Expert Systems With Applications 236 (2024) 121357

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Extracting product design guidance from online reviews: An explainable neural network-based approach

Seyoung Park, Harrison Kim *

Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, 104 S. Mathews Ave., Urbana, IL 61801, USA

ARTICLE INFO

Keywords:
Data mining
Explainable neural networks
Product design

ABSTRACT

With the development of data mining techniques, user-generated data has become a valuable resource in diverse research areas. In product design research, many studies have been utilizing user data to discover implications for new product design. However, previous works focused on analyzing existing features, whereas companies also need strategies for new features. Some studies discovered new feature ideas from user data but did not provide design implications. This paper addresses the above limitation by extracting comprehensive design implications for both features from user data. The method first defines the lists of existing/new features and collects spec data for these features. Then, it constructs customer choice sets based on the online review and spec data. Regarding spec values, this study presents a newness merit function that reflects the changing value of new features and applies it to the choice sets. The final stage trains a neural network model based on choice sets and conducts SHAP (SHapley Additive exPlanations) on the model. The method draws design implications by further analyzing the resultant SHAP values. The suggested methodology was tested on real-world datasets. The result provides design guidance, including strategies for new features and recommended spec ranges for existing features. This article validates the result by showing that the obtained design implications are consistent with previous market research for product features.

1. Introduction

In today's competitive marketplace, companies devise various strategies to improve customer retention and increase sales performance. New product design plays a significant role in differentiating products and thus obtaining a competitive edge in the market (Gemser & Leenders, 2001; Homburg et al., 2015). When a new product is released, companies appeal to customers by emphasizing value-adding changes in the new product. These changes are divided into two categories according to the feature characteristics (Zhou & Nakamoto, 2007). The first category is the enhancement of existing features. The value can be added by increasing the specification performance of these features. For example, manufacturers in the smartphone market keep increasing the size of the display: 5.1'' (Galaxy S7, 2016), 5.8'' (Galaxy S9, 2018), and 6.2'' (Galaxy S20, 2020). The second category is the introduction of new features. This category adds value by providing new emotional and functional experiences to customers. For instance, the security function of the early smartphones was a passcode. Customers unlock the phone screen by entering numbers or patterns. After a few generations of products, biometric functions were introduced into the market as an alternative to the passcode. It provided more convenient usage to customers because they do not have to remember and input the numbers or patterns every time. Instead, users can unlock their devices by just pressing their fingers on the home button or looking at the camera. In industry, the concept of a new product consists of both categories, which become key selling points (KSPs) of the product. Companies advertise these KSPs in various marketing channels. For example, in the official trailer of Galaxy S10+, Samsung emphasizes the enhanced existing features such as CPU and cameras and introduces new features such as infinity display and wireless power share.

To develop successful product concepts, companies conduct research before they begin new product design. Surveys and interviews are the conventional methods companies use (Brace, 2018) to understand customer needs for existing features and discover ideas for new features. The common approach for the existing features is providing multiple spec options to participants and requesting them to select the best one or prioritize the options based on their preferences (Lee et al., 2019). Regarding the new features, the common method is idea generation through group activities (Daly et al., 2012), which adopt ideation techniques such as brainstorming (Osborn, 1953), Analogical thinking (Casakin & Goldschmidt, 1999), and IDEO cards (IDEO, 2002). The above methods help companies identify customer requirements for

* Corresponding author.
E-mail addresses: seyoung7@illinois.edu (S. Park), hmkim@illinois.edu (H. Kim).

https://doi.org/10.1016/j.eswa.2023.121357
Received 27 March 2023; Received in revised form 13 August 2023; Accepted 26 August 2023
Available online 31 August 2023
0957-4174/

Table 1
Literature review.

