

# A Multilabel Text Classification and Association Rule Mining Framework for Data-Driven Skincare Product Name Generation in E-Commerce

**Naffa Nur Fauziah**

Program Study of Information System, Faculty of Engineering, Universitas Negeri Surabaya, Indonesia  
e-mail address: naffa.22097@mhs.unesa.ac.id (corresponding author)

Received: 10 October 2025 | Revised: 25 October 2025 | Accepted: 20 November 2025  
This is an open access article under the [CC BY-SA](#) license.



## ABSTRACT

The rapid expansion of e-commerce has produced large volumes of unstructured product-name data, particularly in the skincare sector, where a single name often conveys multiple functions. This study proposes an integrated framework combining multilabel text classification and association rule mining to extract structured knowledge from skincare product names and generate data-driven naming recommendations. Using product names scraped from Tokopedia, the dataset was preprocessed through text cleaning, TF-IDF vectorization, and multilabel binarization. Three multilabel classification strategies, Binary Relevance, Classifier Chain, and Label Powerset, were evaluated with Support Vector Machine (SVM), Logistic Regression, Random Forest, and K-Nearest Neighbor algorithms under 60:40, 75:25, and 80:20 train-test splits. According to the experimental results, Binary Relevance achieved the best performance across all strategies, with an F1-test score of 0.925 for the SVM model at an 80:20 split, followed by Random Forest (0.912) and Logistic Regression (0.883). Classifier Chain had consistent results across all, still slightly lower, but robust findings, while Label Powerset obtained the lowest results due to label sparsity. The predicted labels were further explored using the Apriori algorithm, which revealed strong co-occurrence patterns, in particular, moisturizers were a dominant component of product bundles. High-confidence rules (0.99) and strong lift values (2.73) provide evidence of a significant association in the dataset. These insights formed the basis for proposing new naming themes such as Daily Radiance, Glow Hydration, UV Protection, and Clean & Fresh. Overall, the study demonstrates that combining multilabel classification with association rule mining is effective for uncovering naming structures in skincare e-commerce data.

*Keywords-Multilabel Classification; Association; Apriori; Skincare Product Name; E-commerce Data Mining*

## I. INTRODUCTION

The rapid development of e-commerce in recent years has driven the growth of text-based product data in very large quantities, particularly in product descriptions and product names [1]. In the skincare industry, product names not only function as an identity but also contain various information such as product type, main ingredients, and benefits. This condition makes product name data become unstructured and contain many meanings within a single text. Therefore, text mining and Machine Learning (ML) based data processing methods are required to manage and utilize the data optimally [2]. The main problem in processing skincare product name data is that one product often has more than one category, for example as “cleanser”, “acne”, and “brightening” within a single product name. This can be affect to the accuary of single-label classification methods. Therefore, a multilabel classification approach is used because it is capable of assigning more than one label to a single data instance simultaneously [3]. Several previous studies have combined the Term Frequency-Inverse Document Frequency (TF-IDF) method with algorithms such as Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and K-Nearest Neighbour (KNN) to perform text classification [4, 5]. In multilabel classification, there are several commonly used strategies, namely Binary Relevance, Classifier Chain, and Label Powerset. In addition, classification results are also highly influenced by model parameter selection, so hyperparameter optimization is required to improve model performance [6].

Previous studies have shown that hyperparameter optimization can significantly increase model accuracy and performance, especially when evaluated using the F1-score metric [7].

In addition to classification, association analysis is also widely used in product data processing to discover relationships between items [8]. One of the most frequently used algorithms is Apriori, which aims to find patterns of item co-occurrence or association rules [9]. Additionally, association rule mining can be used to identify relationships between words that frequently appear together in product names, allowing the interrelated patterns among terms in product naming to be analyzed systematically. However, most studies that use the Apriori algorithm still focus on sales transaction data, not on product names as the main object of analysis [10]. In fact, product names play a very important role in shaping consumer perception and attracting purchase interest. Until now, research that utilizes the Apriori algorithm to extract word patterns from product names as a basis for innovating new product naming, particularly in the skincare e-commerce category, remains very limited.

The purpose of this research is to provide data-driven ideas for new skincare product names by utilizing association analysis using the Apriori algorithm on existing product names. In addition, this research also constructs a text-based multilabel classification model using TF-IDF representation combined with the Binary Relevance, Classifier Chain, and Label Powerset strategies, as well as the RF, LR, SVM, KNN algorithms as supporting stages to understand the characteristics of skincare products. To obtain optimal model performance, this research also applies hyperparameter optimization using GridSearchCV. This research combines multilabel classification methods with association analysis using the Apriori algorithm on skincare product name data to generate new product name ideas. Multilabel classification is also used to help the interpretation of product features with multiple categories under a single name. Meanwhile, the Apriori algorithm is applied to identify associations among words that often occur together in product names. These findings are further analyzed and used to generate skincare product name suggestions that better align with market needs.

## II. METHOD

This research use the Knowledge Discovery in Databases (KDD) process model as shown in Figure 1. There are several steps that must be carried out systematically and sequentially in the data analysis process using the KDD method [11]. Starts with a review of the literature in the area to learn fundamental principles, related methodologies, and previously reported work. The next step is collecting data product of skincare products on Tokopedia. Once data is collected, the work continues with cleaning and transforming it therefore can be used for analysis. Label transformation is conducted based on a multilabel form to transform product names into category representations, which are further translated in binary terms by those generally adopted in multilabel studies [12]. The stage of data mining is unawarely run on Apriori to reveal associations between product categories. This technique is capable of generating frequent itemsets and association rules that can be measured by the support confidence and lift [13].

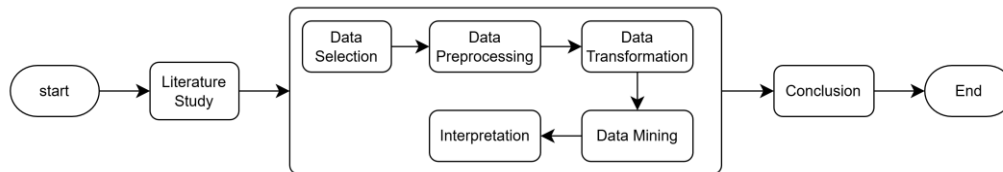


Figure. 1. Research Framework using KDD

### A. Data Selection

The dataset obtained from a web scraping on Tokopedia with the skincare product category. Tokopedia it is one of the largest e-commerce platforms in Indonesia, with a diverse number of products, sellers, and user reviews [14]. So that, this work using data from Tokopedia. The data was collected through a web scraping, an automatic system for retrieving information from the web pages. The Tokopaedi library is used for web scraping of data from Tokopedia platform. The scrapped data obtained (Table 1) consists of 500 rows and several variables, however the Product Name as variable is chosen.

