JOURNAL OF DATA SCIENCE | Vol.2022:12
eISSN:2805-5160

Sentiment Analysis on Natural Skincare Products

Fadly [1], Dewi Marlina [1], Tri Basuki Kurniawan [2,*], Mohd Zaki Zakaria [3], Siti Farahnasihah binti Abdullah [3]

[1] Jurusan Farmasi, Politeknik Kesehatan Kemenkes Palembang, Indonesia
[2] Faculty of Computer Science, University of Bina Darma, Palembang, Indonesia
[3] Faculty of Computer & Mathematics Sciences, University Technology Mara, Malaysia

[*]**Email**: tribasukikurniawan@binadarma.ac.id

Abstract

Skincare Industry was increasing rapidly year by year. In contrast, many skincare companies have brought their products and originality to attract many customers. However, due to many controversial cases involving the chemical substance in skincare products, the company switched to something more natural: natural skincare. With much natural skincare in the shop, many customers face the problem of which one to buy. This research helps customers by giving a guideline for the customers to make the decision. Sentiment Analysis is used to analyze the reviews from past customers and create a visualization containing positivity and negativity of all the reviews. Five classifiers were used to produce the best result: Naïve Bayes, KNN, SVM, Decision Tree, and Deep Learning. The reviews were collected from Sephora.com websites, and the tools used in analyzing the reviews are Python and RapidMiner. Reviews collected are 10000 data from a website. The result shows that Deep Learning and Decision Tree are classifiers in sentiment analysis with almost 80% accuracy and 60% F1 measurement. F1 measure is a measure of a test's accuracy. For future enhancements, the data collected can be more than this research, and no data imbalance was created.

Keywords

Sentiment Analysis, Natural Skincare, Python, RapidMiner

Introduction

Good skin is a blessing in life and a sign of healthiness. However, there is always the desire to keep looking younger as time passes and establish eternal youth. The usage of skincare begins back at 6000 years ago (Jhawar, Schoenberg, & Wang, 2018) (Ben-Noun (Nun), 2016). The term cosmetics defines beauty and the outer look and focuses on the human body. The early usage of skincare is believed to have arisen in Egypt and India. Substances that were used earlier in history are castor oil (as a balm), beeswax (as skin cream), olive oil, and rose water (Mawazi et al., 2022).

Submission: 15 August 2022; **Acceptance:** 31 August 2022

Cosmeceuticals are the terms that refer to cosmetic products that have health purposes rather than just a cosmetic focus. It refers to the combination of the principles of cosmetics and drugs (Nanjwade, 2017). The president of the society of cosmetic chemists defined cosmeceutical as "A scientifically designed product intended for external application of the human body and has aesthetic properties. Furthermore, it also must meet the medical, chemical, and physical standards" (Jhawar et al., 2018).

Skincare is any remedy or product used to clean and remove dirt from the face. It also enhances the outer appearance of dermatology problems such as acne, redness, or dryness on the front (Nilforoushzadeh et al., 2018). There are many kinds of skincare on the market. Each one serves a purpose, such as skin care for combination, dry and oily skin. This skincare can cost hundreds of dollars based on the ingredients used. Therefore, the reviews on the products used by the previous consumer are critical and serve the purpose of making the potential customer buy the suitable product for them (Masłowska, Malthouse, & Bernritter, 2017).

Many products in the markets nowadays are full of chemicals. The study demonstrates that the average individual slathers, foams, rubs, and splashes ten unique healthy body items on their body regularly. This manufactured healthy skin may contain more than 130 synthetic substances, which are bound to be retained by our skin as the idea of our skin is probably going to ingest the compound substances as opposed to as a resistance system for our body. Therefore, products on the market are harsh on the skin and have side effects. According to (Atikah et al., 2017), a beauty therapist revealed that many products on the market are harsh on the skin. Hence, cosmetic products made from natural ingredients have become more popular (Krishnan, Safia Amira, Nur Atilla, Syafawani, & Hafiz, 2017).

Moreover, women in Asia, not to mention in Malaysia, are obsessed with being fair. Fairness is an essential element of beauty standards in Asia (Shankar & Palaian, 2007). A New York Times report in 2006 quoted a survey by Synovate, a market research company, that four out of every ten women in Malaysia used whitening products (Krishnan et al., 2017).

Lastly, the customer also finds it hard to make the right decision when purchasing skincare products as every skincare is different and unique in its way. They also do not want to damage any skin barrier by using the wrong products. Another study claimed that some Asians tended to overuse these randomly picked products without thoroughly investigating their usefulness and the possible adverse outcomes of their usage (Osei-Frimpong, Donkor, & Owusu-Frimpong, 2019). Many cosmetics producers, especially local manufacturers, produced a wide range of whitening products, many of which had illegal substances (Krishnan et al., 2017).

