Journal of Intelligence Studies in Business

Journal of Intelligence Studies in Business

Publication details, including instructions for authors and subscription information: https://ojs.hh.se/index.php/JISIB/index

Business intelligence using the fuzzy-Kano model

Soumaya Lamrhari[a]*, Hamid Elghazi[b], Abdellatif El Faker[a]

[a]ENSIAS, Mohammed V University Rabat, Morocco, [b]National Institute of Posts and Telecommunications Rabat, Morocco

*soumaya_lamrhari@um5.ac.ma

To cite this article: Lamrhari, S., Elghazi, H. & El Faker, A. (2019) Business intelligence using the fuzzy-Kano model. *Journal of Intelligence Studies in Business*. 9 (2) 43-58.

Article URL: https://ojs.hh.se/index.php/JISIB/article/view/408

JISIB
Journal of Intelligence Studies in Business
Vol. 9, No. 2 (2019) pp. 43-58
Open Access: Freely available at: https://ojs.hh.se/

Stevan Dedijer
1911-2004

Business intelligence using the fuzzy-Kano model

Soumaya Lamrhari[a]*, Hamid Elghazi[b] and Abdellatif El Faker[a]

[a]*ENSIAS, Mohammed V University Rabat, Morocco*
[b]*National Institute of Posts and Telecommunications Rabat, Morocco*

Corresponding author (*): soumaya_lamrhari@um5.ac.ma

Received 25 September 2019 Accepted 24 October 2019

ABSTRACT Today, understanding customer satisfaction is becoming a difficult and complex task for companies due to the explosive growth of the voice of the customer in online reviews. This has pushed companies to rethink their business strategies and resort to business intelligence techniques in order to help them in analyzing customer requirements and market trends. This paper proposes a decision support framework for dynamically transforming the voice of the customer data into actionable insight. The framework measures the customer satisfaction by extracting key products' aspects along with customers' sentiments from online reviews using a text mining technique: the latent Dirichlet allocation approach. We apply the Fuzzy-Kano model to classify the real customer requirements, then, map them dynamically to the SWOT matrix. The proposed approach is extensively tested on an empirical dataset based on several performance metrics including accuracy, precision, recall, and F-score. The reported results showed that latent Dirichlet allocation approach has correctly extracted aspects with 97.4% accuracy and 92.4 % precision.

KEYWORDS Business intelligence, customer satisfaction, decision support framework, Fuzzy-Kano model, latent Dirichlet allocation, online reviews, text mining, voice of the customer, web intelligence

"The secret of successful retailing is to give your customers what they want."

Sam Walton

1. INTRODUCTION

In today's competitive marketplace, business leaders have realized that customers are the major driving force leading a company to thrive (Carulli et al., 2013) (Lee et al., 2014). In fact, most of the product-based companies require an in-depth understanding of their customers' satisfaction. Thus, they resort to business intelligence (BI) techniques in order to provide competitive products that meet the customer needs and go in line with the current market trend (Sabanovic and Søilen, 2012). The voice of the customer (VOC) is a widely used term in market research that describes the customers' feedback about their expectations and experiences in relation to products and services. This is considered an essential first step in developing a successful product or service (Aguwa et al., 2012). The VOC is usually captured in a variety of ways such as questionnaire surveys, face to face interviews, telephone interviews, and discussion groups (Goodman, 2014) (Rese et al., 2015). However, most of these methods are demanding in terms of time, cost, and their geographic reach (Szolnoki and Hoffmann, 2013). Additionally, the participants' willingness to provide actual input can impact the collected data quality (Reyes, 2016). Besides, the surveys are generally conducted occasionally, which makes the timeliness of the gathered data questionable (Culotta and Cutler, 2016). Consequently, we need to consider other alternative data sources to reveal customer expectations.

The growing popularity of social media and BI in the last decade makes them a valuable digital channel for listening and capturing customers' voices (Gioti et al., 2018). Unlike conventional approaches, the VOC on social media is publicly available, easily accessible anywhere and anytime at low cost. Examples of these VOCs include customer posts, comments, and reviews. Customer reviews can be considered a trustworthy VOC since they hold massive data where customers voluntarily share their experiences about a specific product or service after use or purchase. Unfortunately, these reviews may not explicitly reflect customer needs since they require more advanced data analysis methods. Therefore, most companies have adopted BI techniques (Nyblom et al., 2012), such as text mining, to discover hidden patterns in this large amount of textual data to support the decision making process (Søilen et al., 2017) (Xu and Li, 2016) (Jia, 2018).

Plenty of studies have been conducted to explicitly or implicitly understand customer satisfaction from online review content. For instance, Decker and Trusov (2010) applied an econometric framework based on Poisson regression, binomial regression, and latent class Poisson regression models. The basic potential of using those classification algorithms is to estimate the relative strength of effects resulting from the list of attributes identified through customer reviews about mobile phones. The methodology findings reveal that the negative binomial regression approach provides significant estimation parameters, which quantify the effects that the product attributes have on overall customer satisfaction. Park and Lee (2011) proposed a systematic framework for extracting customer requirements from an online customer center and transforming them into product specifications data. In their approach, customer opinions are collected, then a text mining analysis is conducted on customer complaints to extract meaningful keywords. Based on the extracted VOCs, customers are clustered into different groups with similar needs. Then, the target groups will be carefully selected by the companies. Further, a co-word and a decision tree analysis are used to translate the customer requirements into product specifications. Xiao et al. (2016) established a novel econometric preference measurement model for extracting overall customers' preferences from online product reviews. The model allows a semi-automatic extraction of product features along with the related reviewers' sentiments. Then, aggregate customer preferences are extracted from online product reviews by a modified ordered choice model, which considers the variety of customers' ratings and allows them to assign rating sores with their own thresholds. Furthermore, the identified customer requirements are classified into different categories, e.g. basic, performance, excitement, innovation-needed, reverse and divergent, by using a marginal effect-based Kano model, which is an extension of the classical Kano model that employs the marginal effect information disclosed by the proposed modified ordered choice model.

In addition, other research studies have applied an aspect-based sentiment analysis approach for understanding customers' satisfaction. This approach involves extracting aspects and finding their corresponding sentiments. Latent Dirichlet allocation (LDA) is considered a state-of-the-art modeling tool for extracting products' features in the aspect-based sentiment analysis (Saura et al., 2019). For instance, Farhadloo et al. (2016) proposed a Bayesian approach that models the customer satisfaction based on the individual aspect ratings. First, the study utilizes the aspect-based sentiment analysis method described in (Farhadloo and Rolland, 2013) as a basis to transform unstructured input data into semi-structured data. Then, the Bayesian method enables the extraction of the relative importance of each aspect of the product or service. For consumer-generated content in marketing, Tirunillai and Tellis (2014) proposed a unified framework that extracts the key latent quality dimensions (known as a "topic" in the LDA literature) of consumer satisfaction and the associated sentiments using unsupervised Bayesian learning algorithm based LDA. Moreover, the approach determines the validity, importance, dynamics, and heterogeneity of the extracted dimensions. In another context, Guo et al. (2017) put forward an LDA based approach to identify the most important dimensions of customer service in the hotel sector. Then, they performed a perceptual mapping to represent the key dimensions influencing the visitors' satisfaction and the visitors' perceived ratings in different hotel classification. Qi et al. (2016) proposed an automatic filtering model to mine customers' requirements from online reviews. First, it filters out the reviews that are helpful for product improvement. Then, a lexicon-based sentiment analysis, LDA, and page rank are used to rank the terms based on their frequencies and semantic relationships. In addition, the conjoint analysis and the Kano model are utilized to determine the product attribute weights and categories and evaluate their impact on customer satisfaction.

Despite the contributions made by the aforementioned studies regarding the understanding of customer satisfaction from online reviews, they still have some drawbacks. First, in (Decker and Trusov, 2010), (Farhadloo et al., 2016), (Qi et al., 2016), (Xiao et al., 2016); (Park and Lee, 2011), the authors quantified the effects that customer requirements may have on their satisfaction by using various modeling methods that measure product attributes, e.g. weights and importance. While in (Guo et al., 2017), (Tirunillai and Tellis, 2014), the authors focused only on mining the relevant products' attributes. Second, most of the existing studies that have measured the effects of customer requirements on customer satisfaction have not classified the identified requirements either from the customer or the provider perspectives. Third, our approach bears a close resemblance to the one proposed by Qi et al. (2016), except that in our study, we have incorporated the Fuzzy analysis to the Kano model instead of the conjoint analysis. With Fuzzy analysis, the measurement of each product's attribute is presented in the form of the degree of membership allowing the customers to express their preferences towards multi-attributes at the same time, unlike the conjoint analysis where the customers can only express their preferences for a single attribute.