TABLE 1. PRODUCT NAME

No	Product Name
1.	[BRIGHT GLOW SET] Emina Bundle All in 7in1 Face Wash + Exfo Toner + Toner + Serum + Moisturizer Gel + Brightening Sunscreen+ Low pH Micellar Water Bright & Smooth - Glow 7 Micellar
2.	[BRIGHT GLOW SET] Emina Bundle All in 7in1 Face Wash + Exfo Toner + Toner + Serum + Moisturizer Gel + Brightening Sunscreen+ Low pH Micellar Water Bright & Smooth - Glow 7 Micellar
3.	[Diskon 10% Sekarang!!] Glad2Glow Renew bundle all in 1 7in1 -Brightening Moisturizer+Serum+Cleanser+toner +Peeling solution+Retinol 10%Niacinamide skincare paket pembersih perawatan kulit mencerahkan warna kulit pelembab wajah g2glow official store - [BEST] 7in1-hot sale

B. Preprocessing

The text is preprocessed to clean the unstructured and inconsistent texts into a structured one. Beforehand, the identification of the columns to be considered for the analysis and verification of duplicates to avoid bias from repeated entries [15]. When there is missing data it can be dealt with by deleting or transforming, so that the uncertainties in data do not impact on analysis. After that, it is proceeded by Text Preprocessing which consists of case folding that change all texts to lowercase making all words in the lowercases form and noise removal removing unwanted characters like numbers symbols punctuations URLs or spaces [16]. The cleaning step is necessary as these features may not add textual meaning and could interfere with feature representation [17]. The results of the preprocessing are shown in Table 2.

TABLE 2. DATA PREPROCESSING

Product Name	Case Folding	Remove Noise
♥ YUKI Manado ♥ HANASUI Glow Expert Paket Mencerahkan Skincare Moisturizer Toner Serum Cleanser – GLOW EXPERT PAKET	♥ yuki manado ♥ hanasui glow expert paket mencerahkan skincare moisturizer toner serum cleanser – glow expert paket	yuki manado hanasui glow expert paket mencerahkan skincare moisturizer toner serum cleanser glow expert paket
Shinzui Brightening White Essentials Series   Face Wash Cleanser / Sunscreen / Serum / Moisturizer   Mencerahkan Melembabkan Glowing – FACE WASH – MINI 20 ML	shinzui brightening white essentials series   face wash cleanser / sunscreen / serum / moisturizer   mencerahkan melembabkan glowing – face wash – mini 20 ml	shinzui brightening white essentials series face wash cleanser sunscreen serum moisturizer mencerahkan melembabkan glowing face wash mini ml
(DAY GOLD)+PAKET GLOWING SERIES DAVIENA SKINCARE Normal Kering Kusam   Skincare Kit Perawatan Kulit   berminyak Mencerahkan - -	(day gold)+paket glowing series daviena skincare normal kering kusam   skincare kit perawatan kulit   berminyak mencerahkan - -	day gold paket glowing series daviena skincare normal kering kusam skincare kit perawatan kulit berminyak mencerahkan

After the text cleaned, a category dictionary of the product keywords was constructed (see Table 3), its contain a list of any keywords to assist the next process (initial labeling).

TABLE 3. PRODUCT CATEGORY KEYWORDS

Category	Keyword
sunscreen	“sun protection”, “sunscreen”, “uv”, “sun”, “sun protect”, “sun cream”, “sun battle”, “protection”, “sunrise”, “sun day”, “spf”, “sunblock”
facial wash	“cleanser”, “facial foam”, “creamy wash”, “face wash”, “wash”, “foam”, “white clay”, “cleansing”, “whip”, “face mask”, “clear double”, “make up clear”, “face soap”, “facial soap”, “facial”
toner	“toner”, “air mawar”, “face tonic”, “facial lotion”, “powder lotion”, “toning”, “hydrating glow essence water”, “rose water”, “hydrating”, “tonique”, “hidrasi”, “hydra”, “hydration”
moisturizer	“gel cream”, “moisturizer”, “gel”, “moisturizing”, “moisture”, “antioksidan juice”, “day cream”, “night cream”, “cream”, “moist”, “moisturize”, “night”, “day”, “moisture”, “bright”, “brightening”
cream and lotion	“milk cleanser”, “milk”, “lotion”, “milk cleansing”, “clean”
scrub and exfoliation	“scrub”, “peeling”, “peel”, “exfoliant”, “exfoliator”, “polisher”, “polish”, “exfoliating”, “ekspoliiasi”, “exfoliasi”
make up removal	“micellar”, “cleansing water”, “remover”, “pembersih”, “micellaire”, “makeup remover”
serum	“serum”

Table 4 presents the labeling results. The labels were created by matching keywords to text using regular expressions (Regex) for efficient, uniform labelling, in line with standard practice in rule-based text annotation [18]. Regex offers a versatile way to define text patterns using literal characters, metacharacters, and quantifiers, allowing fine-tuning of behaviour for detecting words or phrases in unstructured text. For this research, the Regex was established to capture the nuances of terms related to skin care, handling variants such as plural forms, spacing errors, and morphological variations, so that important keywords could still be identified even when they were not written consistently across different sellers. This capability is fundamental in e-commerce text data, as product names include formatting inconsistencies, spelling variations, and promotional phrases.

In addition, Regex is used to minimize annotation ambiguity by imposing a deducible, scalable labeling schema, ensuring that a label is always the same across shared patterns. The regex-based tagging is hands-free, requiring minimal human intervention and being readily scalable, with no limit on the number of product names processed automatically. This process makes the initial annotation more trustworthy and provides promising ground-truth data for multilabel classification. To prevent each product name from being mislabeled, this method guarantees that each name points to the correct cluster before implementing the model.

TABLE 4. LABELLING USING REGEX

Product Name	Cleaned	Label
(cod)Glad2glow Pomegranate Niacinamide Brightening Moisturizer Pelembab Wajah Whitening Glowing Moisture Gel Day & Night Cream Skincare g2g Pelembab Wajah Pemutih Wajah Mencerahkan	cod glad glow pomegranate niacinamide brightening moisturizer pelembab wajah whitening glowing moisture gel day night cream skincare g2g pelembab wajah pemutih wajah mencerahkan	[moisturizer]
[3PCS] Azarine Paket Barrier Moisturizer Series with Sunscreen Kulit Berjerawat Pelembab Wajah Instant Moist Skincare Mencerahkan Kulit Toner Paket Lengkap - HYDRATING SET	pcs azarine paket barrier moisturizer series with sunscreen kulit berjerawat pelembab wajah instant moist skincare mencerahkan kulit toner paket lengkap hydrating set	['sunscreen', 'toner', 'moisturizer']
[BPOM] Glad2Glow Deep Jerawat Bundle all in 1- 2/3/4/5 PCS Skincare Paket-Gel cleanser+Acne Moisturizer+Serum Ceramide Kering Membersihkan Wajah Acid Cooling Kotoran Mencerahkan Vitamin Berjerawat Niacinamide Facial Exfoliating Pencerah G2G - [HOT] 5PCS-Acne set2	bpom glad glow deep jerawat bundle all in pcs skincare paket gel cleanser acne moisturizer serum ceramide kering membersihkan wajah acid cooling kotoran mencerahkan vitamin berjerawat niacinamide facial exfoliating pencerah g g hot pcs acne set	['facial_wash', 'moisturizer', 'cream and lotion', 'scrub and exfoliation', 'serum']

This data preprocessing stage ends with calculating the label frequency, Figure 2 is presents the distribution of product categories and identify potential imbalances before data transformation and multilabel modeling stages.