It targets customers who want information and opinions about skincare products before purchasing them. The mobile application lets the customer visualize the range of skincare products in the market. How does the previous customer value it, whether it is a negative or positive opinion? From that information, it can give choices to the customer to purchase the good products and make it work wonder on their face.

Other than that, the project scope for sentiment analysis on natural skincare products using text mining also focuses on the organization that needs to analyze the performance of their products

based on the sentiment analysis conducted. This sentiment analysis will give the organization a good or even lousy representation of their products based on the performance of their products and the previous customer evaluation.

Methodology

Sentiment analysis requires the practice of Natural Language Processing and Text Analysis to get insights into user emotions, feelings, opinions, or attitudes. Most of the opinions are subjective and not facts. They may vary from one another. It is essential to analyze which one of the polarities is heavy, whether on the positive, neutral, or negative sides. According to a study by Zendesk cited by Mike Waldron (Waldron, 2014), 45% of users share bad customer service, and 30% share good customer service experiences via social media. It means the Sentiment Analysis and its application to business analytics is very important. Sentiment analysis involves data collection, pre-processing, feature selection and feature extraction, and lastly, determining the sentiment of the reviews.

This research has applied eight phases: a preliminary study, data preparation, data pre-processing, data analysis, system architecture design, interface design, system development, and system testing and evaluation, as shown in Figure 1.

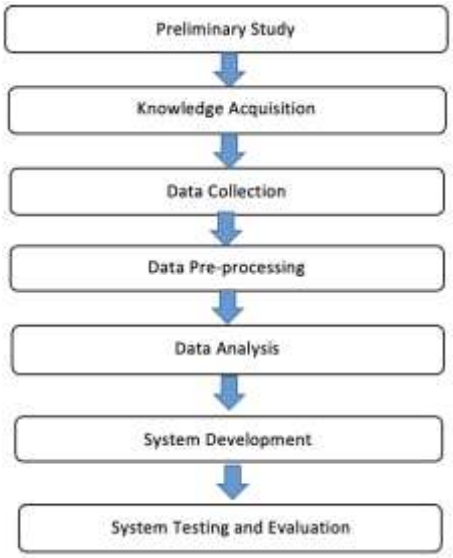

Figure 1. Overview of the Methodology

Figure 1 covers the literature review, development of the research framework, clarification of the research problem and objective, and data collection and preparation for analysis. Finding the research problem is one of the key tasks that need to be carried out in any research. It starts by identifying the problem area and performing an in-depth investigation of its problems or issues. In this study, the issue of customers' reviews on natural skin care products is addressed using the positivity and negativity of the reviews on the internet.

The collected information helped in defining the research problem more clearly by reviewing the literature; such as the study of research works that were related to the sentiment analysis on products review, the rise of natural-based skincare products in the market, techniques used in sentiment analysis and classification methods to classify the reviews based on the dominant positivity and negativity in the language used in the reviews.

Data collection involves the collection and selection of important data sets which can be used for the research evaluation. There are two primary datasets: testing and training data sets used to evaluate the proposed classification techniques. The technique used for collecting data is scraping using a web crawler, as shown in Figures 2 dan 3.

Figure 2. Sephora.com website home page

Figure 3. Ranting and reviews in Sephora.com

Figure 2 shows that the Sephora.com website home page was chosen for the data collection. Figure 3 shows the reviews displayed on Sephora.com. This information will then be collected.

Reviews collected are based on the most helpful reviews acknowledged by the user who read them previously. These reviews are the most reliable to collect.

Data miners used to do data scraping from Sephora.com. Data miner is a free application to extract any data from the website. The procedure for using these tools is shown in Figures 4 until Figure 9. It requires downloading and comes with an extension.

Figure 4. Data miner homepage

Figure 5. Extracting reviews from Data miner, create a new recipe

Figure 6. Select data required from the website

Figures 4, 5 dan 6 show that in this step, a few requirements must be filled up and considered, such as the type of page, rows, columns, navigations, and file format. There are no restrictions on how it is made as long as the desired information is collected. In this case, the information was collected only from reviews from the customer of Sephora.com. Users need to select which rows and columns need to be extracted by the DataMiner.

