

# 大数据技术应用职业技术认证简介

## 一、证书简介

大数据技术应用职业技术认证是由工业和信息化部教育与考试中心推出一套专业化，科学化，系统化的人才考核标准，涉及在互联网、零售、金融、电信、医学、旅游、新闻媒体等行业专门从事数据采集、数据分析、机器学习、人工智能并能制作业务报告、提供决策的新型数据分析人才所需要的技能。

“大数据技术应用职业技术认证”是对在数据分析、机器学习、深度学习、人工智能领域中使用大数据技术作为主要方法从业者的全面认证，分为初级，中级，高级三个等级。

## 二、能力标准

以下是大数据技术应用职业技术认证能力标准概要及每个等级对应的学习课程，每个等级工程师分别从技能要求、应用方法要求、业务应用能力三个方面进行要求，只有同时满足各个方面要求才是一名优秀的大数据技术应用工程师。

	初级	中级	高级
技能要求	Python 编程基础、Python 常用数据分析与处理工具 numpy、pandas、 scikit-learn	在初级基础上掌握数据可视化、 数据分析预处理、机器学习、常 规案例应用	在中级基础上掌握多行业多 领域大数据技术应用、数据 分析、深度学习、自然语言 处理、网络爬虫、人工智能 应用
应用方法要求	掌握基本大数据技术应用流 程、基本数据预处理方法及 数据分析常用模型（线性回 归、聚类分析、逻辑回归、 时间序列等）	除掌握初级大数据技术与应用 方法外，还应掌握高级应用方 法、数据分析应用及数据挖掘方 法：决策树、神经网络、支持向 量机、集成学习、特征工程、朴 素贝叶斯、关联规则等，及常用 数据可视化应用	在中级基础上，还需掌握机 器学习，深度学习、人工智 能等常见应用。熟知大数据 技术在不同行业应用价值及 对应应用方法、流程及案例， 熟悉不同分析方法的使用场 景
业务应用能力	能够理解业务目标，并能将 业务目标初步转化为大数据 技术与应用问题，能结合具 体技术进行简单的目标实现	能够将业务目标准确转化为大 数据技术与应用目标，能将业务 目标拆解成不同任务并能独立 完成对应任务，提升工作效率	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
对应课程	<b>必选课程：</b>  1. Python 编程基 础	<b>必选课程：</b>  1. Python 编程基础 2. Python 数据分析与	<b>必选课程：</b>  1. Python 编程基础 2. Python 数据分析

	<p>2. Python 数据分析与应用</p>	<p>应用</p> <p>3. 数据分析预处理实战</p> <p>4. 机器学习实战</p> <p><b>四选一课程：</b></p> <p>1. 网络爬虫实战</p> <p>2. TensorFlow 实战</p> <p>3. 深度学习原理及编程实现</p> <p>4. 神经网络</p> <p><b>十选三案例：</b></p> <p>1. 基于水色图像的水质识别</p> <p>2. 电子商务网站用户行为分析及服务推荐</p> <p>3. Python 爬虫实践：《流浪地球》豆瓣影评分析</p> <p>4. 市财政收入分析及预测</p> <p>5. 广电大数据营销推荐项目实战</p> <p>6. 航空公司客户价值分析</p> <p>7. 基于文本内容的垃圾短信分类</p> <p>8. 基于卷积神经网络（CNN）的人脸识别系统</p> <p>9. 京东商城美的空调评论数据情感分析</p> <p>10. 大数据岗位人才招聘信息的分析与挖掘</p>	<p>与应用</p> <p>3. 数据可视化</p> <p>4. 网络爬虫实战</p> <p>5. 数据分析预处理实战</p> <p>6. 机器学习实战</p> <p>7. TensorFlow 实战</p> <p>8. 深度学习原理及编程实现</p> <p>9. 神经网络</p> <p>10. 深度神经网络</p> <p>11. 自然语言处理</p> <p><b>十选八案例：</b></p> <p>1. 基于水色图像的水质识别</p> <p>2. 电子商务网站用户行为分析及服务推荐</p> <p>3. Python 爬虫实践：《流浪地球》豆瓣影评分析</p> <p>4. 市财政收入分析及预测</p> <p>5. 广电大数据营销推荐项目实战</p> <p>6. 航空公司客户价值分析</p> <p>7. 基于文本内容的垃圾短信分类</p> <p>8. 基于卷积神经网络（CNN）的人脸识别系统</p> <p>9. 京东商城美的空调评论数据情感分析</p> <p>10. 大数据岗位人才招聘信息的分析与挖掘</p>
--	--------------------------	--	--

大数据技术应用职业技术认证（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大数据技术与应用问题，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。

大数据技术应用职业技术认证（中级）一年以上大数据分析应用工作经验，或通过大技术应用工程师（初级）认证。掌握 Python 编程基础；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大数据技术与应用目标，能够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大数据分析应用的人员。

