgk>

提取码：2rgk

文件存储要求

1. 在 D 盘创建 Task02 文件夹。
2. 文件夹中包含：
 - 2.1 "Original" 文件夹：存放数据集文件。
 - 2.2 "Final" 文件夹：存放最终提交的文件。
3. 所有文件命名使用英文，条理清晰。

任务内容

技术规格要求：

- 数据集：food101（选取15个类别）
- 数据路径：
 - 训练集：./train_set/
 - 测试集：./test_set/
- 预处理要求：
 - 去除纯色图像
 - 图像去重
 - 调整图像分辨率
- 输出路径：
 - 清洗后训练集：./preprocessed_trainset/
 - 预测结果：./predictions/
- 模型要求：
 - 配置正确的深度学习环境
 - 使用清洗后数据训练分类模型
 - 对测试集进行预测并归类至对应类别文件夹

你必须提交：

- 最终提交至 "Final" 文件夹：

- preprocessed_trainset/ 文件夹：包含所有清洗后的训练图像，保持原始目录结构
- predictions/ 文件夹：包含测试集所有图像的分类结果，按类别文件夹组织
- 将文件夹压缩为ZIP格式提交
- 命名规范：高职组-队伍编号.zip（包含Final文件夹与两个子文件夹）

评分细则：

- 1 数据预处理（50分）：
 - 1.1 纯色图像清除率（20分）
 - 1.2 去重正确性（20分）
 - 1.3 图像分辨率一致性（10分）
- 2 分类准确率（50分）：测试集宏观准确率