Sentiment Analysis of Customer Review Behavior in the Online Market toward Health Supplement Products

Phennapa Joomponpong[1]
[1]Faculty of Management Science, Thepsatri Rajabhat University, Thailand
(Received: February 19, 2025; Revised: May 22, 2025; Accepted: June 10, 2025)

Abstract

This research analyzed customer sentiment through review behaviors towards health supplement products in the online market by utilizing natural language processing techniques. Review data were collected from 10 brands of health supplements, comprising 6,284 reviews in Thai, on the Shopee platform from July 2023 to March 2024. The analysis results revealed that positive sentiment scores higher than 0.6 clearly reflect customer satisfaction. K-Means and Fuzzy C-Means were employed for clustering models to enhance accuracy. To determine the optimal number of clusters, Elbow Method and Silhouette Score were utilized. Ultimately, it was found that customers were divided into two clusters (K=2), with key focus on words such as "efficacy," "product," "good," "taste," "consume," and "quality," which are critical factors influencing purchasing decisions. Furthermore, an evaluation of clustering quality using the K-Means and FCM models showed no significant differences. As indicated by the Silhouette Score (0.5319) and Davies-Bouldin Index (0.6991), both models held comparable clustering performance. This study offered important business implications, including customer segmentation, behavior analysis, marketing strategy development, and product/service improvement to help strengthen customer relationships and enhance competitiveness in the rapidly changing online market. Consequently, the findings of this research served as a vital tool for forecasting trends, developing effective strategies, and meeting future customer needs.

Keyword: 1) Sentiment Analysis 2) Customer Review Behavior 3) Health Supplement Products 4) Online Market 5) Purchasing Decisions

[1] Lecturer, Department Business Administration; E-mail : phennapa.j@lawasri.tru.ac.th

Introduction

According to the 2023 Global Economic Monitoring Report on Health by the Global Wellness Institute (GWI), the health market, particularly in mental wellness, has grown rapidly post COVID-19. The market value was estimated at USD 180.5 billion, with expectations of reaching USD 330 billion by 2027 (GlobalWellnessInstitute, 2024). Mental health aids, such as brain training, cognitive enhancement, sleep aids, vitamins, supplements, and cannabis-infused items, have shown strong growth as consumers seek ways to cope with stress and daily life challenges (Thansettakij, 2023).

The GWI further predicted that the global health economy will reach a value of approximately USD 7 trillion by 2025 (GlobalWellnessInstitute, 2024). Regarding the dietary supplements market in Thailand, Euromonitor reported it was valued at THB 53,810 million in 2016 with projection to grow to THB 74,247 million (Packhai, 2023). Based on the survey in 2024, the result found that Thai working adults are increasingly health-conscious, with 65% prioritizing healthy ingredients and 41% paying attention to nutritional labels. Online shopping has surged, reaching a value of THB 700 billion in 2023 (Meeprasert and Rattakan, 2021, pp. 6–18). The growth of the health market and changes in Thai consumer behavior underscore the importance of health awareness and online shopping. Increased purchasing through online platforms compels businesses to prioritize customer feedback and sentiment to understand the Customer Journey and User Experience from the initial interaction to product usage (NoonInch, 2022). Notably, 95% of buyers read online reviews before making purchasing decisions, as reviews, particularly positive ones, significantly enhance credibility (SMEThailand, 2023).

This research focused on analyzing customer sentiments derived from reviews of health supplement products in the online market. The objectives of this study are twofold: (1) to analyze customer sentiment from online reviews of health supplement products using natural language processing (NLP), enabling businesses to gain deep insights into customer needs and opinions while identifying hidden structures and patterns for clustering reviews into positive, negative, and neutral categories (Boonchote, 2022), and (2) to develop a customer sentiment analysis model using Natural Language Processing to extract insights from unstructured text. Although the dietary supplement market and consumer behavior that increasingly values online reviews have continued to grow, there remain research limitations, particularly in the deep analysis of customer sentiments using Natural Language Processing (NLP) in conjunction with Unsupervised Learning models. These techniques have not yet been widely applied within the Thai language context (Gupta and Patel, 2021, pp. 511-517; Cathcart, 2022). Previous studies have primarily focused on quantitative indicators such as star ratings, lacking in-depth interpretation of emotions and user experiences expressed in textual reviews (Dellarocas, Zhang and Awad, 2007, pp. 23-45; SMEThailand, 2023). Applying Unsupervised Learning models enables the discovery of hidden patterns and relationships in review data (Gupta and Patel, 2021, pp. 511-517) without requiring pre-labeled data. This approach reduces the complexity of handling large datasets, processes information quickly, and generates reports automatically, saving time in data labeling. The findings can help businesses promptly adjust marketing strategies and develop products to stay competitive in the market (Sukheja, Chopra and Vijayalakshmi, 2020, pp. 1-4).

Literature Review

1. Customer reviews on Online platform

In the digital era, customer reviews are essential feedback that influence purchasing decisions and help businesses improve product and service quality. Posted on online platforms, reviews provide insights into product features and usage, building consumer confidence. Positive reviews enhance seller credibility, leading to increased sales and business growth. They help turn casual readers into buyers by making purchase decisions easier. As a result, online businesses expand rapidly and gain wider brand recognition (Packhai, 2023).

Currently, customer reviews are also an important data source for businesses to improve their products or services and develop marketing strategies. Positive reviews build trust and drive sales, while negative reviews can provide feedback for improving the store's quality. Therefore, reviews serve as a critical tool in building relationships between businesses and customers, enhancing confidence, and increasing the likelihood of purchases from the store (TKPon, 2020). During the COVID-19 pandemic, the health supplement industry experienced significant growth and continues to be a strong trend. Manufacturers have expanded product lines to meet health and beauty demands, including products for physical health, vitamins supporting the nervous system and brain, and mental wellness products. A 2023 Mintel survey found that 86% of younger consumers read reviews before making a purchase. This study explores factors influencing Generation Y’s decisions to purchase dietary supplements. Key factors include marketing, price, attitudes, current usage, health interest, and consumption duration. The findings support strategic marketing tailored to this group’s behavior (Kanchanawiwin, 2023, pp. 1-94). This highlights that reading reviews is a critical step for younger consumers, enabling confident and informed decisions when purchasing health supplements.

Furthermore, customer reviews play a crucial role in building trust for products in competitive markets, especially in the health supplement industry. A high rate of positive reviews not only make a product stand out from competitors but also increase sales and build customer loyalty. (Park and Lee, 2009, pp. 61-67). Conversely, negative reviews have a significant impact on the product's image and sales. Products receiving many negative reviews for reasons such as not meeting advertised performance or unmentioned side effects may quickly lose popularity and credibility (Lu, Ba and Huan, 2013, pp. 596-612). Customer reviews also allow consumers to contribute to shaping the overall perception of products in the online market. This goes beyond merely helping others make decisions; it also affects the marketing strategies of manufacturers. (Dellarocas, Zhang and Awad, 2007, pp. 23-45).