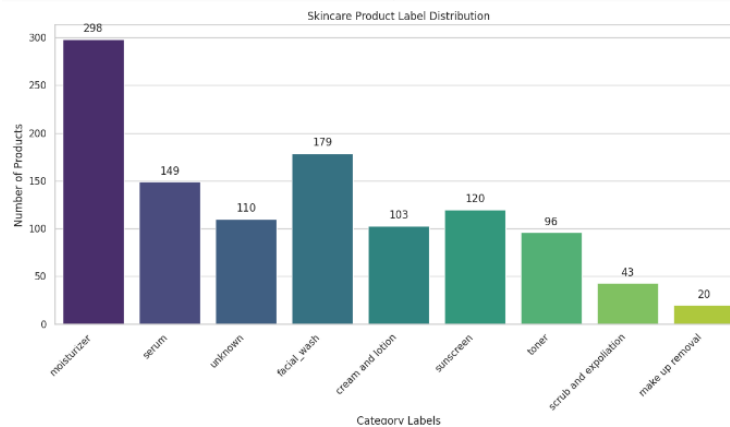


Figure. 2. Distribution Label of Skincare Product Name

### C. Transformation

The data transformation aims to convert the preprocessed text from preprocessing stage into a numerical representation then processed by ML algorithms. The first step is feature extraction; the document content features are extracted using Term Frequency-Inverse Document Frequency (TF-IDF), which assigns a weight to each term based on both its frequency in a document and its distribution across the corpus. We chose to use TF-IDF because it effectively highlights informative words for text classification. This transformation results in a high-dimensional feature matrix, where each text of the skincare product is represented as a numerical vector [19]. The result of feature extraction using TF-IDF are shown in Table 5.

TABLE 5. TF-IDF

Term	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
ac	0	0	0	0	0	0	0	0	0	0
ace	0	0	0	0	0	0	0	0	0	0
achromaxyl	0	0	0	0	0	0	0	0	0	0
acid	0	0	0	0	0	0	0	0	0	0
acidiating	0	0	0	0	0	0	0	0	0	0
acids	0	0	0	0	0	0	0	0	0	0
acnaway	0	0	0	0	0	0	0	0	0	0
acne	0	0	0	0	0,18	0	0	0,20	0,21	0
acnederm	0	0	0	0	0	0	0	0	0	0
acnegenic	0	0	0	0	0	0	0	0	0	0
acneplus	0	0	0	0	0	0	0	0	0	0
acner	0	0	0	0	0	0	0	0	0	0,30

Next, label encoding is performed using MultiLabelBinarizer (Table 6), which converts each product category into a binary vector so that one product can have more than one label [20]. This approach is suitable for multilabel classification modeling, where each instance can be associated with multiple classes. After the encoding process, the dataset is divided into training data and testing data using several proportion scenarios, such as 80:20, 75:25, and 60:40, to evaluate the consistency of model performance across various data configurations. The features and labels should be an optimal format before modeling in the data mining stage.

TABLE 6. MULTILABEL BINARIZER

Product Name	cream and lotion	facial_wash	make up removal	moisturizer	scrub and exfoliation	serum	sunscreen	toner	unknown
cod glad glow pomegranate niacinamide brightening moisturizer pelembab wajah whitening glowing moisture gel day night cream skincare g g pelembab wajah pemutih wajah mencerahkan	0	0	0	0	0	1	0	0	0
dapat pcs serum glowing memutihkan cepat super ampuh mr beauty care	0	0	0	1	0	0	0	0	0
day gold paket glowing series daviena skincare normal kering kusam skincare kit perawatan kulit berminyak mencerahkan	0	0	0	1	0	0	0	0	0

D. Data Mining

At the data mining stage, a modeling approach is to perform multilabel classification on text describing skincare products. For a product, there may be more than one category; therefore, a multilabel model is selected [21]. It is a widely used approach for solving problems with complex label structures [22]. Three problem transformation methods are employed in this study, i.e., Binary Relevance, Classifier Chain, and Label Powerset. Each multilabel approach is applied independently using four base classifiers (RF, LR, SVM, KNN). These algorithms were chosen because they have shown promising performance in document classification, particularly with the TF-IDF feature representation. Moreover, using GridSearchCV, a parameter search is performed to generate combinations of these hyperparameters exhaustively, and the best-performing models are selected.

E. Interpretation

The interpret phase in the KDD process is the final stage, in which meaning is assigned to results from modeling. In this study, the output of the multilabel classification model is used as input for association analysis using the Apriori algorithm. Apriori is an algorithm for finding frequent patterns (frequent itemsets) in a database that meet a specific minimum support threshold [23]. Finally, information on item frequency is used to perform transaction scanning to identify products that might appear and satisfy the minimum support and confidence thresholds [24]. The resulting association rules also yield insights into the most frequently co-occurring labels, which could shed light on the characteristics of skincare products or suggest product category development.

III. PROPOSED METHOD

Figure 3 presents the method proposed in this research for processing skincare product name data to obtain relevant information for further analysis. There are three main stages in this approach: text preprocessing, multilabel classification, and Apriori algorithm-based association analysis. Data preprocessing cleans and normalizes the data for further use. Then, a multilabel classification model is employed to predict product categories from cleaned product names. These prediction results are then binarized and form the basis for further association analysis to uncover patterns of relationships among categories.

