Determining Packaging Design Concepts Using K-Means Cluster Genetic Algorithm Methods Through a Kansei Engineering Approach

Fahri Fadhillah[1], *Novi Purnama Sari*[1,*], *Wiwi Prastiwinarti*[1], *Adita Evalina Fitria Utami*[1]

[1] Jakarta State Polytechnic, Packaging Print Industry Technology, Indonesia
* novi.purnamasari@grafika.pnj.ac.id

Abstract. Indonesia was the fourth-largest cocoa producer in 2023. However, locally produced specialty chocolate bars have not been able to compete optimally with imported products. One main contributing factor is the lack of appealing and consumer-preferred packaging design. A packaging development approach based on Kansei Engineering (KE) is considered appropriate, as it enables the identification of consumers' emotional responses to create products aligned with their preferences. The packaging design concept plays a crucial role in the development process. Therefore, this study aims to determine the most optimal packaging design concept using the K-Means method. The Term Frequency–Inverse Document Frequency (TF-IDF) technique was employed to extract consumer emotional responses (Kansei words), while K-Means, K-Medoids, and K-Means Genetic Algorithm methods were applied to cluster the packaging design concepts. The TF-IDF results showed that the Kansei term "chocolate design" had the highest weight (0.5334). All three clustering methods produced three clusters; however, the K-Means Genetic Algorithm demonstrated the best performance, achieving the silhouette score (0.5264), Davies-Bouldin Index of (0.3844), and a Calinski-Harabasz Index of (23.726). Based on the K-Means Genetic Algorithm clustering, the selected design concept is "Usable." This concept will serve as a reference in determining the packaging design elements.

Keywords: Clustering, Kansei Engineering, K-Means GA, Design Concept, TFIDF.

1 Introduction

Cocoa is a key agricultural commodity in Indonesia's plantation sector and contributes substantially to the national economy. According to the International Cocoa Organization (2023), Indonesia is the fourth-largest cocoa producer in the world. Among various cocoa-based products, chocolate bars are especially favoured due to their practicality and ease of consumption (1). Based on observations, there are 31 chocolate bar brands available on the market, consisting of approximately 35 % specialty products and 65 % imported products. This is notable considering that Indonesia has several local chocolate bar producers, such as those from Aceh, Bali, and Lampung (2). However, current

M. A. Muflikhun et al. (eds.), *Proceedings of the 8th Mechanical and Industrial Engineering Symposium (MIE 2025)*, Atlantis Highlights in Engineering 42,
https://doi.org/10.2991/978-94-6239-687-6_21

specialty chocolate bar products still face challenges in presenting designs that align with consumer preferences. Therefore, an innovative approach to packaging design is essential to enhance the competitiveness of local chocolate bar products against imported brands that currently dominate the market.

Packaging not only serves as a means of protecting the product but also plays a crucial role in attracting consumer attention (3). Attractive and informative design can enhance perceived product quality and stimulate consumer purchasing behaviour (4). According to (5), designing packaging based on consumer preferences in determining the concept can support product innovation and competitiveness. The packaging design concept plays a crucial role as a guideline for designers in developing packaging that aligns with consumers' emotional preferences (6). Packaging design concepts are capable of representing a product's image by consumer preferences (6). This presents a challenge in determining the most appropriate concept based on consumer emotions.

Kansei Engineering is a consumer emotionbased design methodology aimed at understanding consumers' feelings and preferences toward a product (7). According to Johan et al. (8), emotional responses expressed through words (Kansei words) are used as parameters in product development.

Previous studies have stated that consumer emotions (Kansei words) are identified and translated into packaging design concepts, which are then further developed into specific design element specifications (9). Kansei Engineering is capable of uncovering the 'true meaning' within the context of international airport service design (10). Similarly (11), employed the Kansei Engineering method to develop packaging concepts such as "Attractive," "Informative," and "Eco-Friendly" for fried banana products. Meanwhile (12), applied the Kansei Engineering approach in determining the packaging design concept for rendang products by considering consumer emotions and cultural factors from two countries, namely Indonesia and Saudi Arabia. This study resulted in the idea of "Practical-Luxury," derived from the emotional responses elicited, aligned with the identified problems, and target market segments. Accordingly, Kansei Engineering has proven to be an effective strategic method in the development of packaging designs that are adaptive to consumers' emotional aspects.

The Term Frequency-Inverse Document Frequency (TF-IDF) method is employed in the initial stage to extract Kansei words that reflect consumer emotions. This method is used to assign weights to each word, serving as features in text analysis (13). The results of this extraction serve as the foundation for formulating a design concept that aligns with consumer preferences. Research by (14), demonstrated the capability of the TF-IDF method to extract Kansei words simply and efficiently, without the need for training data, while producing reliable and easily interpretable results in document processing. Similarly, (15) showed that the TF-IDF method can effectively extract Kansei words in a language-independent manner, achieving an accuracy rate of 91.3 % without requiring any training data.

According to (9), several clustering methods can be utilized to determine packaging design concepts based on Kansei words, such as Principal Component Analysis, Factor Analysis, and K-Means Clustering. (16), state that the K-Means method is widely used due to its ease of implementation, computational efficiency, and flexibility across various applications. This makes K-Means the first choice for large-scale data analysis.

According to (16), the clustering method using K-means consists of several variants, including K-Medoids, Fuzzy C-Means, X-Means, and K-Means++. Each Kmeans method has different capabilities. In this study, clustering of packaging design concepts is performed by applying K-means, K-Medoids, and K-means optimization with Genetic Algorithm (GA). According to (17), the K-Medoids method is a clustering algorithm similar to K-Means but is more resistant to outliers and noise, making it suitable for data with irregular structures. On the other hand, K-Means optimized with Genetic Algorithms (GA) effectively addresses the issue of empty clusters commonly encountered in standard KMeans, while also producing more stable and accurate cluster formations (18).

