



湖南信息职业技术学院

网络安全学院

大数据技术应用专业 2024 届毕业设计

工作过程性材料

网络安全学院

2024 年 5 月 30 日

## 目 录

一、毕业设计时间流程安排.....	1
1. 毕业设计选题指导阶段.....	1
2. 任务下达阶段.....	1
3. 过程指导阶段.....	2
4. 成果答辩阶段.....	4
5. 资料整理阶段.....	8
6. 质量监控阶段.....	9
二、毕业设计情况分析报告.....	11
1. 选题分析.....	11
2. 成绩分析.....	12
3. 存在的问题.....	12
4. 改进措施.....	13

# 一、毕业设计时间流程安排

## 1. 毕业设计选题指导阶段

学校于 2023 年 9 月 4 日下发毕业设计实施工作方案，我院按照学校要求下达毕业设计工作任务，本专业迅速开展毕业设计工作。

### 1. 9 月 6 日开展毕业设计指导培训会



## 毕业设计



01 毕业设计的目的与要求

02 毕业设计的选题

03 毕业设计成果书撰写要求

04 毕业设计的答辩要求

05 毕业设计的成绩评定



01 毕业设计的目的与要求

1. 毕业设计的教学目的培养和提高学生综合运用所学的基础理论、专业知识和基本技能，分析、解决工程实际问题并进行初步的科学研究;培养学生独立工作能力，创新能力，以及理论联系实际和严谨求实的工作作风。

2. 毕业设计的要求

在毕业设计中，应注重以下几方面能力的培养，并根据专业培养目标与要求以及课题的实际情况有所侧重。

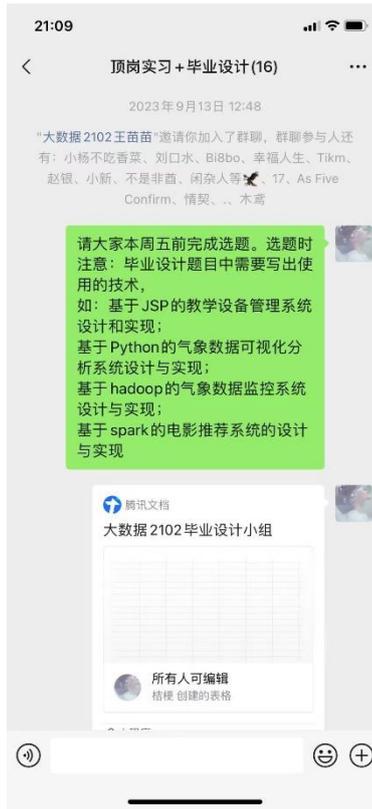
(1) 培养学生运用所学基本理论、基本知识和基本技能，分析与解决问题的能力。

(2) 培养学生严谨认真的科学态度、理论联系实际的工作作风。

(3) 通过完成具有一定实际或理论意义的工程项目开发，培养学生综合运用专业理论与知识分析解决实际问题的能力。

## 2. 任务下达阶段

9 月 13-15 日，给学生下达毕业设计选题任务。

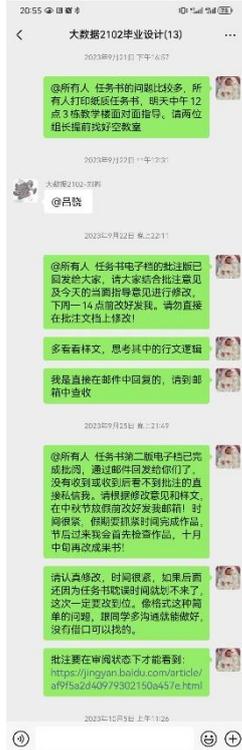


大数据2102毕业设计小组

	B	C	D	E
2	6	基于Python的气象数据可视化分析系统设计与实现	18507416534	天气预报数据采集、数据清洗、数据存储、数据分析、数据
3	12	基于Python的淘宝数据可视化分析系统设计与实现	17369238530	淘宝数据采集、数据清洗、数据分析、数据可视化
4	10	基于Python的网络游戏数据可视化分析	18664066538	网络游戏数据采集、数据清洗、数据分析、数据可视化
5	11	基于python的视频电影数据爬取可视化分析	17773126923	视频数据采集、数据清洗、数据分析、数据显示
6	9	基于Python的旅游攻略数据可视化分析系统设计与实现	17371128843	旅游攻略数据采集、数据清洗、数据分析、数据可视化。
7	2	基于Python的音乐数据可视化分析系统设计与实现	17788924455	音乐数据采集、数据清洗、数据分析、数据可视化。
8	5	基于大数据的NBA球员数据分析与可视化	19158293018	数据采集、数据清洗、数据分析、数据可视化
9	4	基于Python的电影评分数据可视化分析系统设计与实现	19896259941	数据采集、数据清洗、数据分析、数据可视化
10	3	基于python的贝壳上海房价数据分析系统设计与实现	18874109755	房价数据采集、数据清洗、数据分析、数据可视化
11	8	基于springboot的个人理财管理系统	17108329770	用户管理、理财产品管理、贷款管理、个人信息管理
12	202127550213	基于python 实现电器的全面筛查	17508477769	数据采集、
13		基于Python的笔记本电脑销售数据的采集与可视化		数据清洗、数据可视化
14	15	基于Python的网络小说数据可视化分析系统设计与实现	19848198780	数据采集、编辑、开启筛选、查找替换、操作
15	7	基于Python可视化的汉服形制调查	17267311871	数据采集、数据清洗、数据可视化
16				

