

附件 3

编号	
类别	信息技术类

佛山市职业技能提升行动

《大数据应用开发》 培训课程标准

佛山市职业技能培训课程标准委员会办公室

目录

一、培训说明.....	错误!未定义书签。
1. 课程名称.....	错误!未定义书签。
2. 标准定义.....	错误!未定义书签。
3. 培训对象.....	错误!未定义书签。
二、培训目标.....	错误!未定义书签。
1. 职业素养目标.....	错误!未定义书签。
2. 理论知识目标.....	错误!未定义书签。
3. 开发技能目标.....	错误!未定义书签。
三、课时分配.....	错误!未定义书签。
1. 总培训课时.....	错误!未定义书签。
2. 培训课时分配表.....	错误!未定义书签。
四、培训要求与培训内容.....	错误!未定义书签。
五、推荐教材.....	错误!未定义书签。
六、培训实施.....	错误!未定义书签。
1. 培训师资.....	错误!未定义书签。
2. 培训场地.....	错误!未定义书签。
3. 实训设备.....	错误!未定义书签。
4. 培训方法.....	错误!未定义书签。
5. 课程资源.....	错误!未定义书签。
七、考核评价.....	错误!未定义书签。
1. 考核方式.....	错误!未定义书签。
2. 考核内容.....	错误!未定义书签。
3. 考核标准.....	错误!未定义书签。
4. 考核时间.....	错误!未定义书签。
5. 考核设施.....	错误!未定义书签。

《大数据应用开发》课程标准

一、培训说明

1. 课程名称

大数据应用开发

2. 标准定义

运用 Python 程序设计语言，实现大数据采集工具、大数据统计分析、数据预处理及数据挖掘建模，最后应用 Web 技术完成大数据可视化呈现。具备在企事业单位数据分析和数据运营岗位进行数据采集、数据分析和数据可视化呈现的应用开发能力。

3. 培训对象

从事大数据系统工程需求分析及设计的人员、数据分析应用的开发人员、数据运营的工程技术人员。

二、培训目标

掌握大数据采集、大数据分析（描述性和预测性）和大数据可视化呈现的基本技术与工作流程，熟练掌握 Python 程序设计及数据库操作方法，能在企事业单位从事技术管理和咨询服务岗位常规工作。通过培训，学员遵循国家和行业有关标准，按照大数据分析规范化流程，分析行业/企业发展现状与竞争环境，开展业务风险评判与决策支持，利用数据挖掘与建模分析技术充分挖掘大数据的价值，并通过

可视化技术呈现给企事业单位决策者一张清晰、准确、形象且有数据支撑的大数据展示页面；通过培训，使学员熟悉整个大数据分析的生命期，加深理解大数据在行业/企业中的角色和用途，推动大数据技术在企事业单位的应用。

1. 职业素养目标

- (1) 具有积极的人生态度、健康的心理素质、良好的职业道德
- (2) 具有高度责任心、进取心和坚强意志
- (3) 具有良好的人际交往、团队协作能力
- (4) 具有良好的表达能力，善于和协作部门沟通
- (5) 具有良好的人文素养和持续学习的能力
- (6) 具有信息系统安全意识，重视数据存储安全和保护

2. 理论知识目标

- (1) 掌握大数据应用开发基本流程
- (2) 了解数据运营的意义
- (3) 熟悉大数据系统基本架构
- (4) 掌握 HTTP 协议
- (5) 了解网页数据与结构
- (6) 了解结构化数据与非结构化数据特点
- (7) 掌握非结构化数据库基本概念
- (8) 掌握数据质量和数据的一致性概念
- (9) 掌握常用的数理统计概念

- (10) 了解常用的数据分类算法(回归分析、决策树、神经网络、贝叶斯网络、支持向量机等)
- (11) 了解常用的数据聚类算法
- (12) 了解常用的平稳时间序列分析方法
- (13) 了解大数据可视化主流技术
- (14) 掌握 WEB 前端可视化基础语法

