

麦肯锡：优秀数据科学家的5个特征！

Datawhale 2022-02-28 21:00

以下文章来源于数据派THU，作者数据派THU

 **数据派THU**

发布清华大数据相关教学、科研、活动等动态。

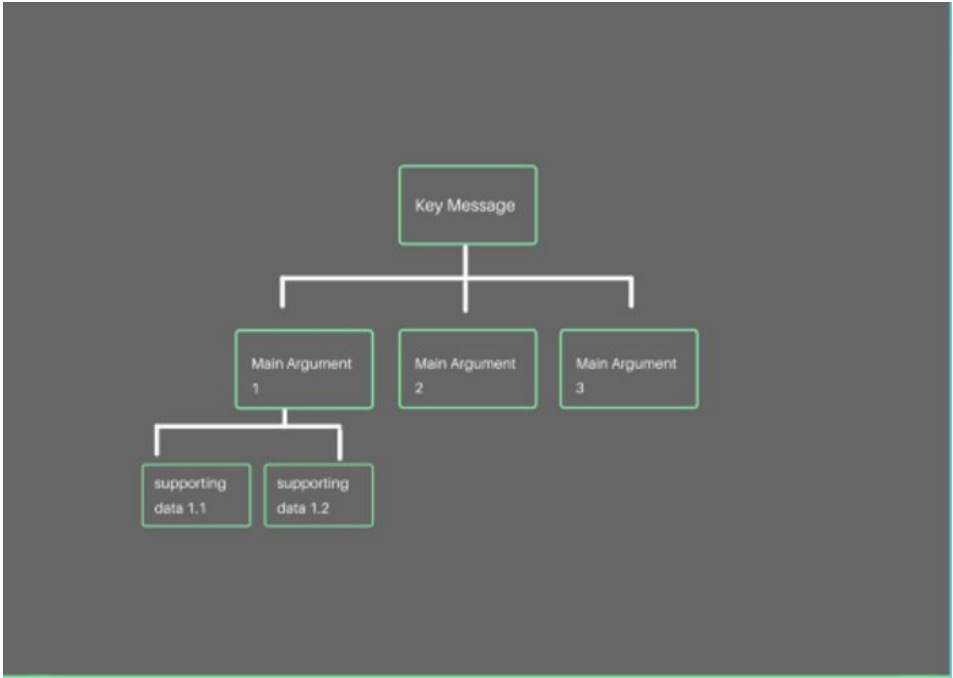
作者：Tessa Xie

本文**约3700字**，建议阅读**5分钟**
本文总结了成为优秀数据科学家的五大关键准则。

近些年来，数据科学家这一岗位已经变得越来越炙手可热，也吸引了大批年轻人涌入渴望在激烈的竞争中抢占一席之地。各个网络平台上都已经有无数千货数据科学行业的简历、求职、面试教程，但是很显然成为一位优秀数据科学家的旅途不会因为求职成功就结束。收到offer仅仅只是第一步。

我在麦肯锡咨询公司的那些年，很荣幸能跟许多厉害的数据科学家以及各行各业的顶尖公司合作。我从中也受益匪浅，并从长期实践和观察中总结出来许多数据科学界的经验和体会。你可能会觉得震惊的一个事实就是：**最优秀的数据科学家不是那些会用各种新奇模型或者代码写得很好的**，事实上，真正优秀的数据科学家是掌握了丰富理论知识的同时还拥有各种“**软知识**”的人。因此，这篇文章总结了我在麦肯锡的学到的成为优秀数据科学家的五大关键准则。

1. 使用金字塔原则沟通是成功的钥匙



巴巴拉·明托的金字塔原则是一项层次性、结构化的思考、沟通技术。这项原理被视为最有效率的信息沟通方式，并且被广泛运用于公司、企业、甚至个人生活中。金字塔原则背后的原理很简单：当你想要交流某个观念或想法的时候，最有效的方法：1. 结论先行 2. 以上统下 3. 逻辑递进

使用金字塔原则进行沟通有以下两个优点：

- 1. 不管你的实际内容有多深奥，使用金字塔原则可以帮助你的听众更好理解你的观点。通常学术论文、或企业报告的开头都会有个摘要，总结整篇文章或材料的中心思想。这样的作法可以确保读者在没办法理解、记住所有文章细节的情况下，也能理解其核心思想。
- 2. 金字塔原则可以帮你适当节省下为不同观众群体准备不同材料（比如演讲ppt）的时间。你可以运用金字塔原则做一份你所要展示的问题的核心思想的演讲材料，针对不同受众群体适当补充他们所感兴趣的细节和论点。

