助教你好！我们是数据科学基础这门课的学生，本次大作业的选题为：计算社会学篇。现阶段对如下方面仍有疑问，希望能够得到解答。

1. 心态词典主要基于一些据有明确情感的词语，比如“太好了”对应于高兴，但是在荔枝网/人民日报等诸多网站上都以官方新闻和正式的新闻为主，文字风格中立，很少能找到这样的词语，对于这类新闻应该怎么建立心态字典呢？
2. 机器学习相关的知识有点多，而且学习起来没有头绪，请问助教能否帮助缩小范围，或者提供一些参考资料甚至指导？
3. 心态字典最开始应该怎么建立？采用什么样的技术手段？
4. “新闻”和“评论”的比例难以达成一种平衡，比较便于获得的信息大部分是新闻，主要代表了官方的意志，能拿到的大众的数据其实很少。现阶段我们能够找到的数据来源中，鲜有“评论”这一方面。而微信朋友圈和QQ空间的说说几乎获取不到，微博内容和b站弹幕的获取难度也较大。怎么保证数据能够代表“大众”？或者说如何定义重点新闻？
5. 如果方便，能否提供些许关于TF-IDF技术的作用和技术指导？
6. 请问“大数据”的数据量是否有具体的量（或者数量级），不知道大概要获取到多少数据才能有效支持我们的分析结果。
7. 课上听陈老师说要给我们一个有关神经网络相关的api，届时是否能有相关使用示范之类？
8. 想了解关于具体如何运用课内知识，对于课内知识点的运用量如果不是很大，是否会对最终结果造成一定影响？（如课本中的：参数估计、假设检验等感觉套不进去，请问能否有个大体方向指导？）