

工信部教考中心《大数据分析师》简介

一、项目简介

大数据分析师是指在不同行业中，专门从事相关数据的收集、整理、分析，并依据数据通过科学算法模型进行行业研究、评估和预测等工作的专项人才。本次课程安排及考试内容符合工信部教考中心大数据分析师等级考试项目要求，适用于大数据产业从业人员和预就业人员，在大数据采集、大数据技术和大数据应用等领域。

课程涉及 Python 编程基础、Python 数据分析与应用、Python 机器学习实战、Python 数据可视化、深度学习原理及编程实现，应用行业涉及互联网、政府、电商和农业等大数据相关行业。

大数据分析师主要分有三个级别，初级、中级和高级，每个等级面向对象和考试内容不一样。初级首要面向事务数据分析，为数据分析领域的初级岗位，与之匹配的岗位为大数据分析师、数据赋能师等。中级分为数据挖掘和大数据方向，为数据分析领域的中级岗位，与之匹配的岗位为数据挖掘工程师、大数据分析师等。高级为数据科学，归于数据分析领域的高级岗位，相匹配的岗位为大数据挖掘工程师、大数据算法工程师等，其相关岗位多为上市、国企的大型企业的招聘岗位。

二、能力标准

1、大数据分析师（初级）

需要掌握 Python 编程基础、Python 常用数据分析及处理工具 numpy、pandas、scikit-learn，能以 Python 作为工具，解决基本的数据分析问题。能够理解业务目标，并能将业务目标初步转化为 Python 技术与应用问题，能结合

具体技术进行初步目标实现。适合政府、金融、电信、零售等行业前端业务及 从事市场、管理、财务、供应、咨询等职位的相关人员。

2、大数据分析师（中级）

一年以上 Python 数据分析应用工作经验，或已获得大数据分析师（初级）证书。掌握 Python 编程基础； Python 常用数据分析及处理工具numpy、pandas、matplotlib、scikit-learn； Python 数据可视化、机器学习、常规案例应用，能够将业务目标准确转化为 Python 技术与应用目标，能将业务 目标拆解成不同任务并找到对应技术实现方法，提升工作价值。适合政府、金融、电信、零售、互联网、电商、医学等行业专门从事 Python 数据分析应用的人员。

3、大数据分析师（高级）

三年以上 Python 数据分析应用岗位工作经验，或已获得大数据分析师（中级）证书。掌握 Python 编程基础； Python 常用数据分析及处理工 具 numpy、pandas、matplotlib、scikit-learn； Python 数据可视化、机器学习、多行业多领域 Python 技术应用、深度学习、典型人工智能应用，能充分理解业务目标，并将业务目标精准转换、拆分为具体技术任务；具有超强数据规划能力，不仅能自主完成各环节任务还能带领数据分析团队完成大型项目，为企业生产赋能。适合政府、金融、电信、零售、互联网、电商、医学等行业 Python 数据分析应用资深人员。

三、课程内容

阶段	课程	知识模块
必备技能	大数据导论	大数据与人工智能
		配套资料下载
	数据分析概述	Python数据分析概述

	Excel数据分析基础与实战	Excel2016概述
		外部数据的获取
		数据处理
		公式应用
		数据透视表 and 透视图
		数据分析与可视化,
	Power BI数据分析与可视化	认识Power BI
		数据获取
		M语言数据预处理
		DAX语言数据建模
		数据分析可视化
职业技术考核	大数据分析师（初级） 职业技术考试	

中级：

阶段	课程	知识模块
基础知识	职业道德	职业道德
	大数据理论基础	大数据概述
		大数据与人工智能
		大数据采集技术
		大数据时代下的存储技术
		大数据与数学
	程序编写基础	准备工作
		代码编写规范与操作
		数据结构
		程序流程控制语句
		函数

		面向对象编程
		文件基础
		内置模块与第三方库
大数据采集	数据获取与采集技术	爬虫环境与爬虫简介
		网页前端基础
		简单静态网页爬取
		常规动态网页爬取
		模拟登录
		网络爬虫框架
大数据分析	大数据处理技术	文件操作基础
		统计分析基础
		数据合并与清洗
		数据变换
	大数据分析技术	机器学习绪论
		性能度量
		回归分析
		决策树
		K近邻算法
		神经网络
		朴素贝叶斯
		聚类分析
		支持向量机
		质量分析与特征分析
		集成算法
		关联规则
		智能推荐
		时间序列

