

# 大数据分析培训招生简章

## 一、 培训背景

国内大数据行业已经进入高速发展时期，2018 年中国大数据产业总体规模为 6200 亿元人民币，同比增长 31.9%，预计到 2020 年，大数据相关产品和服务业务收入将突破 1 万亿元。根据调查，有超过 2/3 的企业已成立数据分析相关部门，企业对数据分析越发重视，有接近六成的企业在大数据领域的投入将持续增加，营销分析、客户分析和内部运营管理是企业大数据应用场景最广泛的三个领域，金融、医疗、商业、交通、公安、政府服务等是大数据运用得较多的行业，大数据应用是实现智能决策、提高运行效率和风险管理能力等方面具有较明显的效果，同时，政策限制、数据资源短缺和大数据人才短缺是限制企业大数据发展最主要的三个因素。在就业人才市场中，大数据分析相关的职位就业岗位多、行业分布广，可供挑选的空间非常大，收入待遇高、薪资涨幅大，待遇远远超过其他岗位。

大数据行业的火热，带来了大数据分析人才的严重短缺与专业能力衡量标准缺失的问题，市场上迫切需要能够对大数据分析专项人才进行培训、能力鉴定的权威项目，同时也需要一个能够将人才有效转化为生产力的专业性服务平台。

## 二、 培训级别

大数据分析培训体系将能力发展通道由低到高划分为三个级别的梯形结构：初级、中级和高级。所有参与培训体系人员均需进行等级评定，并根据工作经验的增加，申请更高等级。

大数据分析培训体系的设计体现了对不同能力等级人员差异化培训的科学理念，有利于明确学习目标、调动学员积极性，鼓励学员能力提升与工作相结合，强调个人能力提升与从事大数据相关职业发展相结合。

大数据分析培训体系设计依据包括以下五个方面：

- 专业技能。如：数据管理、数据处理、数据展示等

- 沟通交流。如：基于数据的内外沟通、数据对业务逻辑的支持等
- 职业道德。如：信息安全、数据安全等
- 项目管理。如：常规项目管理内容，基于数据的业务管理等
- 领导力。如：制定企业级大数据应用基础等

### 三、 培训课程体系

“大数据分析师”培训课程的特点是，以大数据技术为基础，掌握多种分析算法，辅以大量企业实操案例，融会贯通大数据分析思路。

1、大数据相关技术培训：主要介绍大数据相关的技术，包括 HADOOP 核心技术、HIVE 框架、Hbase 框架等。学员在企业中应用大数据分析时，必须要对大数据开发技术有所了解。