对于大部分数据科学家而言，由于他们本身的工作大部分时间都在关注细节、深挖可分析的点，金字塔原则可能看上去并不那么适用。事实上，我也经常看见许多数据科学家在演讲中并未采取金字塔原则，先从细节开始讲起，而这些演讲的效果也并不好，观众们很难快速理解到他们的核

心思想和问题所在。

如何练习/提升：最简单的方法就是在一切最开始前先根据金字塔原则写下自己的结论、论点、和逻辑递进，并在沟通过程中参照自己最开始前写下的内容确保自己没有跑题。时不时地反问自己一些问题（类似于中心结论和问题到底是什么，问题真的被解决了吗）也可以确保你走在正确的道路上。

2. 你是数据的“翻译官”

如果你看麦肯锡的企业蓝图

<https://www.mckinsey.com/industries/financial-services/our-insights/building-an-effective-analytics-organization>，你会发现上面强调了数据的“翻译官”这个角色的重要性。数据的“翻译官”的也可以理解为企业与数据中心之间的沟通桥梁，需要把一个个数据分析的结论和观点翻译成实际可行性方案。我相信一个数据科学家肯定被问过以下问题：可以请你用大白话或清晰易懂且非专业性的语言解释一下你的观点和结论吗？因此，一个数据科学家应该具备把专业性极强的机器学习模型用清晰易懂的语言解释给CEO或者任何非专业背景的听众的能力，并且同时也应该具备把自己的结论观点翻译成企业实际可行性方案的能力。

具备这些能力有以下优点：

1. 不具备专业背景的人很难成为数据的“翻译官”。麦肯锡曾经尝试过很多方法去培训战略顾问成为“翻译官”，但是在我看来，从未成功过。失败的原因很简单：一个人必须具备足够的专业背景和分析能力才能精准地解释那些复杂的数据分析背后隐藏的意义，而丰富的专业背景和杰出的分析能力都不是短期的培训就能培养出的能力。举个例子，如果你并不理解k均值聚类背后的数学逻辑和原理，你没办法清晰解释为何要这么选这个数字作为你的k值而不是其他数字。而作为一个数据科学家，与其花费时间向那些非专业人士解答他们难以理解的专业术语和原理，不如想办法用清晰易懂的话语“翻译”给他们。
2. 如果数据科学家可以自己解释自己的工作成果，可以大幅度避免信息在传播过程中导致的扭曲和失真。我相信大部分人都玩过传话游戏，传播过程越长，涉及的人数越多，信息越容易在传播过程中被扭曲、误传。这种情况也同样适用于数据科学家的日常工作中，如果你通过别人来“翻译”你的工作成果，你很难保证当信息传播到终端用户的时候跟你最开始想要传达的没有一点偏差。

如何练习/提升：找一位身边的朋友（最好是专业人士），向他解释你的模型和数据分析结果。在解释的过程中，如果你发现你不知道如何把一个概念用清晰易懂的方式表达出来，大部分情况下是因为你并没有完全掌握这个概念。因此，这个练习也是一个很好的查漏补缺的机会。

3. 时刻把实际问题放在第一位

这一黄金准则并不只限于数据科学家，事实上，“时刻把实际问题放在第一位”适用于任何行业中的任何职位。发现问题并提出问题的能力尤为重要，而更重要的是有解决问题的能力。如果没有时刻把解决问题放在第一位，开会讨论的时候经常会陷入一个怪圈：过度关注于问题本身而不是想办法去解决问题。

在我看来，大部分顶尖咨询公司都奉行“时刻把实际问题放在第一位”的准则。作为一个数据科学家，你可能经常会因为不具备专业背景的同事提出一些在你看来很奇怪的稀奇古怪的问题或需求而感到懊恼。我见过许多数据科学家们在这种情况下束手无策，直接拒绝他们的需求。比起直接拒绝，更好的做法是运用自己的分析能力和专业背景帮助他们重新定义问题所在并想办法切实地解决他们的问题。

“时刻把实际问题放在第一位”并不意味着你不能在工作中拒绝任何需求。事实上这意味着每当你拒绝后，你可以紧接着提出“或许.....做可以解决你的问题吗？”

如何练习/提升：当你遇到问题时，在跟同事沟通讨论前，先自己花时间想一想如果是你的话你会如何解决。面对问题时，多运用你的创造性思维发掘新的解决方案，不要害怕做那个提出新的解决方案的人。有时候换位思考的能力也很重要，尝试着以整个公司或者其他业务部门的角度去思考他们会如何解决这个问题。从大局出发思考，有时候也会带来意想不到的解决问题的方案。