Literature		Design implications		Evaluation	
		Existing Features	New Features	Manual	Automated
Rathore and Ilavarasan (2020), Kim et al. (2022)	Sentiment change	✓			✓
Suryadi and Kim (2018), Joung and Kim (2022) Çalı and Baykasoğlu (2022), Du et al. (2022)	Feature importance	✓			✓
Hou et al. (2019), Darko and Liang (2022) Park and Kim (2022a), Lipizzi et al. (2015) Jiao and Qu (2019), Bigorra et al. (2019)	Managerial applications	✓			✓
Tuarob and Tucker (2015), Christensen et al. (2017) Zhang et al. (2021), Goldberg and Abrahams (2022)	New feature ideas			✓	
Our study		✓	✓		✓

products and gain ideas for new features. However, these conventional methods have a limitation in that they take much time and cost.

As an alternative, user-generated data has become a popular resource for customer analysis. Compared to surveys and interviews, user-generated data is highly efficient because a large amount of data can be collected in a short time. In product design research, many studies utilize online user data and draw design implications such as features of customer interests (Chung & Tseng, 2012; Yin et al., 2023), changes in customer sentiment for product features (Kim et al., 2022; Rathore & Ilavarasan, 2020), feature importance (Çalı & Baykasoğlu, 2022; Du et al., 2022), and other managerial applications (Darko & Liang, 2022; Hou et al., 2019; Lipizzi et al., 2015). These studies focused on the analysis of existing features, i.e., the features already embedded in commercial products and familiar to users. As a result, they provided design implications proper for the existing features but not for new ones. For instance, one of the methods was to analyze customer satisfaction for features and suggest those to be improved. This approach cannot be applied to new features because measuring customer satisfaction for unreleased features is difficult, and how to improve the new feature is not clearly defined. However, companies need design implications for new features as well as those for existing features. Specifically, product designers want to know new features welcomed by customers and to understand the reasons behind them so that they devise attractive new features for the upcoming product generation. Also, the designers need to know why certain features failed so that they can address the problem and avoid the same failure. But previous studies did not provide answers to these questions. Some studies focused on discovering ideas for new features from user data (Goldberg & Abrahams, 2022; Tuarob & Tucker, 2015). But they neither answered the above questions because they did not provide evaluation for the extracted ideas or suggest design implications for new features.

To address this problem, this study proposes a methodology that extracts design guidance for both feature categories (existing and new features) from online data. The suggested methodology distinguishes between new and existing features and includes them in a neural network model that predicts customer choices among product alternatives. Then, the method draws design implications for two feature categories by interpreting the trained model. The result can help companies design new products by providing necessary information. For existing features, it provides spec ranges preferred by customers. Regarding new features, the result suggests efficient strategies for new features.

2. Literature review

This section reviews previous research in data-driven design based on user-generated data and discusses the gap between research and industry. Table 1 summarizes the studies presented in this section. They draw various design implications by identifying features of customer interests and analyzing customer attitudes toward them. However, most studies focused on the features already embedded in products, thus providing design implications for existing features only. Some papers studied new features using online data but focused only on idea extraction without analysis of design implications. Few studies evaluated extracted ideas by manual process. The details will be explained in the following subsections.

2.1. User-generated data analysis for NPD

Studies in product design proposed various methods to extract design implications for new products from user-generated data. The most basic approach is to identify features of customer interests. There are different methods to extract feature-relevant keywords from user data, such as association rule mining (Chung & Tseng, 2012; Hu & Liu, 2004), word/phrase vector clustering (Joung & Kim, 2023; Park & Kim, 2022b), latent Dirichlet allocation (LDA) (Blei et al., 2003; Darko & Liang, 2022), and Named Entity Recognition (NER) (Han & Moghaddam, 2021; Yin et al., 2023).

After feature extraction, studies analyzed the user data based on the extracted feature words. One of the approaches is to focus on the changes in customer attitudes toward product features. Specifically, these studies measured the sentiment scores for product features, then compared them before and after the events such as the product launch (Rathore & Ilavarasan, 2020) and COVID-19 (Kim et al., 2022). The results indicated the features with decreased sentiment scores and suggested them as features to be improved. Another approach is to focus on the different importance of product features. Relevant studies analyzed the relationship between customers' sentiments toward product attributes and product performance and then drew the importance of each attribute. Suryadi and Kim (2018) presented feature weights in terms of product sales rankings. Joung and Kim (2022) and Çalı and Baykasoğlu (2022) showed the different impacts of product features on customer satisfaction (ratings). Du et al. (2022) analyzed the importance of product attributes in terms of user satisfaction degrees.