Based on the results reported in (Tirunillai and Tellis, 2014), (Qi et al., 2016), (Guo et al., 2017), LDA has demonstrated good stability and satisfactory performance in terms of accurately extracting the key customer requirements from a large volume of online reviews. Therefore, we have selected it as a topic modeling method in our approach. To the best of our knowledge, this is the first attempt to combine LDA, the Fuzzy-Kano model and the SWOT method into one decision support framework for understanding customer satisfaction. Specifically, we will analyze the collected VOC from online reviews, then, extract the actual customers' requirements that have more impact on their experiences with a given product or service.

Such a framework is beneficial for companies since it allows them to deeply understand the customers' needs and proactively adapt their product/service or even their business model accordingly. It is composed of four major modules. The first one consists of collecting and preprocessing data from online customer reviews. The second one extracts the products' aspects and the corresponding customers' sentiments from the preprocessed data using LDA. The third module classifies the real customer needs that affect their satisfaction based on the Fuzzy-Kano model. The fourth module maps the Fuzzy-Kano model's output to a SWOT matrix in order to easily interpret the obtained results. The proposed approach is extensively evaluated using an empirical dataset, which includes mobile phone reviews collected from Amazon. The evaluation is based on several performance metrics including accuracy, precision, recall, and F-score.

The remainder of this paper is organized as follows. Section II provides the theoretical background of the proposed framework. Section III describes our methodology. In Section IV, we evaluate the effectiveness of our method using a real case study. In section V, we draw some conclusions and shed light on further research directions.

2. THEORETICAL BACKGROUND

2.1 Latent Dirichlet Allocation (LDA)

In this paper, we seek a way to map customers' reviews to the topics, without having prior knowledge on what those topics are. This calls into question the unsupervised classification problem on natural language. LDA is an unsupervised topic modeling approach widely applied in natural language processing. The present study deployed LDA (Blei, 2012) instead of other topic model approaches found in the literature because it relies on more comprehensive probabilistic assumptions on the text generation and has shown satisfactory performance and good stability when classifying large data sets (Lu et al., 2011) (Alghamdi and Alfalqi, 2015) (Hofmann, 2017). In LDA, each document consists of a mixture of topics and each topic consists of a collection of words. Given a corpus D consisting of M documents each of length N, each document contains a sequence of W words, each of these words represents the v^{th} word in a vocabulary of V distinct terms and K is the total number of topics. Thus:

- α and β define the prior distribution parameters per-document topic distribution and per-topic word distribution respectively.
- θ_m is the topic distribution for document m.
- φ_k is the word distribution for topic k.
- z_{nm} is the topic for the n^{th} word in document m.
- and w_{mn} is the specific word

Formally, LDA generates a corpus D of M documents according to the following generative process:

- Choose a topic distribution $\theta_i \sim Dir(\alpha)$, where $i \in \{1, \ldots, M\}$, and $Dir(\alpha)$ is a Dirichlet distribution with scaling parameter α which typically is sparse ($\alpha < 1$).
- For each topic $k \in \{1, \ldots, K\}$, Choose $\varphi_k \sim Dir(\beta)$, where β is typically sparse.
- For each of the word positions i, j, where $j \in \{1, \ldots, N_i\}$, and $i \in \{1, \ldots, M\}$:
    - Choose a topic $z_{i,j} \sim Multinomial(\theta_i)$.
    - Choose a word $w_{i,j} \sim Multinomial(\varphi_{z_{i,j}})$.

Moreover, a graphical model can also mirror the generative process of documents. As depicted in Figure 1, the boxes refer to repeated contents where the number of repetitions is presented by the variable at the corner of the corresponding box. The blue node represents the only observed variable (w). The white nodes denote latent variables (φ, θ); Gray nodes represent hyperparameters (α and β). The arrows indicate dependencies among the model parameters.

Practically, the model must determine the hidden variables from the data, namely the document-topic distribution θ, and the topic-word distribution φ. To this end, the Gibbs Sampling algorithm (Darling, 2011) is applied to estimate those two LDA parameters.

2.2 Kano Model

The Kano model (Kano, 1984) is an effective tool used by companies to integrate the VOC into the product and service development

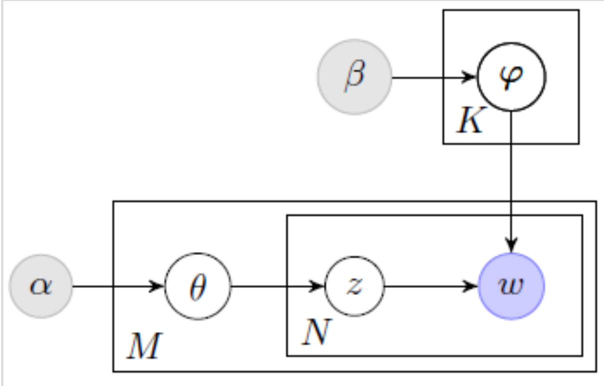

Figure 1 The graphical representation of the LDA model, redrawn from (Blei, 2012)

lifecycle. It is regarded as a nonlinear relationship between product quality and customer satisfaction. It measures customer sentiments to discover which customer requirements have the highest impact on customer satisfaction (Tontini et al., 2013).

The Kano model often carries out surveys and questionnaire investigations on customers to determine the requirements of a particular product or service. For a given product's aspect, a functional question (aspect's presence) and a dysfunctional question (aspect's absence) are asked. Each question form should be answered on a five-point scale such as: like, necessary, neutral, unnecessary, and dislike. Based on a statistical analysis of all the accumulated responses of the survey, each answer pair is aligned with the Kano evaluation (Table 1), forming certain requirements (Ullah and Tamaki, 2011). Table 1 shows that by combining the two answers (functional and dysfunctional), the product's aspects can be classified into six categories of requirement that influence customer satisfaction, including:

- "Must-be" (M) requirement is expected by the customers, its presence does not lead to customer satisfaction, but its absence leads to extreme customer dissatisfaction.

Table 1 The standard Kano evaluation (Ullah and Tamaki, 2011). Nec = necessary; Neu = neutral; Unnec = unnecessary; Dis = dislike.

		Dysfunctional				
		Like	**Nec.**	**Neu**	**Unnec**	**Dis**
Functional	**Like**	Q	A	A	A	O
	Nec	R	I	I	I	M
	Neu	R	I	I	I	M
	Unnec	R	I	I	I	M
	Dis	R	R	R	R	Q

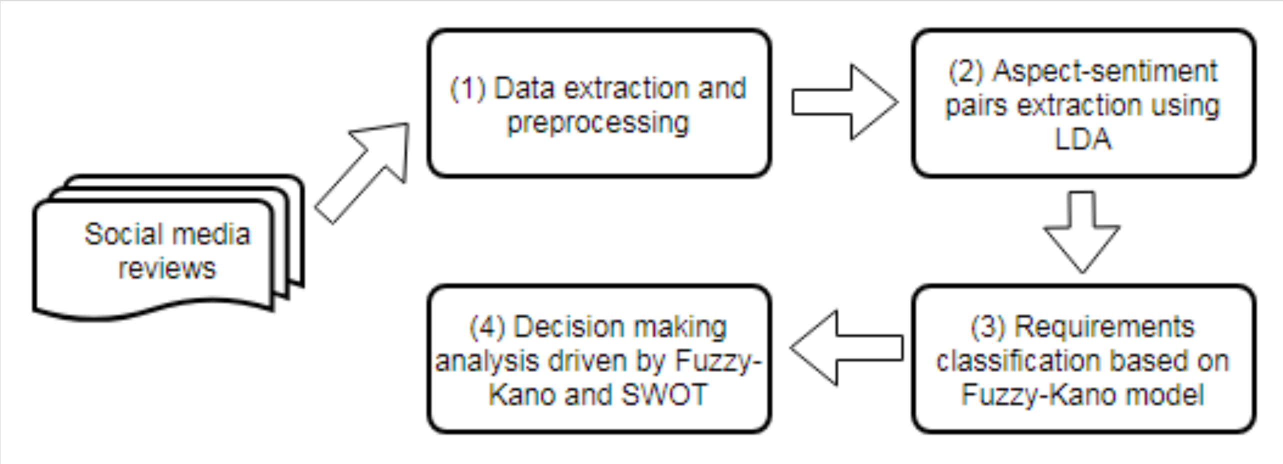

Figure 2 The proposed decision support framework.