Figure 7. Select desired information and run the recipe

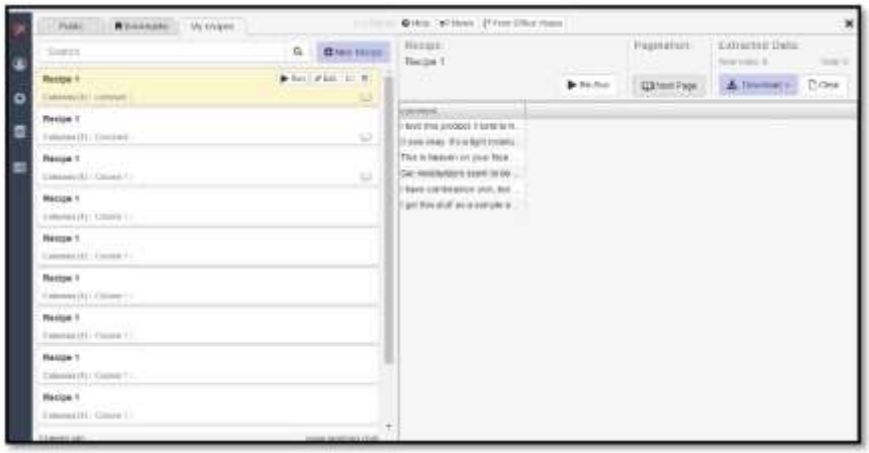

Figure 8. Pagination is where the following information is collected automatically

Figure 7 shows the last steps of recipe creation in Data miner. Make sure information is collected correctly and in order of column. Figure 8 shows the pagination steps. Pagination is essential for the Data miner to know which pages to go next and scrape data needed by the user. Users can set what pages they need to do the scraping.

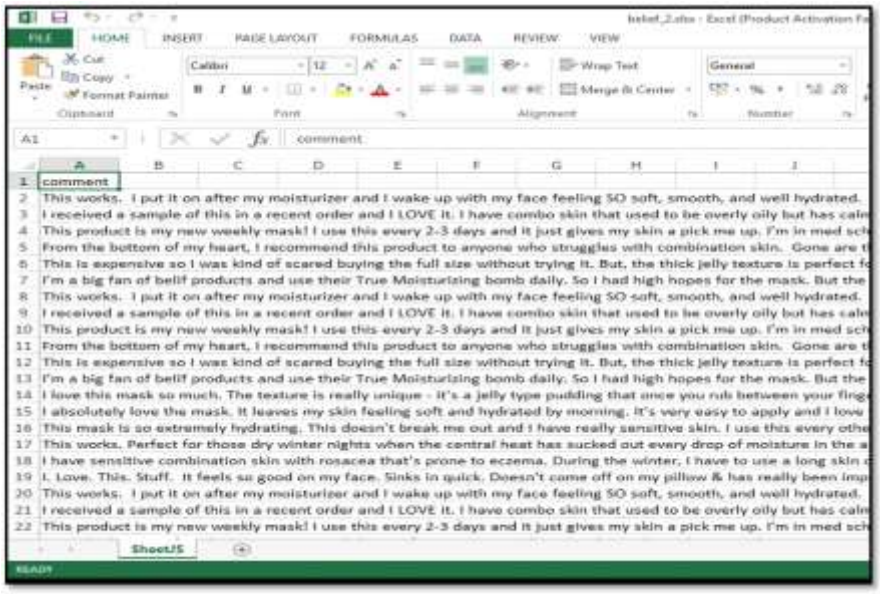

Figure 9. Sample of data collected

Figure 9 shows what happens after the web scraping is done, user have choices to download the reviews according to their preferences, such as in a .csv file or .xlsx file. This feature makes it easy for an analyst to do the next analysis step. In this research, the datasets were downloaded in a .xlsx file. Column names are also crucial for creating coding in the subsequent phases and calling for the datasets.

Data Pre-processing

Data pre-processing is one of the data mining techniques that include the process of transforming data into understandable information. Raw data is usually inconsistent, irrelevant, and contains noisy data. Examples of noisy data are errors and outliers. The data must be cleaned and transformed in sentiment analysis to ease machine learning. Any abbreviation, stop words, punctuation, and stemming are needed to be removed before conducting the sentiment analysis. This phase is important since it can affect the quality of the accurate result.

Data pre-processing deals with data cleaning, sampling, transformation, and attribute selection. The medium that will be used for cleaning the data is Python. Python is used by coding the code to remove redundancy, stop words, and punctuation in the dataset. After that, sampling data is also important for this research. Sampling data focus on users' reviews on Sephora.com, as shown in Figure 10.

Figure 10. Cleaned data after pre-processing phase

Figure 10 shows results after data is cleaned using Python. The differences show where the sentences are in lowercase, and the marks are gone. Cleaning data is important to ensure the machine learning phases can be done without error, as in text analysis, the machine only reads text. When cleaning the datasets, data decreased from 10000 to 9899. The data was reduced because the reviews contained different languages than English and the data was redundant when scraping, as shown in Table 1, and the result complete with labels are shown in Figure 11.