Multiple studies, both in Thailand and internationally, have shown that customer reviews significantly influence purchasing decisions and trust in products. Research by Bharadwaj (2023, pp. 1-34) emphasized the importance of sentiment analysis in online reviews using natural language processing (NLP) and machine learning to analyze positive, negative, and neutral sentiments. Findings revealed that positive sentiments in reviews strongly boosted sales and created a favorable product image, while negative sentiments could quickly diminish a product's popularity. Research by Zhu and Liang (2024, pp. 3915–3928) further indicated that online reviews played a vital role in improving and developing future products. Companies could use insights from reviews of previous products to refine new product designs, ensuring alignment with consumer needs and expectations in competitive markets. In Thailand, research by Panyathorn and Thajang (2020, pp. 121–129) on the attitudes and consumption of health supplements among elderly consumers in Udon Thani province found that older adults had positive attitudes toward dietary supplements enabling them to have healthy and vital lives. However, their purchasing decisions relied more on family and advertisements than on online reviews. This findings contrasted with international research, where online reviews played a more dominant role in building trust and influencing purchasing decisions. Research by Thanam, et al. (2023, pp. 370–386) on consumer satisfaction with GOLD N health supplements in Maha Sarakham province showed high levels of satisfaction, particularly with the product's appearance and benefits, leading to repeat purchases. Comparing internal and external research revealed consistency in the importance of customer reviews in purchasing health supplements.

2. Customer Feeling Analysis on Online Platforms

Customer feeling analysis helps businesses gain detailed and accurate insights into customer emotions and feelings, such as whether customers are satisfied with the products and services or face any issues after using them. Studies show that over 95% of consumers read reviews from other users before deciding to purchase a product. This indicates that each review significantly impacts a brand's sales. To improve products and services for better feedback or reviews, analyzing shared consumer reviews provides straightforward opinions based on actual experiences (Nipa, 2024). Currently, so is building and sustaining customer relationships. Understanding customer feelings about products and services is particularly critical, as this directly determines whether customers are willing to pay, which significantly affects sales. The process involves using computer technology to analyze, process, and identify emotions expressed in text or audio. This can be gathered from surveys, reviews, social media posts, or voice recordings to determine whether the sentiment is positive, negative, or neutral (Pisinee, 2024).

Customer feeling analysis on online platforms involves analyzing sentiments from customer-posted texts on social media. These texts are processed using artificial intelligence (AI) and natural language processing (NLP) to recognize, learn, and interpret words or sentences. This allows businesses to understand whether the customer feedback on social media platforms reflects positive, negative, or neutral opinions. Brands can use sentiment analysis to evaluate customer satisfaction with products or services or track customer feedback to respond more effectively to consumer needs. Sentiment analysis is an essential tool for improving business efficiency in various aspects. Businesses that employ sentiment analysis gain a reliable tool to establish long-term customer relationships and increase brand loyalty (Wisesight, 2023). Sentiment analysis, a computer science technique, aims to understand the meaning of text data using machine learning and NLP. It identifies relationships between words and grammatical accuracy within sentences by analyzing extracted keywords and assigning sentiment scores. This process involves building sentiment analysis models capable of accurately predicting sentiments in unknown data (Mao, Liu and Zhang, 2024)

Numerous studies related to customer sentiment and behavior analysis in Thailand and abroad revealed various techniques to improve the accuracy of sentiment and behavior evaluations. Research by Wen, Liang and Zhu (2023, pp. 1-14) The emotion analysis of hotel reviews using the BERT model helps platforms understand customer needs and recommend suitable hotels. By fine-tuning the BERT model, high classification accuracy is achieved, with the output classified using the softmax function. The ERNIE model, an enhancement of BERT, performs better with stronger accuracy and stability. Research in Thailand by Sirikul and Samphanwattanachai (2023, pp. 53–66) analyzed factors influencing dietary supplement purchase decisions using quantitative methods, finding that the 4P's and customer attitudes played significant roles, though geographical constraints limited the scope. Chen, Hongthong and Suthamdee (2024, pp. 214-229) analyzed linguistic strategies in dietary supplement advertisements, showing that rational and emotional appeals were effective for persuasion, though the study did not encompass all types of advertisements. In contrast employed CNN to detect sentiments in tweets about dietary supplements, revealing predominantly negative sentiments about mental health, though the tweets did not cover all demographic groups. Research by Chanakot and Phoksawat (2024, pp. 193-199) used CFG, TF-IDF, and Deep Learning to analyze sentiments in dietary supplement reviews on Shopee, finding that combining these techniques improved accuracy despite some limitations with specific review types. Arhatia, et al. (2024) analyzed smartphone reviews using VADER and roBERTa, achieving high accuracy in sentiment analysis across multiple languages and understanding implicit meanings, though challenges remained with complex language and multilingual reviews. (Bomma, 2024, pp. 1-5) applied NLP in business intelligence (BI) systems with text mining, NER, and Topic Modeling to extract information from unstructured text, demonstrating that NLP effectively analyzed customer feedback but faced limitations with incomplete BI data. In clustering analysis, Lkotun et al. (2023, pp. 178–210) found that K-means remains popular for its simplicity but has limitations in choosing initial cluster centroids and detecting overlapping clusters. (Thakur, Verma and Tiwari, 2024, pp. 1-9)

Research Methodology

Based on the study of theories and literature related to sentiment analysis from customer review behaviors towards health supplement products in the online market using natural language processing. This conceptual framework employs sentiment analysis theory (Anderson, Sarkar and Kelley, 2024, p. 1-11; Stammback, et al., 2022, pp. 1-14; Yadav and Yadav, 2020, pp. 4335-4385) and clustering theory (Rodriguez, et al., 2019, pp. 1-34) to understand customer sentiments and group similar reviews together.

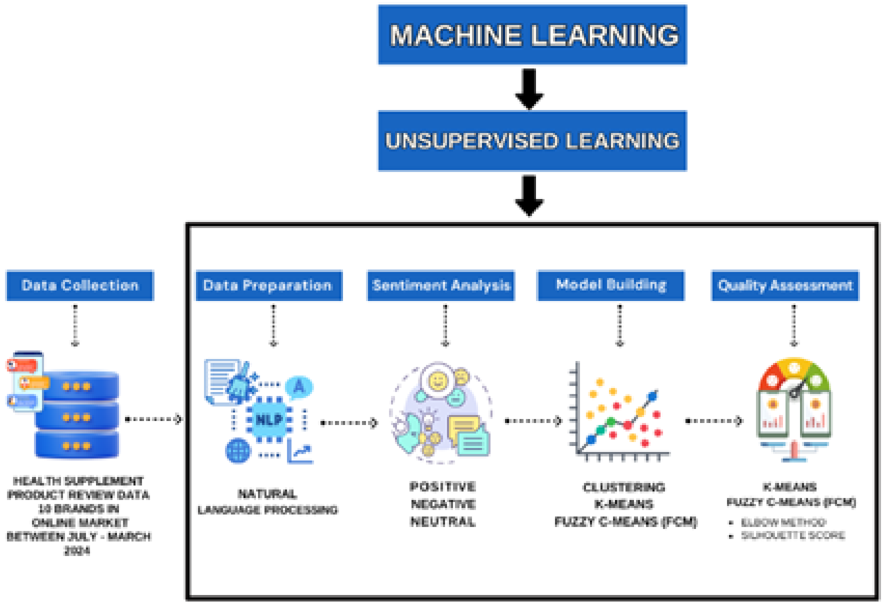

Figure 1 Conceptual Framework of Sentiment Analysis of Customer Review Behavior

This research utilizes natural language processing (NLP) and unsupervised machine learning to analyze customer reviews of health supplement products (Gupta and Jamwal, 2021, pp. 147–156). NLP facilitates the differentiation and analysis of review texts, providing in-depth insights into customer opinions. For unsupervised machine learning, clustering and dimensionality reduction techniques are employed to uncover hidden relationships and structures in the data.K-Means and Fuzzy C-Means clustering methods. This approach helps reveal an overarching view and key factors influencing customer opinions comprehensively and clearly, enabling better interpretation of review data (Thakur, Verma and Tiwari, 2024, pp. 1-9 ; Salman and Alomary, 2024, pp. 465-470). The methodology includes the following steps:

1. Data Collection

This research involved collecting and aggregating data regarding customer sentiment from product review behaviors towards health supplement products with the highest sales on the Shopee platform, which is the most popular online shopping platform in Thailand (Naiyanapakorn, 2024, p.1), as illustrated in Figure 1. The period for collecting customer reviews of health supplement products spanned from July 2023 to March 2024, during which online market growth was consistently observed. The dataset comprises a total of 7,917 publicly available reviews written in Thai, focusing on the 10 most popular health supplement brands during the specified period. The specific data details are not disclosed in this study. All data collected are in textual form.

This research emphasized the application of libraries for web scraping using Python. The obtained data were analyzed to determine positive, negative, and neutral sentiments by employing Natural Language Processing (NLP) technology, combined with unsupervised learning techniques, to transform textual data into structured, actionable data (Pichiyan, et al., 2023, pp. 193-202).

2. Data Preprocessing

Data preprocessing is a crucial step in analyzing customer review sentiments toward health supplement products in the online market. In particular, natural language processing (NLP) is employed in this process to ensure that models or deep learning techniques work accurately and efficiently. The key steps are as follows (Kulkarni and Shivananda, 2019, pp. 139-151):

2.1 Data Cleaning: This involves inspecting and removing missing or incomplete data (Missing Values) from the dataset. It includes eliminating empty text, removing unnecessary special characters, and other irrelevant symbols.

2.2 Tokenization: This step breaks down the text into individual tokens or words using appropriate tools. For example, the PyThaiNLP library in Python is used to split Thai text into words, which serve as the fundamental units for further text processing. For example:

Original text : product this really good help improve health effectively really

Tokenized words : ["ผลิตภัณฑ์", "นี้", "ดีมาก", "ช่วย", "บำรุง", "สุขภาพ", "ได้", "ดี", "จริงๆ"]

2.3 Stopword Removal: Stopwords are common words that carry little semantic meaning, such as polite terms, conjunctions, or frequently used words like "and," "also," "that," etc. Removing stopwords reduces the data size and prepares it for more efficient analysis.

2.4 Additional Text Cleaning: This involves removing numbers, converting text to all lowercase or uppercase letters, eliminating extra spaces, and refining incomplete text segments.

2.5 Sentence Segmentation: This step separates sentences from one another to enable independent sentiment analysis for each sentence.

In addition, duplicate review entries were identified and removed from the dataset to prevent potential bias in the analysis. As a result of these preprocessing steps, the final dataset used for analysis was deemed appropriate and of sufficient quality for further processing using machine learning techniques and sentiment analysis.

3. Text Vectorization

Text vectorization involves converting sequences of words into numerical vectors that can be processed further. Techniques include TF-IDF, which counts word frequency in documents and reduces the weight of commonly occurring words, and Word Embeddings, It is one of the techniques used in text analytics, the TF-IDF feature extraction, which is easy to implement, efficient in computation, and focuses on words that carry more information in the reviews (Lukplamino, 2022). TF-IDF (Term Frequency-Inverse Document Frequency) weighting is a popular technique for measuring the importance of words within a document by comparing the word frequency in a specific document with its frequency across the entire dataset. In sentiment analysis using Machine Learning, the TF-IDF scores of individual words are combined into a feature vector, which can then be used as input data for subsequent machine learning algorithms (Ahmed, et al., 2023, pp. 52-57). If a word is mentioned frequently in reviews, it is likely to be closely related to the key aspects of the sentiment expressed in those reviews. For example, terms such as "Order," "Product," and "Efficacy" may be highly relevant to the overall sentiment of the feedback. TF-IDF is a statistical technique used to evaluate the significance of a word within a document relative to a larger corpus. It integrates term frequency (TF), which measures how frequently a word appears in a specific document, with inverse document frequency (IDF), which quantifies the rarity of the word across all documents. The TF-IDF score is computed as TF × IDF. This technique is based on the bag-of-words model, which considers word frequency while disregarding word order (Tamanna, 2023).

4. Sentiment Analysis

The sentiment analysis process examines customer reviews of health supplement products on the Shopee website. This is achieved through natural language processing (NLP) and unsupervised machine learning, enabling the recognition, learning, and interpretation of words or sentences within text (Kaburuana, Sari and Agustina, 2022, pp. 150-159). This process identifies the sentiment expressed in customer reviews posted on social media, categorizing them into positive, negative, or neutral sentiments. Sentiment analysis helps identify patterns and insights from review data, including understanding customer sentiments regarding products (Broklyn, Olukemi and Bell, 2024 pp. 1-20). It involves coding the data to classify customer reviews into categories such as satisfaction, encountered problems, and recommendations. For example, classifying feedback about health supplement products might include suggestions, potential product issues, and customer critiques. This classification facilitates in-depth analysis by segmenting reviews based on sentiment trends, identified problems, and qualitative analysis. The qualitative analysis interprets data to uncover trends and improvement strategies for products or services. It is also used to identify patterns and groupings in unlabeled data, revealing hidden sentiments within customer feedback (Bowornlertsutee and Paireekreng, 2022, pp. 71-79).

5. Clustering

Clustering is a technique for organizing data into subgroups (clusters) where data within the same group are highly similar, while data in different groups are clearly distinct. This method does not use predefined labels to determine which data belongs to which cluster. The goal of clustering is to uncover hidden patterns in data, enabling a deeper understanding of its structure and relationships.

The most suitable clustering methods include K-Means Clustering and Fuzzy C-Means (Salman and Alomary, 2024, pp. 465-470).

1. Cluster Selection: The Elbow Method and Silhouette Score are used to determine the optimal number of clusters.

2. Sentiment Clustering: Groups sentiments based on the words used in reviews, such as positive, negative, or neutral clusters. For instance, a positive cluster may highlight good performance, while a negative cluster may point out undesirable side effects.

K-Means clustering is an unsupervised learning algorithm that partitions data into K clusters, each represented by a centroid. The algorithm assigns data points to the nearest centroid and iteratively updates the centroids until convergence. Although it is efficient and scalable, K-Means is sensitive to the initial placement of centroids and requires the number of clusters (K) to be predefined (Lkotun, et al., 2023, pp. 178–210). In contrast, Fuzzy C-Means (FCM) allows data points to belong to multiple clusters with varying degrees of membership. Based on fuzzy logic, FCM calculates membership values using the inverse distance to cluster centers, providing a more flexible representation of complex and overlapping data structures (Salman and Alomary, 2024, pp. 465-470).