4. 有时候可以为了模型的可解释性而适当牺牲模型的准确率。

没有人真的想精准预测出结果，所有人都只是在尝试分析、理解结果。

当一个企业想要建立模型去预测出未来客户流失数时，大部分情况下，人们都忘了在第一时间问为什么我们需要预测客户流失数。企业想要预测出某些结果是因为他们可以提前采取措施应对以避免损失。所以当你的模型告诉CEO“在预测客户流失数里网站访问量的立方根是最显著的特征”意味着什么呢？很可能在他看来毫无意义。

作为一个数据科学家，我以前也经常过度追求模型的准确性。但是我逐渐意识到了通过添加一些无法解释、没有意义的变量；或者调参仅仅只是为了把准确率从96%提高至98%对整个公司来说并没有太大的意义（当然，这个观念仅仅只适用于业务导向的数据科学家们，对于其他领域的机

器学习工程师而言，模型准确率可能极其重要）。

对于大部分高管和业务部门来说，如果你无法解释你的模型和模型的结果，那你的模型可能在他们眼里看来并没有那么可靠。模型只是为了辅助他们去更好地做出商业决策，所以模型更大的意义是可解释性而不是准确率。

如何练习/提升：当你在建模或做数据分析时，时刻反问自己“这会如何影响公司呢”。建模过程中，尽量避免放一些随机、意义不大的交互特征。当你的模型或分析有结果后，可以根据结果具有针对性地写下对公司的建设性意见，这样的作法也会帮助你更好的评估自己的模型是否适用。

5. 在开始前，确保自己有个假设场景，但是不要仅局限于一个假设。

在开始任何分析工作之前，设立好自己将要分析的假设场景尤为重要。如果不设立好分析背景，你可能在前期数据处理、EDA 或者是如何选择特征时一头雾水。如果没有任何假设，AB测试也没办法进行。尽管提前设立好自己的假设如此重要，我也见过没有提前设立好假设就开始处理问题的数据科学家。在这种情况下，数据科学家们往往把设立假设这个关键任务丢给了缺乏专业背景的业务部门，而他们往往会因为缺乏一些专业背景和对数据的熟悉设立一些实际工作场景中无法检验的假设。因此，在我看来最好的做法是数据科学家们在一开始就应该参与到设定假设的过程中去指引他们哪些是可尝试的而哪些是无法尝试的。

设立假设尤为重要，因为它是一切的起点，但是这并不代表设立好一个假设后就万事大吉了。我经常看见许多数据集科学家们过于执着于一个假设，尽管这个假设带来了自相矛盾的结果。太过于执着于一个假设可能会导致为了满足这个假设而去人为的更改一些数据或进行一些影响准确率的操作。如果你听过辛普森悖论，那你一定很清楚数据是具备撒谎的能力的。一个好的数据科学家应该把数据可靠性放在第一位，因此不要仅仅只局限于一个假设。

如何练习/提升：设立假设是具备业务思维和敏锐度尤为重要。有了假设之后跟着你的假设去进行数据探索，但是当数据呈现出与你的假设相反的结果时，也不要过度执着于自己一开始的假设。

人们通常把人才分为两种：战略性人才和分析性人才，这样的分法就好像在暗示一个人无法同时拥有这两种能力。但是事实上，一般最好的分析性人才恰恰是那些能理解公司战略性政策和想法，并且懂得如何和业务部门沟通的人。而最好的战略性人才也恰恰是那些具备足够数据分析能力的人。

原文标题：5 Lessons McKinsey Taught Me That Will Make You A Better Data Scientist
原文链接：<https://towardsdatascience.com/5-lessons-mckinsey-taught-me-that-will-make-you-a-better-data-scientist-66cd9cc16aba?gi=2242e432865c>
译者简介：苗雨，美国加州大学洛杉矶分校应用统计学硕士在读，毕业后准备继续在统计学领域读博深造。对机器学习，因果推断很感兴趣，希望能和大家一起成长。
转自：数据派THU 公众号；

AI学习路线和优质资源，在后台回复“培养方案”获取

Datawhale

和学习者一起成长

一个专注于AI的开源组织，让学习不再孤独



长按扫码关注

整理不易，点赞三连！

阅读原文

喜欢此内容的人还喜欢

吴恩达新动作：建立全新机器学习资源Hub，「以数据为中心的AI」大本营
量子位

23个优秀的机器学习数据集
AI蜗牛车

【他山之石】机器学习全面教程 & 半监督学习
人工智能前沿讲习