The aim of this research is to determine the most optimal packaging design concept using the Kmeans method. The method with the highest accuracy will be used as the basis for selecting the packaging concept for specialty chocolate bars. The resulting concept is expected to effectively represent consumer emotions and preferences, thereby enhancing the product's appeal and market value.

2 Methods

The stages of this research are shown in Fig. 1 below:

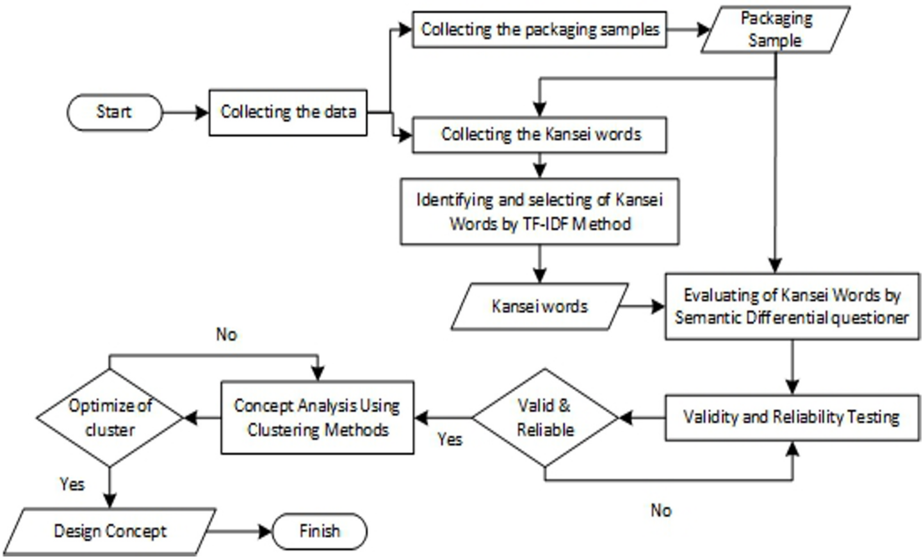

Fig. 1. Flowchart Research

2.1 Collecting the Packaging Samples

The packaging samples were collected from both the market and the internet to provide an overview of packaging design concepts. A minimum of 20-25 samples were collected (19). The selection of samples was based on the Segmenting, Targeting, and Positioning (STP) approach to ensure relevance and accuracy in targeting. The obtained samples will undergo a selection process based on the identification of design elements and significant differences between the samples (20).

2.2 Collecting the Kansei Word

Kansei words were collected through a Google Form questionnaire distributed to 30 loyal respondents, using direct product packaging samples as stimuli. This direct interaction allowed respondents to thoroughly evaluate the packaging and provide relevant emotional feedback. This approach aligns with the principles of Kansei Engineering, which emphasize the importance of emotional aspects in user experience (21).

2.3 Identifying Kansei Word by TF-IDF

The Kansei words obtained from the questionnaires were processed using the Term Frequency-Inverse Document Frequency (TF-IDF) method through Jupyter Notebook. The elimination of Kansei words was carried out by evaluating their importance based on the frequency of their occurrence across various questionnaires (22). Word with the highest frequency is obtained based on emotional consumers.

2.4 Evaluating of Kansei Word

The evaluation of Kansei words was conducted using a Semantic Differential questionnaire to understand respondents' perceptions of the packaging samples. The Semantic Differential method utilizes pairs of positive-negative antonyms on a seven-point scale ranging from -3 to 3. This questionnaire measures the correlation between the packaging samples and the selected Kansei words (23). A score of 3 indicates that the packaging sample strongly aligns with the corresponding Kansei word, whereas a score of -3 reflects the opposite. A previous study conducted by Kittidecha et al. (24), evaluated Kansei words in relation to various types of food wrapping materials using the Semantic Differential method on a seven-point scale, ranging from +3, indicating the highest level of congruence, to –3, indicating the lowest. The evaluation involved 30 respondents, referring to the approach used in previous research (25)(12). Respondent consistency was emphasized to ensure more accurate and targeted results.

2.5 Validity and Reliability Test

Validity and reliability testing were conducted to ensure that the Kansei words obtained through the questionnaire possess high levels of validity and reliability. Reliability was assessed by calculating the

Cronbach's Alpha coefficient, which is expected to exceed 0.60 to ensure data consistency and dependability (26). Thus, the implementation of validity and reliability testing guarantees that the validity and reliability testing guarantees that the collected data accurately represents consumer preferences, providing a strong foundation for developing products that align with market needs.

2.6 Determining Clusters of Kansei Words

The clustering process was carried out using data obtained from a Semantic Differential questionnaire, which had been tested for both validity and reliability. This study employed two types of clustering methods, namely:

a. K-Medoids
   The K-Medoids method is employed to cluster data based on specific characteristics, making it applicable in Kansei word analysis to identify design clusters that align with user preferences. K-Medoids is used to overcome the limitations of the K-Means method, which is sensitive to data outliers (27).
b. K-Means
   Optimize Genetic Algorithm The design concept is formulated through the integration of the K-Means clustering method and a Genetic Algorithm. The incorporation of the Genetic Algorithm into the K-Means process serves to optimize centroid positioning and enhance the overall quality of the clustering results (28). The GA process involves three main stages: selection, crossover, and mutation, which are applied to generate more optimal centroid candidates. Selection is performed to choose the best individuals based on their fitness values, ensuring that high-performing solutions are prioritized for reproduction; this method has been shown to effectively maintain population quality and prevent performance degradation over successive iterations (29). Crossover combines two parent individuals to produce new solutions, allowing for broader exploration of the solution space and increasing the likelihood of finding better centroids (30). Meanwhile, mutation introduces small random changes to individuals in order to preserve diversity and prevent premature convergence, as demonstrated in studies on adaptive mutation in GA–K-Means for unsupervised feature selection (31).