**大数据技术应用职业技术认证（高级）** 三年以上大数据分析应用岗位工作经验，或通过大技术应用工程师（中级）认证。掌握 Python 编程基础；Python 常用数据分析及处理工具 numpy、pandas、matplotlib、scikit-learn；数据可视化、自然语言处理、网络爬虫、人工智能应用、机器学习、多行业多领域大数据技术应用、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、医学等行业数据分析应用资深人员。

### 三、课程大纲

Python 编程基础	Python 数据分析与应用
1 准备工作	1 Python 数据分析概述
1.1 认识 Python	1.1 认识数据分析
1.2 搭建 Python 环境	1.2 熟悉 Python 数据分析的工具
1.3 安装 PyCharm 并创建一个爬虫程序	1.3 安装 Python3 的 Anaconda 发行版
2 Python 基础知识	1.4 掌握 Jupyter Notebook 常用功能
2.1 掌握 Python 固定语法	2 NumPy 数值计算基础
2.2 创建字符串变量并提取里面的数值	2.1 认识 NumPy 数组对象 ndarray
2.3 计算圆形的各参数	2.2 认识 NumPy 矩阵与通用函数
3 Python 数据结构	2.3 利用 NumPy 进行统计分析
3.1 创建一个列表（list）并进行增删改查操作	3 Matplotlib 数据可视化基础
3.2 转换一个元组（tuple）并进行取值操作	3.1 了解绘图基础语法与常用参数
3.3 创建一个字典（dict）并进行增删改查操作	3.2 分析特征间的关系
3.4 将两个列表转换为集合（set）并进行集合运算	3.3 分析特征内部数据分布与分散状况
4 程序流程控制语句	4 Pandas 统计分析基础
4.1 实现考试成绩划分	4.1 读写不同数据源的数据
4.2 实现一组数的连加与连乘	4.2 掌握 DataFrame 的常用操作
4.3 使用冒泡排序法排序	4.3 转换与处理时间序列数据
4.4 实训（猜数字游戏）	4.4 使用分组聚合进行组内计算
5 函数	4.5 创建透视表与交叉表
5.1 自定义函数实现输出方差	5 使用 Pandas 进行数据预处理
5.2 使用匿名函数添加列表元素	5.1 合并数据
5.3 存储并导入函数模块	5.2 清洗数据
6 面向对象	5.3 标准化数据
6.1 认识面向对象编程	5.4 转换数据
6.2 创建 Car 类	6 使用 scikit-learn 构建模型
6.3 创建 Car 对象	6.1 使用 sklearn 转换器处理数据
6.4 迭代 Car 对象	6.2 构建并评价聚类模型

- 6.5 产生 Land\_Rover 对象（子类）
- 7 文件基础
  - 7.1 认识文件
  - 7.2 读取 txt 文件中的数据
  - 7.3 保存数据为 csv 格式文件
  - 7.4 认识 os 模块