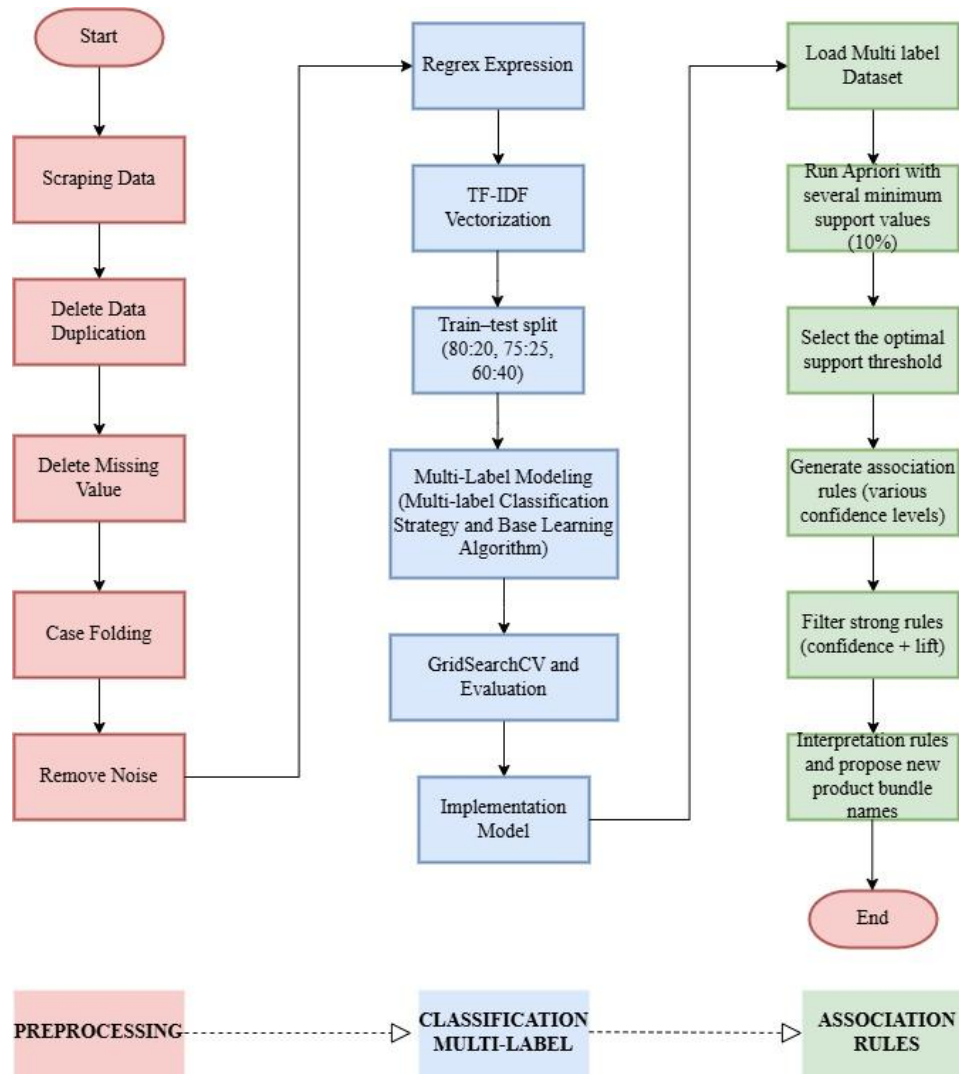


Figure 3. Proposed Research Framework

### A. Multilabel Classification

Multilabel classification is applied after preprocessing and the generation of preliminary labels from the data using regular expressions. The cleaned product names are converted to numerical representations using the Term Frequency-Inverse Document Frequency (TF-IDF) technique, which helps construct numerical inputs for ML models [25]. The transformed output from this step is given to the ML process. Then, using several splitting ratios (80:20, 75:25, and 60:40), the dataset is split into train and test sets to examine the model’s robustness to different proportions of train and test data. The classification phase (Figure 4) is performed by using three popular multilabel techniques (Binary Relevance, Classifier Chain, and Label Powerset). All of these are formulated to convert the multilabel problem into some forms that can be predicted by traditional classification model with different transformation approaches. These strategies are coupled with four base learners, RF, LR, SVM, and KNN. The application of multilabel strategies for each base algorithm yields with many candidate models must be evaluated.

GridSearchCV as the method of hyperparameter optimization to reach its best configuration considering each model combinations. This procedure systematically examines a variety of values for the parameters and applies cross-validation to identify an optimal estimator for each. The best model selected is evaluated through the F1-score (micro) metric on the train and the test data, therefore, it’s performance can be objectively measured in a standardized manner.

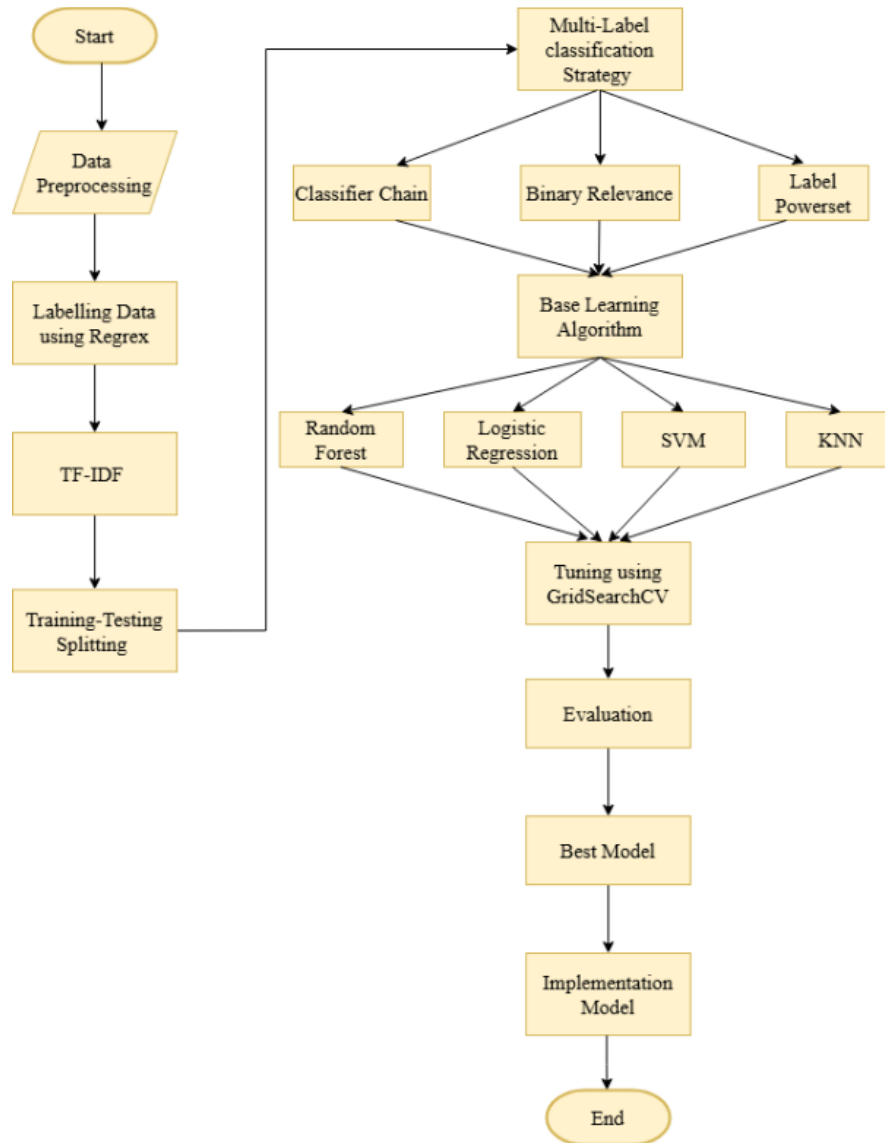


Figure 4. Workflow Classification Multi-label

### B. Association Rules

The association analysis is performed by loading the multilabel dataset, which consist of products and their categories that generated with the multilabel classification process. Then the data is preprocessed, through text cleaning, handling the missing values in data, and extraction of the suitable multilabel categories required for analysis. The multilabel data is then transformed in basket format for applying the Apriori algorithm, making each item a combination of categories with binary values. When the data are processed, we then run the Apriori algorithm using different minimum support levels (10%, 20% and 30%) to find out frequent itemsets, what items, or categories of items commonly combines together. From these tests, the minimum support value that produces the most informative patterns is selected as the best configuration.

Figure 5 shows the next stage, the formation of association rules from the frequent itemsets using several minimum confidence values, such as 50%, 75%, 80%, and 85%. The confidence and lift values are used as the basis for filtering rules that are truly strong and significant. The rules that meet these criteria are then visualized through a network graph to show the relationships between product categories, as well as a heatmap to illustrate the strength of the association between the antecedent and the consequent. After the visualizations are obtained, the rules are interpreted to understand the patterns of relationships among skincare categories. Based on the influence of significant association rules, new product bundling names are recommended, and practical promotion ideas for name design or marketing strategy will be provided. The latter step is completed when the recommendations for bundling are derived from the most significant analysis results.