6. Data Analysis

Data analysis of customer reviews is a crucial step that helps businesses understand customer satisfaction and dissatisfaction with health supplement products. This essential information can be utilized to effectively improve and develop products. The analysis employs natural language processing (NLP) technology to extract in-depth insights from customer reviews. Sentiment analysis of customer reviews regarding health supplement products classifies opinions into positive, negative, or neutral categories (Wharton, 2024). The analysis utilizes review data from 10 health supplement brands, which undergo preprocessing steps such as removing stop words and special characters, as well as tokenization to separate meaningful words or phrases. Sentiment analysis is conducted using unsupervised learning models to uncover patterns and clusters in unlabeled data. In this process, unsupervised learning models employ clustering techniques or dimensionality reduction to identify hidden sentiments within the data. To determine sentiment trends, reviews with sentiment scores greater than 0.6 are counted. This threshold is used to distinguish opinions with greater clarity or weight. Sentiment scores range from 0 to 1, where values closer to 1 indicate clearer sentiments (Mohamed, 2023). Setting the threshold at 0.6 highlights opinions with significant clarity compared to scores closer to 0.5, which are neutral. A threshold of 0.6 helps separate opinions expressing clear emotions from those with uncertainty or neutrality, reducing ambiguity. Reviews with sentiment scores between 0.4 and 0.6 may not reflect clear sentiments, potentially leading to interpretive confusion. For example, if a model assigns a positive sentiment probability of 0.75 and the threshold is set at 0.6, the review is classified as "positive." However, if the score is 0.55, it is categorized as "negative." Setting the threshold at 0.6 filters out uncertain opinions, increasing analysis accuracy (Shah, 2024). This threshold is a standard approach in sentiment analysis to clearly and straightforwardly represent the intensity of sentiments. Using a consistent scoring range simplifies the comparison and analysis of sentiments across different categories. For instance, if positive sentiment scores are close to 0.8 and negative sentiment scores are near 0.2, it indicates a clear overall positive sentiment trend. The range of 0 to 1 is a standard approach in classification within machine learning (Nikhil, 2024).

Results

1. Results of Sentiment Analysis from Customer Reviews on Health Supplement Products in the Online Market Using Natural Language Processing

The collection of customer reviews on health supplement products in the online market yielded a total of 7,917 reviews. After data cleaning, which included removing empty reviews, tokenization, and eliminating reviews unsuitable for analysis, 6,284 reviews remained for sentiment analysis. These reviews were gathered from 10 health supplement brands. To protect the privacy of the brands and maintain neutrality in data analysis, the researcher decided not to disclose the product brand names during the analysis process. Instead, the brands were anonymized using letters and numbers, with brands numbered 1 through 10 represented as "Brand A" to "Brand J," respectively. This anonymization ensures the analysis remains unbiased and focuses solely on customer sentiment data rather than the brand names. As a result, the findings of this research can be broadly applied without bias toward specific product brands.

Table 1 displays the anonymized brand names and the number of reviews for the health supplement products.

Product Brand	Number of reviews
Brand A	914
Brand B	900
Brand C	875
Brand D	696
Brand E	657
Brand F	638
Brand G	597
Brand H	525
Brand I	364
Brand J	118
SUM	**6,284**

From Table 1, the anonymized names of the 10 health supplement product brands and the 6,284 reviews are shown, ranked in order from the brand with the highest number of reviews to the brand with the lowest number of reviews.

Table 2 shows the star ratings from the reviews of the health supplement products.

Star rating	Number of stars	Percentage
5 stars	5,868	93.39
4 stars	351	5.59
3 stars	37	0.59
2 stars	13	0.20
1 stars	15	0.23
Total	**6,284**	**100**

From Table 2, the star ratings from customer reviews of health supplement products are shown without being categorized by product brands. The most frequent star ratings given by customers are, in descending order: 5 stars, 4 stars, 3 stars, 1 star, and 2 stars.

The analysis of customer sentiments toward health supplement products in the online market used a Word Cloud to highlight key terms frequently mentioned in reviews. This visualization helps to understand what aspects customers prioritize or feel strongly about regarding the products. For instance, if words like "efficacy" or "taste" appear prominently in the Word Cloud, it indicates that customers place significant importance on these aspects, as shown in Figure 2.

Figure 2 Word Cloud from Customer Sentiment Reviews on Health Supplement Products

From Figure 2, the prominently displayed keywords represent the topics customers value most. When these keywords are cross-referenced with Figure 3, it is evident that they consistently appear across all products. The Figure keyword the key terms by product brand, clearly showing which issues are most emphasized by customers for each brand, aligning with the insights observed in the Word Cloud.

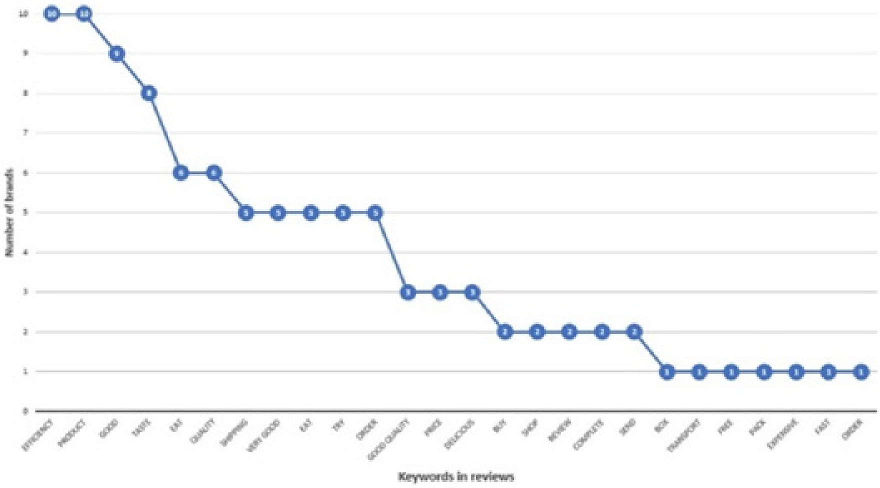

Figure 3 Graph Showing Keyword Frequencies in Reviews (Overall)

From Figure 3, the graph highlights the keywords that frequently appear in reviews of health supplement products across all brands. These keywords reflect customer attributes or expectations for health supplement products in the online market, as follows:

1) The words "efficacy" and "product" are the most frequently mentioned across all brand reviews, indicating that customers place significant importance on the outcomes of using the products and often directly reference the product itself.

2) The word "good" appears in reviews for 9 brands, reflecting overall customer satisfaction with health supplement products, indicating that most customers have a positive perception of these products.

3) The word "taste" appears in reviews for 8 brands, demonstrating that customers care about the flavor of health supplements. This indicates that customers desire health supplements that are not only effective but also palatable.

4) The words "consume" and "quality" appear in reviews for 6 brands, highlighting customer focus on ease of consumption. The term "quality" in these reviews reflects customer satisfaction concerning the efficacy and outcomes of the products.

5) Keywords associated with satisfaction, such as "good quality," "price," and "delicious," collectively reflect key factors customers consider when evaluating health supplement products, suggesting that customers want products that are both effective and enjoyable.

6) Keywords such as "purchase," "store," "review," "neat," "delivery," "box," "shipping," "free gift," "pack," "expensive," "fast," and "order and consume" reflect key elements of the online shopping process.

Table 3 presents dimensions of sentiment expressed in reviews of health supplement product.