There are other studies extracting managerial applications for product design. Hou et al. (2019) summarize customer opinions on product features by defining linguistic patterns for topics of interest, such as product features, affordance, emotion, perception, and usage conditions. Darko and Liang (2022) and Park and Kim (2022a) proposed methods for online customer segmentation. They presented that customers in different segments have different preferences or importance for product features. Lipizzi et al. (2015) compared customer reactions to two competing products in the early stage of product launch. They constructed a concept map that links users and words based on the conversational patterns in social media and extracted words-only networks. The authors showed significant differences in customer perception of the two products by analyzing the keyword map.

Jiao and Qu (2019) constructed a Kansei knowledge tree that shows the connections between product features and user perceptions. Specifically, the authors extracted hierarchical data consisting of product - aspects - attributes - user concerns. Then they analyzed customers' attitudes towards extracted user concerns. The result suggested design adjustments based on users' emotional needs. Bigorra et al. (2019) and Bi et al. (2019) built online data-based Kano models (Kano, 1984).

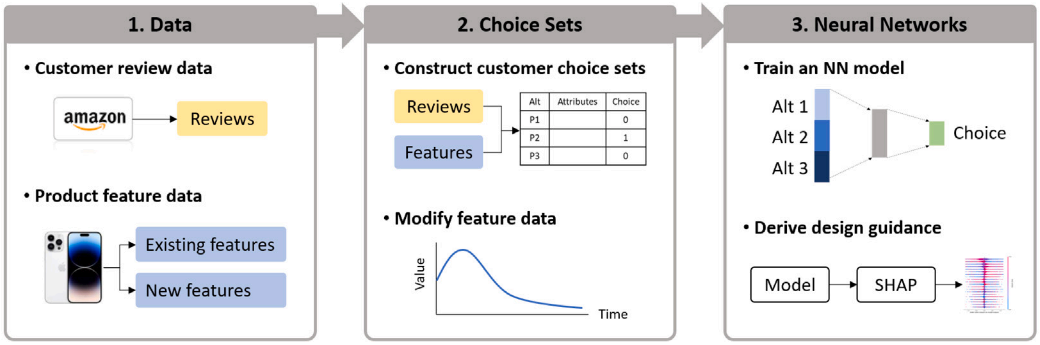


Fig. 1. Overview of the proposed methodology.

Bigorra et al. proposed a mathematical model that classifies product aspects into Kano categories based on customers' sentiments toward them. Bi et al. trained a neural network model that analyzes the relationship between customers' sentiments for product features and their satisfaction (rating) for products. Then they assigned features into Kano categories by analyzing the trained model. The result provided different design strategies for each product feature.

The above studies draw design implications from user-generated data by analyzing features mentioned by customers. Since they focused on the features already on the market, the suggested methods have limitations in that they are appropriate for the existing features but not for the new ones. As mentioned in Section 1, a new product concept consists of both feature categories. Therefore, companies need design implications for new features as well as those for existing features.

2.2. Research for new features

Some studies utilized online user-generated data to discover ideas for new product features. Tuarob and Tucker (2015) suggested mathematical models extracting new features from social media data. First, they defined two product feature sets. One was the set of the ground-truth attributes extracted from product description data, and another was the set of user-discussed features extracted from Twitter data. The authors identified latent features by comparing two datasets and then selected lead users by analyzing Twitter mentions based on the latent features. Then, new feature ideas were obtained by analyzing Twitter mentions of the selected lead users.

Christensen et al. (2017) and Zhang et al. (2021) presented a machine learning-based approach to detect ideas in user-generated data. These papers collected text data from the Lego online community and Amazon reviews respectively. They manually labeled the collected data according to whether the text contains ideas for new products or not. Then classification models were trained based on the labeled dataset. The resultant model is for detecting text data with new feature ideas.

Goldberg and Abrahams (2022) proposed a method to source product innovation ideas from online reviews. The authors adopted and modified the attribute mapping framework to classify reviews into different categories—feature requests, irritators, or complements. After manually dividing review data, they curated smoke terms for innovative ideas and prioritized reviews using these smoke terms. The selected reviews were evaluated by experts, including senior-level managers at the firm.