3. 开发技能目标

- (1) 能使用在线帮助文件获得开发资料
- (2) 能使用 Scrapy 框架编写爬虫工具抓取互联网数据
- (3) 能实现数据库的数据存取
- (4) 能对数据进行结构化和非结构化分析
- (5) 能使用 numpy、scipy、pandas 等 Python 库进行数据预处理
- (6) 能对数据进行数据分析与挖掘
- (7) 能使用可视化工具进行数据展示
- (8) 能编写规范化项目文档

三、课时分配

1. 总培训课时

总培训课时：80

2. 培训课时分配表

具体培训课时分配见下：

培训课时分配表

培训内容	培训课时	总课时
模块一：数据采集技术——网络爬虫程序设计		
1. 爬虫的实现原理与技术	2	24
2. 爬虫框架 Scrapy 及爬虫工具程序实现	8	
3. 网络数据抓取	4	
4. 网站数据解析	4	
5. 非结构化数据库存取	6	
模块二：数据分析技术——基于 Python 的数据分析		
1. numpy 数组计算	2	24
2. scipy 科学计算	2	
3. pandas 统计分析	4	
4. 数据预处理	8	
5. 数据挖掘与建模（数据分类、聚类和预测）	8	
模块三：数据可视化报表设计		
1. 数据可视化工具概述	2	12
2. 数据与图表选择	4	
3. 使用可视化工具设计 BI 报表	6	
模块四：大数据可视化前端技术		
1. HTML+CSS+JavaScript 语法基础	4	20
2. ECharts 基础图表	4	
3. ECharts 交互组件	4	

培训内容	培训课时	总课时
4. ECharts 可视化综合应用	8	
总 计		80

四、培训要求与培训内容

培训内容	培训要求	培训建议
模块一 数据采集技术	<p>操作技能要求：</p> <ol style="list-style-type: none"> 1. 熟练掌握 HTTP 抓包工具 Fiddler 配置与数据抓取 2. 掌握 Scrapy 框架的安装与应用 (Scrapy 网站: https://scrapy.org/) 3. 熟练使用 urllib 包抓取网页 4. 掌握代理服务器设置 5. 熟练使用 requests 类库抓取数据 6. 熟练使用 Xpath 与 lxml 解析库 7. 熟练使用 BeautifulSoup 8. 熟练使用 JsonPath 与 JSON 模块 9. 掌握 MongoDB 的安装与基本使用 10. 熟练掌握 PyMongo 接口开发包基本操作 11. 能按照工作规范保存项目进度文件与终稿，并整理归档。 <p>理论知识要求：</p> <ol style="list-style-type: none"> 1. 理解爬虫的基本工作原理 2. 了解网络数据抓取流程 3. 掌握 HTTP 网络请求原理 4. 掌握 HTTP 抓包工具 Fiddler 工作原理 5. 掌握 urllib 包抓取数据原理 6. 理解代理服务器原理 7. 了解常见的网络异常 8. 了解网页数据与结构 9. 理解正则表达式 10. 掌握 Xpath、lxml、Beautiful Soup、JsonPath 与 JSON 模块等方法原理 11. 理解 MongoDB 功能与特点 12. 理解 Scrapy 架构与运作流程 	<p>重点：</p> <ol style="list-style-type: none"> 1. 使用 urllib 包抓取网络数据 2. 使用 requests 类库抓取网络数据 3. 使用 Scrapy 框架抓取网络数据 <p>难点：</p> <ol style="list-style-type: none"> 1. 使用 Xpath 与 lxml 解析库解析数据 2. 使用 BeautifulSoup 进行数据匹配搜索 3. 使用 JsonPath 与 JSON 模块解析 JSON 数据
模块二	操作技能要求：	重点：