Reviews	Rating Star	Positive	Neutral	Negative
Good quality, good performance, fast delivery. This flavor is delicious. At first, I thought the texture would be a bit floury, but it wasn't that bad. Next time, I'll try buying another flavor to try.	5	0.7519713640213010	0.192119772026060	0.05590891093015670
Good quality, very effective, very good taste, ordered the second jar, meow, feel the muscles increase, effective, gradual, like that it is plant protein, love the environment.	5	0.9702308177948000	0.021790552884340300	0.007978667505085470

From Table 3, the dimensions of sentiment in reviews of health supplement products, assessed through Sentiment Analysis, are divided into three categories: Positive (positive sentiment), Neutral (neutral sentiment), and Negative (negative sentiment).

1) Comment: Reflects customer feedback on the product, such as opinions about taste, quality, service, or experiences related to using the product.

2) Rating star: Indicates the star rating given by customers for the product, ranging from 5 stars,4 stars,3 stars,2 stars,to 1 star in descending order.

3) Positive: Represents the positive sentiment score calculated from customer feedback. This score ranges from 0 to 1, with values closer to 1 indicating a higher positive sentiment.

4) Neutral: Represents the neutral sentiment score (neither positive nor negative), also ranging from 0 to 1.

5) Negative: Represents the negative sentiment score calculated from feedback, ranging from 0 to 1, with values closer to 1 indicating a stronger negative sentiment.

Each row in the table represents a different customer review, encompassing a variety of opinions, such as feedback on product taste, service, delivery, etc., and the majority of ratings are 5 stars.

• The Positive sentiment scores in all rows exceed 0.6, indicating a high level of positive sentiment from customers toward the health supplement products.

• The Neutral sentiment scores and Negative sentiment scores are below 0.2 in all rows, showing that neutral and negative sentiments are minimal compared to positive sentiments.

In summary, the overall results in the table demonstrate that most customers providing reviews are satisfied with the products, as indicated by high star ratings and significantly higher positive sentiment scores compared to negative sentiment scores.

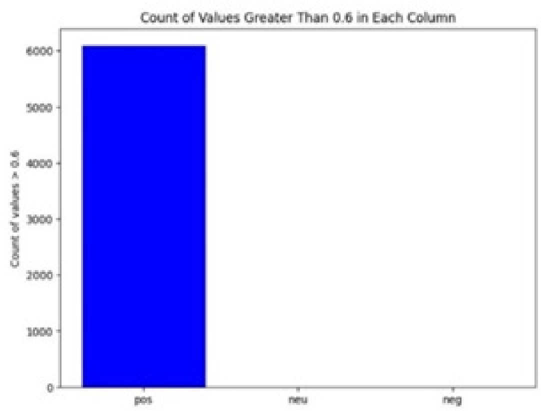

Figure 4 Graph Showing Customer Sentiments Toward Health Supplement Products

From Figure 4, the graph displays the number of sentiments with scores greater than 0.6, categorized into Positive (Positive), Neutral (Neutral), and Negative (Negative). The Y-axis (vertical) represents the number of sentiments with scores above 0.6, while the X-axis (horizontal) shows the three sentiment types: "pos" (Positive), "neu" (Neutral), and "neg" (Negative).

The analysis reveals that the "pos" (Positive) column contains 6,284 sentiments with positive sentiment scores greater than 0.6, indicating a clear dominance of positive sentiments. In contrast, the "neu" (Neutral) and "neg" (Negative) columns have no bars, indicating that there are no sentiments with neutral or negative scores exceeding 0.6.

This graph highlights that the majority of sentiments in the database are strongly positive, with minimal presence of neutral or negative sentiments. The graph is specifically designed to display only sentiments with scores exceeding 0.6, reinforcing the clear trend toward positive sentiment.

2. Result of Developing a Sentiment Analysis Model for Customer Reviews of Health Supplement Products in the Online Market Using Natural Language Processing

Using natural language processing in conjunction with unsupervised learning models helps uncover hidden patterns and relationships in review data without the need for predefined labels. This approach simplifies the management of large datasets while grouping reviews (clustering) into similar categories.

1) Finding the Optimal Number of Clusters for All 10 Products (PLOTTING OPTIMAL CLUSTER NUMBERS):

This involves determining the most suitable number of clusters for grouping all 10 products using statistical methods or machine learning techniques. Clustering analysis techniques such as K-Means Clustering and Fuzzy C-Means were employed (Lkotun, et al., 2023, pp. 178–210; Thakur, Verma and Tiwari, 2024, pp. 1-9).

• Finding K for K-Means Using Silhouette Score:

Silhouette Score is used to evaluate the quality of clustering and determine the optimal number of clusters (K) for analysis using K-Means Clustering. A higher Silhouette Score (approaching 1) on the Y-axis indicates better clustering, and the corresponding X-axis value is selected as the optimal K.

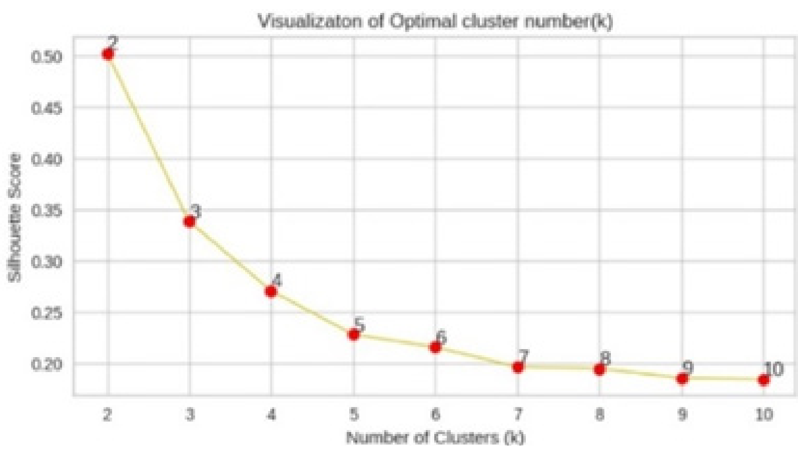

Figure 5 Graph for Determining the Optimal Number of Clusters Using K-Means

From Figure 5, the graph identifies the optimal number of clusters for grouping data using K-Means Clustering and evaluates clustering quality with the Silhouette Score. The X-axis (horizontal) represents the number of clusters (K) ranging from K=2 to K=10, while the Y-axis (vertical) represents the Silhouette Score, with a range from -1 to 1. Higher scores indicate better clustering quality. A yellow line connects red points representing Silhouette Scores for different K values. The results from the graph show that K=2 yields the highest Silhouette Score of approximately 0.50, indicating that dividing the data into 2 clusters provides the best clustering quality compared to other cluster numbers. This reflects optimal grouping, where data within each cluster are more closely related than with data in other clusters.

• **Finding Optimal Clusters for FCM (Fuzzy C-Means):**

Fuzzy C-Means is a clustering technique within unsupervised learning. Unlike K-Means, FCM does not require data points to belong exclusively to a single cluster. Instead, it allows flexibility by assigning membership levels, where a data point can belong to multiple clusters with varying degrees of membership.