### 3. 过程指导阶段

2023年9月-2023年11月指导学生完成毕业设计成果。

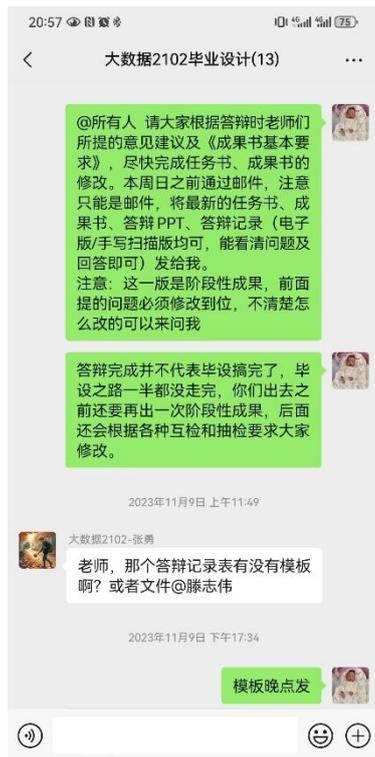


A	B	C	D
姓名	学号	选题	修改建议
杨思成	2	基于Python的音乐数据可视化分析系统设计与实现	无
赵银	3	基于python的贝壳上海房价数据分析系统设计与实现	无
刘亦雄	4	基于Python的电影评分数据可视化分析系统设计与实现	无
刘可	5	基于大数据的NBA球员分析与可视化	基于大数据的NBA球员数据分析与可视化
何琦	6	基于Python的气象数据可视化分析系统设计与实现	无
陈利	7	基于Python可视化的汉服形制调查	基于Python的汉服数据分析系统设计与实现
阳叶群	8	基于springboot的个人理财管理系统	基于springboot的个人理财管理系统设计与实现
何震东	9	基于Python的旅游攻略数据可视化分析系统设计与实现	无
刘旭	10	基于Python的网络游戏数据可视化分析	基于Python的网络游戏数据可视化分析系统设计与实现
陈友明	11	基于python的视频电影数据爬取可视化分析	基于python的电影数据可视化分析系统设计与实现
刘洋	12	基于Python的淘宝数据可视化分析系统设计与实现	无
王苗苗	14	基于python的大众点评网红餐厅评分可视化分析	基于python的大众点评网红餐厅数据可视化分析系统设计与实现
唐安勇	15	基于Python的网络小说数据可视化分析	基于Python的网络小说数据可视化分析系统设计与实现
黄宇	13	基于python 实现电器的全面筛查	基于python 电器数据分析系统设计与实现