The conceptual results of these two methods will be compared with the clustering results obtained using the K-means clustering method. The determination of cluster concepts is carried out through discussions with experts. According to (32), concept determination involves a minimum of four experts in the field of design.

3 Results

3.1 Collection of Packaging Samples

A total of 48 specialty chocolate bar packaging samples, each with distinct shapes and designs, were collected from various sources including Pinterest, Instagram, online marketplaces, supermarkets, and offline chocolate product exhibitions. This number exceeds the minimum sample size recommended by Nagamachi, which ranges from 20 to 25 samples (23). The packaging samples are presented in Fig. 2.

Fig. 2. Packaging Sample

3.2 Collecting the Kansei Word

The process of collecting Kansei words aims to generate packaging designs that reflect consumers' perceptions and preferences toward a product. According to Nagamachi & Lockman, the number of collected Kansei words typically ranges from 50 to 100 words (33). Words with similar meanings are then eliminated to identify those that represent the same concept. A study by (34), involved 30 experienced respondents regarding the product in question to collect Kansei words. Stimulating consumers is essential to gain a deeper understanding of their emotions. According to (32), research explored consumer emotions by providing direct experiences, such as drinking tea and using cups, to the respondents. In this study, the selected participants were sweet food enthusiasts or frequent chocolate bar consumers residing in the JABODETABEK area.

3.3 Kansei Word Extraction with TF-IDF

The Kansei words were obtained through a questionnaire-based interview conducted via Google Forms. In the selection stage using the TF-IDF method, an initial pre-processing phase was carried out to enhance data quality prior to further analysis. This stage aimed to filter, clean, and optimize the word structure to produce more accurate and relevant analytical results (35). The pre-processing was performed using Python in Jupyter Notebook and involved several steps, including case folding, tokenizing, filtering, dan stemming.

a. Case Folding
Case folding is a text pre-processing step aimed at standardizing the writing format by converting all characters to lowercase (36). The results of this pre-processing step are presented in Table 1.

Table 1. Case Folding Process.

Input	Output
Not too sweet, Characteristic aroma of chocolate is more emphasized, Easy to bite texture, looks like chocolate in general, Simple shape and easy to use	*not too sweet, the distinctive aroma of chocolate is emphasized, the texture is easy to bite, the appearance is like chocolate in general, the shape is simple and easy to use,*

b. Tokenizing
The data for the tokenization process is obtained from the case folding results, where the text has been converted to lowercase. Tokenization is the process of breaking down the text into smaller units, such as words or phrases, known as tokens. The results of the tokenization are presented in Table 2.

Table 2. Tokenizing Process.

Input	Output
not too sweet, the distinctive aroma of chocolate is emphasized, the texture is easy to bite, the appearance is like chocolate in general, the shape is simple and easy to use,	*['not', 'too', 'sweet', 'aroma', 'typical', 'chocolate', 'more', 'highlighted', 'texture', 'easy', 'bite', 'look', 'like', 'chocolate', 'on', 'generally', 'shape', 'simple', 'and', 'easy', 'use']*

c. Filtering
Filtering is a text processing technique aimed at removing irrelevant elements, including stop words, to enhance the quality of analysis and accuracy in applications such as text classification (37). The results of the tokenization process, which are translated into consumer expectations regarding specialty chocolate bar packaging, are presented in Table 3.

Table 3. Filtering Process.

Input	Output
[‘not’, ‘too’, ‘sweet’, ‘aroma’, ‘typical’, ‘chocolate’, ‘more’, ‘highlighted’, ‘texture’, ‘easy’, ‘bite’, ‘look’, ‘like’, ‘chocolate’, ‘on’, ‘generally’, ‘shape’, ‘simple’, ‘and’, ‘easy’, ‘use’]	*[‘sweet’, ‘aroma of chocolate’, ‘texture easy to bite’, ‘appearance of chocolate’, ‘simple’, ‘simple’]*

d. Stemming
Stemming is the process of transforming a word into its root form by removing prefixes or suffixes, with the aim of simplifying the variations of a word so that they are recognized as the same word. The results of the stemming process are presented in Table 4.

Table 4. Stemming Process.

Input	Output
[‘sweet’, ‘distinctive of chocolate’, ‘texture easy to bite’, ‘appearance of chocolate’, ‘simple’, ‘simple’]	*[‘sweet’, ‘distinctive aroma of chocolate’, ‘biteable texture’, ‘brown appearance’, ‘simple’, ‘simple’]*

After the pre-processing stage, the Kansei words derived from consumers' emotional responses to the specialty chocolate bar packaging were collected, resulting in 22 terms, as shown in Table 5. The most frequently occurring Kansei words were “Chocolate Design”, “Informative”, “Eye Catching”, “Usable”, “Easy to store”, and “Colourful”. Respondents expressed a preference for packaging with a “Chocolate Design” style, which allows them to see the shape of the chocolate inside before purchasing, serving as an important consideration in their buying decision. The term “Informative' reflects consumers” desire for information about the chocolate’s origin, to explore the diverse flavours of Indonesian chocolate. The phrase “Eye Catching” represents consumers preference for packaging with a balanced composition of colour, typography, and illustrations that can attract their attention. “Usable” refers to consumers need for packaging that is easy to open and can be resealed to maintain the product's quality after opening. Meanwhile, “Easy to store” indicates a preference for practical and portable packaging. Lastly, “Colourful” describes a preference for bright colours on the chocolate bar packaging, as chocolate is a food that can enhance mood.