		主成分与因子分析
	大数据可视化技术	数据可视化概述
		基础绘图
		进阶绘图
		交互式绘图
项目实战	某实习网站招聘信息采集与分析	某实习网站招聘信息采集与分析
职业技术考核	大数据分析师（中级） 职业技术考试	

高级：

阶段	课程	知识模块
基础知识	职业道德	基本知识
	大数据理论	大数据概述
		大数据与人工智能
		大数据采集技术
		大数据时代下的存储技术
		数据仓库
		数据库管理系统
		分布式文件系统
大数据平台管理	软件安装与环境配置	软件安装
		安装虚拟机与Java
		Hadoop集群： 1. 部署模式与固定IP 2. 配置与启动
大数据处理与应用	ETL数据处理	ETL相关知识
	图像数据处理	概述

		读写图像
		图像几何变换
		灰度级修正
		图像二值化与图像平滑
		图像锐化
		视频图像处理练习
		循环读取图片数据
		第三方库读取图像文件
		单张图像增强
		多张图像增强
	文本数据处理	绪论
		语料字符处理
		正则表达式
		正向与逆向最大匹配法
		N元语法模型与隐马尔可夫
		Viterbi算法
		jieba分词与去停用词
		文本向量化-离散表示
		文本向量化： 1. Word2Vec模型 2. Doc2Vec模型
	语音数据处理	语音识别概述与wave操作
		读写音频与录音
		时域、频域与Mel谱图
		降噪与静音处理
		预加重、分帧、加窗与傅里叶变换

		语音特征提取
大数据分析 与 挖掘	机器学习实现	半监督学习
		集成学习
		强化学习
		网络入侵用户自动识别： 1. 背景分析与数据预处理 2. 模型训练与预测
		游客目的地印象分析： 1. 背景与景区酒店的印象分析 2. 景区酒店的综合评价 3. 景区酒店网评文本有效性与特色分析
	深度学习实现	神经网络概述与卷积操作
		池化、全连接与操作实现
		循环神经网络
		手写数字识别
		疫情期间网民情绪识别： 1. 案例背景与数据探索 2. 微博正文数据处理 3. 数据处理与词云图绘制 4. 模型训练与测试
		车牌智能识别： 1. 车牌定位概述与工程构建 2. 车牌图像处理 3. 车牌判断识别与预处理类执行 4. 车牌字符分割 5. 模型构建与测试
		路透社新闻分类：

		1. 相关概念与CBOW词向量训练 2. 数据预处理与词向量训练 3. 模型构建与训练
		语音识别过程与语音标注
		语音识别算法
		中文语音识别： 1. 音频读取与均方根去除静音 2. 音频特征处理 3. CTC模型 4. 批次数据处理与数据集划分 5. 双向GRU模型与Wavenet模型 6. 模型预测
	模型评价与预测	模型评价
		铁路客流量预测： 1. 数据获取 2. 数据合并与预处理 3. 客流量分析 4. 模型构建、定阶、预测与优化 5. 客流量预测
		市财政收入分析及预测： 1. 相关性分析与Lasso回归方法 2. 特征值预测与模型训练
职业技术考核	大数据分析师（高级） 职业技术考试	

四、报考条件

初级：无要求，皆可报考。

中级：（满足其中一个条件即可）

1.获得初级证书。

2.年满 18 周岁，具备高中以上学历，工作年限满 1 年

高级：（满足其中一个条件即可）

1.获得中级证书。

2.年满 20 周岁以上，工作年限满 2 年

3.年满 20 周岁以上，具备大专及以上学历

五、考试方式

考试方式分为线下考试站点或线上考试系统统考两种方式，考试形式为上机答题，闭卷。

考试题型：

初级： 120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

中级： 120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

高级： 120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

六、成绩评分

考试最终成绩满分为100分，成绩80-100分为优秀；成绩60-79分为合格；成绩60分以下为不合格。

七、证书颁发

学员经考核合格，由工业和信息化部教育与考试中心颁发大数据分析师
职业技术证书，证书可登录国家工信部教育与考试中心官网查询。

附：证书样本