2、大数据分析方法培训：主要介绍大数据分析的算法和思路，包括数据预处理、探索性统计分析、分类模型算法、聚类模型算法等。

3、企业实操案例培训：通过实际案例来向学员介绍大数据在企业中是如何应用的。

课程体系的内容主要包括以下内容：

#### 第一部分：理论基础

##### 1. 大数据理论基础

大数据概念及特性

了解：了解大数据概念及 4V 特性，领会大数据与普通数据的区别与联系。

大数据应用领域

了解：了解大数据的分布领域及常见应用，特别是社交媒体数据、金融欺诈大数据应用、零售业大数据应用、电信行业客户生命周期大数据应用等。

大数据发展及前瞻

了解：了解大数据的发展进程及发展趋势，了解大数据与人工智能、机器学习、物联网等相关领域的前瞻研究。

##### 2. 统计学基础

描述性统计分析

熟知：熟练掌握描述性统计方法，熟练掌握数据的计量尺度、集中趋势及分布统计方法，熟悉数据的相似性和相异性度量方法。

#### 抽样估计

了解：了解总体与样本的概念，熟悉随机概率、抽样平均误差等的区间估计方法。

#### 假设检验

了解：了解假设检验的基本概念，熟悉 P 值的含义及计算方法，了解 P 检验、Z 检验、t 检验等统计量的检验步骤。

#### 方差分析

熟知：熟练掌握方法分析的概念，原理及基本步骤。

#### 简单线性回归

掌握：熟练掌握简单线性回归的概念、原理及一般处理步骤，熟悉线性回归的应用领域及场景，能够运用相关工具进行线性回归分析，求解线性方程。

### 3. 数据库基础

#### SQL 概念

了解：了解 SQL 语言的基本原理，数据库及表的概念及原理

#### SQL 一般查询语句

掌握：掌握一般 SQL 查询语言的基本结构，能够熟练应用查询语句进行数据查询处理

#### SQL 连接语句

熟知：熟悉 SQL 连接语句，熟悉 LEFT/RIGHT/INNER JOIN 的用法及注意事项。

#### 数据库高阶

了解：了解表创建、索引创建，了解数据的插入、删除、更新等

### 4. 分析思维基础

#### 金字塔原理

熟知：熟悉金字塔原理分析写作方法论，能够在业务中熟练运用相关知识进行需求沟通、分析框架搭建及分析报告撰写。

### 5. 大数据相关技术

## Hadoop 的基础概念

了解：了解 Hadoop 的架构功能，技术框架涉及的基本概念。了解 MapReduce、HDFS 的基本工作原理，能看懂简单的系统配置。

## HBase

了解：了解 HBase 在 Hadoop 生态圈中的功能，该技术点的特性。

## Hive

了解：了解 Hive 在 Hadoop 生态圈中的功能，该技术点的特性。

## Spark

了解：了解 Spark 在 Hadoop 生态圈中的功能，该技术点的特性。

## Shell/Python

熟知：能运用脚本语言进行数据的分析提取。

## Pig

了解：能跟踪分析 Pig Latin 到 MapReduce 的运算及转换。

## 第二部分：数据处理

### 1. 数据属性

了解：了解数据的属性类型，特别需掌握二元属性、离散属性及连续属性等。

### 2. 数据预处理

#### 数据清洗

熟知：熟悉数据清洗的概念，能够针对不同场景运用不同清洗方法。

#### 数据集成

了解：了解数据集成概念及基本方法。

#### 数据变换

掌握：掌握数据变换的概念，原理；熟练掌握常用的数据变化方法 Z-SCORES, 最大值-最小值方法等。

#### 数据规约

熟知：熟悉数据规约概念，原理，熟悉维规约一般方法及步骤。

### 3. 探索性统计分析

#### 原始数据基础可视化

熟知：熟悉处理阶段统计分析可视化方法，能够根据数据特性选择适当的可视化度量尺度。

#### 衍生变量

了解：了解衍生变量的概念及必要性，熟悉使用衍生变量的场景及常用方法。

### 第三部分：数据挖掘

#### 1. 数据挖掘基础

熟知：熟悉数据挖掘的概念、原理及基本步骤，熟悉数据挖掘的分类及常用算法。

#### 2. 初步数据探索

熟知：熟悉数据探索的方法，能够运用数据挖掘工具进行初步数据探索，可进行变量初步筛选。

#### 3. 分类模型

##### 分类基本概念

熟知：熟悉分类的概念，使用场景及常用算法。

##### 分类一般步骤

熟知：熟悉常用分类算法的一般处理步骤，了解各算法的有点及缺点。

##### 决策树分类挖掘方法

掌握：熟悉常用的决策树算法，能够运用数据挖掘工具进行决策树挖掘模型设计及结果输出。

##### 朴素贝叶斯分类挖掘方法

了解：了解朴素贝叶斯算法的原理、了解贝叶斯算法的注意事项。

##### 逻辑回归分类挖掘方法

掌握：掌握逻辑回归的算法原理、一般步骤及使用场景，能够运用挖掘工具进行逻辑回归建模。

##### 分类器模型评估方法

掌握：掌握重要分类模型的评估方法，能够根据 ROC 曲线等可视化图形判断分类器优劣。

##### 其他分类挖掘方法

了解：了解随机森林、支持向量机等其他分类挖掘算法思想，基本步骤。

#### 4. 聚类模型

##### 聚类基本概念

熟知：熟悉聚类的概念，使用场景及常用算法。

##### 聚类一般步骤

熟知：熟悉常用聚类算法的一般处理步骤，了解各算法的有点及缺点。

##### 距离计算方法

掌握：掌握常用的距离计算方法，了解各种距离的适用场景。

##### 划分聚类挖掘方法

掌握：掌握 K-均值、K-中心点聚类方法的原理、一般步骤，能够运用挖掘工具进行划分聚类。

##### 层次聚类挖掘方法

熟知：熟悉凝聚聚类和分裂聚类的概念、原理及相互关系。

##### 聚类模型评估方法

掌握：掌握聚类模型的评估方法，能够根据评估度量解释模型优劣。

##### 其他聚类挖掘方法

了解：了解基于密度、基于网格等聚类方法及适用场景。

#### 5. 关联规则

##### 关联规则挖掘基本概念

熟知：熟悉关联的概念，使用场景及常用算法，了解算法见的优劣。

##### 关联规则挖掘一般步骤

熟知：熟悉常用关联规则算法的一般处理步骤。

##### Apriori 算法

掌握：掌握 Apriori 算法的原理、一般步骤及处理方法，能够适用挖掘工具进行关联规则挖掘。

##### 关联规则评估方法

掌握：掌握有效关联规则的评估方法，能够根据相关度量对隐藏规则进行解释。

##### 其他关联算法

了解：了解 FP-Tree 算法原理及算法优劣。

## 6. 预测

### 回归基础

熟知：熟知多元回归等回归算法原理、处理步骤，适用场景，能够根据实际问题，计算回归公式。

### 回归诊断

熟知：熟悉回归诊断方法，能够根据诊断结果对回归模型进行解释。

### 时间序列一般步骤

了解：了解时间序列模型的概念、一般流程及使用场景。

### 时序的平滑化和季节性分解

熟知：熟悉时间序列数据的平滑化方法及季节性分解方法。

### 指数预测模型

了解：了解指数预测模型算法原理，能够根据时序数据的特征选择不同的指数预测模型方法。

### ARIMA 模型

了解：了解 ARIMA 预测模型原理。

## 7. 文本挖掘分析

### 文本挖掘简介

熟知：了解文本挖掘工作方式，熟知一般流程。

### 情绪分析

了解：了解情绪分析方法、分析要素及分析流程。

### 文本挖掘应用

了解：了解文本挖掘的一般工具，了解社交媒体、在线分析关键要素。

## 8. 移动分析

### 移动分析简介

了解：了解移动分析概念及类型，了解移动分析的商业价值。

### 执行移动分析

熟知：熟知移动分析的基本流程，了解移动分析的基本工具。

## 第四部分：数据可视化

1. 可视化概念

了解：了解大数据可视化的概念及重要性。

2. 大数据可视化技术、类型及应用

了解：了解大数据可视化相关技术、类型及应用。

## 四、 学员要求

对于希望参加“大数据分析师”培训的学员，需具备一定的要求。

### ■ 学员基本要求

1. 具有中华人民共和国国籍。
2. 年满十八周岁。
3. 具有符合相关职位要求的工作能力。

## 五、 联系方式

咨询电话：010-65677878

服务邮箱：service@topbda.org.cn