3.4 Kansei Word of Evaluation with Sample

The evaluation of the Kansei word is conducted by constructing a questionnaire. The Kansei word is placed on the right side of the scale, while its negation is placed on the left. Respondents are asked to express their opinions using the scale, ranging from “strongly agree” to “strongly disagree” (21). Below is an example of the Semantic Differential questionnaire, in Fig. 3.

Antonim Kata Kansei	-3 Sangat Setuju	-2 Setuju	-1 Cukup Setuju	0 Netral	1 Cukup Setuju	2 Setuju	3 Sangat Setuju	Kata Kansei
Un-informative	☐	☐	☐	☐	☐	☐	☐	Informative
Not Eye-catching	☐	☐	☐	☐	☐	☐	☐	Eyecatching
Un-Useable	☐	☐	☐	☐	☐	☐	☐	Useable
				...				
Non-Warm color	☐	☐	☐	☐	☐	☐	☐	Warm color
Non-Modern Desain	☐	☐	☐	☐	☐	☐	☐	Modern Desain
Non-Chocolate Desain	☐	☐	☐	☐	☐	☐	☐	Chocolate Desain

Fig. 3. Semantic Differential Questionnaire.

A total of 22 Kansei words were used to evaluate 48 packaging samples, as shown in Table 1, in order to assess consumers' emotional perceptions. This evaluation process was conducted by 30 loyal specialty chocolate respondents. The collected data were then analysed using validity and reliability tests to ensure the consistency and accuracy of the measurement results.

3.5 Validity and Reliability Test

The relationship between Kansei words and packaging design was evaluated for validity using IBM SPSS software. In the initial stage, validity testing was conducted on 22 Kansei words with a significance level of 5 % and degrees of freedom (df = N). The criteria for valid data were determined based on the calculated R value being greater than the table R value (38). The results of the initial validity test are presented in Table 5 below.

Based on the validity test results presented in Table 5, 11 Kansei Words were found to be invalid as their calculated R values were lower than the table R value. Among the 11 eliminated Kansei Words, two words with the highest weights, “Chocolate Design” and “Informative”, were also deemed invalid based on the validity test results. This occurred because the results of the first semantic differential questionnaire indicated that respondents tended to assign more negative values than positive ones when evaluating the words "Chocolate Design" and "Informative" in relation to the packaging samples. As a result, the validity test for these two words yielded invalid outcomes. The assessment results for "Chocolate Design" and "Informative" in the first semantic differential questionnaire are presented in Table 6.

The results of the second validity test are consistent with the first stage, where 11 out of 22 Kansei Words were deemed valid. Accordingly, the 11 validated Kansei words presented in Table 7 were utilized to define the design concept through the clustering method.

Table 5. Results of the First Validity Test

No	Kansei Words	TF-IDF Weight		*Correlation*	
		Weight	Table R	Calculated R	Description
1.	Chocolate Desain	0.5334	0.374	-0.144	Invalid
2.	Informative	0.3754	0.374	0.354	Invalid
3.	Eye-catching	0.3359	0.374	0.700	Valid
4.	Useable	0.2371	0.374	0.522	Valid
5.	Easy to Store	0.2173	0.374	0.468	Valid
6.	Colorful	0.1976	0.374	0.381	Valid
7.	Safety	0.1778	0.374	0.537	Valid
8.	Bright Color	0.1778	0.374	0.136	Invalid
9.	Simple	0.1185	0.374	-0.129	Invalid
10.	Sturdy Packaging	0.1185	0.374	0.566	Valid
11.	Easy to Open	0.0790	0.374	0.281	Invalid
12.	Premium	0.0790	0.374	0.693	Valid
13.	Elegant	0.0593	0.374	0.638	Valid
14.	Exclusive	0.0593	0.374	0.772	Valid
15.	Dark Color	0.0593	0.374	-0.018	Invalid
16.	Easy to carry	0.0198	0.374	0.106	Invalid
17.	Practical	0.0198	0.374	0.234	Invalid
18.	Unique	0.0198	0.374	0.653	Valid
19.	Eco-Friendly Material	0.0198	0.374	0.338	Invalid
20.	Soft Color	0.0198	0.374	0.208	Invalid
21.	Warm Color	0.0198	0.374	0.110	Invalid
22.	Modern Design	0.0198	0.374	0.694	Valid

Table 6. Results Semantic Differential 1 “Chocolate Design” and “Informative”

Antonim Kata Kansei	Responden 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Total (+)	Total (-)	Kata Kansei
Un-informa-tive	1	-3	3	-1	2	3	1	-1	1	-1	-1	-2	2	2	-3	-1	1	-2	-1	-1	-1	1	-1	-2	1	-1	3	-3	-1	-1	0.4	-0.5	Informative
Non-Choco-late Design	-1	-3	1	-3	2	3	-2	-1	-2	-1	-3	0	2	-3	-3	-2	-3	1	2	0	-2	-1	1	-2	2	-3	1	-3	-3	-1	0.27	-0,87	Chocolate Design

Table 7. Results of Second Validity Test

No	Kansei Words	Table R	*Correlation* Calculated R	Description
1	Eye-Catching	0.374	0.715	Valid
2	Useable	0.374	0.453	Valid
3	Easy to store	0.374	0.383	Valid
4	Safety	0.374	0.516	Valid
5	Sturdy Packaging	0.374	0.609	Valid
6	Premium	0.374	0.788	Valid
7	Elegant	0.374	0.743	Valid
8	Exclusive	0.374	0.868	Valid
9	Unique	0.374	0.754	Valid
10	Colourful	0.374	0.401	Valid
11	Modern Design	0.374	0.791	Valid

3.6 Determining Clusters of Kansei Words

The clustering analysis was conducted using the K-Medoids. K-Means. and K-Means GA methods with Python programming. The input data consists of validated Kansei words. Fig. 4 shows the clustering results obtained using all three clustering methods.