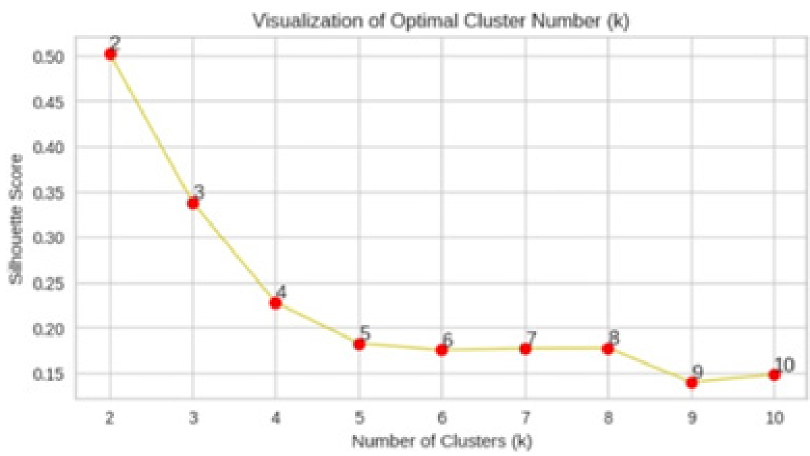

Figure 6 Graph for Determining the Optimal Number of Clusters Using Fuzzy C-Means

From Figure 6, the analysis of data clustering using Fuzzy C-Means with Silhouette Score as an evaluation metric shows that when testing cluster numbers from 2 to 10, the Silhouette Score, representing clustering quality, is displayed as a graph. The X-axis represents the number of clusters (K), and the Y-axis represents the Silhouette Score. A yellow line connects the red points, which indicate the Silhouette Score for each number of clusters. The analysis results show that dividing the data into 2 clusters (K=2) yields the highest Silhouette Score of approximately 0.50. This indicates that grouping the data into 2 clusters is the most appropriate method, as it provides the highest clustering quality compared to other tested numbers of clusters.

• **Determining K Using the Elbow Method:** The Elbow Method is a technique used to select the optimal number of clusters (K) for clustering

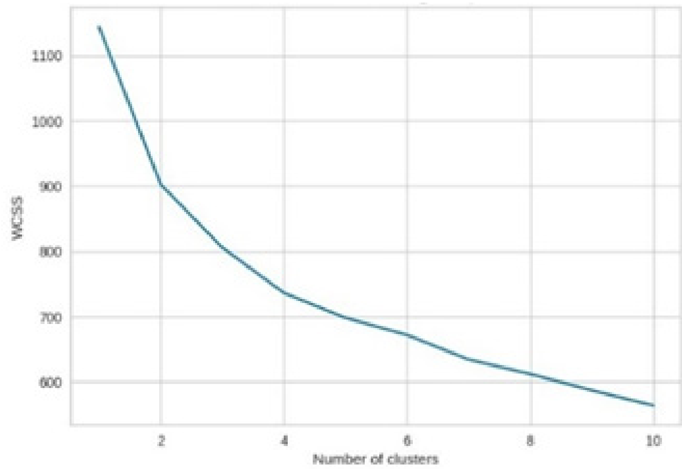

Figure 7 Graph for Determining the Optimal Number of Clusters Using the Elbow Method

From Figure 7, the analysis of selecting the optimal number of clusters for K-Means clustering uses the Within Cluster Sum of Squares (WCSS) (Sulianta, Ulfah and Amalia, 2024, pp. 34-53) as a metric to evaluate clustering quality. The goal is to find the optimal number of clusters by minimizing the WCSS value. The Elbow Method examines the decline in WCSS as the number of clusters (K) increases. The point where WCSS declines significantly and begins to stabilize is often selected as the optimal number of clusters (Elbow Point). The analysis shows that when dividing the data into clusters from K=1 to K=10, the WCSS decreases significantly as K increases, especially from K=1 to K=2. This indicates that the data can be better grouped as K increases. However, the rate of decline becomes less significant after K=2. The point where the graph forms an "elbow" shape is at K=2, indicating that this is the optimal number of clusters, as the WCSS changes less significantly after this point.

Thus, from the analysis using the Elbow Method and confirmation from Silhouette Score in both K-Means and FCM, it is found that K=2 is the most suitable number of clusters for creating the model to group customers.

2) Cluster Profiling:

This is the process of analyzing and interpreting the results of data clustering, with the primary goal of understanding the key characteristics or patterns within each cluster. The analysis profiles customer groups identified by the K-Means model. The first graph represents customer group 1, and the second graph represents customer group 2, analyzing all 10 brands collectively.

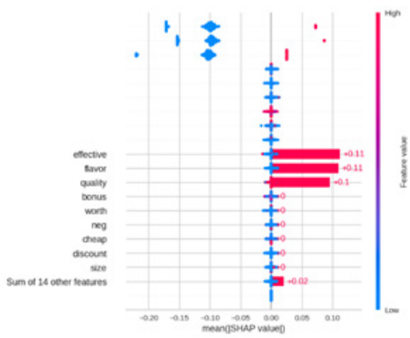

Figure 8 Customer Group 1

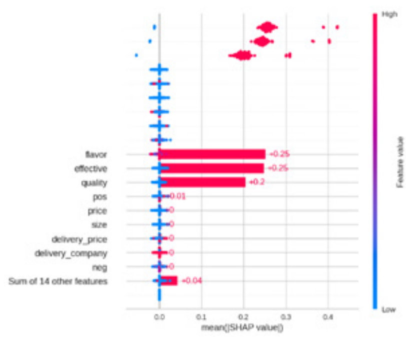

Figure 9 Customer Group 2

From Figure 8 and Figure 9, customers are divided into 2 groups using the K-Means method. The analysis reveals that the 2 customer groups, derived from health supplement products across 10 brands, both prioritize efficacy, product, good, taste, consume, and quality. However, the difference between the two groups lies in the weight given to other factors influencing purchase decisions. Customer group 2 places a weight of 0.04 on other features or factors in their decision-making process, whereas customer group 1 places a weight of 0.02 on these factors. The difference is minimal and does not significantly impact purchasing behavior.

The evaluation of clustering quality from the K-Means and Fuzzy C-Means models uses metrics such as Silhouette Score and Davies-Bouldin Index. The Silhouette Score ranges from -1 to 1, with values closer to 1 indicating high similarity within clusters and clear distinction from other clusters. The Davies-Bouldin Index measures clustering quality, where lower values indicate better clustering performance. The comparison is as follows:

Table 4 Comparison of Clustering Quality Between K-Means and Fuzzy C-Means

Model	Silhouette Score	Davies-Bouldin Index
K-means	0.531932321846374	0.6990938304509099
FCM Model	0.531932321846374	0.6990938304509099

From Table 4, the comparison of clustering quality evaluations between the K-Means and Fuzzy C-Means models shows that both models have the same Silhouette Score of 0.5319, indicating a comparable level of clarity and density within the clusters. A score closer to 1 indicates higher clustering efficiency. Similarly, the Davies-Bouldin Index for both models is equal at 0.6991, which represents the average similarity ratio of each cluster to the most similar cluster. A lower Davies-Bouldin Index signifies better clustering performance. In conclusion, the clustering quality of the K-Means and FCM models shows no significant difference for this dataset. It can be clearly stated that both models are equally capable of clustering data effectively.