The above research presented methods to detect user data containing innovative ideas and discover new feature ideas. However, these studies have a limitation because they focused only on idea extraction. In application aspects, the extracted ideas cannot be reflected in product design just because they are mentioned by customers. It is not guaranteed that these new features will be successful when released in the market. Therefore, further evaluation should be conducted on the extracted new features, but few studies evaluated them based on manual analysis. In addition to the lack of evaluation, design implications for new features have been rarely discussed. It is a gap between research and industry because product designers need strategies for innovation to implement effective and successful new features.

As shown in Table 1, this article bridges this gap by proposing a new methodology that extracts design implications for both feature categories based on an automated evaluation.

3. Methodology

Fig. 1 shows an overview of the proposed methodology consisting of three stages. The first stage collects customer reviews and product feature data. The second stage constructs customer choice sets based on the collected data. Then it modifies the feature data to reflect changing values of new features over time. The final stage trains a neural network (NN) model using the constructed choice sets. This study draws design implications by analyzing SHapley Additive exPlanations (SHAP) values obtained from the NN model.

3.1. Data

The suggested method requires two datasets: (i) customer reviews for target products, (ii) spec values of existing and new features of target products. Fig. 2 illustrates the process of data preparation, and each part is explained in the following subsections.

3.1.1. Customer review data

E-commerce websites such as Amazon and eBay are available resources for customer review data. This study collects reviews from Amazon.com, the No. 1 online shopping website (similarweb, 2023). To capture customers' preferences, this research collects the reviews for the top 100 best-selling items in the target product category. These items become the collection of target products and are used in feature data collection. The collected review data contains the product name, product identification number, review date, and review contents.

3.1.2. Product feature data

Next, the method collects product description data, which contains detailed explanations of product features. This data is available from online sources such as manufacturers' official websites, product manuals, and websites specialized in target products. As mentioned in Section 1, this study aims to provide comprehensive design guidance, including existing and new features. Therefore, it is considered necessary to define how to distinguish between two feature categories. Zhou and Nakamoto (2007) said that enhanced existing features *"enable a new product to claim superiority over competitors on the basis of a common ground"* and defined new features as those *"offering something that other brands lack"*. This paper modifies these definitions for automated categorization. First, we set a time stamp. The existing features are those embedded in the products released before the time stamp. The

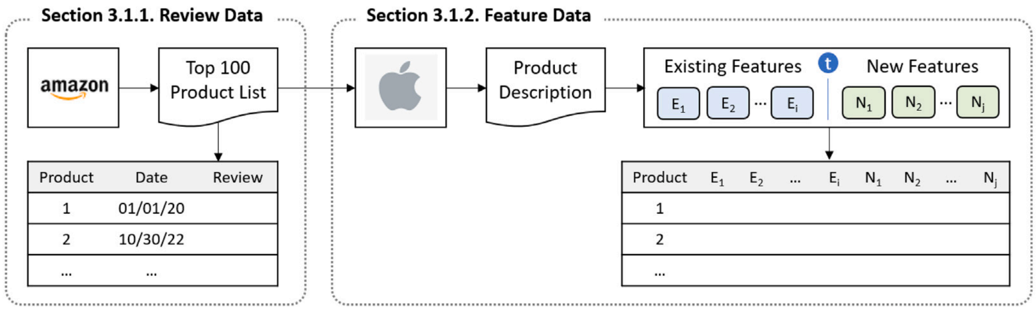


Fig. 2. Data collection and processing.