Table 1. Data collected after the cleaning phase

Brands	Data Collected
Drunk Elephant	1802
Origins	2207
Belif	1800
Laneige	2287
Glam Glow	1803

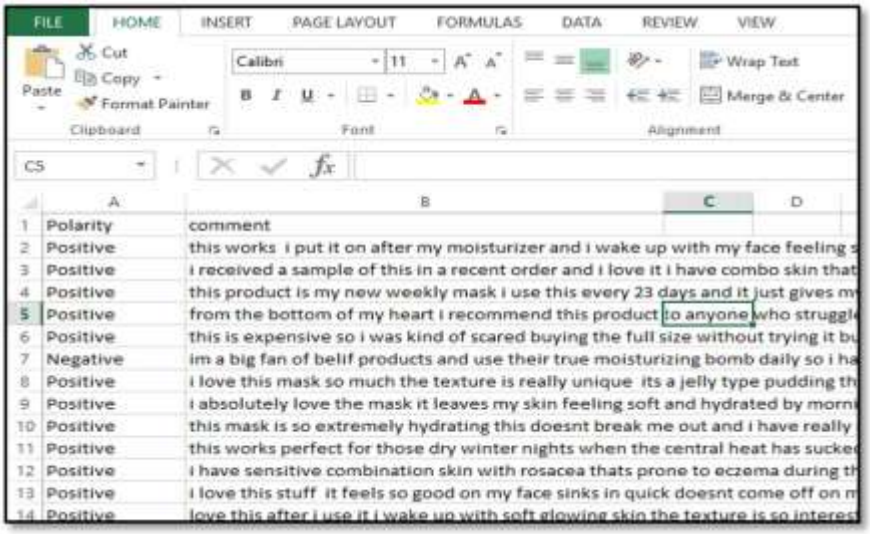

Figure 11. Polarity label for the training dataset

Figure 11 shows the polarity column made to label the reviews into positive and negative labels. This labeling is only done in training datasets for each set. The datasets are randomly divided into 70-30 ratios, whereby 70 is testing datasets, and 30 is training datasets.

Data Analysis

Data analysis is the retrieving and organizing data so a researcher can conclude. Data analysis also allows one to answer questions, solve problems, and derive meaningful information from them. Cleaned datasets obtained from data pre-processing are being analyzed using a machine learning approach to classify the positivity and negativity of the customer review toward natural skincare brands, as shown in Figure 12.

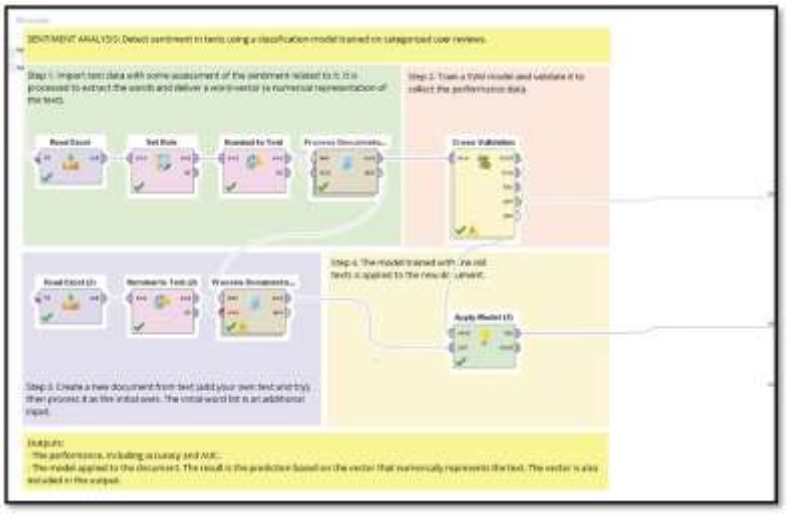

Figure 12. Sentiment Analysis Features

Cleaned data was imported into the RapidMiner for the analysis phase. Sentiment Analysis features are used to analyze the dataset. From Figure 12, the data must first be imported into the RapidMiner using the Read Excel Operator. This operator showed first how the data needed to be in the RapidMiner. The attributes are comment and polarity. The comment is polynomial type; meanwhile, polarity is set to binomial type. Next, Figure 13 shows the process of documents.