- "One-dimensional" (O) requirement is the property of a customer need that increases customer satisfaction when it is fulfilled. Inversely, customer satisfaction decreases when it is not fulfilled.
- "Attractive" (A) requirement is usually uncommon or unexpected by the customers, if included, can truly increase customer satisfaction; if not, there is no feeling of dissatisfaction.
- "Indifferent" (I) requirements are those that the customer does not care about whether they exist or not. That is, these attributes will cause neither the satisfaction nor the dissatisfaction of customers, but that does not mean they do not impact the company's production decisions.
- "Reverse" (R) requirements are those whose presence results in dissatisfaction since not all customers are alike. In other words, what makes one customer satisfied might probably alienate another.
- And the "Questionable" (Q) requirement, which occurs when the customer selects an unclear answer from both functional and dysfunctional sides.

In addition, the Kano questionnaires and surveys allow the users to select only a single option from a set of options. That makes them unable to express their uncertainty toward certain aspects by selecting more than one choice. To address the issue of uncertainty concerning people's satisfaction as well as the vagueness of human thought, our study combines the classical Kano model with the fuzzy analysis to obtain an equivalent Fuzzy-Kano model that classifies the customers' requirements based on fuzzy logic rather than binary logic (Lee and Huang, 2009). The Fuzzy-Kano model allows customers to express multi-feeling, with the help of the different Kano categories, by giving fuzzy satisfactory values to certain aspects. This fuzzy set of values is represented by variable membership degrees ranging from 0 to 1, reflecting the uncertainty, where the sum of elements is equal to 1. Furthermore, this approach automates the building of the Kano model. It incorporates the VOCs into the Fuzzy-Kano model through LDA to obtain much larger scale data with more reliable insights since the classical Kano model, when used alone, cannot directly handle such data.

3. METHODOLOGY

The proposed framework is composed of four modules as illustrated in Figure 2: (1) data extraction and preprocessing; (2) aspect-sentiment pairs extraction using LDA; (3) requirements classification based on the Fuzzy-Kano model; and (4) decision-making analysis driven by Fuzzy-Kano and SWOT. In this section, we describe each of these modules.

3.1 Data Extraction and Preprocessing

The first module consists of gathering online customer reviews as the material for analysis and saving them in the form of a table in which each review denotes a document. Generally, reviews contain emoticons, special characters, punctuation, HTML tags, capital letters and misspelled words. So, it is necessary to apply a set of operations to each review before moving to the next module. These preprocessing operations include:

Tokenization: is the act of breaking up a sequence of textual content into words, phrases, and symbols called tokens. These tokens are used as input data for further processing.

Stop word removal: is the process of filtering out irrelevant words and characters from data, such as prepositions and pronouns.

Part-Of-Speech Tagging (POST): is applied to assign a special label to each token (word) in a text such as a noun, verb, or adjective.

Filtering tokens: is used to filter out all words where the length is out of the range [2-25 characters].

Transforming cases: consists of converting all tokens into lowercase.

Stemming: is applied to discard affixes from each word to obtain their root form.

Additionally, some reviews can be wrapped in a specific electronic file format, such as HTML, XML or JSON, which sometimes requires transformation into another format so as to be easily processed by the next modules. After performing the aforementioned preprocessing operations, a set of valid words is generated by excluding all meaningless words from the token list. Thus, a document-term matrix is produced, which indicates terms and their occurrence frequencies in each document.

3.2 Aspect-Sentiment Pairs Extraction using LDA

In this module, we begin by implementing LDA to reveal all topics being discussed by customers in the reviews. For this, we compute the probability of each word in the review as written in equation 1:

$$p(w|R) = \sum_{i=1}^{K} p(w|T) \times p(T|R_i) \qquad (1)$$

Where $p(w|T)$ is the probability of a word w given a topic T and $p(T|R_i)$ is the probability of a topic T given a review R_i, with K is the total number of reviews in the overall collection.

Then, we extract aspects and sentiments that appear together in the same topic distribution according to the POS tagging process. Words describing sentiments are mainly represented by adjectives and adverbs, meanwhile, a product aspect is mainly represented by nouns or noun phrases (Hu and Liu, 2004a), but not all nouns refer to aspects. Therefore, we select first the most representative nouns as aspect candidates according to their co-occurrence frequencies in the review, as well as their appearance with sentiment words. To identify sentiment word orientation, the Wordnet (Miller, 1995) is used as well as the opinion lexicon provided in (Hu and Liu, 2004b), when the sentiment words are not supported by Wordnet. Next, we use the popular approach of Hu and Liu (2004b) to construct aspect-sentiment pairs, which is based on extracting nearby adjectives to a frequent aspect.

Practically, we define a nearby adjective as the nearest opinion word to a specific aspect considering token distance (measured in the number of words far away from that aspect). The maximum number of the nearest sentiment words is set at two for the simple reason that usually when a third word is found, it was certainly describing another aspect that was ignored during processing. By doing so, we prevent the incorrect attribution of a sentiment word to an aspect. Moreover, we consider that once a sentiment word is assigned to an aspect, it will not be considered in the future attribution.

To compute the final sentiment score for an aspect (positive or negative), we sum up all sentiment word scores related to that aspect as follows:

$$A_i.ss = \sum_j \frac{SW_j.ss}{dist(SW_j, A_i)} \qquad (2)$$

Where $A_i.ss$ is the sentiment score of an aspect A_i, $SW_j.ss$ is the polarity score $\{-1,1\}$ given to the j^{th} sentiment word according to the opinion lexicon, and $dist(SW_j, A_i)$ is the distance between the aspect A_i and the identified sentiment word SW_j. This allows us to identify the opinion words with the highest weight, i.e. the nearest opinion word to the aspect.

3.3 Requirements Classification based on Fuzzy-Kano model

In this module, we use the aspect-sentiment pairs generated previously in combination with the Fuzzy-Kano model to classify the real customer requirements that affect customer satisfaction. In the document collection, each comment is written by a customer, c, to express a sentiment, s, toward several aspects asp of an item, i. By using the quadruplet $\{s, i, asp, c\}$, we form the matrix of aspect and sentiment distribution, denoted as $A = (a_{ij})_{1\le j\le q}^{1\le i\le p}$. For instance, in equation 3, rows represent aspects and columns denote items. The matrix entries represent the customer's sentiment c_{pq} toward the aspect p of the item q. We assign +1 to a positive attitude, -1 to a negative attitude, and 0 to a neutral attitude or no opinion expressed. Then, we construct for each aspect a set of n-dimensional vector distributions. For example, the first row in the matrix indicates that for aspect 1, the customer marks a negative attitude for item 1, neutral or no feeling toward item 2, and a positive attitude for item q. Thus, each row in the matrix constitutes a customer's sentiment vector corresponding to that aspect.

$$A = \begin{bmatrix} -1 & 0 & \cdots & 1 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & 1 \end{bmatrix} \qquad (3)$$

To apply the Fuzzy-Kano, first we calculate for each aspect the customer's degree of preference when the aspect has a functional presence and the customer's degree of dislike when the aspect has a dysfunctional absence or insufficiency. Probability gives real knowledge when the customer feelings are ambiguous or uncertain. So, we calculate such degrees as probabilities of preference and dislike. They are represented, respectively, in equations 4 and 5:

$$preference(c, Asp_i) = \frac{N_s}{p \times q} \times \frac{S_i^+}{S_i} \qquad (4)$$

$$dislike(c, Asp_i) = \frac{N_s}{p \times q} \times \frac{S_i^-}{S_i} \qquad (5)$$

Where $preference(c, Asp_i)$ and $dislike(c, Asp_i)$ represent the probabilities that customer, c, has a positive or negative sentiment, respectively, for aspect Asp_i for a specific item, N_s denotes the number of sentiments either positive or negative expressed by a customer, c, toward some aspects, $p \times q$ refers to the dimension of aspect-sentiment matrix, S_i^+ and S_i^- represent the number of positive and negative sentiments given by c for aspect Asp_i respectively, and S_i is the total number of sentiment attitudes expressed by several customers for the aspect Asp_i.