Based on Fig. 4. it is observed that all three clustering methods yield the same number of clusters. namely three clusters. In Fig. 4 (b). it is evident that the position of the centroids (denoted by stars) shows better alignment compared to the K-Medoids and K-Means methods. This indicates that the GA approach provides a more balanced distribution among the clusters. The polygon shape of the cluster distribution using the GA-optimized K-Means method demonstrates a compact coverage with minimal deviation. suggesting a high-quality separation. According to (39). the K-Means method with GA optimization shows a significant advantage in determining more stable and optimal centroids.

In contrast. the clustering results using the K-Medoids method in Fig. 4 (a) show that the centroids (denoted by stars) tend to be located closer to the data points. resulting in denser clusters. The K-Medoids method demonstrates resilience to outliers or extreme values present in the data (40).

Fig. 4 (c) illustrates the clustering results using the K-Means method. which shows a more widely dispersed polygon. This indicates that the K-Means method is sensitive to data distribution and extreme values (outliers) (25). As a result. the clustering outcomes and data points are suboptimal (39). The words generated from each cluster are presented in Table 8.

As shown in Table 8. most of the clusters produced by the three clustering methods contain similar sets of words. However. in the K-Means method. the word "Unique" is placed in a different cluster from "Eye-catching". "Useable". "Easy to store". "Safety". and "Modern". In contrast. in the K-Medoids and K-Means GA methods. "Unique" is grouped in the same cluster with those five words.

4

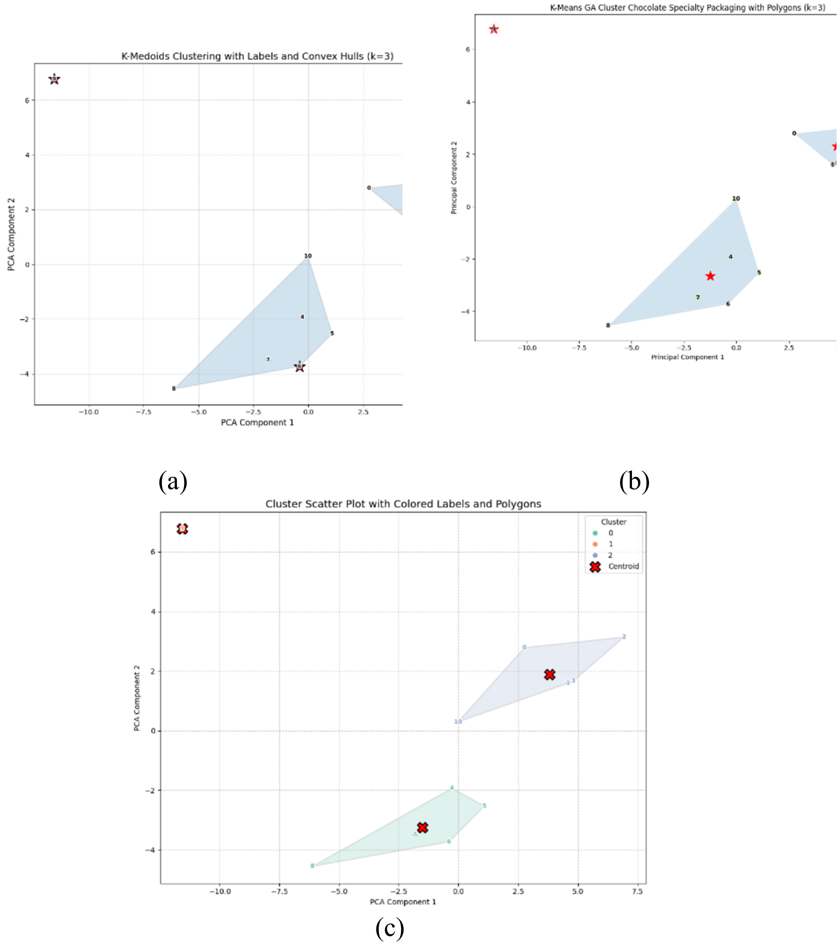

Fig. 4. Cluster Results: (a) K-Medoids. (b) GA optimized K-Means. (c) K-Means

Table 8. Results of Kansei Words Distributin of Each Cluster from Three Methods: (a) K-Medoids. (b) GA optimized K-Means. (c) K-Means

K-Medoids		
Cluster 1	Cluster 2	Cluster 3
Colorful	Eye Catching	Sturdy Packaging
	Useable	Premium
	Easy to Store	Elegant
	Safety	Exclusive
		Unique

(a)

K-Means GA		
Cluster 1	Cluster 2	Cluster 3
Eye Catching	Colorful	Sturdy Packaging
Useable		Premium
Easy to Store		Elegant
Safety		Exclusive
		Unique

(b)

K-Means GA		
Cluster 1	Cluster 2	Cluster 3
Sturdy Packaging	Colorful	Eye Catching
Premium		Useable
Elegant		Easy to Store
Exclusive		Safety
Unique		Modern Design

(c)

Based on the Silhouette Coefficient values obtained from each clustering method. it is evident that the K-Means GA method produces the highest Silhouette value (0.526) compared to the K-Medoids (0.391) and K-Means (0.259) methods. These results are shown in Fig. 5. According to (41). the application of optimization algorithms. such as Genetic Algorithms. results in more optimal cluster (centroid) positions compared to conventional K-Means. Furthermore. the difference in the distribution of points around the centroids reflects the broader exploration of the GA method compared to other K-Means techniques. The variation in Silhouette Coefficient values also indicates differences in the clustering quality (9). According to (18). determining the optimal cluster based on the fitness function evaluation of genetic algorithms leads to a cluster distribution that is more concentrated or minimal.

```
Nilai k optimal berdasarkan Silhouette Score: 3
Silhouette Score maksimum: 0.2888
```

