1. 数据获取
   1. 微博原贴博文采集
   2. 微博一级评论采集
   3. 股价采集
2. 数据预处理
   1. 数据清洗，去重，机械压缩、去符号、去表情包
   2. 分词处理，jieba分词、保留名词、形容词、动词
   3. 情感分析，snownlp模型，情感打分
3. 数据对比
   1. 情感波动，月份正负面占比趋势，查看整体情感倾向
   2. 指标确定，查看指标分布情况，确认指标是否有异常
   3. 指标趋势变化，查看指标按月份变化的趋势
   4. 指标比对，查看情感热度与股价变化影响情况
   5. 指标比对，查看互动量与股价变化影响情况
4. 相关性分析
   1. 利用股价和舆情的数据做相关性分析，查看股价和发帖数量的影响情况
   2. 利用股价和舆情的数据做相关性分析，查看股价和情感热度的影响情况
   3. 在上面的基础上，做线性回归，进一步确认两者对于股价的影响
   4. 再线性回归的基础上，再加上非线性回归，随机森林去确认两者对于股价的影响
   5. 通过上面的两个模型来说明，二者对于股价的影响情况，增加说服力
5. 股价预测
   1. 采用lstm模型来对股价进行预测
   2. 模型的背景说明，为什么要选用这个模型
   3. 模型预测的效果如何，根据给出来的指标说明
   4. 模型预测的值和实际值比较，用可视化的形式展示，使得模型更有说服力