

Analysis of Online Social Network Affecting Making of Choosing Active Ingredient in Facial Moisturizer Products

Panithan Intharawicha¹ and Phisanu Chiawkhun²

¹Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand

²Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

panithan_i@cmu.ac.th

Abstract. This independent study aims to analyze an online social network in a famous website called "Pantip" by presenting the trend of customer interest on choosing the active ingredients in facial moisturizer products. The investigation was done by collecting customer's opinions on the active ingredients from threads in beauty forum of Pantip website by using Python syntax. The data collecting were analyzed by using tokenization and word count. It was found that the top three most popular active ingredients are collagen, retinol, and niacinamide respectively. By collecting keywords referring to these active ingredients, the customers mentioned collagen in terms of moisture, brightness, and resilience similarly, retinol was mentioned with related to anti-aging, pore firming, and moisture. Lastly, niacinamide was commented about moisture, resilience, and pore firming. When data were plotted and presented by word cloud technique. It can be implied that the top five most popular key words showed similarity in term of types but difference in order. However, the statistical analysis by Friedman test showed that the rank of the key words of the active ingredients showed no considerable difference. Therefore, it can be summarized that the customers commented about the top three active ingredients with no significant difference.

Keywords: Data Mining, Social Network Analysis, Word cloud, Active Ingredients

1 Introduction

Nowadays, Thai customers spend more than 40 percent of a day on searching information on the internet. Among that period, they spend 3 hours a day on electronic commerce and social media.[1] In addition, Thai customers have information-searching behaviors. This can be interpreted from increasing rate of using the keyword "review" on search tool of Google.co.th.[2] This characteristic implies that Thai customers always seek important information on the internet before making decision on product purchase. These behaviors of Thai customers make the internet forum called "Pantip.com". The

In this study, the analysis of the internet forum on the Pantip.com was done by presenting the trend of customer interest on choosing the active ingredients in facial moisturizer products. The investigation was done by collecting customers' opinions by using Python, analyzing through tokenization and word count, and presenting the results as a word cloud. The statistical test of hypothesis was also performed, and the results were discussed.

[illegible]

Fig. 1. An example of the collected data in a CSV format.

2.3 Text Cleaning

The collection of contents and comments usually contain different symbols and words which doesn't convey meaning to the analysis. Therefore, the texts in the CSV file were cleaned by removing noisy data and stop words before being used in the next step.

2.4 Tokenization

The texts were spitted into individual words using dictionary-based algorithm via a PyThaiNLP library[5]. The process was repeatedly done on the sentences appeared on the forums related to facial moisturizers.

2.5 Word Counting

The appearance frequency of active ingredients was counted in order to present the most popular mentioned words, which refer to the customer interest.

2.6 Word Cloud Creation

Word clouds were created to illustrate the interest trend of Thai customers. In addition, they also represented the appearance frequency by variation of the word size.

3 Experimental Result

1296 comments related to the active ingredients mixed in facial moisturizer products were collected and analyzed. The active ingredients of 106 words were recorded and there were 488 keywords related to the beneficial effects of the active ingredients. It was found that the "Collagen", "Retinol" and "Niacinamide" were the top three most popularly mentioned active ingredients. Therefore, the word cloud of the active ingredients could be created and illustrated in Fig.2.

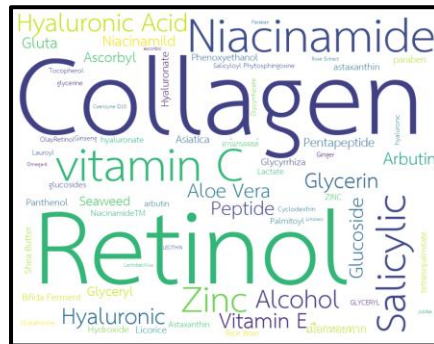


Fig. 2. Word cloud of the active ingredients which were paid attention by the customers.

In addition, the benefits of these active ingredients mentioned by the customers were also analyzed. The most popular mentioned beneficial effects of collagen, retinol, and niacinamide were presented as word clouds individually in Fig.3, Fig.4, and Fig.5, respectively.



Fig. 3. Word clouds presented the most popularly mentioned beneficial effects of collagen, retinol, and niacinamide, respectively.

4 Conclusion

According to the contents and comments regarding the active ingredients mixed in facial moisturizer products posted on the forums of Pantip.com, collagen, retinol, and niacinamide were found to be the most popular mentioned active ingredients. This result implies that customers have paid attention on these active ingredients for choosing facial moisturizer products. Moreover, the customers also commented the beneficial effects of these active ingredients with no considerable difference.

5 Discussion

As seen in Fig.2., the top three most popular active ingredients were collagen, retinol, and niacinamide, respectively. In addition, by analyzing keywords that refers to these active ingredients, the customers mentioned collagen in terms of moisture, brightness, and resilience. Similarly, retinol was mentioned with related to anti-aging, pore firming, and moisture. Lastly, niacinamide was commented about moisture, resilience, and pore firming. It has seen that the top five most popular keywords showed similarity in term of types but difference in order. The statistical test of hypothesis by Friedman test for the rank of keywords indicating beneficial effects showed that there are no different in customers commented about the top three active ingredients at significance level of 0.05

References

- [1] Electronic Transactions Development Agency (Public Organization), Thailand internet user profile 2018, 2561.
- [2] The increased rate of using the keyword review on search tool of Google.co.th [Online]. Available: <https://trends.google.co.th/trends/explore?q=ครีม&geo=TH>. [Accessed: 16-Sep-2020].
- [3] Social Analytic of cosmetic and personal care industry,Thith Zocial. [Online]. Available: <https://positioningmag.com/1139135>. [Accessed: 24-Dec-2019].
- [4] Web Scraping Beautiful Soup function. [Online]. Available: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. [Accessed: 04-Apr-2021].
- [5] Word tokenization with pythainlp. [Online]. Available: https://pythainlp.readthedocs.io/en/stable/pythainlp-1-4-thai/#_1. [Accessed: 04-Apr-2021].