(a)

```
Average Silhouette Score: 0.5264
```

(b)

```
Silhouette Score (k=3): 0.2591
```

(c)

Fig. 5. Comparison of Silhouette Value: (a) K-Medoids. (b) K-Means Optimize GA. (c) K-Means

```
Davies-Bouldin Index (k=3): 0.5562
Calinski-Harabasz Index (k=3): 13.0048
```

(a)

```
Davies-Bouldin Index: 0.38446302041745145
Calinski-Harabasz Index: 23.726806928307578
```

(b)

```
Davies-Bouldin Index: 0.8314
Calinski-Harabasz Index: 7.3320
```

(c)

Fig. 6. Comparison of DBI and CHI Value: (a) K-Medoids. (b) K-Means Optimize GA. (c) K-Means.

In addition to the Silhouette Coefficient. other metrics are also used to evaluate clustering quality. such as the Davies-Bouldin Index (DBI) and the Calinski-Harabasz Index (CHI). The DBI assesses how well-separated the clusters are; lower DBI values indicate higher intra-cluster cohesion and better inter-cluster separation (42). Meanwhile. the CHI measures the ratio between inter-cluster dispersion and intra-cluster compactness; a higher CHI value indicates better clustering performance (43). The DBI and CHI values obtained from the three methods can be seen in Fig. 6.

The evaluation results indicate that the K-Means GA method provides the best clustering performance. The word distribution in each cluster was then analyzed to formulate design concepts through discussions with domain experts (9). Concept determination was carried out in collaboration with at least four experts in the field of design (32). The words appearing in each cluster. as shown in Table 8 (b). were used to identify relevant concepts in consultation with the experts. The results of the concept analysis conducted with the experts are presented in Table 9.

Based on discussions with experts. as presented in Table 9. it was concluded that the first cluster represents the concept of "Useable." the second reflects "Cheerful." and the third generates the concept of "Premium". These three concepts align with the highest-weighted words listed in Table 5. The "Useable" concept is reflected in terms such as

"Useable" and "Easy to store." indicating that ease of use and storage can enhance brand loyalty. The "Cheerful" concept refers to words like "Eye Catching" and "Colourful." suggesting that packaging aesthetics. such as colour and visual background. have a psychological impact on purchasing decisions (44). Meanwhile. the "Premium" concept is represented by terms such as "Chocolate Design" and "Informative." as premium packaging effectively conveys information and visually represents the product content (45).

Based on Fig. 4 (b). it is evident that the third cluster exhibits the highest Fitness Function value. According to (18). his value indicates the optimal cluster. as it has the smallest cluster range. Therefore. the concept of "Useable" that emerges from this cluster is selected as the recommended concept for implementation in specialty chocolate bar packaging. The "Useable" concept represents consumers' emotions in addressing their needs. where they seek chocolate bar packaging that is easy to use. open. carry. and store. This is consistent with the term that carries the highest weight in the Kansei word extraction results presented in Table 5. where "Useable" ranks fourth among the most frequently mentioned terms by consumers when expressing their emotional responses or Kansei words. The word "Useable" has a weight of (0.2371). followed by "Easy to Store" with a weight of (0.2173). According to (46). packaging that is easy to use plays a significant role in influencing consumers' purchasing decisions. This finding is reinforced by market research indicating that modern chocolate bar consumers increasingly prefer packaging that offers ease of use. As stated by Towards Packaging. simple yet effective packaging. such as easy-to-tear wrappers or resealable pouches. provides convenience without compromising the integrity of the product (47). Therefore. the "Useable" design concept identified in this study aligns well with current consumer expectations in the market.

Table 9. Analysis Concepts with Experts

Cluster 1	Cluster 2	Cluster 3
Eye Catching	Colorful	Sturdy Packaging
Useable		Premium
Easy to Store		Elegant
Safety		Exclusive
		Unique
	Expert 1	
Useable	Colorfull	Premium
	Expert 2	
Useable	Cheerful	Premium
	Expert 3	
Functional Design	Playful	Modern Luxury
	Expert 4	
Useable	Cheerful	Premium

4 Conclusion

The weighting results using TF-IDF in the Kansei word extraction process revealed that the words with the highest weight values are "Chocolate Design" (0.5334). Informative (0.3754). Eye Catching (0.3359). Usable (0.2371). "Easy to Store" (0.2173). and "Colourful" (0.1976). These words represent consumer feelings based on their preferences for specialty chocolate bar packaging. Out of the 22 Kansei words obtained from TF-IDF. 11 were deemed valid and used in the clustering process. The results of the three clustering methods contain similar sets of words. However. in the K-Means method. the word "Unique" is placed in a different cluster from "Eye-catching." "Usable." "Easy to Store." "Safety." and "Modern." In contrast. the K-Medoids and K-Means GA methods group "Unique" in the same cluster as those five words. In terms of performance. the K-Means GA method produced better results compared to the K-Medoids and K-Means methods. with a Silhouette Coefficient of (0.5264). a Davies-Bouldin Index of (0.3844). and a Calinski-Harabasz Index of (23.726). The advantage of the K-medoids method is its resistance to extreme data values. while the K-means method is sensitive to such values. On the other hand. the K-Means GA method is more optimal in determining centroids. whereas K-medoids tends to have centroids closer to outliers. The design concepts derived from the word sets in each cluster. as assessed by an expert panel. include "Useable." "Cheerful." and "Premium". Among these. "Usable" was selected as the most appropriate concept due to its optimal cluster distribution based on the Fitness Function value. This concept reflects consumers' needs for chocolate bar packaging that is "Eye-catching." "Easy to Store." "Usable." and "Safe." and is further supported by the highest weight in the Kansei word extraction results. Therefore. this concept offers a solution to current challenges in specialty chocolate bar packaging. This finding demonstrates that an emotion-based approach combined with data analysis can generate design concepts aligned with consumer preferences. Consequently. this approach not only effectively addresses consumers' emotional needs but also serves as a strategic guide for the development of competitive specialty chocolate bar packaging designs in the market.