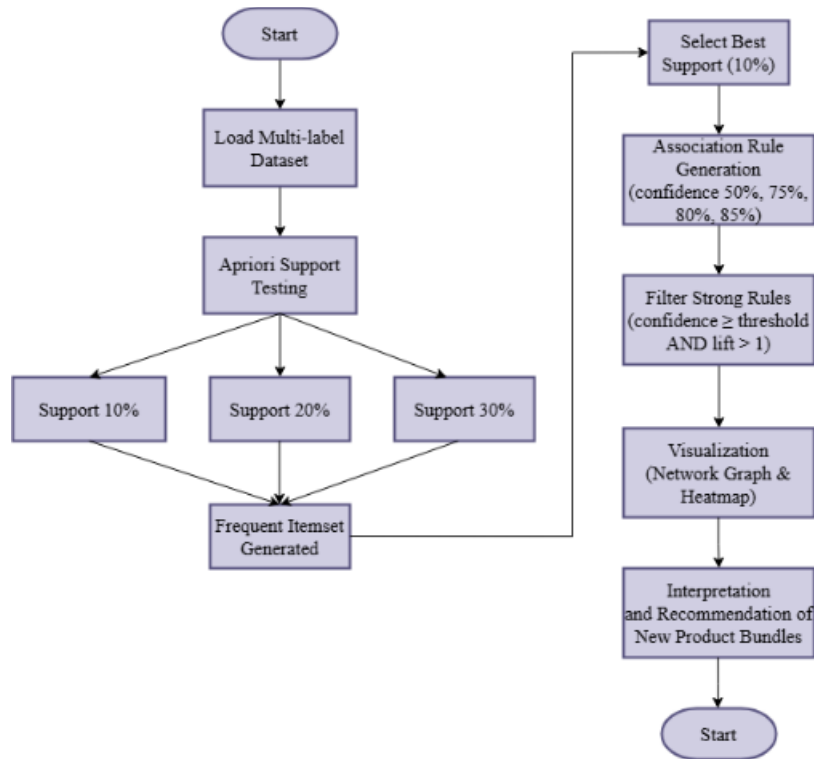


Figure 5. Workflow Association Rules

#### IV. RESULT AND DISCUSSION

##### A. Multilabel Classification

Table 7 presents the evaluation results for Scenario 1 with a 60:40 train–test split. Overall, Binary Relevance is shown to be both the most stable and the best-performing across all the wrappers studied, especially when combined with SVM and RF Classifiers. SVM achieves the best F1-test score of 0.8911 under Binary Relevance, and RF obtains close performance with an F1-test score of 0.8805. LR and KNN also achieve F1-test scores of 0.8641 and 0.6137, respectively. However, KNN's performance is not very good due to its sensitivity to high-dimensional TF-IDF features.

TABLE 7. EVALUATION RESULTS OF SCENARIO 1 (60:40)

60:40			
Wrapper	Base Model	F1 Train	F1 Test
Binary Relevance	K-Nearest Neighbor	0.712135	0.613744
	Logistic Regression	0.999236	0.864088
	Random Forest	0.998471	0.880562
	Support Vector Machine	1.000000	0.891134
Classifier Chain	K-Nearest Neighbor	0.631760	0.587264
	Logistic Regression	0.999236	0.847007
	Random Forest	0.998471	0.859155
	Support Vector Machine	1.000000	0.883616
Label Powerset	K-Nearest Neighbor	0.550984	0.502966
	Logistic Regression	0.984030	0.731092
	Random Forest	1.000000	0.768700
	Support Vector Machine	0.987842	0.701092

The performance of the classifier chain wrapper retains the general trends observed with Binary Relevance but is predominantly lower. The best-performing model based on the F1-test score is SVM (0.8836), followed by RF (0.8592), and LR shows the lowest performance (0.8470). KNN achieves the worst generalization performance (0.5873), indicating that the dependency modeling within Classifier Chain remains insufficient to improve performance. Label Powerset delivers the worst

results as a wrapper. RF and SVM keep relatively high F1-train scores (around or above 0.98), but suffer a decline in their F1-test scores to 0.7687 and 0.7011, respectively. LR also again a significant decrease (0.7311). This degradation reveals that Label Powerset is more affected by label sparsity and class imbalance, which are often found in many-label datasets corresponding to product names. KNN performs worst again with Label Powerset, scoring 0.503, confirming its unsuitability for this problem. In general, Scenario 1 indicates that Binary Relevance with an SVM achieves the best performance across all configurations, followed by RF. These results confirm that a set of independent binary classifiers for each label performs better than either label propagation (Classifier Chain) or learning full joint label combinations (Label Powerset) on this particular data. The results also highlight the appropriateness of classifiers based on margins and trees for high-dimensional, sparse text features, rather than distance-based methods such as KNN.

TABLE 8. EVALUATION RESULTS OF SCENARIO 2 (75:25)

75:25			
Wrapper	Base Model	F1 Train	F1 Test
Binary Relevance	K-Nearest Neighbor	0.713791	0.657895
	Logistic Regression	1.000000	0.887348
	Random Forest	0.998189	0.897059
	Support Vector Machine	1.000000	0.930728
Classifier Chain	K-Nearest Neighbor	0.690695	0.629630
	Logistic Regression	0.998794	0.879725
	Random Forest	1.000000	0.914077
	Support Vector Machine	1.000000	0.921708
Label Powerset	K-Nearest Neighbor	0.591922	0.525692
	Logistic Regression	0.982036	0.758974
	Random Forest	0.998187	0.781982
	Support Vector Machine	1.000000	0.775000

Table 8 shows the evaluation results for Scenario 2 with a train–test split of 75:25. Such as the Scenario 1, the Binary Relevance wrapper remains as the one achieving better results. Under this envelope, SVM holds the best F1-test score at 0.9307 with excellent generalization ability when more data for training is available. RF is a close second with an F1-test of 0.8971, followed by LR with 0.8873. On the other hand, K Nearest Neighbor (KNN) is referred as the weakest classifier again (0.6579) and it is suffered from sensitivity to high dimensional feature vectors of TF-IDF and lack of capability to capture boundaries of text related features.

The Classifier Chain wrapper gives a similar performance, and RF maintains its Binary Relevance F1-test result (0.9141) competitive on the two meta-features as well. SVM also gives a satisfactory result of 0.9217, which again indicates the strength of margin-based classifiers in multi-label text domain. LR has F1-test score of 0.8797 that is not scenario specific. KNN also lags with F1-test score of 0.6296, once more validating that dependency modelling in Classifier Chain is not sufficiently offset to the limitations that KNN faces on sparse text spaces. Label Powerset yields the worst performance in general among the three wrappers, except when using a few algorithms. Within this wrapper, RF has the best F1-test score (0.7819), followed by SVM (0.7750), and LR (0.7589). Performance drop is still not negligible from Binary Relevance and Classifier Chain, which can mainly be attributed to the higher label combination sparsity characterised by the Label Powerset transformation. KNN has the lowest score of 0.5257, which again confirms its invalidity for multi-label text classification with TF-IDF features. The findings based on the f1-test scores in Scenario 2 confirm that Binary Relevance along with SVM always results in achieving superior predictive performance (obtaining an F1-test score = 0.9307, being the maximum over all models analyzed within this setting). RF also remains highly competitive across wrappers, with an increasing ratio of training data from Scenario 1 to Scenario 2, the F1-test score improvement for SVM and RF is quite significant, which indicates that high-dimensional text classification can still benefit from having a lot of training samples. It can be concluded that the margin-based and ensemble-based classification methods are the most promising for multi-label product name classifier, whereas Label Powerset is still not effective in such transformation due to its requirements on sparsity of labels.

Building on the performance trends observed in Scenario 2, Scenario 3 delves deeper into how increasing the training sample ratio to 80% affects the solidity capability, normalized across models under three wrappers. The evaluation results for Scenario 3 (with the highest training ratio among the three scenarios) using an 80:20 train–test split are shown in Table 9. As previously found, the Binary Relevance wrapper yields the best overall results, especially with SVM and RF Classifiers. For Binary Relevance, SVM achieves the highest F1-test score of 0.9255. However, as we will see, the generalization is excellent even for larger training set sizes. RF is also a strong performer (0.9128), with LR achieving an F1 score of 0.8830 afterward. The KNN, meanwhile, produces the poorest performance (0.6506) because it fails to exploit semantic information in TF-IDF features.

