**计算机应用技术专业（大数据技术与应用方向）**

**2018年秋季校外专业综合实训方案**

# **实训概要**

## **实训对象**

大数据3161（42人）、大数据3162（36人）和大数据3163（42人），共120人。

## **实训时间**

2018年12月3日至2018年12月22日，共80个学时（详见附件1）。

## **实训地点**

湖南省长沙市高新区延农路72号自兴人工智能大厦二楼。

## **实训目标**

本次实训是学生在完成大学阶段所有专业课程学习的基础上，综合运用所学知识、所掌握技能围绕一个工程项目进行数据清洗、分析及开发，培养学生分析解决问题能力，具体目标如下：

**1.4.1 知识目标**

1. 进一步熟练掌握Hadoop和Spark生态系统的基本原理；
2. 进一步掌握Python语言在大数据领域中的应用；
3. 进一步掌握大数据应用平台开发的方法；
4. 进一步掌握数据采集、清洗、存储、处理、分析的流程。

**1.4.2 技能目标**

1. 能够理解应用开发具体任务；
2. 能够进行需求调研，并撰写需求说明书；
3. 能够进行概要设计、数据库设计及系统架构设计；
4. 能够理解系统运行需求、硬件参数、部署环境；
5. 能够完成系统搭建并编写系统搭建文档；
6. 能够根据需求进行数据采集、清洗、存储、处理、分析、呈现等工作；
7. 能够对自己开发的应用代码进行调试和整改及发布应用。

**1.4.3 素质目标**

1. 培养学生的团队精神；
2. 培养学生的沟通能力；
3. 培养学生的创新精神。

## **实训任务**

完成一个大数据电商分析平台的数据采集、清洗、存储、处理、分析和呈现等任务。

## **实训流程**

1. 第一周：基础课为主；每天上午安排大数据专业教师集中上课，下午安排助教老师带领学生完成各项任务，晚上安排指导性学习。
2. 第二周：周一安排大数据专业教师和助教老师一起进行立项，包括项目实训整体流程、目标以及相关软件、数据准备等事项；其余时间由助教老师指导学生开展项目实训；实训后期大数据专业教师讲解机器学习算法课程，晚上安排指导性学习。
3. 第三周：由大数据专业教师点评回顾第二周的实训情况，根据学生前面实训完成情况进行分组，助教老师和班级中完成情况好的学生一起指导完成情况不太好的学生完成实训项目；实训后期大数据专业教师讲解机器学习算法实践，晚上安排指导性学习。

# **实训安排**

表 1 实训安排

|  |  |
| --- | --- |
| 日期 | 内容 |
| 2018年12月3日 | 1. 08:30~9:45 大巴接学生从暮云校区到自兴大厦   大巴安排：  龚师傅（18711083058）--湘AE0457  熊师傅（13707497098）--湘A24691  孟平（15873182861）--湘A21593  学校辅导员：  唐茹萍（15116358908）  刘璟（18973383131）   1. 09:45~10:30 住宿安顿 2. 10:30~11:10 实训欢迎仪式（二楼报告厅）   校方：  成奋华（人工智能学院院长）  高登（大数据教研室主任）  企业：  刘赛（自兴人工智能学院校企事业部负责人)  彭伟雄（自兴人工智能集团副总裁）  班主任、助教等其他人员   1. 11:10~11:30 入学物料发放 2. 11:30~12:30 午餐 |
| 2018年12月3日——  2018年12月22日 | 按《大数据专业综合实训流程表》、《课程表》进行实训 |
| 2018年12月22日上午 | 分班项目展示答辩，优秀学生甄选  校方：高登、张群慧、杨灿、刘璟、唐茹萍  企业：主讲教师、助教老师、课程设计人员 |
| 2018年12月22日下午 | 1. 14:00~16:30欢送仪式   校方：张群慧、杨灿、刘璟、唐茹萍  企业：企业相关负责人   1. 16:30 同学们登车回学校   唐茹萍（15116358908）  刘璟（18973383131） |