References

1. Herdhiansyah D. Asriani A. Kajian Proses Pengolahan Cokelat Batangan (Chocolate Bar) Di Pt Xyz Di Kota Kendari - Sulawesi Tenggara. Agritech J Fak Pertan Univ Muhammadiyah Purwokerto. 2022;24(1):28.
2. Farhanandi BW. Indah NK. Karakteristik Morfologi dan Anatomi Tanaman Kakao (Theobroma cacao L.) yang Tumbuh pada Ketinggian Berbeda. LenteraBio Berk Ilm Biol. 2022;11(2):310–25.
3. Khan ZA. Kamran H. Ahmed I. Safdar U. Asim Z. Alwie AB. Consumers' perceptions of quality. packaging. and purchase intentions: an analysis of the perfume market in Oman. Sapienza. 2025;6(1):1–12.
4. Shukla P. Singh J. Wang W. The influence of creative packaging design on customer motivation to process and purchase decisions. J Bus Res [Internet]. 2022;147(April):338–47. Available from: https://doi.org/10.1016/j.jbusres.2022.04.026

5. Lamatinulu. Dahlan M. Mardin. Packaging Design Concepts as Attributes of Product Diversification Using Kansei Engineering Approach in SMEs Scale Cocoa Industry in South Sulawesi. Int J Innov Sci Res Technol. 2021;6(11):1–4.
6. Cahyani A. Singgih ML. Development of Frozen Food Packaging Design Based on Kansei Engineering. Atlantis Press International BV; 2024.
7. Cai W. Wang Z. Wang Y. Zhou M. Research on wheelchair form design based on Kansei engineering and GWO-BP neural network. Sci Rep. 2025;15(1):1–25.
8. Johan VS. Riftyan E. Khairany S. Enhancing Consumer Engagement through Kansei Engineering : A Novel Approach to Sago Rice Packaging Design Peningkatan Daya Tarik Konsumen melalui Kansei Engineering Pendekatan Inovatif dalam Desain Kemasan Beras Sagu. 2024;13(1):36–53.
9. Sari NP. Prastiwinarti W. Pramesti RD. Putri LY. Permana RR. Implementation of the K-Means Genetic Algorithm to Determine the Design Concept for Toast Bread Packaging by Kansei Engineering. Commun Comput Inf Sci. 2024;2313 CCIS:41–52.
10. Hartono M. The modified Kansei Engineering-based application for sustainable service design. Int J Ind Ergon [Internet]. 2020;79(July):102985. Available from: https://doi.org/10.1016/j.ergon.2020.102985
11. Adiyanto O. Jatmiko HA. Erni. Development of food packaging design with kansei engineering approach. Int J Sci Technol Res. 2019;8(12):1778–80.
12. Sari NP. Prastiwinarti W. Lokman AM. Yamin I. Isna A. Yennia L. et al. Developing the Concept of Emotion for Rendang Packaging Design Using Kansei Engineering. Commun Comput Inf Sci. 2024;2313 CCIS:15–27.
13. Wang T. Xu M. Yang L. Zhou M. Sun X. Constructing a MOEA approach for product form Kansei design based on text mining and BPNN. J Intell Fuzzy Syst. 2024;46(4).
14. Li J. A comparative study of keyword extraction algorithms for English texts. J Intell Syst. 2021;30(1):808–15.
15. Hashemzadeh B. Abdolrazzagh-Nezhad M. Improving keyword extraction in multilingual texts. Int J Electr Comput Eng. 2020;10(6):5909–16.
16. Abiodun M. Ikotun. Ezugwu AE. Abualigah L. Abuhaija B. Jia Heming. K-means clustering algorithms: A comprehensive review. variants analysis. and advances in the era of big data. Inf Sci (Ny). 2023;622:178–210.
17. Abdel-Basset MRM. El-Shahat. Sallam DKM. Hezam IM. AbdelAziz NM. Energy-efficient trajectory optimization algorithm based on K-medoids clustering and gradient-based optimizer for multi-UAV-assisted mobile edge computing systems. Sustain Comput Informatics Syst. 2024;44(101045).
18. Malki A Al. Rizk MM. El-Shorbagy MA. Mousa AA. Hybrid Genetic Algorithm with K-Means for Clustering Problems. Open J Optim. 2016;05(02):71–83.
19. Nagamachi M. Lokman AM. Kansei innovation: Practical design applications for product and service development. Kansei Innovation: Practical Design Applications for Product and Service Development. 2015. 1–129 p.
20. Sari NP. Halimatus I. Muryeti M. Perancangan model kemasan distribusi minuman kopi cup menggunakan kansei engineering (studi kasus kopi ruang normal). 2024;18(4):871–81.
21. Verhagen T. van den Hooff B. Meents S. Toward a better use of the semantic differential in IS research: An integrative framework of suggested action. J Assoc Inf Syst. 2015;16(2):108–43.
22. Djatna T. Wrasiati LP. Santosa IBDY. Balinese Aromatherapy Product Development Based on Kansei Engineering and Customer Personality Type. Procedia Manuf. 2015;4(Iess):176–83.