- 6.3 构建并评价分类模型
- 6.4 构建并评价回归模型

数据可视化	网络爬虫实战
<ul style="list-style-type: none"> <li>1 准备工作环境               <ul style="list-style-type: none"> <li>1.1 准备介绍</li> <li>1.2 安装 matplotlib、Numpy 和 Scipy 库</li> <li>1.3 安装图像处理工具：Python 图像库（PIL）</li> <li>1.4 配置 matplotlib 参数</li> </ul> </li> <li>2 绘制并定制化图表               <ul style="list-style-type: none"> <li>2.1 简介</li> <li>2.2 定义图表类型—柱状图、线形图和堆积柱状图</li> <li>2.3 设置坐标轴长度和范围、线型、属性和格式化字符串</li> <li>2.4 添加图例和注解</li> </ul> </li> <li>3 学习更多图表和定制化               <ul style="list-style-type: none"> <li>3.1 简介</li> <li>3.2 设置坐标轴标签的透明度和大小</li> <li>3.3 为图表线条添加阴影</li> <li>3.4 向图表添加数据表</li> <li>3.5 使用 subplots（子区）</li> <li>3.6 定制化网格</li> </ul> </li> <li>4 创建 3D 可视化图表               <ul style="list-style-type: none"> <li>4.1 简介</li> <li>4.2 创建 3D 柱状图</li> <li>4.3 创建 3D 直方图.</li> <li>4.4 在 matplotlib 中创建动画</li> </ul> </li> <li>5 用图像和地图绘制图表               <ul style="list-style-type: none"> <li>5.1 用 PIL 做图像处理</li> <li>5.2 在具有其他图形的图表中显示图像</li> <li>5.3 使用 Basemap 在地图上绘制数据</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>1 Python 爬虫环境与爬虫简介               <ul style="list-style-type: none"> <li>1.1 认识爬虫</li> <li>1.2 认识反爬虫</li> <li>1.3 配置 Python 爬虫环境</li> </ul> </li> <li>2 网页前端基础               <ul style="list-style-type: none"> <li>2.1 认识 Python 网络编程</li> <li>2.2 认识 HTTP 协议</li> </ul> </li> <li>3 简单静态网页爬取               <ul style="list-style-type: none"> <li>3.1 实现 HTTP 请求</li> <li>3.2 解析网页</li> <li>3.3 数据存储</li> </ul> </li> <li>4 常规动态网页爬取               <ul style="list-style-type: none"> <li>4.1 逆向分析爬取动态网页</li> <li>4.2 使用 Selenium 库爬取动态网页</li> <li>4.3 存储数据至 MongoDB 数据库</li> </ul> </li> <li>5 模拟登录               <ul style="list-style-type: none"> <li>5.1 使用表单登录方法实现模拟登录</li> <li>5.2 使用 Cookie 登录方法实现模拟登录</li> </ul> </li> <li>6 终端协议分析               <ul style="list-style-type: none"> <li>6.1 分析 PC 客户端抓包</li> </ul> </li> <li>7 Scrapy 爬虫               <ul style="list-style-type: none"> <li>7.1 认识 Scrapy</li> <li>7.2 通过 Scrapy 爬取文本信息</li> <li>7.3 定制中间件</li> </ul> </li> </ul>
机器学习实战	TensorFlow 实战
<ul style="list-style-type: none"> <li>1 机器学习绪论               <ul style="list-style-type: none"> <li>1.1 引言</li> <li>1.2 基本术语</li> <li>1.3 假设空间</li> <li>1.4 归纳偏好</li> </ul> </li> <li>2 模型评估与选择               <ul style="list-style-type: none"> <li>2.1 经验误差与过拟合</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>1 TensorFlow 与深度学习实战               <ul style="list-style-type: none"> <li>1 TensorFlow 安装与入门                   <ul style="list-style-type: none"> <li>1.1 TensorFlow 环境搭建</li> <li>1.2 TensorFlow 计算模型性:计算图</li> <li>1.3 TensorFlow 数据模型:张量 Tensor</li> <li>1.4 TensorFlow 运行模型:会话</li> </ul> </li> </ul> </li> <li>2 TensorFlow 数据类型</li> </ul>

- 2.2 评估方法
- 2.3 性能度量
- 2.4 偏差与方差
- 3 回归分析(Regression Analysis)
  - 3.1 基本形式
  - 3.2 线性模型
  - 3.3 逻辑回归
- 4 决策树(Decision Tree)
  - 4.1 基本流程
  - 4.2 划分选择
  - 4.3 剪枝
- 5 人工神经网络(Artificial Neural Network)
  - 5.1 神经元模型
  - 5.2 感知机与多层网络
  - 5.3 误差逆传播
  - 5.4 BP 神经网络
- 6 最近邻算法 (KNN)
- 7 朴素贝叶斯分类(Naive Bayesian)
- 8 聚类分析(Cluster Analysis)
  - 8.1 聚类任务
  - 8.2 性能度量
  - 8.3 距离计算
  - 8.4 常用聚类算法
- 9 支持向量机(Support Vector Machine)
  - 9.1 间隔与支持向量
  - 9.2 对偶问题
  - 9.3 核函数
  - 9.4 软间隔与正则化

- 2.1 常量、变量及其构建
- 2.2 TensorFlow 实现线性回归模型
- 2.3 操作: TensorFlow 实现鸢尾花分类
- 3 TensorFlow 实现多层神经网络
  - 3.1 BP 神经网络模型(Back Propagation)
  - 3.2 操作: 利用 TensorFlow 构建 BP 网络模型实现鸢尾花分类
- 4 mnist 手写数字识别
  - 4.1 图片预处理
  - 4.2 占位符:placeholder
  - 4.3 操作: 利用 TensorFlow 实现 SoftMax 网络对手写数字分类
- 5 TensorFlow 动态学习速率
  - 5.1 TensorFlow 动态学习速率设置
  - 5.2 TensorFlow 动态学习速率使用
- 6 TensorFlow 模型保存与调用
  - 6.1 TensorFlow 模型保存操作
  - 6.2 TensorFlow 模型加载与调用
  - 6.3 操作: softmax 网络保存与调用实例

## 人工智能-深度神经网络

- 1 卷积神经网络 CNN
  - 1.1 卷积神经网络(CNN)简介
  - 1.2 CNN 关键结构:卷积层与池化层
  - 1.3 经典卷积网络模型: LeNet-5
  - 1.4 图像数据处理
  - 1.5 操作: CNN 的 TensorFlow2.0 实现
- 2 循环神经网络 RNN
  - 2.1 循环神经网络(RNN)简介
  - 2.2 RNN 网络关键结构: 隐层互联
  - 2.3 经典 RNN 模型: LSTM
  - 2.4 时序数据处理
  - 2.5 操作: RNN 的 TensorFlow2.0 实现