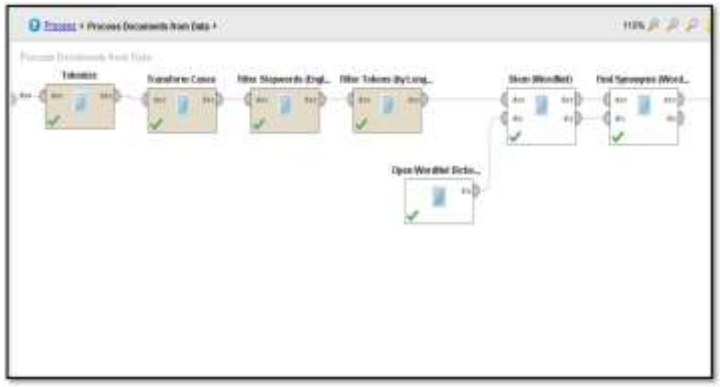

Figure 13. Process data from the document

Figure 13, the Process Documents from Data Operator contained Tokenize, Transform Cases, Filter Stop words, Filter Tokens (by Length), Stem (WordNet), and Find Synonyms operators. WordNet is the electric database of English. It contained nouns, verbs, adjectives, and adverbs grouped into sets of cognitive synonyms (synsets). WordNet also acts like a thesaurus by grouping words into their meaning. WordNet was added to make the sentiment analysis more accurate by finding the synonym in each row to predict the label. It will then compare the number of synonym words in the testing dataset.

The cross-validation then splits the data by the holdout method. In this step, the 10-fold cross-validation is used. 10-fold cross-validation means the data is divided into ten (10) divisions or parts, one for testing and 9 for training in the first run. In a second run, a different part is used for testing, and nine (9) parts, including the oat used for testing in run one (1), are used in training.

Absolute accuracy is the average of the accuracies obtained in 10 runs—other operators in cross-validation, including classifier and performance operator, to calculate the classifier's performance.

Classifier Algorithms

Support Vector Machine Classifier

The Support Vector Machines Classifier (SVM), a discriminative classifier formally described by a separating hyperplane, created procedures for an ideal hyperplane using training datasets. SVM is frequently employed to solve classification and regression issues. The data points closest to the hyperplane are known as support vectors, as shown in Figure 14.

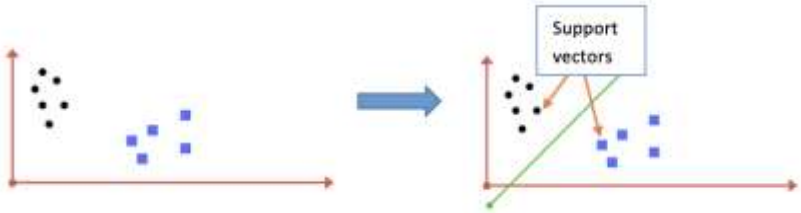

Figure 14. How Linear Classification Works

K-NN Classifier

The following classification techniques used are K-nearest neighbors. KNN can be used both in regression and classification problems. The class of a red star is not decided yet. Make a circle first be the red star as the center. Assuming that K=3, the three closest points to the red star are all blue circles; hence, it can be decided that the red star is in a class of blue circles, as shown in Figure 15.

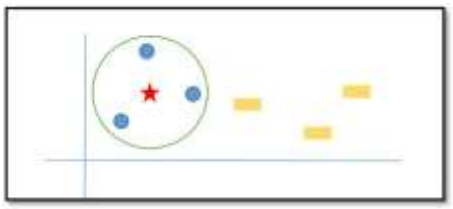

Figure 15. KNN Classification

Decision Tree Classifier
Decision Tree are supervised learning also used for classification and regression. Decision trees learn from data to approximate a sine curve with a set of if-then-else decision rules. The deeper the tree, the more complex the decision rules, and the fitter the model.

Naïve Bayes Classifier
The Bayes Theorem, based on Naive Bayes, assumes that all attributes, regardless of any relationships, independently contribute to the datasets' probability. In other words, the Naïve Bayes algorithm predicts a class given a set of features using probability.

Deep Learning Classifier
Deep Learning is based on a multi-layer feed-forward artificial neural network trained with stochastic gradient descent using back-propagation. The network can contain many hidden layers consisting of neurons with tanh, rectifier, and max-out activation functions.

Each computes node trains a copy of the global model parameters on its local data with multi-threading (asynchronously) and contributes periodically to the global model via model averaging across the network.

Results and Discussion

The visualization used in this research was Word Cloud. Word Cloud is an image composed of words used in a particular text or subject, in which the size of each word indicates its frequency or importance. The bigger text of words indicates that the word is mentioned numerous times in the dataset. Word cloud has been popular these days as it was a simple visualization yet important in describing one dataset.

For Belif Skincare, the result is shown in Figures 16 and 17.