Discussion and Conclusion

In an era where online information plays a significant role in consumer purchasing decisions, sentiment analysis from customer review behaviors toward health supplement products in the online market using natural language processing (NLP) is essential. This is especially true for the growing health supplement market. From analyzing 6,284 reviews from 10 brands (anonymized as "Brand A" to "Brand J" for neutrality and to avoid bias), the study provided insights into factors influencing customer sentiment and purchasing behavior. Key terms identified through Word Cloud analysis include "efficacy," "product," "good," "taste," "consume," "quality," "delivery," "price," and "delicious." These terms reflect customer expectations for products, particularly regarding efficacy, taste, quality, and value for money—factors crucial in purchasing decisions. Sentiment analysis revealed that most reviews expressed positive sentiments, with positive sentiment scores exceeding 0.6 across all reviews, while neutral and negative sentiments were minimal. This indicates that the majority of customers reviewing health supplements were satisfied with the products.

This research supports and aligns with related studies on sentiment analysis of customer behaviors toward health supplement products in the online market. Specifically, it highlights the use of NLP techniques, which are widely recognized in consumer behavior studies (Gupta and Patel, 2021, pp. 511-517). The study emphasizes the importance of data cleaning and anonymizing brand names to ensure unbiased analysis by replacing brand names with letters and numbers. The use of Word Cloud to display frequently mentioned keywords provides insights into key customer sentiments, such as "efficacy" and "taste."

These findings align with studies showing that customers prioritize product quality and efficacy as key factors in purchasing decisions (Rewnark, 2023, p. 65). Additionally, categorizing sentiments into positive, negative, and neutral aligns with sentiment analysis methodologies used in various studies aimed at clearly grouping and interpreting customer sentiments (Thakur, Verma and Tiwari, 2024, pp. 1-9). The sentiment analysis of health supplement product reviews reveals word usage that reflects satisfaction or dissatisfaction, similar to findings in (Zaghloul, Barakat and Rezk, 2024, p. 79), which studied customer satisfaction predictions in e-commerce. These studies highlight the efficiency and precision of handling complex data, consistent with trends in leveraging advanced technology for consumer behavior analysis. By incorporating these insights, businesses can enhance their understanding of consumer preferences, leading to more effective strategies for product improvement and innovation.

Although the statistical results did not reveal a significant difference between the models, a comparative analysis of their characteristics suggests that Fuzzy C-Means provides greater flexibility in handling overlapping opinions, whereas K-Means is more suitable for clear-cut segmentation, aligning well with practical marketing applications. Moreover, while both customer groups emphasized similar core terms such as “efficacy,” “product,” and “quality” a deeper analysis identified subtle differences in the weighting of secondary factors like “taste” and “consume.” These distinctions reflect nuanced consumer preferences and can inform more precise product development and communication strategies tailored to each segment. The application of both the Elbow Method and Silhouette Score to determine the optimal number of clusters reinforced the reliability of the findings, with both methods consistently indicating that K=2 yields the best clustering structure. The analysis results indicate that the majority of customer reviews tend to be positive. Nevertheless, negative feedback despite appearing in a smaller proportion constitutes a valuable source of information that should not be overlooked. Such feedback often highlights critical issues or limitations of the product or service, including dissatisfaction with taste, lack of perceived effectiveness, or substandard delivery experiences. These factors can significantly influence overall customer satisfaction. Therefore, conducting a systematic and in-depth analysis of negative reviews has the potential to yield insightful information that can inform targeted improvements and product or service development aligned with customer needs. Furthermore, such analysis may contribute to reducing the recurrence of similar complaints in the future and enhancing the organization's competitiveness in a rapidly evolving online marketplace.

Future research may consider adopting more sophisticated clustering techniques or integrating supplementary data sources such as metadata or demographic attributes to improve segmentation accuracy and support more strategic applications in the context of an increasingly dynamic online market.

Recommendations

The research findings can guide businesses in improving competitiveness and operations, leading to the development of products and marketing strategies that better meet customer needs. The recommendations are as follows 1) Customer Segmentation : Applying K-Means and Fuzzy C-Means clustering to group customers by similar behaviors or characteristics. 2) Customer Behavior Analysis : The clustering results can be used to analyze customer behavior within each group, particularly identifying trends and preferences. 3) Marketing Strategy Development : Data from clustering can be used to design marketing strategies tailored to high-potential customer groups. 4) Product/Service Improvement : Through data clustering, businesses can identify the specific needs of different customer groups and refine their products to better meet these needs. 5) Forecasting Future Trends : Clustering analysis helps businesses identify future trends in customer behavior or market responses in the online space. 6) Strategic Decision Making : Sentiment analysis supports decision-making in business planning.

Sentiment analysis of customer review behaviors toward health supplement products in the online market using NLP enhances the efficiency of text-based data processing. Computers can interpret complex text, understand the intent and sentiment of messages, and categorize words and phrases effectively.

Acknowledgment

The authors would like to express their sincere gratitude to Thepsatri Rajabhat University, the Faculty of Management Science and particularly the Language Center for their invaluable support in providing the necessary resources, knowledge, and data required for the successful completion of this research.

Bibliography

Ahmed, W., Semary, N. A. E.-M., Amin, K. M. and Hammad, M. (2023). Sentiment analysis on Twitter using machine learning techniques and TF-IDF feature extraction: A comparative study. **International Journal of Computers and Information,** 10(3), 52-57.

Anderson, T., Sarkar, S. and Kelley, R. (2024). Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques. **Natural Language Processing Journal,** 8(100097), 1-11.

Bharadwaj, L. (2023). Sentiment analysis in online product reviews: Mining customer opinions for sentiment classification. **International Journal for Multidisciplinary Research,** 5(5), 1-34.

Bomma, H. P. (2024). Natural language processing (NLP) in business intelligence. **International Journal of Leading Research Publication,** 5(9), 1-5

Bowornlertsutee, P. and Paireekreng, W. (2022). The model of sentiment analysis for classifying the online shopping reviews. **Journal of Engineering and Digital Technology (JEDT),** 10(1), 71-79.

Broklyn, P., Olukemi, A. and Bell, C. (2024). Social media sentiment analysis for brand reputation management. **SSRN,** 1-20.

Boonchote, T. (2022). **Business intelligence insight.** Retrieved April 29, 2024, from https://cathcarttechnology.co.th/insights/business-intelligence-insight/

Chanakot, B. and Phoksawat, K. (2024). Analysis of consumer sentiments towards online shopping using context-free grammar and deep learning. **Revue d'Intelligence Artificielle,** 38(1), 193-199.

Chen, Z., Hongthong, T. and Suthamdee, A. (2024). Strategies of language use and persuasion in supplementary product for beautiful skin advertisement. **Journal of MCU Philosophy Review,** 7(1), 214–229.

Dellarocas, C., Zhang, X. and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. **Journal of Interactive marketing,** 21(4), 23-45.

GlobalWellnessInstitute. (2024). **Understanding GWI's mental wellness economy data. Global Wellness Institute.** Retrieved June 18, 2024, from https://globalwellnessinstitute.org/global-wellness-institute-blog/2024/06/25/understanding-gwis-mental-wellness-economy-data-a-market-that-will-grow-from-181-billion-in-2022-to-330-billion-by-2027/

Gupta, H. and Patel, M. (2021). Method of text summarization using LSA and sentence-based topic modelling with BERT. In **The International Conference on Artificial Intelligence and Smart Systems (ICAIS)** (pp. 511–517). Coimbatore, India.

Gupta, P. and Jamwal, S. S. (2021). **Designing and development of stemmer of Dogri using unsupervised learning.** In N. Marriwala, C. C. Tripathi, S. Jain, & S. Mathapathi (Eds.), Soft computing for intelligent systems (pp. 147–156). Springer.