The Classifier Chain wrapper performs well, with the RF achieving an F1-test score of 0.9206 and the SVM taking second place with 0.9193. These findings indicate that aggregating dependencies across labels in Classifier Chain provides no

significant improvement over Binary Relevance for high-performing models. Notably, LR maintains competitive performance (0.8811) and, in the case of KNN, still underperforms (0.6238), demonstrating poor performance in both setups and across all wrappers en todos los escenarios y involucrados. Label Powerset performs the worst overall, as reported earlier. Although RF and SVM still achieve reasonable scores (0.8219, 0.7735 on the F1-test), they are still much lower than what we have seen under Binary Relevance and Classifier Chain. LR yields an F1-score of 0.7666, whereas KNN is once again the worst (0.4851). The poor performance of Label Powerset further confirms that it is particularly susceptible to label sparsity. The transformation generates many ungeneralizable label combinations in small sample sets. Overall, Scenario 3 supports the previous conclusion that BR combined with SVM is the best model, achieving the highest F1-test score across all scenarios (0.9255). RF is also consistently strong across wrappers and train–test splits. Top-performing algorithms demonstrate a second, moderate absolute increase in F1-test score when the proportion of training data is increased from 60%, 75%, to 80%, with the increase being significantly greater for SVM and RF. These results emphasize the need for a sufficient sample size in multi-label text classification and show that margin-based (SVM) and ensemble methods (RF) are distinctly better than distance-based algorithms (KNN) and better than the Label Powerset transformation.

TABLE 9. EVALUATION RESULTS OF SCENARIO 3 (80:20)

80:20			
Wrapper	Base Model	F1 Train	F1 Test
Binary Relevance	K-Nearest Neighbor	0.724490	0.650602
	Logistic Regression	1.000000	0.883002
	Random Forest	0.998878	0.912844
	Support Vector Machine	1.000000	0.925508
Classifier Chain	K-Nearest Neighbor	0.686513	0.623810
	Logistic Regression	0.998878	0.881057
	Random Forest	1.000000	0.920635
	Support Vector Machine	1.000000	0.919283
Label Powerset	K-Nearest Neighbor	0.597353	0.485149
	Logistic Regression	0.982730	0.766595
	Random Forest	0.997753	0.821918
	Support Vector Machine	1.000000	0.773455

Across all three experimental settings, we observe a general trend in the efficacy of multilabel classification wrappers and base models. We generally observe that Binary Relevance consistently exhibits the best generalization capabilities. At the same time, SVM has the highest F1-test scores across all train–test set settings (from 0.8913 (60:40), to 0.9307 (75:25), and 0.9255 (80:20)). RF also shows stable across all cases, always being the second best performer. When treating label dependencies with the Classifier Chain, results are still on par, especially for SVM and RF, but slightly lower than when using Binary Relevance, which suggests that explicit modeling of dependencies has little added value on this dataset. In contrast, Label Powerset consistently performs poorly due to the sparsity introduced by a large number of label combinations. KNN has the lowest generalization in all cases and exhibits larger drops than others.

Viewed collectively, the progressive increase in training data from Scenario 1 to Scenario 3 leads to modest but consistent improvements in the performance of top-performing models, reaffirming the importance of larger training samples in high-dimensional TF-IDF feature spaces. These findings highlight the superiority of margin-based learners (SVM) and ensemble-based methods (RF) for multi-label skincare product name classification, while also demonstrating the limitations of distance-based algorithms and the Label Powerset transformation under sparse label conditions. Overall, the three scenarios confirm that Binary Relevance with SVM provides the most reliable and stable approach for multi-label text classification in this domain.

TABLE 10. HYPERPARAMETER TUNING

Model	Parameter	Value 1	Value 2	Value 3
Support Vector Machine	C	0.1	1	10
Random Forest	n_estimators	100	200	300
	max_depth	None	10	20
	min_samples_split	2	5	10
Logistic Regression	C	0.1	1	10
K-Nearest Neighbor	n_neighbors	3	5	7

In addition to model evaluation, Table 10 shows a process of searching for the best hyperparameters was also carried out. From the tuning results table, it can be seen that optimum SVM’s performance is seen at C = 10 and RF doesn’t seem to suffer from overfitting no matter how I tune n\_estimators (100 – 300), max\_depth (None – 20) and min\_samples\_split (2 – 5). LR’s best performance also occurs at C = 10 with the KNN performer optimally about k = 5 or k = 7. These findings are consistent

with the theory that regularized models achieve high performance when their complexity is adjusted according to the characteristics of the data.

Figure 6 shows a pattern that is consistent with the evaluation table. The graph of the average F1-Test value per wrapper confirms that Binary Relevance is the best overall approach, followed by Classifier Chain and finally Label Powerset. The base classifier's average graph, on the other hand, presents that RF is the top performing BT algorithm followed by SVM and LR while KNN always ends up as underperforming model. These findings further confirm that algorithms with strong generalization ability and appropriately model complexity could be more applicable to the text-based multilabel tasks. Overall, the best combination based on all evaluation results is Binary Relevance with SVM, which consistently produces the highest F1-Test across the three testing scenarios. This model is then used as the final model for the implementation process and becomes the basis for generating prediction labels on the complete dataset before entering the Apriori association analysis stage.

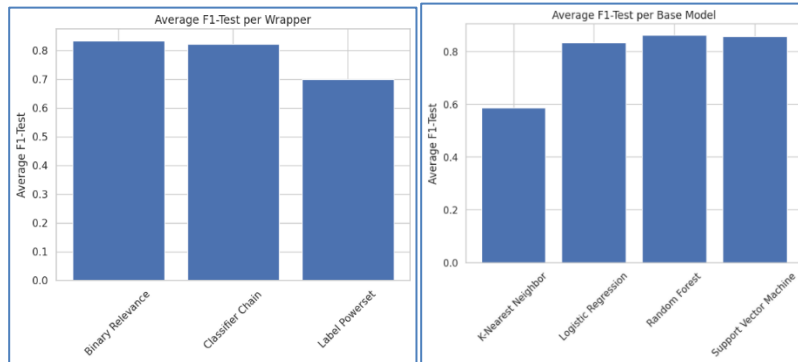


Figure. 6. Average Wrapper and Base Model

### B. Association Rules

The association analysis results presented in Table 11 (R1–R8) show patterns of relationships between skincare product labels based on their co-occurrence frequency within a single product name, obtained from the predicted results of the multilabel model. The association rules describe product categories that frequently appear together in skincare product names on the marketplace. Each rule consists of an antecedent (a combination of labels that appear together in a product), a consequent (another label that tends to co-occur within the same product), as well as the support, confidence, and lift values that reflect the strength of the relationships between labels in the context of co-occurrence within product name texts.

TABLE 11. ASSOCIATION RULES RESULTS

Rules	Antecedent	Consequents	Support	Confidence	Lift
R1	(cream and lotion)	(facial_wash)	0.206061	0.990291	2.738515
R2	(moisturizer, cream and lotion)	(facial_wash)	0.171717	0.988372	2.733208
R3	(serum, facial wash, toner)	(moisturizer)	0.101010	0.980392	1.628504
R4	(facial wash, toner)	(moisturizer)	0.127273	0.969231	1.609964
R5	(serum, toner)	(moisturizer)	0.119192	0.951613	1.580699
R6	(serum, facial wash)	(moisturizer)	0.169697	0.943820	1.567755
R7	(sunscreen, facial wash)	(moisturizer)	0.123232	0.924242	1.535235
R8	(toner)	(moisturizer)	0.171717	0.885417	1.470742

Table 11 shows the label combinations that most frequently appear together in a single product entry. Rules R1 and R2, for example, indicate that the labels cream & lotion as well as the combination of moisturizer, cream, and lotion have a strong tendency to be followed by the label facial\_wash within a single product name. The very high confidence values (0.99 and 0.98) and lift values above 2.7 indicate that these labels are indeed very frequently found together within a series of skincare product descriptions sold in the marketplace. This pattern does not describe the order of product usage by consumers, but rather shows the tendency that producers include these terms together in a single product package or certain skincare set.