Second, each of the obtained preference and dislike values refers to a fuzzy set, which contains elements that have varying degrees of membership in the set. These degrees correspond to the five Kano's standard answers ('like', 'necessary', 'neutral', 'unnecessary', and 'dislike'). They are determined using the membership functions where each element of the fuzzy set is mapped to a value ranging from 0 to 1. In particular, we employ in this paper the triangular membership function because of its simplicity in determining the input parameter values, namely the $preference$ and $dislike$ in our case (Umoh and Isong, 2013). According to the triangular membership method, the five Kano's standard answers are represented as five triangular fuzzy numbers between $\tilde{0}$ and $\tilde{1}$, as follows:

- Dislike: $(0, 0, 0.25)$
$$\mu_R(x) = \begin{cases} 0.25 - x & 0 \le x \le 0.25 \\ 0 & otherwise \end{cases}$$

- Unnecessary: $(0, 0.25, 0.5)$
$$\mu_R(x) = \begin{cases} x & 0 \le x \le 0.25 \\ 0.5 - x & 0 \le x \le 0.5 \\ 0 & otherwise \end{cases}$$

- Neutral: $(0.25, 0.5, 0.75)$
$$\mu_R(x) = \begin{cases} x - 0.25 & 0.25 \le x \le 0.5 \\ 0.75 - x & 0.5 \le x \le 0.75 \\ 0 & otherwise \end{cases}$$

- Necessary: $(0.5, 0.75, 1)$
$$\mu_R(x) = \begin{cases} x - 0.5 & 0.5 \le x \le 0.75 \\ 1 - x & 0.75 \le x \le 1 \\ 0 & otherwise \end{cases}$$

- Like: $(0.75, 1, 1)$
$$\mu_R(x) = \begin{cases} x - 0.75 & 0.75 \le x \le 1 \\ 0 & otherwise \end{cases}$$

Where x is the fuzzy set represented by the degree of preference/dislike, and $\mu_R(x)$ is its triangular membership function.

Figure 3 illustrates the graphic presentation of the triangular membership function. The closer the value of preference/dislike degree to a Kano's standard

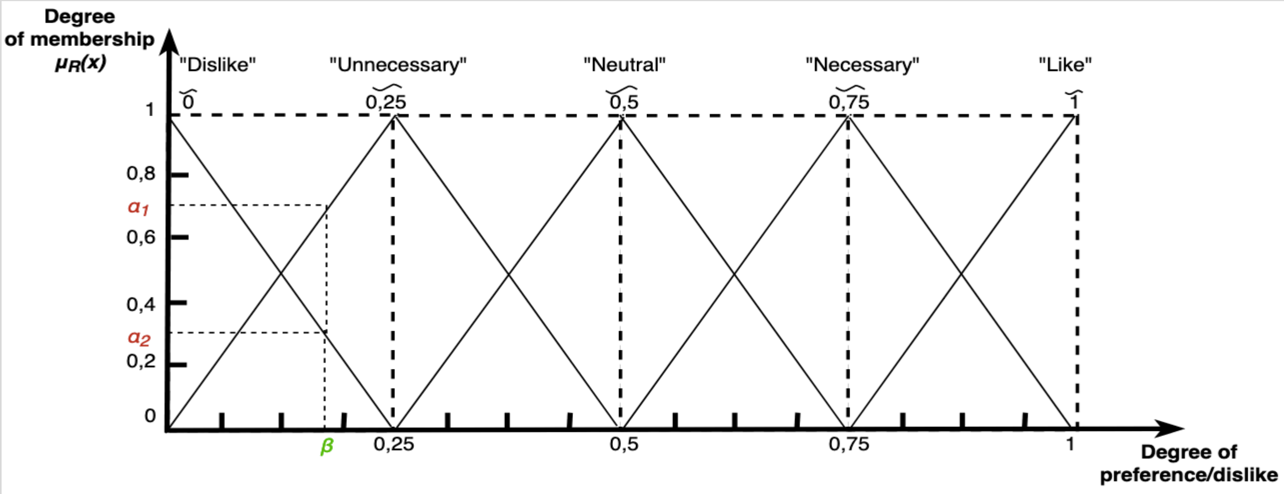

Figure 3 The triangular membership function of the degree of preference/dislike to the Kano standard answers.

answers, the higher the membership degree to it. For instance, while a $preference$ value is located between 0 and 0.25, namely β, the membership degrees to "dislike" and "unnecessary" are α_1 and α_2 respectively.

In Table 2, we illustrate an example of a customer's membership degrees of preference and dislike for aspect 1 in topic 0. Using Table 2 only, it is difficult to determine the proper classification of the customer requirements. Therefore, the customer's membership degrees of preference and dislike can be transformed into two five-vector representations, namely $Pre = \{0.75, 0.21, 0.04, 0, 0\}$ and $Dis = \{0, 0, 0, 0.91, 0.09\}$ as defined in (Lee and Huang, 2009). Then, using a matrix multiplication $Pre^T \otimes Dis$, a 5×5 Kano's two-dimensional Fuzzy relation matrix 'MS' is obtained as:

$$MS = Pre^T \otimes Dis = \begin{bmatrix} 0 & 0 & 0 & 0.68 & 0.06 \\ 0 & 0 & 0 & 0.19 & 0.01 \\ 0 & 0 & 0 & 0.03 & 0.003 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (6)$$

Relative to Table 1 stated in the literature, the customer requirements can also be written as a two-dimensional 5×5 matrix 'ME' as:

$$ME = \begin{bmatrix} Q & A & A & A & O \\ R & I & I & I & M \\ R & I & I & I & M \\ R & I & I & I & M \\ R & R & R & R & Q \end{bmatrix} \quad (7)$$

After 'MS' being obtained, we sum the values of the 'MS' matrix entries with each other if they belong to the same cell in the evaluation matrix 'ME'. As a result, the classification of the customer requirements can be acquired as follows:

$$R = \left\{\frac{0.68}{A}, \frac{0.013}{M}, \frac{0.06}{O}, \frac{0.22}{I}, \frac{0}{R}, \frac{0}{Q}\right\} \quad (8)$$

As mentioned earlier, the Kano model's classification of requirements is qualitative and judged to be ineffective in the quantitative evaluation of customer satisfaction. Therefore, Berger et al. (1993) proposed customer satisfaction coefficients to provide quantitative values of satisfaction and dissatisfaction in case of fulfillment or non-fulfillment of a customer requirement, as given in equations 9 and 10:

$$CS_i^+ = \frac{A_i + O_i}{A_i + O_i + M_i + I_i} \quad (9)$$

$$CD_i^- = -\frac{O_i + M_i}{A_i + O_i + M_i + I_i} \quad (10)$$

Table 2 An example of a customer's membership degree to Kano's standard answers for aspect 1 in Topic 0. S = standard answers; M = membership degrees; Nec = necessary; Neu = neutral; Unnec = unnecessary; Dis = dislike.

S / M	Like	Nec	Neu	Unnec	Dis
Preference	75%	21%	4%		
Dislike				91%	9%

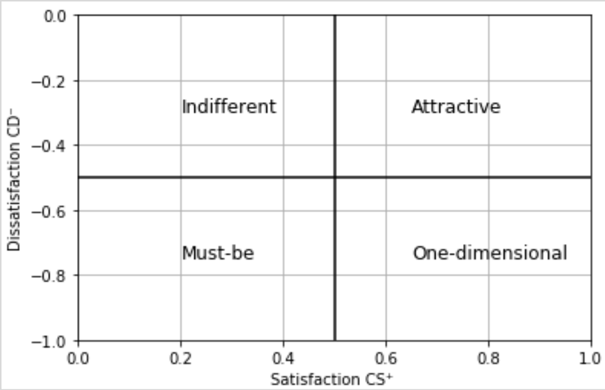

Figure 4 The Kano requirements classification according to customer satisfaction coefficients.

