

## 你为什么要学数据挖掘？

“数据驱动”早已被各大互联公司奉为圭臬，并且很多大公司都建有成熟的数据仓库、数据分析平台、数据挖掘平台、数据运营平台，数据驱动业务的做法已经成为常态，数据挖掘的重要性不言而喻。

**今日头条**借助数据挖掘为用户推荐对他们更有价值、更加个性化的新闻；**滴滴**依靠数据挖掘算法来规划、寻找最优路线；**美团外卖**凭借数据挖掘算法进行热度预估和订单分派，等等。几乎每一个你所用到的 App 背后都有数据挖掘的身影。

而且近些年，数据挖掘在许多传统型公司也得到极高重视，它们纷纷成立数字化转型部门，期望搭上时代的数字化列车 为业务赋能。

- **银行**，期望借助数据挖掘来对贷款、信用卡等业务进行风险控制；
- **金融基金公司**，通过数据挖掘来提高基金及金融产品的收益；
- **大型商超公司**，希望用数据挖掘来寻找最佳的进货和销售配比；
- **能源电力公司**，希望借助数据挖掘来分析风力发电站建设的位置以及机组的安全评估，等等。

这样的例子不胜枚举，金融、零售、医药、制造、交通、通信、能源等，几乎所有的行业都会应用到数据挖掘，**就业市场对数据挖掘人才的需求远远没有饱和**。

职业咨询公司的调查显示，1 年经验的数据挖掘岗位平均月薪可达 20K；超过 50% 的数据挖掘岗位在职人员月薪在 25K 以上：

<b>数据挖掘工程师 [北京-海淀区]</b> 13:46发布	<b>字节跳动</b>
20k-40k 经验1-3年 / 本科	文娱   内容 / C轮 / 2000人以上
大数据 移动互联网 MySQL 数据挖掘 Linux 数据仓库	“六险一金 带薪年假 房补免费食堂”
<b>数据挖掘算法工程师 [北京-朝阳区]</b> 11:46发布	<b>美团点评</b>
25k-50k 经验3-5年 / 本科	消费生活 / 上市公司 / 2000人以上
数据挖掘	“大平台 主要项目 广阔空间”

随着技术的发展，了解数据挖掘已经成为一种趋势。**发掘数据隐藏价值的能力已经不只属于数据挖掘专家或者数据科学家，每一位数据分析师、业务经理都应该具备这样的能力。**

## 学习数据挖掘过程中遇到的困惑

但学习数据挖掘并不是一蹴而就的、简单的事情。我多年的数据挖掘工作，也是一个不断学习、不断深化认识的过程。

最初，在学校里学习数据挖掘课程，虽然也看过一些书，但说起来那时我对数据挖掘的理解始终停留在公式、算法层面，虽然浑浑噩噩地通过了学校的考试，找到了工作，却不能理解数据挖掘到底有什么用。通过工作中的实际案例，我才慢慢理解了数据挖掘的意义，才真正找到了通往数据挖掘殿堂的大门。

在这个过程中我算是走了不少弯路，现在想想，其实就是学习不得其法所造成的，总结下来我认为主要有以下几个原因：

- **纯粹的算法原理详解。** 一些课程和图书是从技术角度讲解数据挖掘技术的，很少与实际的业务场景相结合，再加上艰深的数学概念、令人头痛的严密的数学推理，即便你学完了可能仍然云里雾里：我知道有这么一个算法，但是该怎么运用到实际工作中呢？
- **过于单一、局限的例子。** 还有部分资料直接从例子出发，通过经典案例教授内容当然是好的，但是数据挖掘与业务极度相关，例子可能与实际场景存在很大的区别。这些例子通常有着完美整齐的数据，虽然有助于理解场景，但不能帮我们形成合理的数据挖掘思维，无法在实际工作中举一反三。
- **缺乏对数据挖掘全流程的讲解。** 数据挖掘应该有一套完整的方法论来支撑，然而大多数资料只是在技术方面进行了详细地讲解，而忽略了方法论，以至于你在很长一段时间内难以对数据挖掘建立一个合理、全貌上的认知。

