

阿里云天池牛年读书会

《数据分析通识》的故事

分享嘉宾：途索（贾建超）
蚂蚁集团算法专家

天池读书会

TIANCHI 天池

人民邮电出版社
POSTS & TELECOM PRESS

异步社区
人民邮电出版社

《数据分析通识》

数据分析是链接技术认知、理论认知的桥梁。

直播嘉宾：途索

直播时间：3月28日20:00 ~ 21:00



扫码领取
读书会配套学习资源

bookclub





途索（贾建超）

- 2008-2015：毕业于北京邮电大学与电信科学技术研究院
- 2015至今，先后在高德地图、阿里云、蚂蚁集团担任算法工程师的工作，从事多个TB与PB级别的LBS大数据挖掘、企业政府数字化建设与大型商业推荐、广告等数据智能系统建设。
- 数十项发明专利。热衷技术分享，参与公开演讲与分享数十次，著有《数据分析通识》《我是一个程序员》两本书。

关于《数据分析通识》

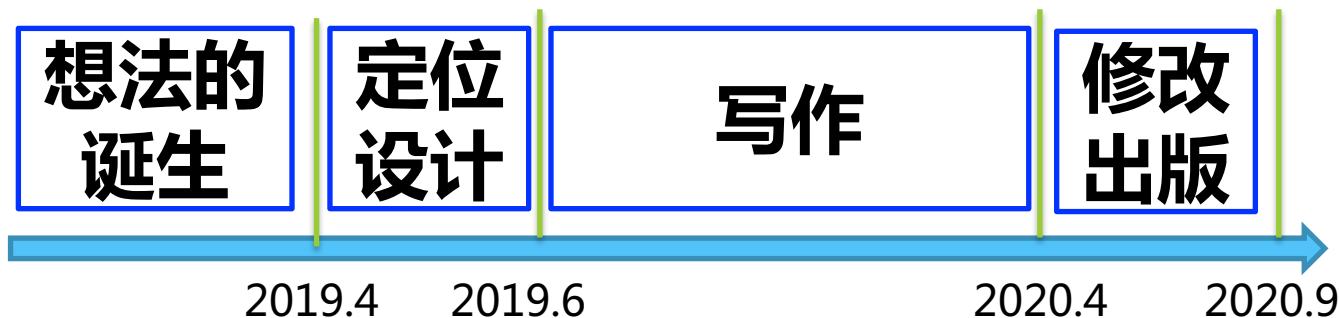
TIANCHI 天池



定位

- 研究数据科学的第一本书
- 研究数据科学逃不过的一本书
- 让新人友好地融入，让老人有所收获

创作时间轴



《数据分析通识》的结构

TIANCHI 天池

序

第0章 技术与业务

第1章 数据处理的基本流程与数据系统的有机组成

第2章 数据获取

第3章 探索性数据分析与目的性数据分析

第4章 数据可视化

第5章 特征工程

第6章 模型

第7章 结果评价

第8章 数据应用与人工智能

第9章 未来的数据与数据的未来

后记：拥抱不确定性的美好

总览

技术与业务的分工区别与数据工作者的定位
数据分析的含义与流程

业务分析

数据分析的本质
业务分析的方法

挖掘与建模

特征工程的架构与方法
模型的内涵、作用、举例与评价

应用与展望

数据应用系统
数据科学与人工智能未来发展的展望

《数据分析通识》-总览

TIANCHI天池

第0章 技术与业务

第1章 数据处理的基本流程与数据系统的有机组成

什么是技术？
什么是业务？

数据分析？
数据开发？
数据算法？

广义数据分析
狭义数据分析

数据处理的一般环节：
数据获取、探索分析、
特征工程、建模、评估

To B 与 To C

《数据分析通识》-业务分析

TIANCHI 天池

第2章 数据获取

第3章 探索性数据分析与目的性数据分析

第4章 数据可视化

数据获取的一般方法
采样数据的陷阱

可视化的常用图表
图表与分析
可视化设计
可视化工程评估

数据分析的本质
和谁比，比什么，怎么比
方法：

假设检验
多维分析与钻取分析
交叉分析
秩次分析
相关分析
回归分析
分类与聚类

《数据分析通识》-挖掘与建模

TIANCHI天池

第5章 特征工程

第6章 模型

第7章 结果评价

变量、字段、属性、
维度、特征

特征获取
特征处理与提取（特
征选择、特征变换、
特征抽取、特征衍生）
特征监控

模型的含义（为人服务还是
为数服务）
一些函数模型（分类、回归、
聚类、关联、半监督等）
模型技巧（交叉验证、过拟
合、迁移学习等）

评价指标：混淆矩阵、
ROC/AUC、M ? (P)E、轮
廓系数、“三”度

《数据分析通识》-应用与展望

TIANCHI天池

第8章 数据应用与人工智能

第9章 未来的数据与数据的未来

数据应用的常见产出形式：
指标、表格、图表、
模型、报告

人工智能与智能域

常见的数据应用系统：
离线挖掘
实时任务
推荐系统
搜索引擎
Feed流

数字化
打通与融合
计算、算力与算法

《数据分析通识》-序与后记

	A1	A2	B	C1	C2	C3	D1	D2	D3	E
0	★★★★	★★★★★	★★	★★★★	★★★★	★★★				
1	★★★★	★★★★★	★★	★★★★	★★★★	★★	★	★	★	
2	★★★★	★★★★		★★★★	★★★★	★	★	★	★	
3	★★★★	★★★★★	★★★★	★★★★★	★★★★	★★★★	★★★★	★★★★	★★	★
4	★★★★	★★★★	★★★	★★★★	★★★	★	★★	★★	★	
5	★★★★	★★★★★	★★	★★★★	★★★★★	★	★★	★★	★	★
6	★★★★	★★★★★	★	★★	★★★★★	★	★	★★	★	
7	★★★★	★★★★★	★★★★	★★★★	★★★★★	★	★★	★	★	
8	★★★	★★★	★★	★★★★	★★★	★★	★★	★	★	
9	★★★	★★★	★★	★★★★	★★★	★★★	★★	★	★	

好的数据胜过好的特征，好的特征胜过好的算法。

回归好比是生活，每人都有顺风顺水、春风得意的时候，每人也都有跌入谷底、遭遇挫折的时候，但生活总是会回归到它本应该有的样子。

召回率与精确率不可得兼，激进勇敢的冒险者获得高召回，而冷静保守的理性人获得高精确。

生活与智能一样，是一个系统。

数据科学是

研究确定性的科学

Thank You

Q & A

学习资料、直播回放交流

大家可以使用电脑访问下方地址或者扫下方二维码进入天池读书会页面，获取读书会相关学习资料、项目实践代码和分享PPT等资源，还可以提前预约其他读书会直播，查看回放。

<https://tianchi.aliyun.com/specials/promotion/activity/bookclub>



扫码领取读书会相关
学习资料