Figure 16. Word cloud for Belif Skincare

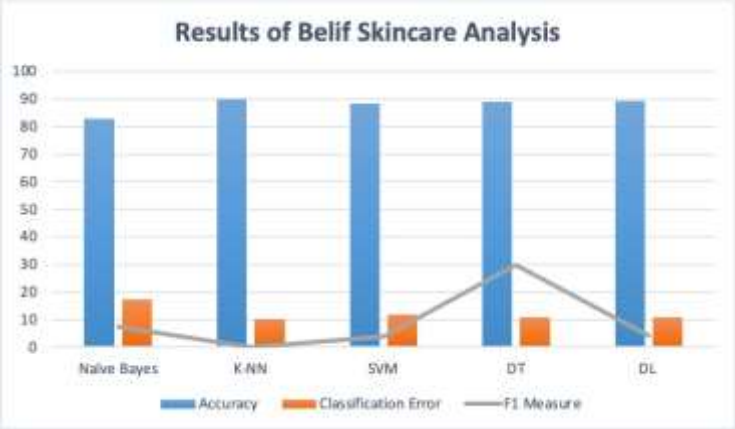

Figure 17. Belif skincare analysis summary chart

Figure 16 shows the word cloud of Belif Skincare. It can be summarized that the customers talk about their skin. Several positive words such as amazing, great, good, work well, and excellent is also mentioned numerous times. It indicates that the dataset of Belif Skincare received positive reviews.

Based on the charts in Figure 17, Naïve Bayes classifier accuracy is 82.76%, the lowest accuracy obtained among the classifier. Meanwhile, the higher accuracy obtained is 89.98% by the KNN classifier. Meanwhile, SVM, Decision Tree, and Deep Learning classifiers had an accuracy of 88.36%, 89.06%, and 89.29%. The classification error highest is from Naïve Bayes with 17.24%.

For the Drunk Elephant Skincare dataset, the result is shown in Figures 18 and 19.

Figure 18. Word cloud for Drunk Elephant Skincare

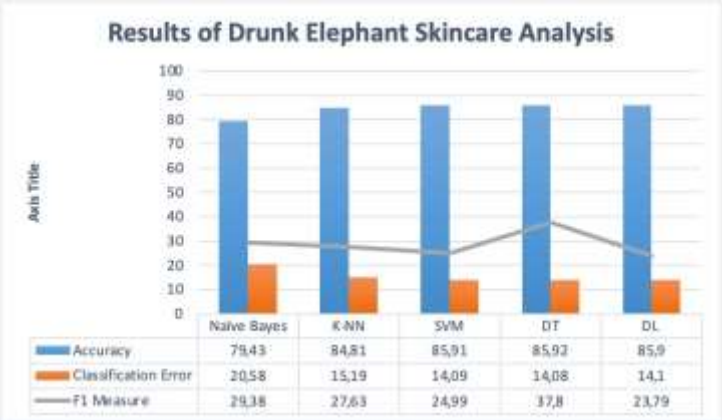

Figure 19. Drunk Elephant analysis summary chart

Figure 18 shows the word cloud of Drunk Elephant Skincare. It can be summarized that the customers talk about their skin. In Comparison to Belif Skincare, Drunk Elephant uses fewer positive words in smaller font sizes.

Figure 19 shows the summary of the analysis conducted. Based on the charts, naïve Bayes classifier accuracy is 79.43%, the lowest accuracy obtained among the classifier. Meanwhile, the higher accuracy obtained is 85.92% by the Decision Tree classifier. Even though SVM, Decision Tree, and Deep Learning classifiers all achieved the same accuracy, variations can be seen in the F1 measure, where Decision Tree obtained the most balanced precision of 37.8% and recall of just 24.99% and 23.9% for the other two classifiers.

The following dataset is Glam Glow Skincare; the result is shown in Figures 20 and 21.

Figure 20. Word cloud for Glam Glow Skincare

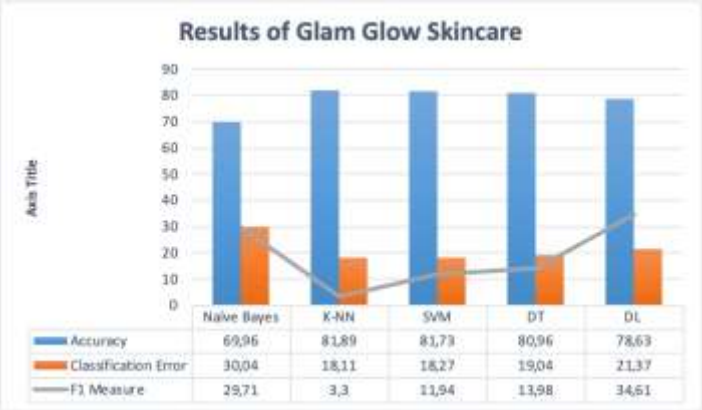

Figure 21. Glam Glow analysis summary chart

Figure 20 summarises the analysis conducted for Glam Glow skincare using Word cloud. Based on the charts in Figure 21, the KNN classifier shows the highest accuracy at 89.89%, while Naïve Bayes has the lowest accuracy at 69.96%. SVM, Decision Tree, and Deep Learning score 81.73%,80.96%, and 78.63%, respectively. Naïve Bayes has the highest classification error with 30%. Meanwhile, the F1 measure for all classifiers variates with Naïve Bayes (29.71%), KNN (3.3%), SVM (11.94%), Decision Tree (13.98%), and Deep Learning (34.61%). Therefore, Deep Learning has the highest F1 measure compared to other classifiers.