# **考核方式**

实训考核内容由作业检查、技能抽查、项目评审、项目答辩四部分组成。

**3.1 作业检查**

(1) 任课老师每周会对所带班级的作业完成情况进行≥3次的全面检查。

(2) 平时会对作业完成情况进行抽查，每班的作业抽查率将到达30%。

**作业标准:**

|  |  |  |
| --- | --- | --- |
| **级别** | **控制台或数据库作业** | **界面程序** |
| **优秀** | 1) 编码规范，代码可读性强；  2) 控制台程序有对算法或设计模式的运用；  3) 测试数据规范；  4) 数据库操作考虑到了执行性能最佳；  5）解决过程创新思路。 | 1) 界面设计美观；  2) 功能可操作性强；  3) 功能实现方式有创新；  4) 代码结构清晰，编码规范，可读性强。 |
| **良** | 1) 功能基本实现；  2) 代码有注释，编码规范欠缺；  3) 测试数据规范；  4) 数据库操作。 | 1) 功能基本实现；  2) 界面美观较差，不友善；  3) 代码结构清晰，编码规范，可读性强。 |
| **合格** | 1) 作业要求的主要功能基本实现；  2) 代码无注释，编码规范不规范；  3) 无测试数据。 | 1) 作业要求的主要功能基本实现；  2) 界面美观较差，不友善；  3) 代码结构较混乱，可读性差，编码不规范。 |
| **不合格** | 程序无法运行，但编程思路已经体现，代码可读性较高。 | |
| **差** | 程序无法运行，编程思路错误或无法体现，代码可读性差。 | |
| **缺交** | 无作业记录或作业抄袭他人。 | |

（注：作业完成情况---以任课老师上交的抽查结果为准）

**3.2技能抽查**

技能抽查是指专业老师会针对前一天所讲解的课程内容进行抽查，以了解学生对知识的掌握程度。通常会使用点名答问，现场编码等方式进行抽查，再根据学生回答情况进行综合评分，并记录在学生实训手册中。

**3.3 项目评审**

(1) 考核内容：项目完成情况，程序质量，文档，团队表现，个人表现。

(2) 项目启动：由指导老师在班级组织进行学生项目分组，并拟定开发项目。

(3) 项目抽检：由专业老师及助教老师对学生项目进行评审，从项目完成情况，程序质量，文档，团队表现，个人表现等几个方面考核学生项目情况。

其中项目完成情况包括:项目功能完整，界面交互性，代码规范程度。

(4) 项目评分：

项目成绩 = 项目完成情况\*30% + 程序质量\*20% + 文档\*10% + 团队表现\*20% + 个人表现\*20%

**3.4 项目答辩**

12月21日进行项目展示答辩，答辩小组成员由学校老师和企业老师组成。

3.5 综合评价

根据学生平时作业完成情况(20%)、技能抽检(20%)、项目评审(40%)和项目答辩情况(20%)，对每一位学生进行综合评价并打分，将成绩记录在实训日志上，并对学生在整个实训过程的表现，进行客观的评价并记录，甄选优秀学生。