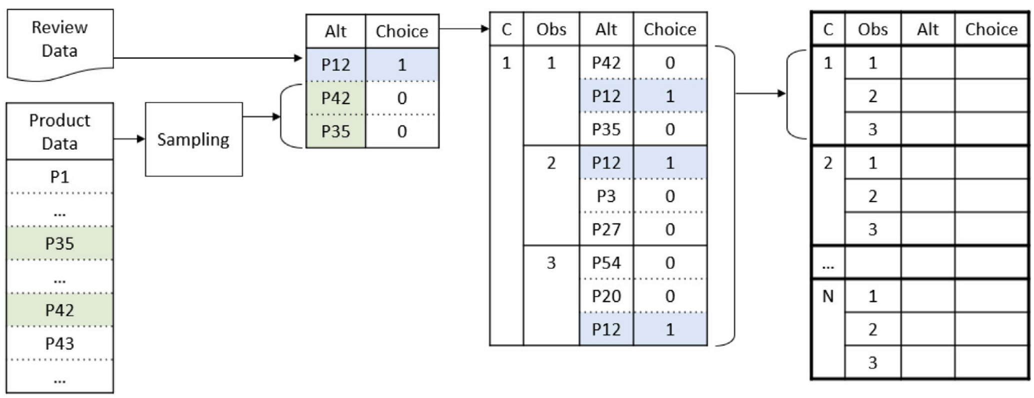


Fig. 3. Constructing choice sets.

new features are those not commercialized before the time stamp. For example, if we set a time stamp on April 2014, the 4G network (launched June 2010) is classified as an existing feature, and the 5G network (released March 2019) belongs to new features. The product release dates are available online, so we can easily collect the data and categorize product features into new/existing ones. In the end, the method establishes two feature lists—one for the existing feature category and another for the new feature category.

After defining feature lists, the method collects spec data of the target products, determined in the previous subsection. The data can be collected from the same resources mentioned above (e-commerce websites, manufacturer homepages, and websites specialized in the target products). Two feature categories have different types of data. For the existing features, discrete spec values are collected. For example, the smartphone memory (ROM) has spec values of {16, 32, 64, 128, 256, 512} gigabytes (GB). On the other hand, the new features have binary values indicating whether the product supports the function or not. For instance, the value of the wireless charging feature is 0 for iPhone 7 and 1 for iPhone 8.

3.2. Choice sets

In the second stage, the method automatically constructs customer choice sets based on the collected review and feature data. Each product alternative in a choice set has existing and new features. The values of existing features are normalized, and those of new features are converted from binary to continuous by a newness merit function (NMF).

3.2.1. Constructing choice sets

There have been a few studies about online data-based customer choice sets. Wang and Chen (2015) and Suryadi and Kim (2019) used random sampling as a baseline method. It assumes a uniform distribution where all products have an equal probability of being selected as an alternative. This study uses the random sampling method and constructs more than one choice sets for each customer referring to survey inquiries (Ryan et al., 2012).

Fig. 3 shows the process of choice set construction. First, the product purchased by a reviewer ($P12$) is assigned to the first alternative in a choice set. Next, the remaining alternatives are randomly selected from product candidates. The candidates include all products in the review data, excluding the products with the same spec configuration with $P12$. In Fig. 3, $P42$ and $P35$ are chosen, completing the choice set. After filling all the alternatives in, the method assigns choice values in the last column. The choice value for the purchased product is 1, and the values for the other options are 0. Since the choice always goes to the first alternative in this process, the completed choice set should be shuffled. The method constructs more than one choice set for each customer, so the final choice sets are managed by customer ID (C) and observation ID (Obs). The process is repeated for all reviewers.

3.2.2. Modifying feature data

After constructing choice sets, the method modifies feature data in them. Since the spec values of existing features have different scales, they are normalized ranging [0, 1]. For example, the largest battery capacity among the entire product becomes 1, and the smallest battery capacity becomes 0. Regarding new features, the method evaluates their values at the review date and converts the binary indicator to continuous values. To reflect their changing values over time (Fabijan et al., 2016; Thölke et al., 2001), this study goes through two steps: (i) calculate the time elapsed after the introduction of the target feature; (ii) define a newness merit functions (NMF) for the target feature and compute its value based on the time from (i).