姓名	学号	选题	成果书修改意见
李添福	1	基于SpringBoot的图书系统的设计与实现	1、目录中出现了任务人员管理模块等与课题无关的内容。2、图4-1 系统功能结构图中的用户界面，建议修改为用户。3、4.2系统详细设计缺少图书分类管理的设计。4、5 系统实现缺少图书分类管理的实现。
赵银	3	基于python的贝壳上海房价数据分析系统设计与实现	1、4.1系统详细设计中的系统整体架构内容和图调整到第2张2.2节处。2、第5章多次提到的基于Python的杭州房价与课题内容不符。3、横轴比例建议采用饼图来可视化。房屋类型数量对比情况建议采用柱状图可视化。
何琦	6	基于Python的气象数据可视化分析系统设计与实现	1、成果书没有封面。2、目录中的页码显示有问题，第一章从页码3开始。3、2.2节技术架构需增加技术架构图。4、成果书格式不对，请按照样例中的要求设置文字字体和段落等格式。5、标题5.4.5(少)每周最高最低温度中的饼图，看不懂展示了什么内容。6、6.3测试结果分析提到豆瓣读书数据采集分析与可视化系统，也与课题无关。7、部分参考文献时间太老，建议选择近3年的参考文献。
何震东	9	基于Python的旅游攻略数据可视化分析系统设计与实现	1、目录中的页码显示有问题，第一章从页码1开始。2、1.2课题目标参考样例图片修改。3、2.2节技术架构需增加技术架构图。4、请按照样例修改章节内容。缺少系统设计，第3、4、5章建议整合到系统实现中。5、参考文献建议选择近3年的文献。
刘旭	10	基于Python的网络游戏数据可视化分析系统设计与实现	1、课题题目建议修改为：基于Python的网络游戏数据可视化分析系统设计与实现。2、目录中的页码显示有问题，第二章从也页码1开始。3、字数不足8000。4、成果书格式不对，请按照样例中的要求设置格式。
陈友明	11	基于python的电影数据可视化分析系统设计与实现	1、5.1中提到对豆瓣书籍网站进行数据获取，与课题内容不符。2、5.3数据可视化模块中，多张图展示的年份截止到2018年，建议展示数据的年份要到2022年。因为5年前的数据分析结果意义不大。3、图5-18每年图书出版量，图的名称与课题内容不符。4、总结中，毕业设计从4月份到7月份，时间说明有误。
刘洋	12	基于python的京东商品评论数据可视化分析系统设计与实现	1、目录中的页码显示有问题，第一章从页码1开始。2、成果书格式不对，请按照样例中的要求设置格式。3、2.2 技术架构内容为空。4、请按照样例修改章节内容。5、第五章:情感分析，6、6.1 数据可视化结果展示，除了图片，还需要文字阐述此图具体可视化什么数据，得出什么结论。7、可视化只有3张图，分析的内容太少。至少要从5个方面进行数据展示和分析。8、6.2 情感分析结果展示提到了基于机器学习的情感分析模型，可是成果书中并没有具体介绍该模型。而采用的词云图展示结果，这个与基于机器学习的情感分析模型有什么关系？9、商品的种类很多，请从某种产品的评论来完成你的毕业设计。

## 4. 成果答辩阶段

11月初完成毕业设计答辩工作。



湖南信息职业技术学院 2024 届学生毕业设计成绩汇总表

序号	姓名	班级	学号	毕业设计 课题名称	成果 成绩	答辩 成绩	总成 绩	等级
1	梁翔恺	大数据 2103	2021275 50302	基于大数据技术的Boss直 聘网页数据分析与可视化	72	80	76	良
2	张磊	大数据 2103	2021275 50304	基于仿真电商数据构建实 时数据库与可视化平台	70	80	75	良
3	刘志豪	大数据 2103	2021275 50306	基于大数据技术的离线数 据仓库设计与实现	70	80	75	良
4	何超群	大数据 2103	2021275 50308	基于Python的影视大全数 据分析与可视化	65	70	65	合格
5	周宇星	大数据 2103	2021275 50310	基于大数据技术的当当网 图书数据分析与可视化	65	67	66	合格
6	王志	大数据 2103	2021275 50312	基于Python的晋江文学网 数据分析与可视化	67	67	67	合格
7	邓远芳	大数据 2103	2021275 50314	笔趣阁小说数据的分析与 可视化	67	67	67	合格
8	何珺	大数据 2103	2021275 50316	链家网成都二手房数据的 采集与可视化	65	67	66	合格
9	刘帅	大数据 2103	2021275 50318	基于飞卢小说网的数据采 集与可视化	65	67	66	合格
10	汪凯文	大数据 2103	2021275 50320	基于Python的江山文学网 数据采集与可视化	67	65	66	合格
11	肖芳琪	大数据 2103	2021275 50322	基于Python的汽车之家二 手车数据分析与可视化	67	67	67	合格
12	李艳芝	大数据 2103	2021275 50324	基于大数据技术的豆瓣图 书数据分析与可视化	65	70	65	合格
13	吴庆荣	大数据 2103	2021275 50326	基于Python的搜狐二手车 网数据分析与可视化	65	67	66	合格
14	张良文	大数据 2103	2021275 50328	茶杯狐网页电影数据分 析与可视化	65	67	66	合格
15	蒲蓉蓉	大数据 2103	2021275 50330	基于Python的北京新发地 数据分析与可视化	65	65	65	合格

二级学院 (盖章):

填报时间: 2024 年 5月15日

注: 成绩评定分四个等级: 优、良、合格、不合格。优: 总成绩 $\geq 85$ 分; 良:  $70 \leq$ 总成绩 $< 85$ 分; 合格:  
 $60 \leq$ 总成绩 $< 70$ 分; 不合格: 总成绩 $< 60$ 分。

