Dear Sir/Madam,

We are very honored to have this opportunity to show you and your company the results of our mining and analysis of the data of hair\_dryer, microwave and pacifier. Here we will give you a comprehensive report about our research in the data.

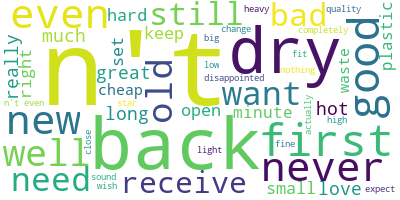
First of all, through the preliminary observations provide three data sets, we find that each product in the marketplace in the US, and product\_category respectively within each data set is the same, we think it useless for analysis, so to be deleted. Similarly, found product\_parent data with product\_id data redundancy, then we delete the product\_parent column data. At the same time, do other data preprocessing.

After emotional analysis in natural language processing, we will not quantitative reviews is mapped to the numerical space, using the numerical range of [1, 1] to comment contains specific emotions. The higher the number, the stronger the positive emotion. After the emotional analysis, we carried out feature engineering. A good feature engineering can make the model better grasp the data features, better discover the internal patterns in the data, and improve the accuracy of the model. Through feature engineering, a new feature rating rate was established and some abnormal conditions were removed. For example, a 1-star comment: "I have used the dryer several times and it works great. I had questions which were answered by other customers which was helpful in making my decision. Definitely recommend.". However, this should be a 5-star favorable rating, so the data with opposite star rating and rating were deleted. Fortunately, that's not a lot of data, with only 27 of the three data sets. Then, we divided the data into the training set and the test set according to the ratio of 8:2, and took the ones with a rating greater than or equal to 4 stars as the positive example, and the ones with a rating less than 4 stars as the negative example, and used the machine learning algorithm LightGBM to conduct separate data modeling for each data set. Finally, the F1 value of the model on the three data sets was 0.889, 0.874 and 0.939, which showed that our model was very successful! Not only that, we also learned from the model that review\_body, review\_date, rate and total\_votes are the most important to the model. Your company should mainly focus on the changes of these four types of data, and try to pay little or no attention to other data with low importance such as reviewers. Once your company plans to sell the three products online, you can use the above model to evaluate the sales data of the products, and compare the predicted results of the model with the actual scores of users to get the result of whether the two are consistent. If it is consistent, it will seize the evaluation characteristics of such users and make targeted improvements to the product to achieve better sales. If it is inconsistent, you can focus on whether such users are fake or malicious users and so on.

We also use entropy weight method to obtain product scores from comprehensive analysis of rating and evaluation, which can measure product reputation and other relevant information. In order to evaluate the potential success or failure of a product, we combined the results of the established machine learning model LighGBM and entropy weight analysis, selected important indicators related to text and rating for fuzzy comprehensive evaluation modeling, and finally concluded that the potential success of a product can be indicated by calculating specific scores.

Finally, we use the correlation analysis technique to obtain that certain ratings will trigger certain types of reviews, that is, users' reviews may be influenced by previous ratings. For example, "-do NOT BUY - the 5 star reviews are fake".

At the same time, we set up a specific vocabulary list and drew the following two word cloud maps for the word frequency statistics of all comment data. The font size of the words in the word cloud map is related to how often they appear in the comments. The higher the frequency, the larger the font. We can see from the word cloud map that the one-star word cloud map is mostly negative words, such as “n't” “like”, “never”, “bad”, while the five-star word cloud map is mostly ‘great’, ‘love’, ‘well’, ‘perfect’, etc. Although there is “n't” in the word cloud image of five stars, we find that it mainly represents positive emotions through specific analysis, such as “like n't allow finger print” and so on.

One star word cloud map Five star word cloud map

Our data mining results are presented, I believe you will have a lot of harvest after you see, finally, I hope your company's products can be a great success!