



Kaggle入门，看这一篇就够了

 a2Mia姐 · 4 个月前

这次酝酿了很久想给大家讲一些关于Kaggle那点儿事，帮助对数据科学(Data Science)有兴趣的同学们更好的了解这个平台，最好能亲身参与进来，体会一下学校所学的东西和想要解决一个实际的问题所需要的能力的差距。虽然不是Data Science出身，但本着严谨的科研态度，在进行了大量的调研、学习以及对相关经验者的访谈之后，决定写下这篇专栏，一方面让那些对数据科学（Data Science）有兴趣的求职者和申请者能真正了解这个项目，另一方面也让那些有想法参与进来的人get到正确的打开方式。

注：此文章只允许ApplySquare（申请方）的相关媒体转账，其他不给予转载权。

什么是Kaggle？

Kaggle成立于2010年，是一个进行数据发掘和预测竞赛的在线平台。从公司的角度来讲，可以提供一些数据，进而提出一个实际需要解决的问题；从参赛者的角度来讲，他们将组队参与项目，针对其中一个问题提出解决方案，最终由公司选出的最佳方案可以获得5K-10K美金的奖金。

除此之外，Kaggle官方每年还会举办一次大规模的竞赛，奖金高达一百万美金，吸引了广大的数据科学爱好者参与其中。从某种角度来讲，大家可以把它理解为一个众包平台，类似国内的猪八戒。但是不同于传统的低层次劳动力需求，Kaggle一直致力于解决业界难题，因此也创造了一种全新的劳动力市场——不再以学历和工作经验作为唯一的人才评判标准，而是着眼于个人技能，为顶尖人才和公司之间搭建了一座桥梁。

这里有一篇对Kaggle首席科学家Jeremy Howard的采访，介绍了Kaggle的创建初衷及运营模式，即任用最聪明的人解决世界上最棘手的问题；同时，任何公司和组织都可以受益于机器学习的发展进步，感兴趣的同学可以戳这里：[原文链接](#)，来进一步了解。

Kaggle的竞赛模式是什么样的？

Kaggle上的竞赛有各种分类，例如奖金极高竞争激烈的的“Featured”，相对平民化的“Research”等等。但他们整体的项目模式是一样的，就是通过出题方给予的训练集建立模型，再利用测试集算出结果用来评比。同时，每个进行中的竞赛项目都会显示剩余时间、参与的队伍数量以及奖金金额，并且还会实时更新选手排位。在截止日期之前，所有队伍都可以自由加入竞赛，或者对已经提交的方案进行完善，因此排名也会不断变动，不到最后一刻谁都不知道花落谁家。