## 4 校外实训中期检查

### 4.1检查小组成员

校方：成奋华、许彪、李尹、高登、刘璟、周志化、方芳、杨灿、张群慧、周茜

### 4.2检查方式

12月3日至12月21日，企业方每天在双方微信群中以照片形式展示学生生活、学习情况。学生通过微信群等向专业负责人、校内辅导员反馈学习、实训以及生活情况。

12月12日下午，由成奋华院长带队到企业了解学生生活状况，检查实训完成情况。

12月16日下午，毕业设计指导老师到企业了解学生实训项目完成情况，指导毕业设计。

# 附件1：大数据专业综合实训详细安排表

表 1 大数据专业综合实训详细安排表

|  |  |  |
| --- | --- | --- |
| 日期 | 星期 | 内容 |
| 2018年12月3日 | 一 | 下午：集中上课   1. 虚拟化 2. linux增强 3. shell编程 4. Zookeeper 5. Zookeeper 数据模型 6. Zookeeper shell 7. Zookeeper Watch 8. Zookeeper 选举机制 9. Zookeeper 典型应用 10. Zookeeper 客户端API |
| 2018年12月4日 | 二 | 1. 上午： 2. 虚拟机克隆练习 3. SSH分对称免密登陆练习 4. 文件上传下载sftp和lrzsz练习 5. linux查找、su、sudo、挂载、系统服务、网络管理、crontab配置练习 6. shell编程（流程控制、运算符、参数传递）练习 7. zookeeper shell增删改查节点练习 8. zookeeper watch练习 9. zookeeper 集群搭建 10. 下午： 11. Hadoop国内外应用 12. Hadoop集群-版本介绍 13. Hadoop集群介绍 14. Hadoop集群-环境准备 15. Hadoop编译 16. Hadoop集群 17. Hadoop集群体验 18. HDFS简介 19. HDFS特点 20. HDFS基本操作 21. HDFS基本原理 22. HDFS应用开发 |
| 2018年12月5日 | 三 | 1. 上午： 2. hadoop编译练习 3. hadoop集群练习 4. HDFS shell命令练习 5. HDFS java API 文件操作练习 6. 下午： 7. MapReduce计算模型思想 8. MapReduce计算模型设计思想 9. MapReduce编程规范 10. MapReduce 统计字符示例 11. MapReduce程序运行模式 12. MapReduce输入输出 13. MapReduce数据分区 14. MapReduce处理流程解析 15. MapReduce的序列化 16. MapReduce手机流量统计示例 17. MapReduce combiner组件 18. Scala 19. Spark core 20. Spark RDD 21. Spark 集群 |
| 2018年12月6日 | 四 | 1. 上午： 2. MapReduce 统计编程练习1 3. MapReduce 统计编程练习2 4. Spark 统计编程练习 5. 下午： 6. Hive安装部署 7. Hive交互方式 8. Hive数据模型 9. Hive基本操作DDL 10. Hive基本操作DML 11. Hive参数配置 12. Hive函数 13. Hive特殊分隔符 14. Hive DDL操作 15. Hive DML操作 16. Hbase Rowkey设计 17. Hbase版本选择 18. HBase架构体系、工作原理、Shell用法 19. HBase物理存储-Htable、HRegion、Hfile详解 20. HBase表、RowKey、列族、时间戳、Cell详解 21. Hbase-微博淘宝RowKey设计案例 22. Hbase-二级索引概念及案例 23. Hbase-API介绍和开发 24. Hbase调优 25. Hbase-Hive接口表 26. 电商网站使用Hbase场景 27. HBase流程算法详解，Region定位、RS和Master 28. Hbase-API开发 |
| 2018年12月7日 | 五 | 1. 上午： 2. flume spooldir-hdfs采集练习 3. flume tail-hdfs采集练习 4. flume load-balance采集练习 5. flume failover采集练习 6. flume扇入采集练习 7. flume扇出采集练习 8. sqoop mysql hdfs hbase 导入导出 9. 下午： 10. Flume运行机制 11. Flume采集系统结构图 12. Flume安装部署 13. Flume采集数据案例 14. Flume load-balance 15. Flume failover 16. Flume实战案例 17. Flume自定义拦截器 18. flume taildir 19. sqoop-从mysql导出到hdfs 20. sqoop-从hdfs导出到mysql 21. sqoop-从mysql导出数据到hbase 22. sqoop-从hbase导出数据到mysql |
| 2018年12月8日 | 六 | 1. 上午： 2. 机器学习概述 3. 分类、回归 4. 下午： 5. 聚类 6. 推荐算法 7. 模型评估与优化 |
| 2018年12月9日 | 日 | 放假 |
| 2018年12月10日 | 一 | 1. 上午：   企业教师进行立项，包括项目实训整体流程、目标以及相关软件、数据准备等事项   1. 下午：   大数据电商分析平台项目需求分析、功能设计 |
| 2018年12月11日 | 二 | 1. 上午：   数据采集（js sdk；java SDK）   1. 下午：   数据采集（Nginx网站服务搭建） |
| 2018年12月12日 | 三 | 1. 上午：   数据采集（flume采集数据，导入HDFS）   1. 下午：   数据采集（flume采集数据，导入HDFS） |
| 2018年12月13日 | 四 | 1. 上午：   数据迁移（mysql与hdfs、hive、hbase之间数据传输）   1. 下午：   数据迁移（mysql与hdfs、hive、hbase之间数据传输） |
| 2018年12月14日 | 五 | 1. 上午：   数据存储（HDFS日志原始数据存储、中间数据存储）   1. 下午：   数据存储（HDFS日志原始数据存储、中间数据存储） |
| 2018年12月15日 | 六 | 1. 上午：   数据清洗、分析（mapreduce日志数据清洗、ETL、批量数据分析）   1. 下午：   数据清洗、分析（mapreduce日志数据清洗、ETL、批量数据分析） |
| 2018年12月16日 | 日 | 放假 |
| 2018年12月17日 | 一 | 1. 上午：   数据清洗、分析（mapreduce日志数据清洗、ETL、批量数据分析）   1. 下午：   数据清洗、分析（mapreduce日志数据清洗、ETL、批量数据分析） |
| 2018年12月18日 | 二 | 1. 上午：   数据分析（hive统计分析）   1. 下午：   数据分析（hive统计分析） |
| 2018年12月19日 | 三 | 1. 上午：   数据存储（中间结果存储）、数据存储（mysql最终存储）   1. 下午：   数据存储（中间结果存储）、数据存储（mysql最终存储） |
| 2018年12月20日 | 四 | 1. 上午：（机器学习）   聚类（kmeans、DBSCAN等）   1. 下午：（机器学习）   推荐  模型评估与优化 |
| 2018年12月21日 | 五 | 大数据电商分析平台呈现 |
| 2018年12月22日 | 六 | 1. 上午   分班项目展示与答辩 |

