

Homework 1

Name: Ye Juelin

Student ID:1801212963

I. Description of Problem and Variables

With the proliferation of mobile network and micro-blog big data marketing application, it becomes a critical issue for enterprise operators to achieve accurate and efficient online marketing to improve the market competitiveness of enterprises and the share of sales.

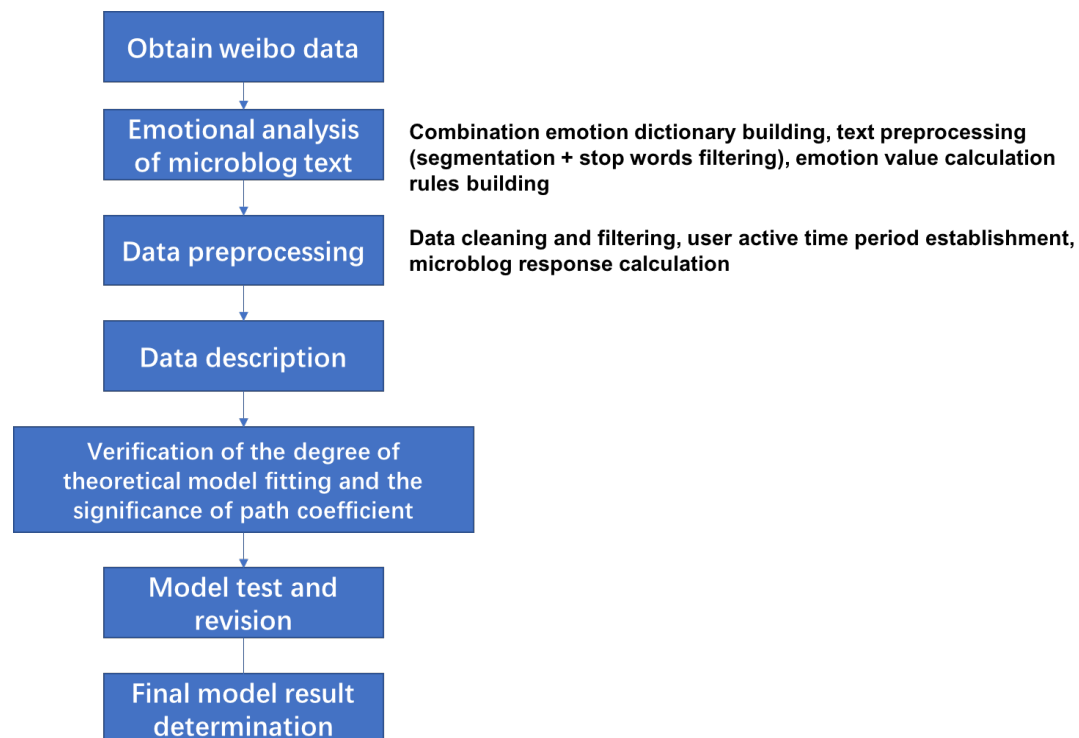
Through crawling the microblog marketing data of beauty enterprises (taking example of Baiqueling), this project explores how the characteristic variables of online media marketing, including the *number of mentioning others '@', pictures, videos, hashtags ('#'), the posting time in users' active period, the text length of marketing micro-blog and micro-blog sentiment intensity*, affect the degree of reaction of marketing micro-blogs.

Model variables	Variable attributes
Times of mentioning others ('@')	independent variable
Number of pictures included	independent variable
Text length	independent variable
Number of external URLs included	independent variable
Number of hashtags	independent variable

Score of emotional intensity	independent variable
Whether the release time is in the active period	independent variable
Weibo epidemic response, including comments, likes and repos	dependent variable

After exploring the micro-blog dissemination pattern and mechanism, this project could provide with practical basis for the improvement of online marketing strategies in the beauty industry, so as to enhance the micro-blog marketing efficiency and effectiveness of beauty companies to some certain extent.

II. Workflow



III. Database type selection

1. Relational database

Weibo characteristic value: to store the characteristic data of each Weibo. These data will not be changed after Weibo was published, and they are structural and suitable to be stored in a relational database.

Weibo epidemic response: to store repos, likes, comments of each weibo, which changes with time and needs to be updated daily.

2. Non-relational database

Microblog text data: considering that microblog is a platform with 140 characters as the main expression carrier, the analysis of microblog text is very important. Since the texts of microblog are unstructured data, they are suitable to be stored in non- relational database for further text sentiment analysis.