Rules R3 to R7 show that label combinations such as serum, facial\_wash, and toner tend to appear together with the label moisturizer. This indicates that many skincare products marketed as a “complete care set” contain these terms within the product name. With lift values above 1.5, this association reflects product packaging patterns or naming strategies commonly used by sellers in e-commerce to attract consumers. Rule R7 also shows that the labels sunscreen and facial\_wash often appear together with moisturizer, which again reinforces the tendency of sellers to offer several products at once in a single package, rather than individual purchasing behavior.

Rule R8, namely the relationship between toner and moisturizer, has a lower confidence (0.885) but still has a lift above 1, indicating a sufficiently strong association in the context of product naming. Many skincare products are packaged as a toner and moisturizer set, causing this relationship to appear consistently in the dataset. From all the rules, it is evident that the label moisturizer is the category that most frequently appears as the consequent. This indicates that moisturizer is an element that is often included in the series of skincare product names marketed in bundling or multi-product packages.

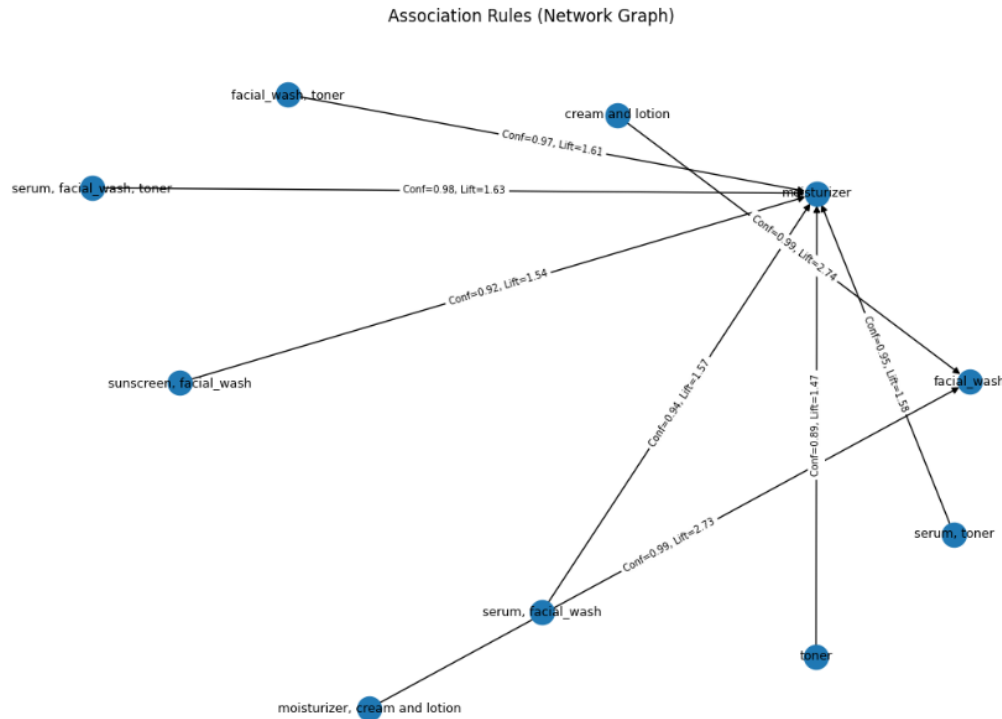


Figure 7. Association Rules (Network Graphh)

The network graph in Figure7 provides a more easily understandable visual depiction of the relationships between product labels. Nodes that are connected to each other indicate categories that frequently appear together in a single product name. It is observed that moisturizer has the highest number of connections, in accordance with the table results showing that this category often appears as the final part of the label combinations. This indicates that moisturizer is a component commonly added in product bundling. Important relationships such as the combination of cream & lotion with facial wash as well as serum with facial wash and moisturizer are also clearly visible. This pattern shows groups of categories that are frequently used together in the naming or packaging of skincare products. Overall, the network graph supports the table analysis results and provides a concise overview of the main relationships between labels.

Based on the strongest association rules, the bundling recommendations are combined according to the fitting several main themes (Table 12). The product variants this time come from the Daily Radiance line of products, a skin brightening and everyday care themed combination of facial wash cream & lotion, and moisturizer. The Glow Hydration Theme is all about the Tandem of Moisturizer with Serum and simply adding the freebies like toner or face wash to boost the hydration. The UV Protection series brings the elements of moisturizer and sunscreen together to form a complete "sun protection package, while Clean & Fresh bring facial wash, moisturizer, and/or serum and toner indicating emphasis on cleansing and skin freshening functions.

TABLE 12. PRODUCT NAME RECOMMENDATIONS

Theme	Bundling Name	Bundling contents
Daily Radiance	Daily Radiance Duo	Facial Wash & Moisturizer
	Daily Radiance Duo	Cream & Lotion & Facial Wash
	Daily Radiance Duo	Moisturizer & Toner
	Daily Radiance Duo	Cream & Lotion & Moisturizer
	Daily Radiance Set	Cream & Lotion & Facial Wash & Moisturizer
Glow Hydration	Glow Hydration Duo	Moisturizer & Serum
	Glow Hydration Set	Facial Wash & Moisturizer & Serum
	Glow Hydration Set	Moisturizer & Serum & Toner
	Glow Hydration Bundle	Facial Wash & Moisturizer & Serum & Toner
UV Protection	UV Protection Duo	Moisturizer & Sunscreen
	UV Protection Set	Facial Wash & Moisturizer & Sunscreen
Clean & Fresh	Clean & Fresh Set	Facial Wash & Moisturizer & Toner
	Clean & Fresh Set	Facial Wash & Serum & Toner

## V. CONCLUSION

This paper demonstrates the usefulness of integrating multilabel classification and association rule mining to handle unstructured skincare product names while consolidating structured knowledge. Across all three train–test split scenarios, Binary Relevance consistently achieved better classification performance, with Support Vector Machine achieving the highest F1-test score of 0.925 on an 80:20 split. K-Nearest Neighbour resulted in lower performance, while other F1-test scores ranged from 0.883 to 0.912 using Random Forest and Logistic Regression. Classifier Chain achieved competitive F1-test scores, albeit with some degradation, while Label Powerset yielded relatively low F1-scores owing to the unmanageable number of possible label subsets. An Apriori-based association rule analysis also confirmed these results, revealing strong co-occurrences among skincare categories. Perhaps more interestingly, the rules for the moisturizer category showed high confidence (0.88–0.99) and a lift above 1.5, suggesting an important role for moisturizers in product bundling in the marketplace. Based on these patterns, practical naming guidelines for four themes: Daily Radiance, Glow Hydration, UV Protection, and Clean & Fresh were developed. In summary, the combination of high-performing multilabel classification (with F1-test score up to 0.925) combined with strong association rules provides a powerful data-driven approach for advancing product naming in e-commerce. In future work, more features, such as price, ratings, aesthetic attributes, or embedding-based text representations, could be added to improve the model’s performance and semantic recognition.