Where $CS_i^+$and CD_i^- are respectively the customer satisfaction and dissatisfaction coefficients of the i^{th} customer requirements, and A_i,O_i,M_i and I_i represent the probability distributions obtained according to the Kano's evaluation for the requirement i. Reverse and questionable requirements were ignored. Note that the minus sign in equation 10 emphasizes the negative impact on customer satisfaction, which will be decreased if these (one-dimensional and must-be) requirements are not included. On the other hand, the value of CS_i^+ is usually positive, indicating that customer satisfaction will be increased by providing these (attractive and one-dimensional) requirements.

A positive satisfaction coefficient ranges from 0 to 1, while a negative satisfaction coefficient runs from 0 to -1. A value of zero implies no impact on customer satisfaction whether the requirement is met or not. The closer CS_i^+ is to 1, the higher the influence of meeting the requirement is on the customer satisfaction, and the closer CD_i^- is to -1, the greater the influence of not meeting the requirement is on the customer dissatisfaction. In this way, all evaluated requirements can be represented graphically through a scatterplot, which is divided into four quadrants according to the satisfaction coefficient values. The X-axis is for CS^+ and the Y-axis is for CD^-. Each customer requirement could be assigned to different quadrants of the scatterplot based on the Kano requirements. As shown in Figure 4, the first quadrant stands for the one-dimensional requirements, the second quadrant stands for the attractive requirements, the third quadrant stands for the indifferent requirements and the fourth quadrant stands for the must-be requirements. Therefore, in designing new products/services, priority should be given to the higher CS^+ and the lower CD^- i.e. Attractive requirements, and when improving an existing product/service, more focus should be given to the high CS^+ value and the high CD^-value, i.e. one-dimensional requirements. This rule guides the decision-maker's team of a company when deciding on which customer requirement has more impact on the company's quality production process.

3.4 Decision Making Analysis driven by Fuzzy-Kano and SWOT

In this module, we propose a bi-layered matrix that maps the Fuzzy-Kano outputs into the SWOT matrix in order to interpret the requirements from the customer and the provider perspectives, as shown in Figure 5. The upper matrix lists the requirements from the customer's perspective. Its horizontal axis represents the fulfillment level of a requirement deducted from the customer satisfaction and dissatisfaction coefficients previously calculated, while the other axis refers to the Fuzzy-Kano requirement's classification. The upper matrix results are mapped into the SWOT matrix (lower matrix). SWOT is used as an analysis tool to provide insights about products by identifying their strengths and weaknesses (i.e. internal factors) along with potential opportunities and threats (i.e. external factors) (Phadermrod et al., 2019).

As can be seen from Figure 5, the upper matrix includes six zones ranging from (a) to (f). Zone (a) contains unfulfilled must-be requirements. The product's provider needs to fulfill these requirements in order to guarantee the minimum quality of the product. Zone (b) includes fulfilled must-be requirements which

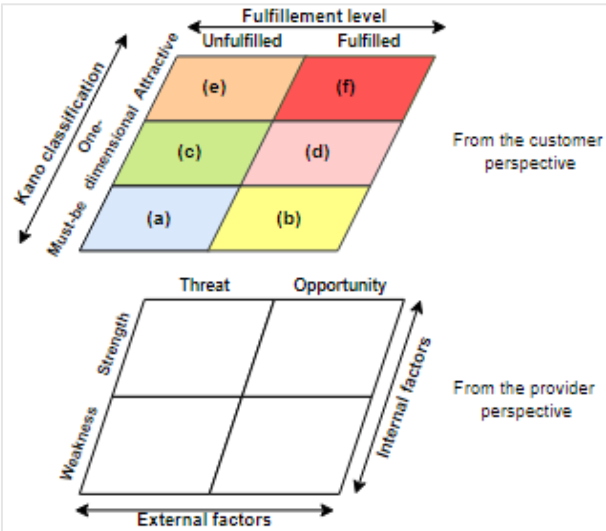

Figure 5 The KANO and SWOT bi-layered matrix.

means that the product already retains a minimum of quality. Zone (c) includes unfulfilled one-dimensional requirements. The product's provider should invest more in improving these requirements in order to avoid customer dissatisfaction and increase customer satisfaction. Zone (e) contains unfulfilled attractive requirements. Even though these requirements will not cause the customer dissatisfaction since they are not expected by the customers, they create a product with a novel attractive aspect that can achieve unexpectedly positive effects. Zones (d) and (f) hold fulfilled/one-dimensional and fulfilled/attractive requirements, respectively. The product's provider does not need to modify the product since those requirements are already at a high level of satisfaction. However, if they make more effective improvements, this can dramatically raise customer satisfaction. The improvements to be made in both zones are different. In (f), improvements are more innovative, while in (d) they are more realistic.

In the lower matrix, the aforementioned zones are mapped to the SWOT matrix. Zones (a) and (c) include unfulfilled/must-be and unfulfilled/one-dimensional requirements which can be regarded as a weakness of the product or even a potential threat for the provider. Therefore, zones (a) and (c) can be put in the W-T cell. Zone (e) holds unfulfilled attractive requirements that can be interpreted differently depending on the studied case. They can be considered as weaknesses that the product's provider can minimize by improving further the product quality and turn those weaknesses into an opportunity. In this case, zone (e) can be put in the W-O cell. On the other hand, those requirements can be considered strengths if the provider includes them in the product and they were not expected by the customers. However, if these requirements do not meet the customers' expectations, then they can become a potential threat. In this case, zone (e) can be put in the S-T cell. Zones (b), (d), and (f) respectively include the fulfilled/must-be, fulfilled/one-dimensional, and fulfilled/ attractive requirements that can be considered strengths since they can be easily fulfilled. In addition, adding new features to the product can be an opportunity to create a new market related to these features. Thus, these zones are put in the S-O cell.

Note that the indifferent requirements are not considered in the bi-layered matrix, simply because they are of little or no consequence to the customer. So, the provider can ignore them to save time, cost, and resources.

4. EXPERIMENTS AND RESULTS

In this section, we conduct a case study to evaluate the effectiveness and feasibility of the proposed framework using online mobile phone reviews collected from Amazon. In the following, we describe our dataset and show potential results.

4.1 Dataset

4.1.1 Preprocessing

In order to evaluate the effectiveness and feasibility of the proposed framework, the first phase consists of collecting and preprocessing the required dataset. In this paper, a dataset of unlocked mobile phone reviews has been selected. This dataset was acquired from Amazon using ("PromptCloud"). It includes 400,000 mobile phone reviews, containing product and customer information, ratings and plaintext reviews. In this study, we conducted the experiments on a subsample of the original dataset, which contains approximately 2000 reviews.

Table 3 Partial demonstration of experimental dataset.

Review	Price	Rating
I feel so LUCKY to have found this used (phone to us & not used hard at all), phone on line from someone who upgraded and sold this one. My Son liked his old one that finally fell apart after 2.5+ yea...	199.99	5.0
It's battery life is great. It's very responsive to touch. The only issue is that sometimes the screen goes black and you have to press the top button several times to get the screen to re-illuminate.	199.99	3.0