培训内容	培训要求	培训建议
数据分析技术	<ol style="list-style-type: none"> 1. 熟练使用 numpy、scipy 和 pandas 等 Python 科学计算库 2. 掌握 Python 科学计算库网站帮助文件的查阅技巧（ numpy 网站：https://numpy.org/， scipy 网站：https://www.scipy.org/， pandas 网站：https://pandas.pydata.org/） 3. 熟练操作数组，进行数值计算 4. 熟练使用常用的插补方法和异常值处理方法进行缺失值和异常值分析与处理 5. 熟练掌握几种常用的一致性分析方法，并进行数据汇总 6. 熟练使用几种常用的数据变换方法 7. 熟练使用 Python 相关函数实现分布分析、对比分析、统计量分析、周期性分析、贡献度分析、相关性分析 8. 熟练使用 Python 主要数据探索函数，包括基本统计特征函数、拓展统计特征函数、统计作图函数 9. 熟练应用回归分析、决策树、神经网络、贝叶斯网络、支持向量机等相关 Python 函数实现数据分类 10. 熟练应用 k-means 算法相关函数实现数据聚类分析 11. 熟练应用相关 Python 函数实现平稳时间序列分析与预测 12. 能按照工作规范保存项目进度文件与终稿，并整理归档。 <p>理论知识要求：</p> <ol style="list-style-type: none"> 13. 掌握常用的插补方法原理 14. 掌握常用的异常值处理技术 15. 掌握一致性分析方法原理 16. 掌握数据变换方法（简单的函数变换、规范化、连续属性离散化、属性构造） 17. 掌握数据规约要求（属性规约、数值规约） 18. 掌握几种常用数据统计特征分析原理 19. 了解回归分析、决策树、神经网络、贝叶斯网络、支持向量机、k-means 等分类和聚类算法基本原理 20. 了解常用的时间序列算法原理 	<ol style="list-style-type: none"> 1. 应用回归分析、决策树、支持向量机等相关 Python 函数实现数据分类 <p>难点：</p> <ol style="list-style-type: none"> 1. 应用 k-means 算法相关函数实现数据聚类分析

培训内容	培训要求	培训建议
<p>模块三 数据可视化报表设计</p>	<p>操作技能要求：</p> <ol style="list-style-type: none"> 1. 能根据业务需求，选择合适的可视化工具 2. 能根据业务需求，针对不同数据指标选择不同的图表表现形式 3. 能使用可视化工具，设计可视化方案 4. 能根据业务需求，优化可视化报表 5. 能按照工作规范保存项目进度文件与终稿，并整理归档。 <p>理论知识要求：</p> <ol style="list-style-type: none"> 1. 熟悉数据可视化基本工作流程 2. 熟悉主流可视化工具及语言的特点 3. 掌握常用可视化图表的功能 4. 掌握数据与图形的映射关系 5. 掌握常用可视化图表的构成 	<p>重点：</p> <ol style="list-style-type: none"> 1. 能根据数据特点选用可视化图形 2. 能使用可视化工具,设计可视化方案 <p>难点：</p> <ol style="list-style-type: none"> 1. 数据与图表的映射关系
<p>模块四 大数据可视化前端技术</p>	<p>操作技能要求：</p> <ol style="list-style-type: none"> 1. 会安装与使用 ECharts 2. 会绘制常见可视化图表 3. 能实现图表数据异步加载与更新 4. 能在图表中加入交互组件 5. 会优化图表辅助组件参数 6. 能按照工作规范保存项目进度文件与终稿，并整理归档。 <p>理论知识要求：</p> <ol style="list-style-type: none"> 1. 掌握 WEB 前端基础语法 2. 掌握 ECharts 语法规则 3. 掌握 ECharts 常用图表配置 4. 掌握 ECharts 组件配置 	<p>重点：</p> <ol style="list-style-type: none"> 1. ECharts 可视化图表绘制 2. ECharts 组件参数配置 <p>难点：</p> <ol style="list-style-type: none"> 1. ECharts 组件参数配置

五、推荐教材

1. 《解析 Python 网络爬虫：核心技术、Scrapy 框架、分布式爬

虫》，黑马程序员 编著，中国铁道出版社，2018 年 8 月第 1 版，ISBN 9787113246785

2. 《Python 数据分析与挖掘实战》，张良均等主编，机械工业出版社，2019 年 11 月第 2 版，ISBN 9787111640028

3. 《大数据可视化技术与应用》，黄源，蒋文豪，徐受蓉主编，清华大学出版社，2020 年 6 月第 1 版，ISBN: 9787302547501

六、培训实施

1. 培训师资

要求主讲教师毕业于计算机或应用数学专业，具有本职业技师及以上职业资格证书或本专业中级及以上专业技术职务任职资格，掌握计算机应用专业基础知识，熟悉信息技术行业工作，热爱教育事业。