For Laneige Skincare, the result is shown in Figures 22 and 23.

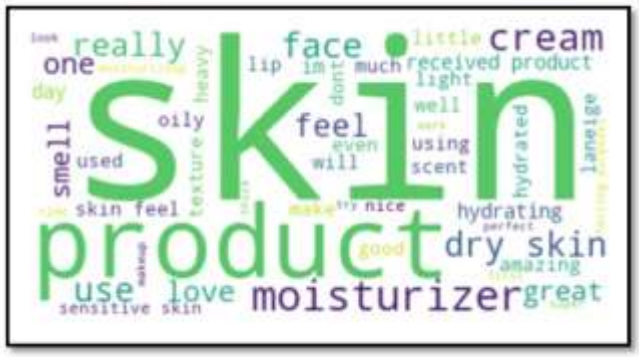

Figure 22. Word cloud for Laneige Skincare

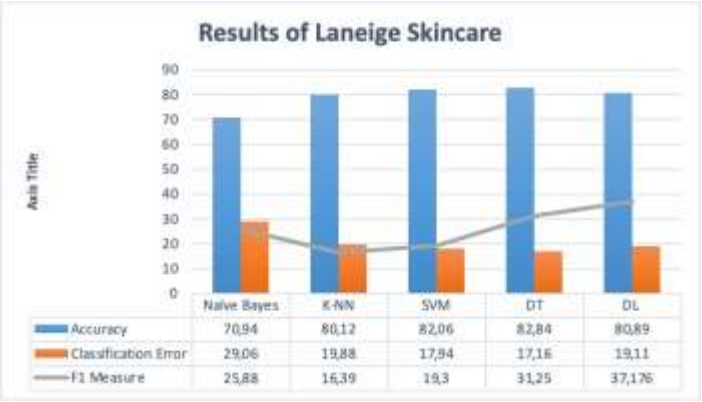

Figure 23. Laneige skincare analysis summary chart

Figure 22 shows the summary of the analysis conducted for Laneige skincare. Based on the charts shown in Figure 23, the Decision Tree classifier has the highest accuracy, 82.84%, while Naïve Bayes has the lowest accuracy, 70.94%. KNN, SVM, and Deep Learning score 80.12%, 82.06%, and 80.89%, respectively. Naïve Bayes has the highest classification error with 29.06%. Meanwhile, the F1 measure for all classifiers variates with Naïve Bayes (25.88%), KNN (16.39%), SVM (19.3%), Decision Tree (31.25%), and Deep Learning (37.17%). Therefore, Deep Learning has the highest F1 measure compared to other classifiers.

For the last dataset, Origin Skincare, the result is shown in Figures 24 and 25.

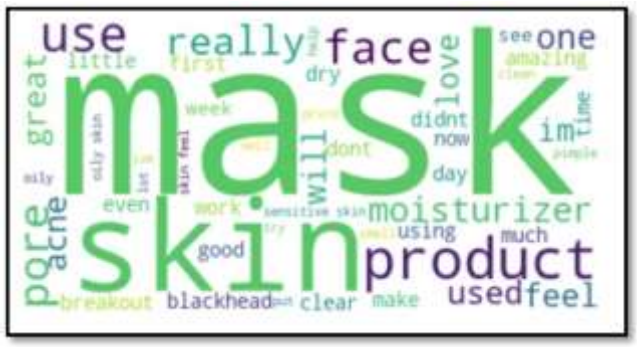

Figure 24. Word cloud for Origins Skincare

Figure 25. Origins skincare analysis chart

Figure 24 shows the summary of the analysis conducted for Origins skincare. Based on the charts shown in Figure 25, the SVM classifier has the highest accuracy, 75.17%, while Naïve Bayes has the lowest accuracy, 61.21%. KNN, Decision Tree, and Deep Learning scored 69.31%, 71.38%, and 73.62%, respectively. Naïve Bayes has the highest classification error with 38.79%. Meanwhile, the F1 measure for all classifiers variates with Naïve Bayes (49.43%), KNN (52.15%), SVM (58.38%), Decision Tree (41.15%), and Deep Learning (62.95%). Therefore, Deep Learning has the highest F1 measure compared to other classifiers.