湖南信息职业技术学院 2024 届学生毕业设计成绩汇总表

序号	姓名	班级	学号	毕业设计 课题名称	成果 成绩	答辩 成绩	总成绩	等级
1	曾志超	大数据 2103	202127 550301	基于Python的房天下数据 采集与可视化	66	66	66	合格
2	胡朵怡	大数据 2103	202127 550303	基于Python的湖南省天气 分析与可视化	68	70	69	合格
3	孟粤晨	大数据 2103	202127 550305	基于网易云音乐数据的分 析和可视化	65	67	66	合格
4	朱炎森	大数据 2103	202127 550307	基于Python的城市租房数 据分析与可视化	65	70	65	合格
5	许琦	大数据 2103	202127 550309	基于Python的哔哩哔哩视 频数据分析可视化	65	67	66	合格
6	周鹏程	大数据 2103	202127 550311	基于Python的新华书店的 书籍数据采集和可视化	67	67	67	合格
7	杨威	大数据 2103	202127 550313	基于Python的旅游景点数 据分析可视化	67	67	67	合格
8	熊日一	大数据 2103	202127 550315	基于python爬虫爬取17k 小说网站评论数据的可视 化与分析	65	67	66	合格
9	罗鑫武	大数据 2103	202127 550317	基于大数据技术的二手房 数据分析与可视化	66	68	67	合格
10	曹东	大数据 2103	202127 550319	基于拉勾招聘数据的分析 与可视化	67	65	66	合格
11	吴锦华	大数据 2103	202127 550321	基于孔夫子旧书网的数据 分析和可视化	60	60	60	合格
12	朱柳	大数据 2103	202127 550323	基于IMDB影视数据的采集 与可视化	65	70	65	合格
13	卿逸	大数据 2103	202127 550325	基于Python的当当网服装 数据分析与可视化	65	67	66	合格
14	邹志	大数据 2103	202127 550327	推书君网页小说数据分析 可视化	65	67	66	合格
15	伍亚鹏	大数据 2103	202127 550329	基于大数据下对AcFun的 数据进行分析 and 可视化	65	65	65	合格

二级学院(盖章):

填报时间: 2024 年 5月15 日

注: 成绩评定分四个等级: 优、良、合格、不合格。优: 总成绩 $\geq 85$ 分; 良:  $70 \leq$ 总成绩 $< 85$ 分; 合格:  
 $60 \leq$ 总成绩 $< 70$ 分; 不合格: 总成绩 $< 60$ 分。

**湖南信息职业技术学院 2024 届学生毕业设计  
评阅、答辩及成绩评定表**

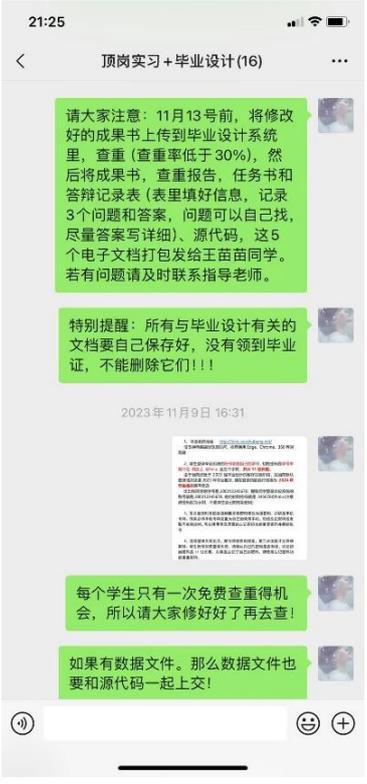
课题名称	基于Python的音乐数据可视化分析系统设计与实现						
姓 名	杨思成	系别	网络空间安全学院	专业 / 班级	大数据2102	学号	202127550202
指导教师	邓慈云	所在部门	网络空间安全学院	职务	教师	职称	讲师
答 辩 与 会 人	姓名	职务 (职称)	姓名	职务 (职称)	姓名	职务 (职称)	
	邓慈云	讲师	蒋亮	讲师	滕志伟	讲师	
<p>指导教师评语 (主要对学生毕业设计的工作态度、研究内容与方法、工作量、文献应用、创新性、实用性、科学性及存在的不足等进行综合评价) :</p> <p>成果书结构合理, 逻辑思路清晰, 格式规范, 总体上达到了毕业设计要求。</p> <p>成绩: <u>66</u>                  指导教师签名: 邓慈云                  2023年 11月 7 日</p>							
<p>答辩记录:</p> <p>1.关于可视化: 你的可视化展示非常清晰, 特别是通过柱状图呈现了网易云音乐流行歌单的播放 TOP10 和收藏 TOP10。这些图表很生动地展示了用户对歌曲的喜好和收藏情况。你是否考虑过在这些图表中加入一些建议或者见解, 以便更好地解释数据背后的趋势?</p> <p>关于可视化: 感谢您的建议! 我确实考虑过在图表中加入一些见解或建议, 以更好地解释数据的趋势。例如, 可以添加一些标注, 说明某些歌曲为什么在播放和收藏方面表现出色。可能是因为艺术家的知名度、歌曲风格或者其他特定因素。这样的补充信息将使图表更具信息量。</p> <p>2.数据清洗: 在第一个代码块中, 你使用了 pd.to_numeric 将 "play" 列转换为数值类型, 这是一个很好的处理方法。但在第二个代码块中, 对 "collection" 列的清洗是手动进行的。是否考虑过使用类似的自动化方法, 以便更好地适应未来可能的数据变化?</p> <p>数据清洗: 您提到的是一个很好的观点。对于 "collection" 列的手动清洗确实略显繁琐, 可以考虑使用类似的自动化方法, 以适应未来可能的数据变化。比如, 使用 Pandas 的 apply 函数结合一个自定义的清洗函数, 可以更灵活地处理不同情况。这样的话, 可以更容易地应对未来可能的数据变化。</p> <p>3.数据库存储: 目前你选择将数据保存在 CSV 文件中, 这是一个合理的选择。不过, 考虑到你的数据可能会持续增长, 是否考虑过将数据存储在一个数据库中, 以便更好地管理和查询大量数据?</p> <p>数据库存储: 目前选择将数据保存在 CSV 文件中是合理的, 但随着数据的增长, 数据库可能是更好的选择。MySQL 已经是一个不错的选项, 它提供了稳定、高效的数据存储和查询功能。将数据存储数据库中可以更好地管理和查询大量数据, 尤其是当数据量逐渐增大时。</p> <p>会议主持人: 蒋亮                  记录人: 邓慈云                  2023年 11月 7 日</p>							
<p>答辩小组意见: 答辩合格</p> <p>评语:</p> <p>符合毕业设计要求。</p> <p>成绩: <u>66</u>                  答辩小组 (组长) 签名: 蒋亮 2023 年11月 7 日</p>				<p>评定成绩: <u>66</u></p> <p>评定等级: <u>及格</u></p> <p>答辩委员会 (主任): 刘彪 2023 年11月 7 日</p>			