# 附件2：管理制度

《自兴人工智能学院实训学生管理制度》

已阅签名表

备注：

1. 本制度包含宿舍管理规定、课堂管理规定、实训考核规定三部分内容。
2. 阅读并理解管理制度后，每位学生在下面对应栏签名。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **一班** | **签名** | **二班** | **签名** | **三班** | **签名** |
|  | 蒋振宇 |  | 谭逸霏 |  | 刘欢 |  |
|  | 杨明 |  | 彭程翔 |  | 吴卓锋 |  |
|  | 林之栖 |  | 唐柏蔚 |  | 李悍宇 |  |
|  | 彭智 |  | 陈亮 |  | 何明 |  |
|  | 唐肖南 |  | 荣安 |  | 陈园 |  |
|  | 张琪 |  | 陈磊 |  | 李宇翔 |  |
|  | 刘杰 |  | 肖佶达 |  | 郭浩 |  |
|  | 蒋延 |  | 胡祺 |  | 黎广 |  |
|  | 刘海量 |  | 吴将飞 |  | 彭宸 |  |
|  | 李炳志 |  | 刘博韬 |  | 周思京 |  |
|  | 陈慧 |  | 赵小蝶 |  | 雷熙宇 |  |
|  | 夏润泽 |  | 吴俊 |  | 周智勇 |  |
|  | 尹文杰 |  | 李彪 |  | 邹颖 |  |
|  | 李富祥 |  | 刘向阳 |  | 张文玉 |  |
|  | 刘铸 |  | 成磊 |  | 王毅 |  |
|  | 何峰 |  | 吴可 |  | 宋菲 |  |
|  | 何佳华 |  | 葛涓江 |  | 唐华 |  |
|  | 梁丽华 |  | 吴庆林 |  | 周佳军 |  |
|  | 刘媛媛 |  | 匡林东 |  | 蒋三峰 |  |
|  | 唐松芳 |  | 邓禹 |  | 范志科 |  |
|  | 郭沐霖 |  | 曹永胜 |  | 王涵 |  |
|  | 胡翼芳 |  | 周启宏 |  | 许贺添 |  |
|  | 张成明 |  | 王思晨 |  | 张英杰 |  |
|  | 陈志文 |  | 伍粤璋 |  | 汤国栋 |  |
|  | 钟著源 |  | 曾洲 |  | 何强 |  |
|  | 蔡豪 |  | 廖明德 |  | 罗慧 |  |
|  | 王敏雪 |  | 刘伟杰 |  | 覃泓宇 |  |
|  | 程子炎 |  | 陈轩云 |  | 许艳峰 |  |
|  | 张义麟 |  | 许海涛 |  | 邓子超 |  |
|  | 谢卓钊 |  | 张中科 |  | 喻银凤 |  |
|  | 吴伟 |  | 孙致远 |  | 王亨祥 |  |
|  | 张作武 |  | 邓垚 |  | 赵寅 |  |
|  | 王鹏瑞 |  | 汪岳晨 |  | 钟连金 |  |
|  | 黄邦龙 |  | 李南平 |  | 陈一星 |  |
|  | 廖金涛 |  | 廖湘玲 |  | 曾德辉 |  |
|  | 李天资 |  | 舒嘉琪 |  | 杨岳 |  |
|  | 郭莲 |  |  |  | 张冬枚 |  |
|  | 陈锡喜 |  |  |  | 鲁选勇 |  |
|  | 黎威盛 |  |  |  | 关永来 |  |
|  | 方思 |  |  |  | 肖鹏 |  |
|  | 周泽鑫 |  |  |  | 陈祺 |  |
|  | 黎亮 |  |  |  | 唐清勇 |  |