23. Sari NP. Isna A. Maharani D. Fadhillah F. Implementasi Kansei Engineering dalam Menentukan Konsep Pengembangan Kemasan Rujak Buah Potong. J INTECH Tek Ind Univ Serang Raya. 2024;10(1):9–18.
24. Kittidecha C. Marasinghe AC. Koichi Y. Application of Kansei Engineering to Tactile Sense in the Thai Food Wrapping Materials. J Appl Packag Res. 2016;8(2):2.
25. Restuputri DP. Fridawati A. Masudin I. Customer Perception on Last-Mile Delivery Services Using Kansei Engineering and Conjoint Analysis: A Case Study of Indonesian Logistics Providers. Logistics. 2022;6(2).
26. Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ. 2018;48(6):1273–96.
27. Rahman F. Ridho II. Muflih M. Pratama S. Raharjo MR. Windarto AP. Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country. IOP Conf Ser Mater Sci Eng. 2020;835(1).
28. Mustafi D. Sahoo G. A hybrid approach using genetic algorithm and the differential evolution heuristic for enhanced initialization of the k-means algorithm with applications in text clustering. Soft Comput. 2019;23(15):6361–78.
29. McNicholas SM. McNicholas PD. Ashlock DA. An Evolutionary Algorithm with Crossover and Mutation for Model-Based Clustering. J Classif. 2021;38(2):264–79.
30. Tinós R. Zhao L. Chicano F. Whitley D. NK Hybrid Genetic Algorithm for Clustering. IEEE Trans Evol Comput. 2018;22(5):748–61.
31. Bennaceur H. Almutairy M. Alhussain N. Genetic Algorithm Combined with the K-Means Algorithm: A Hybrid Technique for Unsupervised Feature Selection. Intell Autom Soft Comput. 2023;37(3):2687–706.
32. EROL A. BASAR DL. Expert Designer's Perceptions About Designing Using Kansei Engineering Results. In: MARCO-ALMAGRO L. LÉVY P. YANAGISAWA H. LOKMAN AM. editors. 9th Kansei Engineering and Emotion Research Conference KEER2022 Barcelona. 6-8 September 2022. Barcelona; 2022. p. 309–14.
33. Sari NP. Akkili VA. Muryeti M. Penerapan jaringan syaraf tiruan untuk menentukan elemen desain kemasan Numany rempeyek berbasis kansei engineering. Agrointek. 2024;18(3):742–52.
34. Nur R. Ushada M. Ainuri M. Affan M. Falah F. Modeling Consumer Overall Acceptance for Traditional Spice-Based Ready-to-Drink Using Artificial Neural Network and Kansei Engineering. 2024;14(4):1178–84.
35. AlShammari AF. Implementation of Keyword Extraction using Term Frequency-Inverse Document Frequency (TF-IDF) in Python. Int J Comput Appl. 2023;185(35):9–14.
36. Hasanah U. Astuti T. Wahyudi R. Rifai Z. Pambudi RA. An experimental study of text preprocessing techniques for automatic short answer grading in Indonesian. Proc - 2018 3rd Int Conf Inf Technol Inf Syst Electr Eng ICITISEE 2018. 2018;230–4.
37. Palomino MA. Evaluating-the-Effectiveness-of-Text-PreProcessing-in-Sentiment-AnalysisApplied-Sciences-Switzerland.pdf. 2022;
38. Da Costa Pinto D. Pintauli S. Rahardjo A. Sibarani R. Validity and Reliability of the Timor Leste Version Instruments for Measuring Oral Health-Related Quality of Life. J Int Dent Med Res. 2024;17(2):777–82.
39. Tang Z. Improved K-means Clustering Algorithm based on Genetic Algorithm. TELKOMNIKA Indones J Electr Eng. 2014;12(3):1917–23.
40. Rahmawati O. Fauzan A. Provincial Clustering Based on Education Indicators: K-Medoids Application and K-Medoids Outlier Handling. BAREKENG J Ilmu Mat dan Terap. 2024;18(2):1167–78.

41. Agarwal N. Sikka G. Awasthi LK. WGSDMM+GA: A genetic algorithm-based service clustering methodology assimilating dirichlet multinomial mixture model with word embedding. Futur Gener Comput Syst. 2023;145:245–66.
42. Dewi DAIC. Pramita DAK. Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. Matrix J Manaj Teknol dan Inform. 2019;9(3):102–9.
43. Bunga Mardhotillah. Amril Fadli. Edi Elisa Z. Indeks Calinski Harabasz Analisis Fuzzy C Means dan K Means Cluster Kabupaten/Kota di Provinsi Jambi Menurut Potensi Pertambangan. Penggalian. Pengadaan Listrik. dan Gas. J Stat Univ Jambi. 2023;2(1):16–22.
44. Liu C. Psychological impacts of packaging aesthetics of chocolate and its effects on consumer purchasing decision. Highlights Business. Econ Manag. 2024;41:238–51.
45. Gunaratne NM. Fuentes S. Gunaratne TM. Torrico DD. Francis C. Ashman H. et al. Effects of packaging design on sensory liking and willingness to purchase: A study using novel chocolate packaging. Heliyon. 2019;5(6):e01696.
46. Sitopu JW. Firdaus A. Analyzing the Impact of Packaging Design on Consumer Purchasing Decisions in the Cosmetics Industry. J Econ Manag Bus Technol. 2024;3(1):1–8.
47. Packaging T. press release. OpenPR. 2025. How convenience and sustainability are redefining chocolate bar packaging. Available from: https://www.openpr.com/news/3821490/how-convenience-and-sustainability-are-redefining-chocolate