湖南信息职业技术学院 2024 届学生毕业设计  
评阅、答辩及成绩评定表

课题名称	基于大数据技术的Boss直聘网页数据分析与可视化						
姓名	梁栩恺	系别	网络空间安全学院	专业/班级	大数据技术/2103	学号	02
指导教师	唐紫璐	所在部门	网络空间安全学院	职务	无	职称	讲师
答辩与会人员	姓名	职务(职称)	姓名	职务(职称)	姓名	职务(职称)	
	刘彪	讲师	唐紫璐	讲师	滕志伟	讲师	
该生毕业设计展现出积极态度，工作量大且深入细致，具备一定的实用性，但仍有提升空间。							
成绩： 72		指导教师签名：唐紫璐			2023年11月7日		
<p>答辩记录：</p> <p>问题1：一共有多少条数据，在处理之后有多少条数据，用到了什么来处理。 答：我一共爬取了八万多条数据，在处理之后共有67640条数据，使用的是SparkSQL来处理数据，对数据进行了去重去空的操作。</p> <p>问题2：画了多少张图，用了多少个字段。 答：一共画了18张图，在爬到的数据中，爬了20个字段，并且也用到了20个字段</p> <p>问题3：对于薪资分布这张图，你做了什么。 答：首先，我随机抽取了20条数据，然后讲薪资分为了0-10k，10-20k，20-30k，30-40k，然后将这20条数据放在上面进行可视化</p>							
会议主持人：唐紫璐		记录人：滕志伟			2023年11月7日		
<p>答辩小组意见：</p> <p>评语： 答辩合格</p> <p>成绩： 80 答辩小组(组长)签名：刘彪</p> <p style="text-align: right;">2023年11月14日</p>				<p>评定成绩： 76</p> <p>评定等级： 良好</p> <p>答辩委员会(主任)：刘彪</p> <p style="text-align: right;">2023年11月14日</p>			

## 5. 资料整理阶段

由学生将毕业设计相关成果上传到毕业设计平台，并安排相关负责人收集毕业设计最终资料。



## 6. 质量监控阶段

学院开展了多次校内互检和交叉互检。

搜索学生:  查重状态:  全部  已查重  未查重 是否合格:  全部  合格  不合格

序号	学生基本信息			文档基本信息			文档查重		文档格式检测		文档检测	
	所在班级	姓名	学号	文件名称	字符数	参考文献数	总体相似度(%)	查重报告	错误数	检测报告	提交方式	检测状态
1	大数...	李添福	201927550407	07李添福-成果书-...	16585	10	15.94%	<a href="#">下载</a>			导师提交	检测完成
2	大数...	杨思成	202127550202	2杨思成-成果书-基...	22328	12	9.26%	<a href="#">下载</a>			导师提交	检测完成
3	大数...	赵银	202127550203	03赵银成果书基于...	11060	9	10.22%	<a href="#">下载</a>			导师提交	检测完成
4	大数...	刘亦雄	202127550204	04刘亦雄成果书基...	14887	11	9.82%	<a href="#">下载</a>			导师提交	检测完成
5	大数...	刘可	202127550205	05刘可基于大数据...	16542	9	11.21%	<a href="#">下载</a>			导师提交	检测完成
6	大数...	何琦	202127550206	06何琦成果书基于...	15813	11	5.65%	<a href="#">下载</a>			导师提交	检测完成

