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

Pantip.com is so popular because this website provide an internet discussion forum and a source of specialized knowledge. [3] It also influences the customer's buying decision, especially on choosing a cosmetic product. Therefore, the collection of forum titles, information and comments from the social network through Pantip.com will lead to understanding of the customer interests.

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 world cloud. The statistical test of hypothesis was also performed, and the results were discussed.

2 The Proposed System

The analysis of this study is divided into six sections: (1) purposive sampling, (2) web scraping, (3) text cleaning, (4) tokenization, (5) word counting, and (6) word cloud creation. The details of each section are explained as follows.

2.1 Purposive Sampling

This process was done by selecting the forums based on the following criteria: (a) discussion topics were in the To-Kruang-Pang forums of the Pantip.com; (b) the topics were posted during October 2019 to January 2020; (c) the discussion is based on active ingredient mixed in facial moisturizer products; and (4) at least one comment was posted in the forums.

2.2 Web Scraping

In Python, a selenium command was used to access and collect the contents and comments via BeautifulSoup function[4]. The data collecting was stored in a CSV file. Fig.1. illustrates an example of the stored contents and comments.

279	1277	39445125 สวัสดีค่า วันนี้เราจะมาบอกเล่าความรู้สึกหลังจากที่โดโช้ BIODERMA Sensibio กันค่า ถ้าโครโม่อยากอ่านตัวหนังสือ เราท่าเป็นะ 1962304	12/17/2019 23:09
085	1278	39445125 Hello ผม Touch MANMAN REVIEW จากปอง Double Bear ครับผม สวัสดีครับสาวๆหนุ่มๆทุกคน ทัศแมนแบบริวิว มาแล้วววววว 2967792	12/18/2019 7:27
281	1279	39445125 รีวิวหลังจากได้รับผลิตภัณฑ์ Bioderma Sensibio ขนาดทดลองสำหรับผิวแพ้ง่าย มาเป็นที่เรียบร้อยแล้วและหลังจากทดลองใช้ 959629	12/18/2019 9:43
282	1280	39445125 ขอบคุณกิจกรรมดีของ พันทิป วันนี้เราจะมารีวิว BIODERMA Sensibio ขนาดทดลอง - ใบโอเตอร์มา เช็นชิบิโอ เอชทุโอ (BIODE 3104157	12/18/2019 10:05
283	1281	39445125 สวัสดีค่ะเพื่อนๆ กลับมาพบกับโอซอีกแล้ว ^0^ และเนื่องจากโอซมีใด้โอกาสใช่ผลิตภัณฑ์ดีๆ ของ BIODERMA Sensibio (สำหรั 1456522	12/18/2019 12:51
284	1282	39445125 รีวิว ความรู้สึกหลังการใช้ bioderma เชตทตลอง หลังจากได้รับผลิตภัณฑ์มา สิ่งแรกที่ลองเลยคือ ผลิตภัณฑ์ทำความสะอาดผิ 2243363	12/18/2019 12:54
285	1283	39445125 สวัสดีค่ะ วันนี้ เราจะมารีวิว ผลิตภัณฑ์ หลังจากที่ได้ใช้ Bioderma Sensibio ผลิตภัณฑ์ที่เราได้รับมา จะมีทั้งหมด 3 ปั้น ดังภาพ 1877555	12/18/2019 13:35
86	1284	39445125 ขอรีวิวแบบบ้านๆนี้แหละเด้อ เป็นอี่ห้อที่สังเกตเห็นตลอดตอนแว่บเข้าไปขึ้นเวชสาอางค์ แต่ก็มองข้ามไปทุกที่สีน่า อาจจะเพราะ 200617	12/18/2019 16:10
87	1285	39445125 หลังจากใต้รับผลิตภัณฑ์ Bioderma Sensibio และทดลองใช้เกือบ ๆ เดือน เท่าที่สังเกตได้ชัดจากผิวหน้าคือ ผดผื่นได้ลดลงอย่ 600140	12/18/2019 16:23
88	1286	39445125 รีวิวการทดลองใช้ผลัดภัณฑ์ Bioderma Laboratoire Dermatologique ได้รับและทดลองใช้ผลัดภัณฑ์ที่ 3 ชนิดเรียบร้อยแล้วค่ะ 187326	12/18/2019 16:47
189	1287	39445125 รีวีว BIODERMA SENSIBIO สวัสดีเพื่อบุขาวพันที่ปทุกคนนะคะ วันนี้เรามารีวิว Bioderma Sensibio เป็นskinecareสำหรับผิวบอน: 4590820	12/18/2019 17:10
90	1288	39445125 หลังจากได้ ผลิตภัณฑ์ Bioderma Sensibio จึงให้ทดลองใช้คือตามลำดับขั้งตอบการทำความสะอาดและการบำรุงผิวเขตนี้ ล่าะ 1274991	
191	1289	39445125 จากใต้รับผลิตภัณฑ์ขนาดทดลองไปจากไปผลิตภัณฑ์ Bioderma ซึ่งเป็นกิจกรรมของเวป Pantip นี่คือรีวิว หลังจากใต้รับผลิตภัณ	
192	1290	39445125 สวัสดีค่ะเพื่อนๆ ช่วงนี้คงเป็นช่วงที่หนักใจสำหรับสาวๆที่ผิวแห่งและมีผิวที่ระดายเดืองง่ายหลายคนนะคะ เพราะสภาพอากาศใเ 656290	12/18/2019 20:05
93	1291	39445125 สวัสดีคร้าบ ผม ชันนี่คนเดิม เพิ่มเดิมคือวันนี้ผิวไส แข็งแรง!! เพราะว่า ได้ไบโอเดอร์มา มาใช้ มาแบบมิสโดล์ ในกูงชีป นี้คิดอยู่ว่ 4919039	12/18/2019 20:09
94	1292	39445125 มาแล้นๆๆๆ สวัสดีคำาา พรนะคะ ขอร่วมรีวิว ผลิตภัณฑ์ Bioderma Sensibio ขนาดทดลองตัวยนะคำา ใครที่มีผิวบอบบาง ระคาย 2063805	12/18/2019 20:41
95	1293	39445125 และแล้ววันสุดท้ายของการรีวิก็มาถึง สายร่วมกิจกรรมอย่างเรา ก็ไม่พลาดที่จะมารีวิวแห่นอน และก่อนอื่น ต้องขอขอบคุณกิจกร_818148	12/18/2019 21:04
96	1294		12/18/2019 21:11
97	1295	39445125 โบเป็นคนที่บอบบาง แพ้ง่าย โบเคยแพ้โพ่มล้างหน้าที่มีสารเคมี ทำให้หน้าโบพัง ในตอนนั้นโบใต้รักษาสิวกับหมอจนหาย และเ 2755168	12/18/2019 21:40
98	1296	39445125 เว็มทดลองคลีหชั่งต่ะกลิ่นดีไม่ฉุนไม่แสบหน้า 8/10เพราะ ผิวไม่ดีอาจจะเกิดจากการล้างหน้าผิดวิธีก็ได้ เพราะผิวหน้าของเราเจะ 339569	12/18/2019 21:43
00	-		

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 works before being used in the next step.

2.4 Tokenization

The texts were spitted into individual words using dictionary-based algorithm via a Py-ThaiNLP 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 beneficials 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=ີ່າ ວິວ&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 tokenizeation with pythainlp. [Online]. Available: https://pythainlp.readthedocs.io/en/stable/pythainlp-1-4-thai/#_1. [Accessed: 04-Apr-2021].