Table 3 illustrates some samples from the dataset. Each single review includes a considerable amount of unnecessary data, which must be cleaned to reduce noisy data and extract insightful information such as aspects and sentiments. The preprocessing operations applied in this work include tokenization, stop word removal, transform cases, stemming, and non-alphanumeric character removal. All the preprocessing operations were conducted using the Python NLTK toolkit (version 3.7). In addition, we grouped synonyms to reduce dimensionality by using a manually entered list including the most common synonyms e.g. the words "cellphone", "smartphone", "phones" are all transformed into "phone". Negation

```
Topic: 0 Word: 0.021*"bad" + 0.018*"worth" + 0.016*"turn" + 0.016*"purchase" + 0.013*"plastic" + 0.012*"problem" + 0.012*"dat
a" + 0.012*"may" + 0.011*"products" + 0.011*"battery" + 0.010*"could" + 0.010*"quality" + 0.010*"overall" + 0.010*"model" +
0.010*"problems" + 0.010*"mobile" + 0.010*"using" + 0.009*"price" + 0.009*"running" + 0.009*"loves"

Topic: 1 Word: 0.029*"super" + 0.022*"screen" + 0.020*"fast" + 0.018*"works_fine" + 0.018*"without" + 0.018*"love" + 0.018*"b
ig" + 0.017*"ive" + 0.016*"thin" + 0.015*"box" + 0.015*"us" + 0.014*"dont_NEG" + 0.014*"battery_NEG" + 0.014*"battery_life" +
0.014*"great" + 0.013*"performance" + 0.013*"cant" + 0.013*"away" + 0.012*"happy" + 0.012*"little"

Topic: 2 Word: 0.064*"nice" + 0.023*"unlocked" + 0.020*"well" + 0.019*"good" + 0.018*"times" + 0.017*"bought" + 0.016*"home"
+ 0.015*"grm" + 0.015*"calls" + 0.014*"likes" + 0.014*"husband" + 0.014*"work" + 0.014*"Verizon" + 0.013*"seems" + 0.013*"fee
l" + 0.012*"nd" + 0.012*"android" + 0.012*"like" + 0.011*"person_NEG" + 0.011*"money"

Topic: 3 Word: 0.027*"far" + 0.025*"got" + 0.024*"sim_card" + 0.023*"great" + 0.022*"around" + 0.022*"havent" + 0.020*"androi
d" + 0.018*"update" + 0.018*"amazing" + 0.017*"work_NEG" + 0.017*"straight" + 0.016*"doesnt" + 0.016*"wasnt" + 0.015*"connect
ed" + 0.015*"available" + 0.015*"ok" + 0.014*"constantly" + 0.014*"week" + 0.014*"ago" + 0.014*"never"
```

Figure 6 List of top 20 keywords for the first four topics.

handling is quite important in this study, it assists in improving sentiment analysis accuracy. Therefore, we used the simplest approach proposed in (Das et al., 2001), which is based on appending a negation tag *"_NEG"* to every word found between a negation and the first punctuation mark following it, so as to reverse the polarity of all these words while computing their scores. Misspelling is also taken into consideration since the reviews are usually hand-typed. Some predefined functions from the *"autocorrect package"* are used to deal with misspellings. The POS tagging is used to find adjectives that are considered sentiment words, as well as products' aspects where nouns (NN) and noun phrases (NNP) are considered potential aspect candidates.

Table 4 Setting values for running LDA.

Parameter settings	Values
Number of documents (M)	1593
Number of topics (K)	20
Number of iterations	50
$\alpha = 1/K$	1/20
$\beta = 1/K$	1/20

Table 5 List of aspects along with their sentiment polarity and scores for topic ID = 5.

Aspect(s)	Polarity	Sentiment score
Battery safety	-1	-0.72
Booting time	-1	-0.14
Price	1	0.53
Speakers quality	1	0.83
Battery life	-1	-0.57
Shipping	1	0.33
Screen size	-1	-0.92
Internet speed	-1	-0.10
weight	1	0.69
Camera resolution	1	0.86

Moreover, we applied certain filtering operations, such as: excluding reviews without an adjective POS tag, since sentiments are mainly identified from adjectives; pruning words that are not recognized by the opinion lexicon or Wordnet; and keeping reviews in which an aspect appeared at least once. In the end, the final list was made up of 1763 reviews, which was split into 1593 reviews intended for training and 170 reviews for testing. The testing reviews were chosen randomly, and a new column was added, including aspects and the relative sentiments' polarity.

4.1.2 Extracting Topics and Constructing Aspect-Sentiment Pairs

Before proceeding with the LDA application, we prepared the data for phrase modeling, which consisted of grouping common words that often get a special meaning when they are used together. That is, we built bi-gram phrases from the reviews. Then, using the "*GENSIM*" library, we built our LDA model over the parameters cited in Table 4. The number of topics K was set at 20 to avoid producing a general result with a lack of details. Moreover, a larger number of topics may take longer to converge. For the other parameters, *GENSIM* default values were used.

Through the LDA model, we obtained the first output, namely, the word-topic matrix. It included 20 meaningful topics each represented as a weighted list of words in descending order. Figure 6 indicates the first four topics with the top 20 most frequent words. Topics were inspected by a specific index. Instead, topic names can be defined manually by inferring topics from relevant words' meanings. For instance, looking at topic 1 keywords, we can summarize it to "phone screen and battery performance". The second output generated by LDA was the document-topic matrix. An example of topic allocation to the five first documents (reviews) is illustrated in Figure 7.

By extracting numerous aspects that customers are reviewing and their corresponding sentiments along with the accumulated sentiment scores calculated using equation 2, we gain insights into what negatively or positively impacts product reviews, as well as what the customers like or dislike about the product. Table 5 shows a partial list of such aspects along with their polarity classes and sentiment scores grouping by topic ID 5.

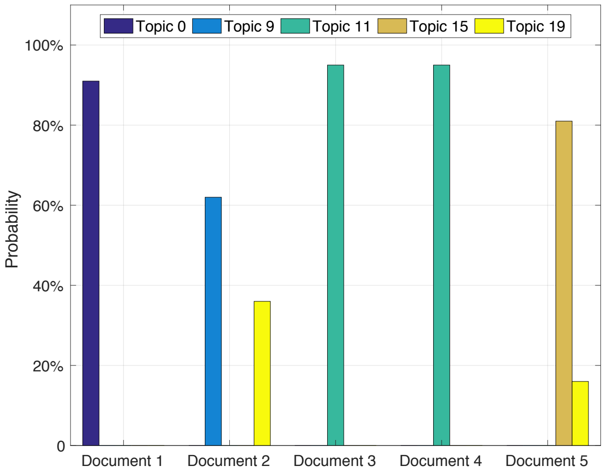

Figure 7 Topic distribution for the first 5 documents.

4.2 Evaluation and Results

4.2.1 Results of the Extracting Aspect-Sentiment Pairs

To evaluate how the extracting aspect-sentiment pairs approach performed, two set of experiments were conducted: (i) measure the effectiveness of the aspects extraction and (ii) measure the effectiveness of the sentiments assignment to the corrected aspects extracted. In this regard, four performance metrics were used: accuracy (Acc), precision (P), recall (R), and F1-score (F_1). Accuracy means how often our model is correct but when used alone, it cannot be trusted to select a well-performing model. Therefore, we used the three other metrics to give more detailed insights into the performance characteristics of our method. Precision refers to the percentage of the relevant data. A higher precision indicates more true positives and less false positives. On the other hand, recall expresses the proportion of all relevant results correctly classified by our model. High recall means less false negatives and high true positives. According to the confusion matrix notations (Ting, 2017), the accuracy, precision, and recall are computed respectively by the following equations:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$P = \frac{TP}{TP + FP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

Where TP is true positives, TN is true negatives, FP is false positives, FN is false negatives. The F1-score combines precision and recall and gives an overall view of the accuracy of the approach. The F1-score is given by:

$$F_1 = 2 * \frac{P \times R}{P + R} \tag{14}$$

In the experiment set (i), TPs refer to the correctly extracted aspects. TNs are the aspects that were discarded by the model and did not appear in the test data either. FPs are words that the model classified as aspects but are not actually aspects. FNs are the aspects that the model labeled as not being aspects when they were actually aspects. In the experiment set (ii), TPs refer to the aspects correctly classified with positive scores. FPs are the aspects incorrectly classified with positive scores. FNs are the aspects incorrectly classified with negative scores.

Table 6 Performance results. Acc = accuracy; Pre = precision.

Experiments set	**Acc.**	**Pre**	**Recall**	**F1-score**
(i) Aspects extraction	97.4%	92.4 %	84.5%	88.27%
(ii) Sentiments assignment	89,8%	90.7%	94.7%	92.6%

Table 6 depicts the accuracy, precision, recall, and F1-score of the proposed aspect-sentiment pairs approach in the experiments set (i) and (ii). As one can see, in (i), the model reports a high precision value (92.4%) meaning that most of the actual aspects are correctly classified with low FP values. The recall rate is 84.5%, suggesting that the most returned aspects are correctly labeled with low FN values. The F1-score is relatively high, meaning that the model represents insightful results in terms of extracting the most discussed aspects of specific products. In (ii), the results are significantly different than the first experiment set. In particular, the F1-score is 92.6%, which indicates that assigning correct sentiments' polarity performs fairly well compared to the aspects' extraction, which reports 88.27%. These results suggest that the extraction of aspect-sentiment pairs performs efficiently in identifying accurate aspects and assigning appropriate sentiments to them. This will help in feeding the Fuzzy-Kano model with accurate inputs, consequently providing valuable business insights.