Table 2 shows the summary of the best result obtained by the algorithms.

Table 2. Summary Table of Skincare Analysis by Brand

Skincare Brands	Selected Classifier after Comparison	Positive Prediction (%)	Negative Prediction (%)	Recommended (✓)
Drunk Elephant	Decision Tree	96	4	
Belif	Decision Tree	96	4	
Glam Glow	Deep Learning	78	22	✓
Origins	Deep Learning	67	33	✓
Laneige	Deep Learning	84	16	✓

Table 2 shows the summary table of the Skincare Analysis by Brands. The table indicates that Deep Learning Classifier is the preferable classifier for accurate analysis. Deep Learning can learn the negative and positive reviews more than the other classifier. Based on the table also indicates that Glam Glow, Origins, and Laneige are preferable brands chosen by past users.

Although the negative reviews of these brands are higher than those Belif and Drunk Elephant obtained, it does not mean that these brands are lesser than the others. Choosing which class label they belong to could lead to a classification error. It is also due to data imbalance, and the most reliable datasets are from Glam Glow, Origins, and Laneige.

Conclusion

Several steps and phases have been done to achieve all the objectives in this research. The first objective was to investigate reviews from past users about natural skincare products. The second objective was to analyze the positivity and negativity of products reviewed using text mining.

The first objective is to investigate reviews from past users about natural skincare that had been fulfilled in the architecture design phase. Numerous studies have been done to achieve the first goal. It involved using a variety of websites, apps for mobile devices, and blogs. Since Sephora.com is acceptable for web scraping, it has been chosen to capture the data.

The second objective is to analyze the positivity and negativity of products reviewed using sentiment analysis. This objective requires many studies in past literature on a machine learning algorithm. In order to increase the variety of the results received, many brands are employed to find the best classifier. Five classifiers—Naive Bayes, KNN, SVM, Decision Tree, and Deep Learning—were utilized in this study, and results were obtained and compared.

References

Atikah, N., Wavi, N., Suhaida, N., Fahmi, M., Yusof, M., Iskandar, M., … Rahim, M. (2017). *THE FACTORS THAT INFLUENCE THE CONSUMPTION OF LOCAL COSMETIC PRODUCTS AMONG CONSUMERS IN KOTA BHARU, KELANTAN*.

Ben-Noun (Nun), L. (2016). *BEAUTY OF HUMANS*.

Jhawar, N., Schoenberg, E., & Wang, J. V. (2018). US CR. *Clinics in Dermatology #pagerange#. Johnsy, G., S. Davidson Sargunam, and V. Kaviyarasan.*, *4*, 1.

Krishnan, S., Safia Amira, N., Nur Atilla, U., Syafawani, S., & Hafiz, M. (2017). The Usage of Cosmetic in Malaysia: Understanding the Major Factors that Affect the Users. *Management*, *2017*(1), 48–51. Retrieved from http://journal.sapub.org/mm

Masłowska, E., Malthouse, E., & Bernritter, S. (2017). *The Effect of Online Customer Reviews' Characteristics on Sales*. https://doi.org/10.1007/978-3-658-15220-8_8

Mawazi, S., Ann, J., Othman, N., Khan, J., Alolayan, S., thagfan, S., & Kaleemullah, M. (2022). A Review of MoisturizersMoisturizers; History, Preparation, CharacterizationCharacterization and Applications. *Cosmetics*, *9*, 61. https://doi.org/10.3390/cosmetics9030061

Nanjwade, B. (2017). DEVELOPMENT OF COSMECEUTICALS. *World Journal of Pharmacy and Pharmaceutical Sciences*, 643–691. https://doi.org/10.20959/wjpps20174-8927

Nilforoushzadeh, M., Amirkhani, M., Zarrintaj, P., Moghaddam, A., Mehrabi, T., Alavi, S., & Mollapour, M. (2018). Skin care and rejuvenation by cosmeceutical facial mask. *Journal of Cosmetic Dermatology*, *17*. https://doi.org/10.1111/jocd.12730

Osei-Frimpong, K., Donkor, G., & Owusu-Frimpong, N. (2019). The Impact of Celebrity Endorsement on Consumer Purchase Intention: An Emerging Market Perspective. *The Journal of Marketing Theory and Practice*, *27*, 103–121. https://doi.org/10.1080/10696679.2018.1534070

Shankar, P. R., & Palaian, S. (2007). Fair skin in South Asia: An obsession? *Journal of Pakistan Association of Dermatologists*, *17*.

Waldron, M. (2014). Why is Sentiment Analysis important from a business perspective? Retrieved from http://blog.aylien.com/why-is-sentiment-analysis-important-from-a/