The initial task for the elapsed time calculation is to determine the release date of new features. This study selects the top brands in the market (US smartphone market for this study) and analyzes released products from 2014 to 2022 to identify the first release date of the new features defined in Section 3.1.2. Once the release date is determined, the elapsed time for each review is calculated. For example, the wireless charging function was presented in the smartphone market on April

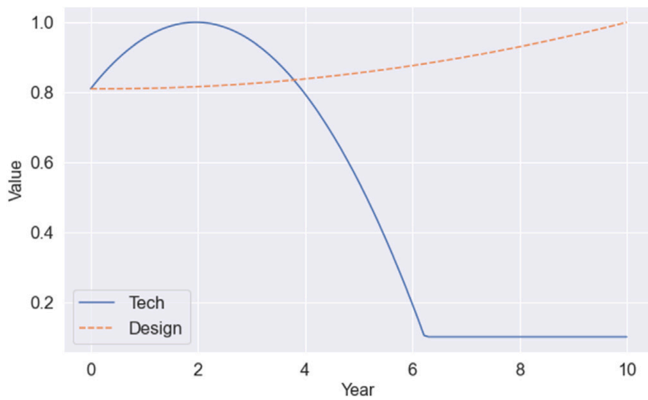


Fig. 4. Newness merit functions for new features.

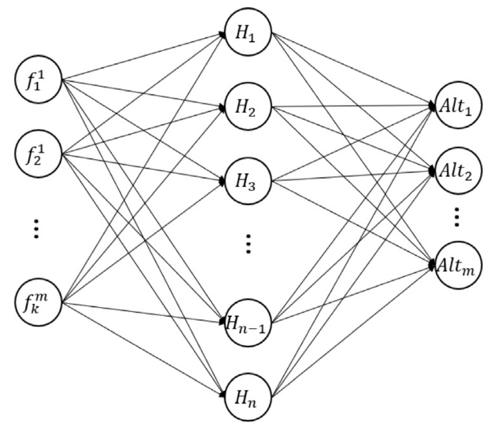


Fig. 5. Neural network structure.

1, 2015. Let us assume a customer wrote a review for the product supporting wireless charging on June 1, 2020. Then the elapsed time is 62 months.

The next task is to define NMF. Talke et al. (2009) suggested an explicit formula relevant to new feature valuation. The authors estimated the effect of the new design and technology on the yearly car sales, as shown in Eq. (1) where t indicates the time since product release. $S_{i,t}$ is the sales of model i during year t. $DN_{i,t}$ represents the design newness of model i in year t, and $TN_{i,t}$ means the technical newness of model i in year t.

$$\begin{aligned} S_{i,t} &= \beta_0 + \beta_1 t + \beta_2 t^2 + (\beta_3 + \beta_4 t + \beta_5 t^2)DN_{i,t} \\ &+ (\beta_6 + \beta_7 t + \beta_8 t^2)TN_{i,t} + \beta_9(DN_{i,t} \times TN_{i,t}) \\ &+ f(Price_{i,t}, CompetitiveIntensityi_{i,t}, t) \end{aligned} \tag{1}$$

This study draws NMF from Eq. (1), assuming a similar valuation for new features across the industry. The NMF for new tech features is shown in Eq. (2). The original equation in Talke et al. (2009) is scaled so that the maximum value becomes 1, the same scale as the existing features. Also, The minimum value is restricted to 0.1 so that installation of a new function can be distinguished from non-installation (value = 0). The unit for the elapsed time stays the same, a year. In other words, when a review is written six months after the feature release, t = 0.5. Eq. (3) is the NMF for new designs with the same value range [0.1, 1]. The original formula is offset so that all new features have the same initial value (≈ 0.8).

$$f_{tech}(t) = max\{3.8 \times (0.213 + 0.051t - 0.013t^2), 0.1\} \tag{2}$$

$$f_{design}(t) = min\{(0.8094 - 0.001t + 0.002t^2), 1\} \tag{3}$$

In Fig. 4, NMF shows how feature values change over time. The value of a new technology increases during the first two years after introduction and keeps decreasing since then. The graph shows a flat line after six years because new feature values are limited to between 0.1 and 1. The value of the new design – exterior or outer shape – does not fluctuate much and remains significantly different from zero.