### 禁用词检测结果

2843	网络安全学院	大数据技术	大数据2101班	罗永坚	邓思怡	202127550132	《邓思怡32号任务书(1).doc》文档内题目与毕设系统内指导老师给学生指定的题目不一致
2844	网络安全学院	大数据技术	大数据2101班	罗永坚	范芳招	202127550112	
2845	网络安全学院	大数据技术	大数据2101班	罗永坚	魏嘉婷	202127550126	
2846	网络安全学院	大数据技术	大数据2101班	罗永坚	黎金博	202127550124	《24黎金博任务书.doc》文档内题目与毕设系统内指导老师给学生指定的题目不一致
2847	网络安全学院	大数据技术	大数据2101班	罗永坚	李臣超	202127550116	
2848	网络安全学院	大数据技术	大数据2101班	罗永坚	李季	202127550130	
2849	网络安全学院	大数据技术	大数据2101班	罗永坚	彭标	202127550138	
2850	网络安全学院	大数据技术	大数据2101班	罗永坚	任伟	202127550102	《基于python的欣欣旅游数据分析与可视化任务书.doc》有5处禁用词“论文”
2851	网络安全学院	大数据技术	大数据2101班	罗永坚	余曾栋	202127550140	
2852	网络安全学院	大数据技术	大数据2101班	罗永坚	肖锦菊	202127550114	
2853	网络安全学院	大数据技术	大数据2101班	罗永坚	谢隆源	201927550127	
2854	网络安全学院	大数据技术	大数据2101班	罗永坚	谢美晶	202127550108	
2855	网络安全学院	大数据技术	大数据2101班	罗永坚	严康	202127550136	
2856	网络安全学院	大数据技术	大数据2101班	罗永坚	杨园丽	202127550110	《杨园丽毕业设计任务书.docx》文档内题目与毕设系统内指导老师给学生指定的题目不一致
2857	网络安全学院	大数据技术	大数据2101班	罗永坚	杨志民	202127550128	《杨志民任务书.doc》文档内题目与毕设系统内指导老师给学生指定的题目不一致
2858	网络安全学院	大数据技术	大数据2101班	罗永坚	余远志	202127550120	
2859	网络安全学院	大数据技术	大数据2101班	罗永坚	曾全	202127550134	
2860	网络安全学院	大数据技术	大数据2101班	罗永坚	湛高峰	202127550118	
2861	网络安全学院	大数据技术	大数据2101班	罗永坚	钟博文	202127550122	《大数据2101-钟博文(任务书)1.doc》有1处禁用词“论文”，文档内题目与毕设系统内指导老师给学生指定的题目不一致
2862	网络安全学院	大数据技术	大数据2101班	罗永坚	周舒婷	202127550106	