4.2.2 Results of the Fuzzy-Kano Model

The Fuzzy-Kano model classified the ten aspects previously extracted into must-be, one-dimensional, attractive, and indifferent requirements by calculating their degrees of preference and dislike. Table 7 highlights the findings of the assessed requirements' classification along with their impact on customer satisfaction.

According to the customer satisfaction coefficient (CS^+/CD^-) reported in Table 7, we can represent all the classified requirements via a scatterplot, as shown in Figure 8.

Table 7 Fuzzy-Kano classification and customer satisfaction coefficients results. R.No. = requirement number; A. Req. = assessed requirements; Kano Class = Kano Classification.

R. No.	**A. Req.**	**Kano Class**	**CS^+**	**CD^-**
R_0	Battery safety	Must-be	0.29	-0.83
R_1	Booting time	One-dimensional	0.78	-0.62
R_2	Price	Indifferent	0.06	-0.05
R_3	Speakers quality	One-dimensional	0.54	-0.58
R_4	Battery life	Must-be	0.46	-0.89
R_5	Shipping	Indifferent	0.42	-0.12
R_6	Screen size	Attractive	0.83	-0.36
R_7	Internet speed	One-dimensional	0.60	-0.70
R_8	Weight	Attractive	0.57	-0.32
R_9	Camera resolution	Attractive	0.71	-0.49

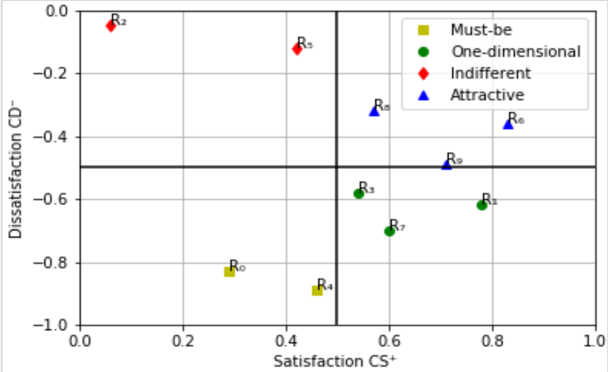

Figure 8 The representation of the Fuzzy-Kano classification results according to CS^+ and CD^-.

From Figure 8 and Table 7, the findings indicate that all the must-be requirements are battery-related, namely, R_0 and R_4 since they have a higher level of dissatisfaction among the customers compared to other requirements. Furthermore, R_1, R_3, and R_7 are all one-dimensional requirements, which implies that customers expect the companies to improve the performance of this product requirement. On the other hand, the attractive requirements such as R_6 and R_9 have a greater impact on satisfaction if fulfilled while R_8 has a relatively lower impact on customer satisfaction when compared to R_1. The indifferent attributes, R_2 and R_5 reflect a low impact on customer satisfaction and dissatisfaction, thus, they should be the last to be focused on over the three other requirements.

4.2.3 Fuzzy-Kano and SWOT Mapping and Analysis Results

In this section, the identified requirements are mapped to the bi-layered matrix. First, they are classified according to the Fuzzy-Kano model from the customer's perspective, then, classified according to the SWOT method from the provider's perspective. The results of the mapping are shown in Figure 9.

Considering the aforementioned results and the analysis reported in the fourth module of our proposed framework, R_0 and R_4 must be fulfilled to guarantee the minimum quality of the product and meet the customers' requirements. These requirements are headed to W-T, which motivate the provider to improve the battery performance, including safety and durability. In addition, internet speed (R_7) is considered W-O from the provider's perspective. Therefore, further enhancements of R_7 will not only lead to increased customer satisfaction but also decrease its dissatisfaction. Requirements in the zones (d)

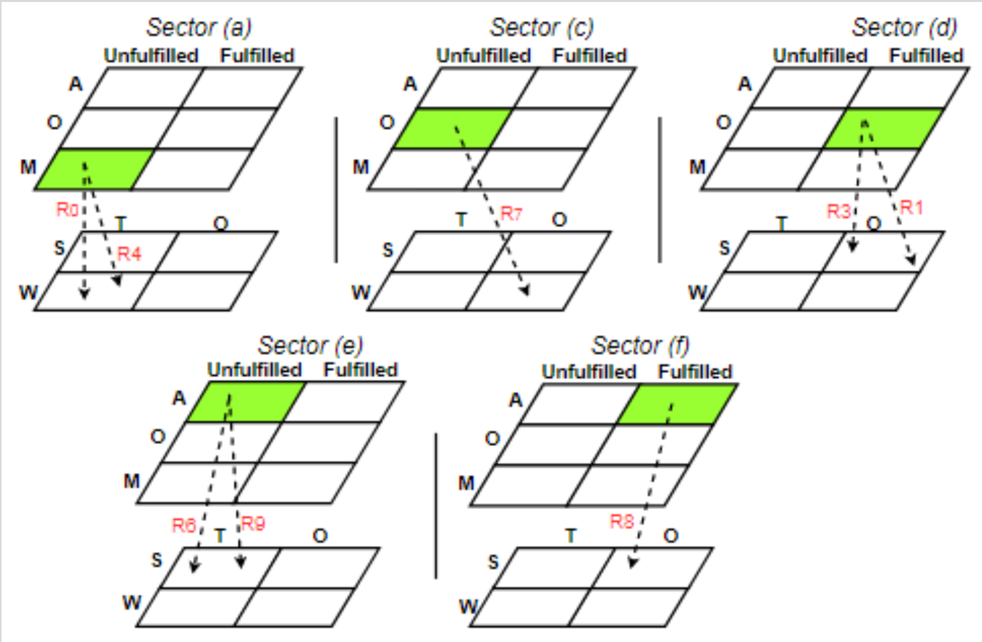

Figure 9 Requirements mapping results.

and (f) such as booting time (R_1), loudspeaker quality (R_3), and weight (R_8) are included in S-O, which means that those requirements are easy to fulfill, and when the provider makes more improvements on them, this will lead to a higher level of customer satisfaction than the current level. The requirements in zone (e) are related to S-T. Even though (R_9) and (R_6) are not expected by the customers, the provider should be able to assess the customers' preferences and overcome the current threat by adding a new value to the product, e.g. improve the camera resolution.

5. CONCLUSION

A good understanding of customer satisfaction is important for the survival of any company in today's competitive market. No business can deny the critical role of the customers' voices in increasing customer satisfaction. However, drawing insights from a huge amount of VOC data is challenging. Thus, companies resort to BI methods and tools to extract actionable information for improving their products and meeting their customers' needs.

This study proposes a decision-making framework for assisting companies in understanding their customers' satisfaction through extracting meaningful insights from online VOC data. The proposed framework consists of four main modules: data extraction and preprocessing, aspect-sentiment pairs extraction using LDA, requirement classification based on the Fuzzy-Kano model, and decision-making analysis driven by Fuzzy-Kano and SWOT.

A case study including online reviews of mobile phones is considered to evaluate the performance of the aspect-sentiment pair extraction module based on several metrics including the accuracy, precision, recall, and F-score. The results showed that the aspects were correctly extracted with a value of 97.4% in accuracy and 92.4 % in precision. Additionally, the sentiments were accurately assigned to the extracted aspects with a value of 89.8% and a precision value of 90.7%. These results constitute an accurate VOC input to feed the Fuzzy-Kano model. They allow us to classify the customer requirements that affect their satisfaction into four main categories: must-be, one-dimensional, attractive, and indifferent. Then, we can map them dynamically to the SWOT matrix in order to provide valuable and interpretable insights for companies.

This framework has some potential limitations that serve as a direction for future work. First, the study is conducted on online reviews which are assumed to be hand-typed and written by honest reviewers (i.e. not fake). However, if these reviews have been maliciously manipulated, they may impact the analysis process and result in biased decisions. An efficient spam review detection technique would be needed to identify whether the reviews are real or fake.