**自兴人工智能学院**

**2018年12月3日**

**自兴人工智能学院学生宿舍管理规定**

为进一步规范我院学生宿舍的各项工作，不断提高学生宿舍的管理服务水平，强化学生宿舍的育人功能，依据《自兴人工智能学院学生手册》等有关要求，结合我院实际情况，特制订本规定。

**第一条** 学生应遵守治安法律、法规，维护学生宿舍的安全。不得从事非法集会、游行，不得观看、下载、传播非法信息等破坏安定团结、危害国家安全的行为；

**第二条** 学生应提高警惕，外出随手关好寝室门窗，抵制上门推销，提高网络转账等防骗意识，自觉做好防盗、防骗、防火、防破坏、防自然灾害等工作，发现安全隐患和违法犯罪行为，及时向有关部门报告。

**第三条** 爱护学校寝室的公共财产（桌椅、柜子、门窗、床等），如有损坏照价赔偿。私人物品需自行保管好，并锁好放在柜子里，如因学生自身原因导致的财物损失由学生自行承担。上课、晚自习、休假期间或离开寝室并且寝室无人时，寝室成员必须关灯关水并锁好门窗。如有发现因学生自己出入未关门关窗而导致的财务损失，由学生自行承担。

**第四条** 严禁在校生出现偷盗公共物品财物和学生物品财物等行为，若情节严重涉及违法犯纪行为，则依法办理。

**第五条** 学生应自觉维护自身和他人人身财产安全。不得翻越宿舍围墙、栏杆；不得在宿舍内燃放烟花爆竹、焚烧物品和烧煮食物；不得携带危险刀具等危险物品（钢管、刀具、枪支等）、易燃、易爆、各种有毒危险物品或其他违规物品进入宿舍；不得有其它危及他人人身安全或有可能引发公共卫生事件的行为。