### 毕业设计校内互检结果

序号	院系名称	班级名称	指导老师	学生姓名	学生学号	查重文档名称	查重状态	查重结果	最大单篇相似度(%)
1	网络安全学院	大数据2101班	罗永坚	谢隆源	201927550127	"毕业设计成果书-谢隆源.doc"	检测完成	通过	6.49%
2	网络安全学院	大数据2101班	罗永坚	任伟	202127550102	"基于python的欣欣旅游数据分析与可视化成果书.doc"	检测完成	通过	8.21%
3	网络安全学院	大数据2101班	罗永坚	周舒婷	202127550106	"基于JAVA的校园二手交易平台的设计与实现.doc"	检测完成	通过	7.49%
4	网络安全学院	大数据2101班	罗永坚	谢美晶	202127550108	"成果书-谢美晶-基于Spark的母婴玩具数据分析系统的设计与实现.doc"	检测完成	通过	4.14%
5	网络安全学院	大数据2101班	罗永坚	杨园丽	202127550110	"杨园丽毕业设计成果书.docx"	检测完成	通过	2.78%
6	网络安全学院	大数据2101班	罗永坚	范芳招	202127550112	"大数据2101范芳招-成果书.doc"	检测完成	通过	9.03%
7	网络安全学院	大数据2101班	罗永坚	肖锦菊	202127550114	"基于java的小型论坛管理系统设计与实现成果书.doc"	检测完成	通过	18.93%
8	网络安全学院	大数据2101班	罗永坚	李臣超	202127550116	"大数据2101-李臣超(成果书)1.doc"	检测完成	通过	3.22%
9	网络安全学院	大数据2101班	罗永坚	湛高峰	202127550118	"2023-7-6-成果书-湛高峰-基于spark的房价数据分析系统的设计与实现(1).doc"	检测完成	通过	8.56%
10	网络安全学院	大数据2101班	罗永坚	余远志	202127550120	"2023届德核-毕业设计成果(数据分析与可视化方向)(2).doc"	检测完成	通过	2.64%
11	网络安全学院	大数据2101班	罗永坚	钟博文	202127550122	"钟博文-游戏体验(1)+++20240422.docx"	检测完成	通过	5.29%
12	网络安全学院	大数据2101班	罗永坚	黎金博	202127550124	"24黎金博成果书(1)-20240424.docx"	检测完成	通过	3.48%
13	网络安全学院	大数据2101班	罗永坚	魏嘉婷	202127550126	"基于python的茶叶销售数据采真与可视化成果书.doc"	检测完成	通过	6.73%
14	网络安全学院	大数据2101班	罗永坚	杨志民	202127550128	"基于Python的淘宝电器销售数据分析与可视化.doc"	检测完成	通过	10.81%
15	网络安全学院	大数据2101班	罗永坚	李季	202127550130	"毕业设计成果书1.docx"	检测完成	通过	2.87%
16	网络安全学院	大数据2101班	罗永坚	邓思怡	202127550132	"邓思怡32号成果书(2).doc"	检测完成	通过	3.75%
17	网络安全学院	大数据2101班	罗永坚	曾全	202127550134	"基于Python的长沙房屋出租数据分析与可视化成果书.doc"	检测完成	通过	6.22%
18	网络安全学院	大数据2101班	罗永坚	严康	202127550136	"毕业设计成果(数据分析与可视化方向)大数据2101严康36.doc"	检测完成	通过	11.97%
19	网络安全学院	大数据2101班	罗永坚	彭标	202127550138	"大数据2101-彭标(成果书).doc"	检测完成	通过	3.88%
20	网络安全学院	大数据2101班	罗永坚	余曾栋	202127550140	"大数据2101-余曾栋(成果书)2.doc"	检测完成	通过	4.93%
21	网络安全学院	大数据2101班	彭成辉	彭喻新	202127550101	"基于Python的民宿数据分析与可视化系统-(1).doc"	检测完成	通过	11.12%
22	网络安全学院	大数据2101班	彭成辉	马恩泽	202127550103	"大数据2101马恩泽成果书.doc"	检测完成	通过	3.23%
23	网络安全学院	大数据2101班	彭成辉	陈定福	202127550105	"大数据2101陈定福成果书.doc"	检测完成	通过	3.83%
24	网络安全学院	大数据2101班	彭成辉	龙骥般	202127550107	"大数据2101龙骥般成果书.doc"	检测完成	通过	1.20%

校内抽检名单

学号	学生姓名	选题名称	链接地址	指导老师	结论 (合格/不合格)	是否上传毕 业设计任务 书	是否上传 毕业设计 成果	是否存在选题名 、学生姓名、指 导教师等信息	毕业设计任务书是否 明确毕业设计任务及 要求、进程安排	其它问题 (有/无, 如有记 录下具体记录)	检查人	复核人
02127520325	李雨婷	基于springboot框架海洋生态知识管理系统的设计与实现	<a href="http://hniu.xueshubang.net/s/10139">http://hniu.xueshubang.net/s/10139</a>	邓慧云	合格	是	是	否	是	无	彭丽玲	姚亚琳
02127550107	龙瑞毅	基于Spark的保险索赔数据分析与可视化	<a href="http://hniu.xueshubang.net/s/8687">http://hniu.xueshubang.net/s/8687</a>	彭成辉	合格	是	是	否	任务书排版乱了	任务书委员会不规范	张冬梅	陈东升
02127550133	黄芝蔚	用户行为数据分析与可视化	<a href="http://hniu.xueshubang.net/s/8700">http://hniu.xueshubang.net/s/8700</a>	彭成辉	合格	是	是	成果书封面下划线没有对齐,完成时间和任务书不一致	任务书的字体格式行间距不一致		张媛侠	陈东升
02127550204	刘亦维	基于Python的电影数据分析系统设计及实现	<a href="http://hniu.xueshubang.net/s/10095">http://hniu.xueshubang.net/s/10095</a>	邓慧云	合格	是	是	无	任务书中的“课题简介”要缩进两格		李凤娥	陈东升
02127550210	刘旭	基于Python的网络游戏数据可视化分析系统设计及实现	<a href="http://hniu.xueshubang.net/s/10101">http://hniu.xueshubang.net/s/10101</a>	邓慧云	合格	是	是	学生单独任务书和成果不一致	任务书时间2023年	封面没有对齐,参考文献格式有问题	陈丽萍	陈东升
02127550214	王茜茜	基于python的笔记本电脑销售数据的采集与可视化	<a href="http://hniu.xueshubang.net/s/10105">http://hniu.xueshubang.net/s/10105</a>	邓慧云	合格	是	是	查重报告不是同一课题	任务书时间只到2023年	成果书封面没对齐,参考文献格式有问题	陈丽萍	陈东升
02127550215	唐安勇	基于python的网络小说可视化分析系统设计及实现	<a href="http://hniu.xueshubang.net/s/10106">http://hniu.xueshubang.net/s/10106</a>	邓慧云	合格	是	是	无	任务书时间只到2023年,任务书中课题目标不完整;成果封面没对齐;任务书和成果中的参考文献格式不对		柏雷	陈东升
02127550217	陈利	基于Python的热门歌曲数据分析系统	<a href="http://hniu.xueshubang.net/s/10347">http://hniu.xueshubang.net/s/10347</a>	蒋亮	合格	是	是	任务书课题名称和成果名称不一致,少了两个字	任务书时间只到2023年,任务书没有做研究生和院部盖章和签名;成果封面没对齐;目录格式不对		柏雷	陈东升
02127550218	方诗鑫	基于Python的新能源汽车数据可视化分析系统	<a href="http://hniu.xueshubang.net/s/10348">http://hniu.xueshubang.net/s/10348</a>	蒋亮	合格	是	是	任务书格式需要调整,空白太多	任务书时间只到2023年	无	赵玉婧	陈东升
02127550220	刘迎春	基于Python爬虫的多平台小说爬虫系统	<a href="http://hniu.xueshubang.net/s/10350">http://hniu.xueshubang.net/s/10350</a>	蒋亮	合格	是	是	成果封面格式需要调整	任务书时间只到2023年	无	赵玉婧	陈东升

