现在大众在进行网购之前都会先看下相关商品的评论，包括好评与差评，再综合衡量，最后才决定是否会购买相关的物品。

根据商品的好评与差评，提取出关键词，可以快速了解一件商品的好与不好的地方。

对于一个关心该产品使用性、安全性等的消费者来说，他希望从评论中了解到他想了解的问题，即已经购买使用过的用户认同该商品的哪些优点，不认同该商品的哪些弱点或缺点。

使用LDA主题模型

如果一条评论有一个中心思想或主题，那么一些特定词语会更频繁地出现。一条评论通常可以包含多种主题，每个主题所占比例各不相同。对于不同主题的评论而言，其不同的主题实际上构成了一个分布。

一条评论有主题，对于所有的评论，我们也可以分析其主题的分布，借此可以看出消费者对于产品主要关心的方面是什么。

在 LDA 中，主题的数目没有一个固定的最优解。模型训练时，需要事先设置主题数，根据训练出来的结果，手动调参，优化主题数目，进而优化文本分类结果。

LDA主题模型本质上是一个词袋模型，不考虑词的顺序。给定一条评论，可以看到都是由哪些词构成。

LDA模型假设每条评论都是在讲述若干个主题，至于具体几个主题，可以认为预先设定k值。

通过LDA模型，可以求出这k个主题（latent）中，每个主题到底长什么样子。

LDA优点：

（1）无监督：不用对训练数据打标签

（2）预处理简单：分词，去除停用词

（3）参数少：一般只需要设定主题数k

在此场景下，使用LDA模型的目标输出

（1）商品好评的若干个topic中，其关键词各是什么，以及每个关键词的权重。

（2）商品差评的若干个topic中，其关键词各是什么，以及每个关键词的权重。