

# 电影大数据

电影评分对于观众选择电影有很重要的参考建议，如今比较热门的电影评分网站包括：豆瓣，IMDB，猫眼电影，Rotten Tomatoes 等等，每个网站都有自己不同的评分模型和规则，参见[1]。目前的评分模型有直接求平均分的，但这样可能会存在恶意刷分的现象；也有给不同的用户分配不一样的权重的，大 V 的权重较高，甚至某些网站只允许大 V 影评人评分；还有通过贝叶斯统计算法来进行综合评分的。每种评分模型都有各自的优缺点，大部分网站也没有确切给出评分模型的具体细节。

而对于电影出品方、投资方来说，更关注的不是评分，而是电影的票房。如果能够在电影出品前期准确预测该电影的票房，会对电影出品方、投资方提供更好的参考。

如何建立合理的电影评分体系以及如何预测电影的票房走势，都是如今大数据时代的需求，请你们团队完成如下几个任务：

1. 请分析比较现有的多家网站的电影评分模型，在此基础上给出你们团队认为合理的电影评分模型；
2. 豆瓣 TOP50 电影和 IMDB 的 TOP50 电影有很多相同也有很多不同的电影，请基于你们的分析，给出影响两者排名不同的具体因素，并且给出你们的电影评分模型下的 TOP50 电影；
3. 请你们团队建立电影票房预测模型，并预测今年如下两部电影的全球票房：Nolan 的《信条 Tenet》和娄烨的《兰心大剧院》（可假设完全不受疫情影响，电影正常上映，电影院正常开放）；
4. 依据你们团队的票房预测模型，分析影响电影票房的最重要的因素。

参考：

[1] [豆瓣、猫眼和 IMDB 的电影评分是怎么得出来的？](#)