**第六条** 不得损坏、挪用或者擅自拆除、停用消防设施、器材；不得埋压、圈占、遮挡消火栓；不得占用、堵塞、封闭疏散通道、安全出口；不得擅自拆除、迁移、自增宿舍内的供电线路及设施；不得在寝室存放或使用热得快、电热杯、电饭煲、电热毯、烘鞋器、取暖器、电磁炉、紫砂锅、豆浆机、咖啡机、直板夹、卷发棒、电炉、电火锅、微波炉、煮蛋器、加湿器等违规电器和禁用的大功率电器。因煎药等特殊原因可持医院相关证明到所在宿舍管理员处办理使用手续。对未在规定内的新兴电器，不论功率大小，一律严格按照要求在规定期间内进行审批，审批同意方可使用，未经审批或审批未通过一律不能使用并不得存放在寝室。

**第七条** 学生自觉遵守请假制度和作息制度。除学院统一安排的集体活动外，学生因回家等原因，不能返院就寝者(包括双休日和节假日)，应事先向学院教质部各班班主任处办理好请假手续，并告知学院宿舍管理员，获准后方可离院。

**第八条** 学院门禁时间为：6:30-23:00。周五、周六晚上，学生必须在（每晚）23:00以前回寝室就寝，其它时间必须在（每晚）22:30前回寝室就寝，不得夜不归宿。因特殊原因迟归晚出者，须向宿舍管理员说明情况，并报给该学生所属班级班主任，宿舍管理员进行登记后方可进出宿舍。

**第九条** 寝室做到人走灯关，人离水断，养成随手关灯、关水的良好习惯。杜绝“长流水”“长明灯”的现象。

**第十条** 严禁在校生或者工作人员携带外来人员留宿。

**第十一条** 学生须遵守公共秩序，不得随地吐痰、乱丢垃圾、不得在学生宿舍区域内哄闹、影响他人正常学习生活。不得有赌博、酗酒、吸烟、吸毒、结伙滋事、打架斗殴及其它危及他人人身安全等违法、违规、违纪活动。

**第十二条** 学生不得损坏和拆卸学生宿舍内的公共设施和公共财物，不得擅自在墙壁和家具上刻写涂画，损坏者须照价赔偿。若学生存在故意破坏学院公共设施的行为，除照价赔偿外。

**第十三条** 宿舍所有区域禁烟，并设有烟感报警系统和手动报警系统。所有入住人员（包括学生、宿舍管理员和其他工作人员）以及被允许进入宿舍管理区域的所有外来人员，绝对不允许在宿舍任何区域出现抽烟、焚烧物品、烹饪食品等容易引发烟感报警的行为。

**第十四条** 各寝室自觉做好寝室日常内务卫生工作，寝室长负责安排和督促值日工作，寝室成员有义务保持自己物品的整齐和卫生，寝室卫生标准如下（后附安全卫生检查标准）：

1. 物品摆放整齐、床和柜子等设施不随意搬动；
2. 寝室内不能出现烟头、烟盒、酒瓶、违纪电器等；
3. 不乱挂衣物，乱扯电线；
4. 不在喷水系统上挂放重物；
5. 卫生间洗漱台干净整洁、杯子摆放整齐；
6. 蹲坑无明显污渍，厕所无异味。厕所垃圾要放在垃圾桶里。

**第十五条** 宿舍热水供应时间为每晚19:00-21:00，请学生在合理用水、节约用水的前提下安排好洗漱时间。

**第十六条** 本规定由自兴人工智能学院负责解释。本规定自发布之日起执行。

**学生课堂管理规定**

1. 按时上课，不迟到、早退和无故旷课。

2. 上课认真积极，严禁玩游戏、看视频、打游戏、聊天等,不做与课堂无关之事。

3. 上课期间不得随意出入教室。

4. 上课期间手机调至振动或者静音，严禁在课堂上接打电话和玩手机，并在上课期间把手机上交班主任或者放至手机袋中。

5. 尊重师长和同学，言辞文明，见面主动问好。

6. 爱护公共卫生，不随地吐痰，乱扔垃圾。

7. 不在教室和休息区吃饭，学生不允许带早餐进教室。

8. 不在学校内任何区域吸烟，不在走道等公共场合追逐打闹。

9．进入教学区域，每位学生必须佩戴学生证方可入校。

10.每天上完课后需要填写实训日志，每周上交老师检阅。