In addition, the aspect-sentiment pairs extraction module deals only with the explicit aspects but does not tackle the implicit ones. For example, in the following sentence *"The battery of this phone is pretty good"*, the aspect *"battery"* appears explicitly. However, in the sentence *"The phone lasts all day"*, the aspect *"battery"* is implicit because it is not stated directly, but only inferred from the meaning of the sentence.

Furthermore, the dynamics of the Fuzzy-Kano model are not included. It considers the evolution of the customer requirements over time. e.g., current attractive requirements can be transformed into must-be requirements in the coming years.

6. REFERENCES

Aguwa, C.C., Monplaisir, L., Turgut, O., 2012. Voice of the customer: Customer satisfaction ratio based analysis. Expert Systems with Applications 39, 10112–10119. https://doi.org/10.1016/j.eswa.2012.02.071

Alghamdi, R., Alfalqi, K., 2015. A survey of topic modeling in text mining. Int. J. Adv. Comput. Sci. Appl.(IJACSA) 6.

Berger, C.C., Blauth, R.E., Boger, D., 1993. kano's methods for understanding customer-defined quality.

Blei, D.M., 2012. Probabilistic Topic Models. Commun. ACM 55, 77–84. https://doi.org/10.1145/2133806.2133826

Carulli, M., Bordegoni, M., Cugini, U., 2013. An approach for capturing the Voice of the Customer based on Virtual Prototyping. J Intell Manuf 24, 887–903. https://doi.org/10.1007/s10845-012-0662-5

Culotta, A., Cutler, J., 2016. Mining Brand Perceptions from Twitter Social Networks. Marketing Science 35, 343–362. https://doi.org/10.1287/mksc.2015.0968

Darling, W.M., 2011. A theoretical and practical implementation tutorial on topic modeling and gibbs sampling, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. pp. 642–647.

Das, S.R., Chen, M.Y., Agarwal, T.V., Brooks, C., Chan, Y., Gibson, D., Leinweber, D., Martinez-jerez, A., Raghubir, P., Rajagopalan, S., Ranade, A., Rubinstein, M., Tufano, P., 2001. Yahoo! for amazon: Sentiment extraction from small talk on the web, in: 8th Asia Pacific Finance Association Annual Conference.

Decker, R., Trusov, M., 2010. Estimating aggregate consumer preferences from online product reviews. International Journal of Research in Marketing 27, 293–307. https://doi.org/10.1016/j.ijresmar.2010.09.001

Farhadloo, M., Patterson, R.A., Rolland, E., 2016. Modeling customer satisfaction from unstructured data using a Bayesian approach. Decision Support Systems 90, 1–11. https://doi.org/10.1016/j.dss.2016.06.010

Farhadloo, M., Rolland, E., 2013. Multi-Class Sentiment Analysis with Clustering and Score Representation, in: 2013 IEEE 13th International Conference on Data Mining Workshops. Presented at the 2013 IEEE 13th International Conference on Data Mining Workshops, pp. 904–912. https://doi.org/10.1109/ICDMW.2013.63

Gioti, H., Ponis, S.T., Panayiotou, N., 2018. Social business intelligence: Review and research directions. Journal of Intelligence Studies in Business 8.

Goodman, J., 2014. Customer experience 3.0: High-profit strategies in the age of techno service. Amacom.

Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. Tourism Management 59, 467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Hofmann, T., 2017. Probabilistic Latent Semantic Indexing. SIGIR Forum 51, 211–218. https://doi.org/10.1145/3130348.3130370

Hu, M., Liu, B., 2004a. Mining Opinion Features in Customer Reviews, in: AAAI.

Hu, M., Liu, B., 2004b. Mining and Summarizing Customer Reviews, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04. ACM, New York, NY, USA, pp. 168–177. https://doi.org/10.1145/1014052.1014073

Jia, S.S., 2018. Leisure Motivation and Satisfaction: A Text Mining of Yoga Centres, Yoga Consumers, and Their Interactions. Sustainability 10, 4458.

KANO, N., 1984. Attractive quality and must-be quality. Hinshitsu (Quality, the Journal of Japanese Society for Quality Control) 14, 39–48.

Lee, H., Han, J., Suh, Y., 2014. Gift or threat? An examination of voice of the customer: The case of MyStarbucksIdea. com. Electronic Commerce Research and Applications 13, 205–219.

Lee, Y.-C., Huang, S.-Y., 2009. A new fuzzy concept approach for Kano's model. Expert Systems with Applications 36, 4479–4484. https://doi.org/10.1016/j.eswa.2008.05.034

Lu, Y., Mei, Q., Zhai, C., 2011. Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA. Inf Retrieval 14, 178–203. https://doi.org/10.1007/s10791-010-9141-9

Miller, G.A., 1995. WordNet: a lexical database for English. Communications of the ACM 38, 39–41.

Nyblom, M., Behrami, J., Nikkilä, T., Solberg Søilen, K., 2012. An evaluation of Business Intelligence Software systems in SMEs-a case study. Journal of Intelligence Studies in Business 2, 51–57.

Park, Y., Lee, S., 2011. How to design and utilize online customer center to support new product concept generation. Expert Systems with Applications 38, 10638–10647. https://doi.org/10.1016/j.eswa.2011.02.125

Phadermrod, B., Crowder, R.M., Wills, G.B., 2019. Importance-Performance Analysis based SWOT analysis. International Journal of Information Management 44, 194–203. https://doi.org/10.1016/j.ijinfomgt.2016.03.009

PromptCloud: Fully Managed Web Scraping Service, n.d. URL https://www.promptcloud.com/ (accessed 9.24.19).

Qi, J., Zhang, Z., Jeon, S., Zhou, Y., 2016. Mining customer requirements from online reviews: A product improvement perspective. Information & Management, Big Data Commerce 53, 951–963. https://doi.org/10.1016/j.im.2016.06.002

Rese, A., Sänn, A., Homfeldt, F., 2015. Customer integration and voice–of–customer methods in the German automotive industry. International Journal of Automotive Technology and Management.

Reyes, G., 2016. Understanding non response rates: insights from 600,000 opinion surveys.

Sabanovic, A., Søilen, K.S., 2012. Customers' Expectations and Needs in the Business Intelligence Software Market. Journal of Intelligence Studies in Business 2.

Saura, J.R., Palos-Sanchez, P., Grilo, A., 2019. Detecting indicators for startup business success: Sentiment analysis using text data mining. Sustainability 11, 917.

Søilen, K.S., Tontini, G., Aagerup, U., 2017. The perception of useful information derived from Twitter: A survey of professionals. Journal of Intelligence Studies in Business, 7(3).

Szolnoki, G., Hoffmann, D., 2013. Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. Wine Economics and Policy 2, 57–66. https://doi.org/10.1016/j.wep.2013.10.001

Ting, K.M., 2017. Confusion Matrix, in: Sammut, C., Webb, G.I. (Eds.), Encyclopedia of Machine Learning and Data Mining. Springer US, Boston, MA, pp. 260–260. https://doi.org/10.1007/978-1-4899-7687-1_50

Tirunillai, S., Tellis, G.J., 2014. Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation. Journal of Marketing Research 51, 463–479. https://doi.org/10.1509/jmr.12.0106

Tontini, G., Solberg Søilen, K., Silveira, A., 2013. How interactions of service attributes affect customer satisfaction: A study of the Kano model's attributes. Total Quality Management & Business Excellence 24, 1253–1271.

Ullah, A.M.M.S., Tamaki, J., 2011. Analysis of Kano-model-based customer needs for product development. Systems Engineering 14, 154–172. https://doi.org/10.1002/sys.20168

Umoh, U.A., Isong, B.E., 2013. Fuzzy logic based decision making for customer loyalty analysis and relationship management. International Journal on Computer Science and Engineering 5, 919.

Xiao, S., Wei, C.-P., Dong, M., 2016. Crowd intelligence: Analyzing online product reviews for preference measurement. Information & Management 53, 169–182. https://doi.org/10.1016/j.im.2015.09.010

Xu, X., Li, Y., 2016. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. International Journal of Hospitality Management 55, 57–69. https://doi.org/10.1016/j.ijhm.2016.03.003