## REFERENCES

- [1] M. Viu-Roig and E. J. Alvarez-Palau, “The Impact of E-Commerce-Related Last-Mile Logistics on Cities: A Systematic Literature Review,” *Sustainability*, vol. 12, no. 16, Aug. 2020, doi: 10.3390/su12166492.
- [2] H. J. Bermudez-Sosa, J. Olarte-Henao, and S. Rojas-Berrio, “Sentiment and Emotion Analysis From Textual Information: A Systematic Literature Review,” *Journal of Information Science*, Jul. 2025, doi: 10.1177/01655515251353170.
- [3] O. Ouda, E. AbdelMaksoud, A. A. Abd El-Aziz, and M. Elmogy, “Multiple Ocular Disease Diagnosis Using Fundus Images Based on Multilabel Deep Learning Classification,” *Electronics*, vol. 11, no. 13, Jul. 2022, doi: 10.3390/electronics11131966.
- [4] S. N. Alsubari, S. N. Deshmukh, A. A. Alqarni, N. Alsharif, T. H. H. Aldhyani, F. W. Alsaade, and O. I. Khalaf, “Data Analytics for the Identification of Fake Reviews Using Supervised Learning,” *Computers, Materials & Continua*, vol. 70, no. 2, doi: 10.32604/cmc.2022.019625.
- [5] M. AbdelHamid, A. Jafar, and Y. Rahal, “Levantine Hate Speech Detection in Twitter,” *Social Network Analysis and Mining*, vol. 12, no. 1, Dec. 2022, doi: 10.1007/s13278-022-00950-4.
- [6] Q. A. Hidayaturohman and E. Hanada, “A Comparative Analysis Of Hyper-Parameter Optimization Methods For Predicting Heart Failure Outcomes,” *Applied Sciences*, vol. 15, no. 6, 2025, doi: 10.3390/app15063393.
- [7] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A.-L. Boulesteix, D. Deng, and M. Lindauer, “Hyperparameter Optimization: Foundations, Algorithms, Best Practices, And Open Challenges,” *WIREs Data Mining and Knowledge Discovery*, vol. 13, no. 2, Apr. 2023, doi: 10.1002/widm.1484.
- [8] J. Llivisaca and J. Avilés-González, “Customer Segmentation In Food Retail Sector: An Approach From Customer Behavior And Product Association Rules,” in *Applied Technologies. ICAT 2022*, M. Botto-Tobar, M. Zambrano Vizuete, S. Montes León, P. Torres-Carrión, and B. Durakovic, Eds. Cham: Springer, vol. 1755, pp240-254, 2023, doi: 10.1007/978-3-031-24985-3\_18.
- [9] O. F. Althwaynee, A. Aydda, I.-T. Hwang, Y.-K. Lee, S.-W. Kim, H.-J. Park, M.-S. Lee, and Y. Park, “Uncertainty Reduction Of Unlabeled Features In Landslide Inventory Using Machine Learning T-SNE Clustering And Data Mining Apriori Association Rule Algorithms,” *Applied Sciences*, vol. 11, no. 2, 2021. doi: 10.3390/app11020556.
- [10] N. Verma, D. Malhotra, and J. Singh, “Big Data Analytics For Retail Industry Using Mapreduce-Apriori Framework,” *Journal of Management Analytics*, vol. 7, no. 3, pp424-442, 2020, doi: 10.1080/23270012.2020.1728403.
- [11] N. Singhal, C. R. (Chhotu Ram), and H. S. Sirohi, “A Review on Knowledge Discovery from Databases,” in *Electronic Systems and Intelligent Computing*, Lecture Notes in Electrical Engineering, vol. 860, pp. 457–464, Jan. 2022, doi: 10.1007/978-981-16-9488-2\_43.

- [12] J. Bogatinovski, L. Todorovski, S. Džeroski, and D. Kocev, "Comprehensive Comparative Study of Multilabel Classification Methods," *Expert Systems with Applications*, vol. 203, Nov. 2022, doi: 10.1016/j.eswa.2022.117215.
- [13] I. D. Hunyadi, N. Constantinescu, and O.-A. Țicleanu, "Efficient Discovery of Association Rules in E-Commerce: Comparing Candidate Generation and Pattern Growth Techniques," *Applied Sciences*, vol. 15, no. 10, May 2025, doi: 10.3390/app15105498.
- [14] M. R. Setiawan, H. Pudjoprastyono, and N. Hariyana, "The Influence of Security, Convenience, and Customer Trust on Purchase Decision in Tokopedia Marketplace in Surabaya City," *Indonesian Interdisciplinary Journal of Sharia Economics (IJSE)*, vol. 8, no. 3, 2025, doi: 10.31538/ijse.v8i3.6866.
- [15] M. Arslan and C. Cruz, "Business Text Classification with Imbalanced Data and Moderately Large Label Spaces for Digital Transformation," *Applied Network Science*, vol. 9, no. 11, 2024, doi: 10.1007/s41109-024-00623-5.
- [16] A. Akundi, D. Euresiti, S. Luna, W. Ankobiah, A. Lopes, and I. Edinbarough, "State of Industry 5.0—Analysis and Identification of Current Research Trends," *Appl. Syst. Innov.*, vol. 5, no. 1, Feb. 2022, doi: 10.3390/asi5010027.
- [17] M. A. Palomino and F. Aider "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis", *Applied Sciences*, vol. 12, no. 17, 2022, doi: 10.3390/app12178765.
- [18] P. Sharma and B. Nagpal, "Regex: An Experimental Approach For Searching In Cyber Forensic," *Int. J. Inf. Technol.*, vol. 12, no. 2, pp339-343, 2020, doi: 10.1007/s41870-019-00401-y.
- [19] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. M. U. Din, "Machine Learning Based Approaches for Detecting COVID-19 Using Clinical Text Data," *International Journal of Information Technology*, vol. 12, no. 3, pp731-739, Jun. 2020, doi: 10.1007/s41870-020-00495-9.
- [20] Z. Deng, W.-T. Chen, L. Chen, and P. S. Yu, "AE-smnMLC: Multilabel Classification with Semantic Matching and Negative Label Sampling for Product Attribute Value Extraction," *arXiv*, 2023, doi: 10.48550/arXiv.2310.07137.
- [21] W. Qian, J. Huang, F. Xu, W. Shu, and W. Ding, "A Survey on Multilabel Feature Selection from Perspectives of Label Fusion," *Information Fusion*, vol. 100, Dec. 2023, doi: 10.1016/j.inffus.2023.101948.
- [22] E. Deniz, H. Erbay, and M. Coşar, "Multilabel Classification of E-Commerce Customer Reviews via Machine Learning," *Axioms*, vol. 11, no. 9, Aug. 2022, doi: 10.3390/axioms11090436.
- [23] P-H. Lu, J-L. Keng, K-L Kuo, Y-F. Wang, Y-C. Tai, C-Y. Kuo, "An Apriori Algorithm-Based Association Rule Analysis to Identify Herb Combinations for Treating Uremic Pruritus Using Chinese Herbal Bath Therapy," *Evidence-Based Complementary and Alternative Medicine*, vol. 2020, no. 1, 2020, doi: 10.1155/2020/8854772.
- [24] Z. Zhao, Z. Jian, G. S. Gaba, R. Alroobaea, M. Masud, and S. Rubaiee, "An Improved Association Rule Mining Algorithm for Large Data," *Journal of Intelligent Systems*, vol. 30, pp. 750–762, 2021, doi: 10.1515/jisys-2020-0121.
- [25] A. Sharma, S. Rani, D. K. Sah, Z. Khan, and W. Boulila, "HOMLC—Hyperparameter Optimization for Multilabel Classification of Intrusion Detection Data for Internet of Things network," *Sensors*, vol. 23, no. 19, doi: 10.3390/s23198333.