在工作上我也遇到、听到过许多迷茫的声音，比如：

- **没有接触过，却突然接到数据挖掘项目。** 越来越多的公司或部门，期望能够通过数据挖掘技术来解决一些业务问题，数据分析师等部分数据相关从业者，会因此突然接到数据挖掘项目，但没有学习过这方面内容的他们往往不知从何入手。
- **没有接触过完整的数据挖掘项目。** 有一些同学，虽然身处数据挖掘相关部门，也在做相关的事情，却只接触过其中一个环节，从来没有窥探过数据挖掘的全貌，对整个流程没有系统性的认识，影响了面试和上升的机会。
- **目标效果不好，不知如何处理。** 有些与数据打交道的人，如数据挖掘工程师、数据分析师等，已经开始行动，在自己的工作中尝试使用数据挖掘的方法来解决业务问题，却发现巨大挑战不在于算法本身，而是数据如何清洗、特征如何设计、效果如何优化，这些与业务极度相关的问题。

## 课程设计

基于以上原因，我将从构建数据挖掘思维的角度出发，为你详解数据挖掘，具体分为三大部分：

- **基础知识准备（模块一）。** 带你建立对数据挖掘的全局认知，学习课程中可能会涉及的一些基础知识（包括必备的 Python 语言知识，以及如何搭建 Python 环境），帮你快速进入状态。
- **数据挖掘过程（模块二）。** 授人以鱼不如授人以渔，重点是让你能够在思想和行为上都做足准备，全面细致地了解数据挖掘方法的实施过程。同时，我也会讲到数据挖掘过程中的每一个步骤都会遇到哪些坑，帮你加强思考、少走弯路，明白从理论到实战，“数据挖掘”这个词是如何一步步变具体的。
- **算法详解（模块三~模块七）。** 涉及数据挖掘的分类、聚类、回归、关联分析这四大问题，以及一些自然语言处理的知识。重点介绍每个算法的理念、优缺点、应用场景，配合一些简短的代码来说明该如何使用，让你能够快速上手应用。每个模块的最后一个课时，我会通过为本课程设计的实践案例，来带你处理一个实际问题。

课程最后还有一个彩蛋，我会整理一些数据挖掘的开源工具和学习资源：如果你不会写代码，也可以先使用这些工具来进行数据挖掘；如果你希望在数据挖掘方面有更加深入的学习和理解，那么这些资源也会帮到你。

23 word2vec：让文字可以进行逻辑运算，女人+王冠=女王

24 实践 5：使用 fastText 进行新闻文本分类

彩蛋：数据挖掘工程师如何进阶

结语：培养数据挖掘思维，终身学习

@拉勾教育

## 课程适合谁学

- **数据挖掘从业者。** 对于初入行业的新人，我可以帮助你建立数据挖掘思维，以及了解数据挖掘的整体框架，迅速进入工作的状态。对于已经在这个行业摸爬滚打的选手，我也希望能够给你带来一点新的想法，让你可以在实践之上又充实了理论，提纲挈领、融会贯通，形成知识的螺旋式上升。
- **数据分析人员：** 数据分析师、金融分析师和其他有数据分析需求的人，过去可能更多的是依赖数据平台和 Excel 进行数据分析工作，大大限制了处理数据的规模和维度，同时缺乏灵活性。这门课可以帮助你了解数据挖掘的一般流程，以及如何使用算法来进行数据分析，让你在工作中更加得心应手。
- **产品、运营、业务负责人，以及其他数据挖掘爱好者：** 数据已经成为很多公司的立命之本，产品设计、运营策略、业务决策几乎都要依托于数据去进行。了解数据挖掘方法，才能制定、提出合理的需求，设定靠谱的策略，做出正确的决策，学习数据挖掘将使你具备优秀的数据思维，同时也能更好地与数据挖掘工程师沟通协作。

## 讲师寄语

虽然数据挖掘所涉及的内容非常广泛，运用的公式又显得晦涩难懂，但请不要担心，本课程中没有让人头疼的算法数学公式推导，也没有各种晦涩的算法理论与实现细节，我会结合工作中遇到的有趣案例来帮你理解每一个算法，如使用 XGB 实现酒店信息消歧、用关联分析找到景点与玩法的关系，等等。

数据挖掘是一种面向应用的方法，在不同的场景下有不同的解决方式，希望这个课程能让你建立解决业务需求的思维，而不仅仅是会一个算法、了解一个场景。

最后，欢迎留言区和我分享你所遇到的困难与挑战，还有你学习成长过程中的小成就，这里会有很多同行的人与你一起精进。

加油！话不多说，我们马上进入课程的学习。