2. 培训场地

培训场地应具有满足教学需要的面积80平米以上、且能安排40个以上机位的计算机机房1间，要求配有3~4台数据库服务器、投影设备、互联网接口和教学用广播软件等。

3. 实训设备

技能培训的实习工具与设备表

序号	设备及用品名称	数量	备注
1	专业开发软件	40 套	部署客户端开发环境，一机一套

序号	设备及用品名称	数量	备注
2	服务器	4 台	用于大数据存储和数据处理
3	大数据案例库	3 套	用于应用开发案例教学
4	电脑	40 台	学员使用
5	大屏幕显示器	2 台	用于课堂教学和大数据可视化呈现

4. 培训方法

推荐采用案例分析法、实训操作法、小组学习法等。

5. 课程资源

本课程提供以下基本资源：PPT 课件、实训指导书、作业解答、项目案例及录像、微课、行业资讯等。此外，可使用以下公开网站：

- 智慧职教

https://www.icve.com.cn/portalproject/themes/default/efn3aj6nk75bryjc0daqpq/sta_page/index.html?projectId=efn3aj6nk75bryjc0daqpq

主要内容包括：大数据技术与应用专业核心课程资源，包括多门课程的 PPT 课件、实训指导书、作业解答、项目案例及录像、微课等。

- 亿信社区：<http://bbs.esensoft.com/>

主要内容包括：数据分析学习交流社区，包括交流活动、问答、学习认证、产品演示、案例等。

- 数据分析网：<https://www.afenxi.com/>

主要内容包括：大数据相关的最新的资讯、业内人物的点评文章、数据分析相关的技术文章、大量的数据分析的相关免费的学习资源以及相关的线下线上活动。

- 爱统计网：<http://www.itongji.cn/>

主要内容包括：数据运营、数据平台、大数据技术、数据分析、数据挖掘、数据百科等。

七、考核评价

1. 考核方式

本课程着重考核学员掌握所学的基本技能，并能综合运用所学知识和技能分析、解决实际问题的能力。采用理论+实操考核方式。考核成绩均实行百分制，成绩皆达 60 分及以上者为合格。

2. 考核内容

(1) 理论知识权重表

项目 \ 技能等级		比重 (%)
基本要求	职业道德	5
	程序设计基础知识	25
相关知识要求	HTTP 数据传输原理	10
	数据采集常用技术	15
	数据统计分析方法	20
	常用可视化图表功能	10
	WEB 前端基础知识	15
合计		100

(2) 技能要求权重表

项目 \ 技能等级		比重 (%)
技能要求	规范化项目文档的编写	10
	网页数据抓取与存储	20
	Python 科学计算库的使用	30
	数据可视化工具使用	20

	ECharts 图表绘制	20
合计		100

3. 考核标准

(1) 理论知识考核

题型	考核标准	总分值
客观题	考核学员对考核内容的基本概念的理解与掌握。 题型包括单选题、多选题、填空题、判断题等。	40 分
主观题	考核学员对考核内容相关知识的分析和综合应用能力。 题型包括：简答题、问答题、画图题等。	60 分
总分		100 分

(2) 技能操作考核

考核环节	考核标准	总分值
程序编写	考核各课程模块核心技术的代码编写及调试能力。 能完成功能并成功运行，该项满分； 无法成功运行，但基本功能编写完善，根据代码内容酌情扣分；	70 分
文档编写	考核学员能否按照规范完成项目文档编写。	10 分
工作规范	考核学员能否按照企业工作规范进行项目实操，此项考核贯穿技能操作考核全程。	20 分
总分		100 分

4. 考核时间

理论知识考核时间不少于 60 分钟；技能操作考核时间不少于 180 分钟。

5. 考核设施

理论知识考核在标准教室进行；技能操作考核在安装有 Python 程序设计软件和 Visual Studio Code 软件的计算机机房内进行。