Kaburuana, E. R., Sari, Y. S. and Agustina, I. (2022). Sentiment analysis on product reviews from Shopee marketplace using the Naïve Bayes classifier. **Lontar Komputer: Jurnal Ilmiah Teknologi Informasi,** 13(3), 150-159.

Kanchanawiwin, A. (2023). **Factors affecting consumer's purchasing decision of dietary supplement: An empirical study on Generation Y.** Master's Independent study, Thammasat University, Bangkok.

Kulkarni, A. and Shivananda, A. (2019). **Natural language processing recipes.** India: Apress.

Lkotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B. and Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. **Information Sciences,** 622, 178–210.

Lukplamino. (2022). **What is TF-IDF? Explained with the simplest example Medium.** Retrieved June 18, 2024, from https://shorturl.asia/Lbxp6/

Lu, X., Ba, S., Huang, L. and Feng, Y. (2013). Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. **Information Systems Research,** 24(3), 596-612.

Mao, Y., Liu, Q. and Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. **Journal of King Saud University - Computer and Information Sciences,** 36(4), 1-16.

Meeprasert, V. and Rattakan, E. (2021). An analysis of customer opinions on Shopee Thailand through Twitter. **Journal of Information Systems in Business (JISB),** 7(3), 6–18.

Mohamed, A. (2023). **Sentiment score: A comprehensive guide. AIM Technologies.** Retrieved April 7, 2024, from https://www.aimtechnologies.co/2023/09/

Nikhil.(2024). **Guide to sentiment analysis using natural language processing.** Retrieved April 26 2024, from https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment-analysis/

Nipa. (2024). **Sentiment analysis: A marketing tool every business should know.** Retrieved June 19, 2024, from https://nipa.co.th/th/article/digital-marketing/tracking-customer-journey-with-sentiment-analysis

Naiyanapakorn, P. (2023). **Understanding online consumer behavior. Trade policy and strategy office.** Retrieved April 24, 2024, from https://tpso.go.th/news/2402-0000000019

NoonInch. (2022). **According to recent statistics on online product reviews for marketing purposes, 63% of consumers check Google Reviews before making a purchase decision. Everyday Marketing.** Retrieved June 19, 2024, from https://www.everydaymarketing.co/trend-insight/stat-รีวิวสินค้า-online-for-marketing-plan/

Panyathorn, K. and Thajang, S. (2020). Opinions and consumption behaviors of supplementary food products among the elderly in Nongbu village, Sam Phrao Sub-district, Udon Thani Province. **UdonThani Hospital Medical Journal,** 28(2), 121–129.

Park, C. and Lee, T. M. (2009). Information direction, website reputation and eWOM effect: A moderating role of product type. **Journal of Business research,** 62(1), 61-67.

Packhai. (2023). **8 product review techniques for online stores to attract customers and boost purchase decisions.** Retrieved April 18, 2024, from https://packhai.com/how-to-do-product-review/

Pichiyan, V., Muthulingam, S., Sathar, G., Nalajala, S., Akhil, C. and Das, M. N. (2023). Web scraping using natural language processing: Exploiting unstructured text for data extraction and analysis. **Procedia Computer Science,** 230, 193–202.

Pisinee. (2024). **Sentiment analysis: When customer emotions can be understood through text.** Retrieved October 26, 2024, from https://contentshifu.com/blog/what-is-senti ment-analysis

Rewnark, K. (2023). **Factors affecting the decision making on purchasing dietary supplements through mobile applications.** Master's independent study, Thammasat University. Bangkok.

Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F. and Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. **PLOS ONE,** 14(1), 1-34.

Salman, Z. and Alomary, A. (2024). Performance of the K-means and fuzzy C-means algorithms in big data analytics. **International Journal of Information Technology,** 16(1), 465-470.

Shah, R. (2024). **Different methods for calculating sentiment of text.** Retrieved October 26, 2024, form https://www.analyticsvidhya.com/blog/2021/12/different-methods-for-calcu lating-sentiment-score-of-text/

Sirikul, C. and Samphanwattanachai, B. (2023). Factors affecting the purchase decision of sleep and stress support supplements among Thai consumers in Bangkok. **Rajapark Journal,** 17(53), 53–66.

SMEThailand. (2023). **Online reviews as a powerful strategy: Which generation is most influenced by them?.** Retrieved October 26, 2024, form https://www.smethailandclub.com/startup-digitalmarketing/8959.html

Stammback, D., Webersinke, N., Bingler, J. A., Kraus, M. and Leippold, M. (2022). A dataset for detecting real-world environmental claims. **Center for Law and Economics Working Paper Series,** 7(2022), (pp. 1-14). Gunupur, India.

Sukheja, S., Chopra, S. and Vijayalakshmi, M. (2020). Sentiment analysis using deep learning – A survey. In **The International Conference on Computer Science, Engineering and Applications** (pp. 1-4). Gunupur, India.

Sulianta, F., Ulfah, K. and Amalia, E. (2024). Revealing consumer preferences in the fashion industry using K-means clustering. **International Journal of Engineering Continuity,** 3(2), 34-53.

Tamanna. (2023). **Why embeddings usually outperform TF-IDF: Exploring the power of NLP.** Retrieved October 30, 2024, form https://medium.com/@tam.tamanna18/exploring-the-power-of-nlp-why-embeddings-usually-outperform-tf-idf-98742e7b0bce/

Thanam, Y., Sirigonganjana, V., Thanam, P., Panomratanasak, J. and Chulakasate, S. (2023). The consumer attitudes and preferences of gold N health supplements. **Journal of Roi Kaensarn Academi,** 8(12), 370–386.

Thansettakij. (2023). **Global spotlight: 11 key health market segments projected to reach USD 1.1 trillion in 2023.** Retrieved October 30, 2024, form https://www.thansettakij.com/business/marketing/581011/

Thakur, P. S., Verma, R. K. and Tiwari, R. (2024). Analysis of time complexity of K-means and fuzzy C-means clustering algorithm. **Engineering Mathematics Letters,** 4(2024), 1-9.

TKPon. (2020). **Customer reviews. Takra online.** Retrieved October 3, 2024, form https://shorturl.asia/jnq29

Wen, Y., Liang, Y. and Zhu, X. (2023). Sentiment analysis of hotel online reviews using the BERT model and ERNIE model data from China. **PLoS ONE,** 18(3),1-14.

Wharton, B. (2024). **Natural language processing: Transforming large data into strategic business insights.** Retrieved April 26, 2024, from https://inmoment.com/blog/nlp-natural-language-processing/

Wisesight. (2023). **What is sentiment analysis and how it enhances customer understanding and business performance.** Retrieved October 30, 2024, form https://wisesight.com/news/what-is-sentiment-analysis/

Yadav, A. and Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. **Artificial Intelligence Review,** 53(6), 4335-4385.

Zaghloul, M., Barakat, S. and Rezk, A. (2024). Predicting E-commerce customer satisfaction: Traditional machine learning vs. deep learning approaches. **Journal of Retailing and Consumer Services,** 79, 103865.

Zhu, Y. and Liang, G. (2024). Designing product upgrades in the presence of online consumer reviews. **Managerial and Decision Economics,** 45(6), 3915–3928.