## 二、毕业设计情况分析报告

本次 2024 届大数据技术专业毕业班 3 个,毕业生 114 人,毕业设计全部合格。大数据技术专业毕业设计包括了毕业设计任务书、毕业设计成果书和毕业设计源代码等材料。在整个过程中,学生们积极准备,认真投入,从确定毕业设计选题、撰写任务书,设计和开发系统,毕业设计答辩等。指导教师们也认真负责,全程给予指导、督促和反馈,确保毕业设计的顺利完成。通过完成毕业设计,提高了学生综合运用所学的基础理论、专业知识和基本技能,分析、解决实际问题,培养学生独立工作能力,创新能力,以及理论联系实际和严谨求实的工作作风。

### 1. 选题分析

结合本专业的培养计划,2021 级学生的毕业设计选题主要涵盖了大数据专业的相关技术,如数据采集,数据清洗、数据分析和可视化。学生可以选择研究分布式计算框架(如 Hadoop、Spark 等)的原

理与实践，或者运用数据清洗、数据分析和可视化等方面的技术，具有一定的多样性。选题贴近生活实际，富有创意和教育价值。例如，基于 Spark 的电影数据分析与可视化系统的设计与实现，基于 python 的京东电子产品销售数据分析与可视化，基于 Python 的新能源汽车数据分析与可视化。然而，也有部分选题涉及范围太宽泛或实践性不强问题，需要指导教师帮助他们重新调整选题。

## 2. 成绩分析

毕业设计成绩由指导老师评定的成绩、答辩成绩共同折合成总成绩，并以等级的形式标定。从最后结果来看，全部学生毕业设计都合格，毕业设计在各个方面基本达到要求，但在系统功能方面还有完善和提升的空间。毕业设计等级为良好的数量只有 9 人，可以看出毕业设计的创新性和作品细节设计等方面还有提升的空间。

## 3. 存在的问题

### （一）毕业设计任务书中存在的问题

毕业设计任务书时间分配不合理。

### （二）毕业设计作品中存在的问题

（1）格式错误：毕业设计成果书中目录页面的设置、图片和表格的编号等名称错误。

（2）动力不足或拖延，少数学生对毕业设计的重视程度不够，存在完成任务时拖拖拉拉的现象。

(3) 实验或数据分析困难。虽然花费了大量的时间和精力，但是实验进展缓慢。

(4) 写作困难，毕业设计成果书的写作对很多学生来说是一个挑战。

#### 4. 改进措施

(1) 对照学校的格式指南，反复多次检查成果书，仔细核查图片和表格等名称的命名来避免这些错误。

(2) 加强对毕业生的《毕业设计》课程重要性的提醒，提高学生对毕业设计的重视程度，严格要求学生认真完成毕业设计任务。设定小目标，定期检查进度，并奖励自己达成小成就，可以帮助保持动力。

(3) 实验设计和数据分析是项目的核心。当遇到技术问题时，寻求专业人士的帮助，如指导老师，是解决这一问题的关键。

(4) 克服协作困难这一困难的方法包括提前开始写作，分阶段完成，并多次审阅和修改。此外，利用写作辅导资源和软件，如文献管理工具，也非常有帮助。

总之，本次大数据技术专业毕业设计整体情况良好，取得了一定的成果和收获。但也存在一些问题。通过对这些问题的分析和总结，我们及时采取相应的改进措施，不断提高毕业设计的质量和水平，努力提高大数据技术专业的培养质量。