## 深度学习原理及编程实现

## 人工智能-自然语言处理

- 1 自然语言处理基础
  - 1.1 自然语言处理简介
  - 1.2 中文分词
  - 1.3 NLP 概率图
  - 1.4 文本分类
- 2 词表征 (Word Representation)
  - 2.1 词表征的背景与应用
  - 2.2 one-hot 编码
  - 2.3 tf-idf 权值向量
  - 2.4 word2vec 原理及实现
- 3 深度学习挖掘语义特征
  - 3.1 词向量组合相似度
  - 3.2 基于循环神经网络 (RNN) 编码的相似度计算

## 10 大项目案例

- 1 卷积神经网络 CNN
  - 1.1 卷积神经网络(CNN)简介
  - 1.2 CNN 关键结构:卷积层与池化层
  - 1.3 经典卷积网络模型: LeNet-5
  - 1.4 图像数据处理
  - 1.5 操作: CNN 的 TensorFlow 实现
- 2 循环神经网络 RNN
  - 2.1 循环神经网络(RNN)简介
  - 2.2 RNN 网络关键结构: 隐层互联
  - 2.3 经典 RNN 模型: LSTM
  - 2.4 时序数据处理
  - 2.5 操作: RNN 的 TensorFlow 实现
- 3 自然语言处理基础
  - 3.1 自然语言处理简介
  - 3.2 中文分词
  - 3.3 NLP 概率图
  - 3.4 文本分类
- 4 词表征 (Word Representation)
  - 词表征的背景与应用
  - 编码
  - 原理及实现
- 5 深度学习挖掘语义特征
  - 5.1 词向量组合相似度
  - 5.2 基于循环神经网络 (RNN) 编码的相似度计算

## 基础案例

- 1 数据分析案例: 基于水色图像的水质识别
- 2 数据分析案例: 电子商务网站用户行为分析及服务推荐
- 3 数据采集和分析案例: Python 爬虫实践:《流浪地球》豆瓣影评分析
- 4 数据分析案例: 市财政收入分析及预测
- 5 数据分析案例: 广电大数据营销推荐项目实战
- 6 数据分析和可视化案例: 航空公司客户价值分析
- 7 文本挖掘和可视化案例: 基于文本内容的垃圾短信分类

## 人工智能案例

基于卷积神经网络 (CNN) 的人脸识别系统

## 综合案例

- 1 京东商城美的空调评论数据情感分析
- 2 大数据岗位人才招聘信息的分析与挖掘

## 四、报考条件

**初级:** 无要求, 皆可报考。

**中级:** (满足以下之一皆可报名)

- 1. 获得初级认证证书。
- 2. 本科及以上学历需从事数据分析相关工作 1 年以上
- 3. 本科以下学历需从事数据分析相关工作 2 年以上

**高级:** (满足以下之一皆可报名)

- 1. 获得中级认证证书。
- 2. 本科及以上学历需从事数据分析相关工作 3 年以上
- 3. 本科以下学历需从事数据分析相关工作 4 年以上

(注: 上述数据分析相关工作不限制行业, 工作可涉及统计, 数据分析, 数据挖掘, 数据库, 数据管理, 大数据架构等内容。)

## 五、考试方式

初级和中级为线下和线上统考, 上机答题, 闭卷。高级分为线上、线下上机考试 (闭卷) + 线上答辩 (开卷) 两个部分。

## 六、考试内容

**初级：** 120 分钟，客观题（单选+多选），上机答题。

**中级：** 90 分钟，客观题（单选+多选），上机答题；120 分钟，案例操作，自行携带电脑操作

**高级：** 第一阶段：150 分钟，客观题+主观题，上机答题。  
第二阶段：项目案例操作，提交项目结果；60 分钟，线上答辩面试。

## 七、考试安排

**考试时间：**

**初级：** 详见准考证信息。

**中级：** 详见准考证信息。

**高级：** 详见准考证信息。

### 3. 考试地点

各地授权考试中心

## 八、报考流程

进入考试系统进行报名，考试后 15 日登录系统查询成绩，通过者获取证书（证书将在考试后 30 日内寄送）。

## 九、成绩评分

考试最终成绩分为 A,B,C,D 四个层次，A,B,C 皆为通过考试并获得认证证书，D 为不通过。

## 十、证书样本

学员经考核合格，将颁发由工业和信息化部教育与考试中心颁发大数据技术应用职业技术认证证书，证书可登录国家工信部教育与考试中心官网查询。