As explained earlier, new features in choice sets have binary values indicating the existence of the feature. This binary indicator has limitations because it does not provide any information about elapsed time since launch, making it impossible for product design studies to extract customer preferences over time. Therefore, the new feature data in choice sets are modified by Eq. (4), where $I(k)$ is an indicator of feature k. $f_k(t)$ is the NMF corresponding to feature k, and $V_k(t)$ represents the value of feature k at time t. The resulting values contain time information, so design studies can analyze at what point customers adopt new features and how their preferences change.

$$V_k(t) = I(k) \times f_k(t) \tag{4}$$

3.3. Neural networks

In the final stage, the method trains a neural network (NN) model that predicts customers' choices among product alternatives. Subsequently, the method conducts SHapley Additive Explanations (SHAP) on the trained NN model. The resultant SHAP values are further analyzed to derive design guidance for product features.

3.3.1. Training an NN model

The constructed choice sets need to be transformed so that they can be input to an NN model. Fig. 5 shows a structure of the NN model with one hidden layer, where m represents the set size, i.e., the number of alternatives in a choice set. k means the number of product attributes, and n is the dimension of the hidden layer. The input data is an array of specs of all m alternatives, so product attributes in a choice set are merged into a one-dimensional array. Specifically, $m \times k$ matrix is transformed to $1 \times mk$ array. The NN output is the customer's decision in a choice set, which is also a one-dimensional array. For example, the choice result of Obs 1 in Fig. 3 is transformed to [0, 1, 0].

This study tests various NN structures and adopts a shallow neural network (SNN) with one hidden layer because other structures do not improve prediction accuracy, as shown in Appendix. A deep neural network (DNN) does not show better performance for this study, and similar results are found in relevant research (Joung & Kim, 2022; Lee et al., 2018; Park et al., 2023). Regarding sample size, Alwosheel et al. (2018) suggested the minimum sample size for discrete choice analysis, which is fifty times the number of weights in the NN model. The performance of a trained model is evaluated by prediction accuracy, and results will be discussed in Section 6.

3.3.2. Deriving design guidance

After training, the method conducts SHAP (Lundberg & Lee, 2017), which interprets the NN model by analyzing the effect of each input value on the model output. It is based on Shapley values in a game theory introduced by Shapley (Roth, 1988) and uses the additive feature attribution method shown in Eq. (5). The actual output $f(x)$ is approximated by $g(x')$, where x' is a binary variable that maps to the original input x. M is the total number of input features, and weight ϕ_i represents the impact of x_i on the output $g(x')$. Therefore, the output $f(x)$ is approximated by attributing an effect ϕ_i to each feature x'_i and summing all influences.

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x'_i \tag{5}$$

Based on the above method, Lundberg and Lee (2017) proposed SHAP values in Eq. (6), where z' is a subset of input variables x'. It compares

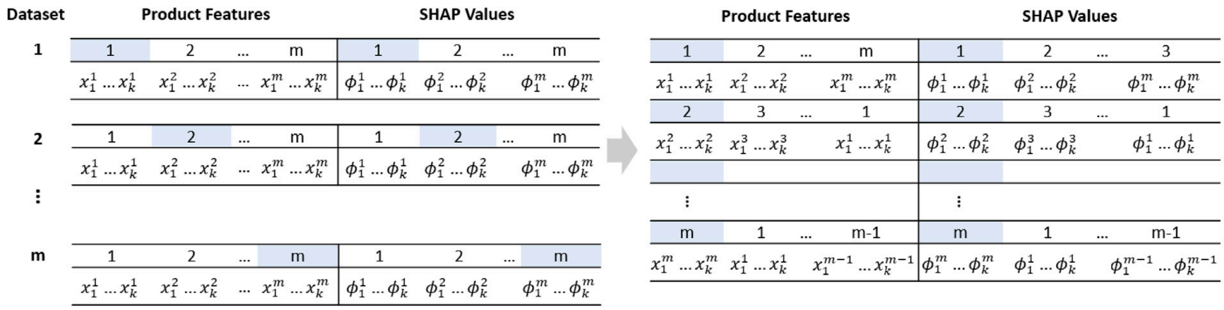


Fig. 6. Transforming SHAP result.

Table 2
Feature lists.

Existing features		New features	
Item	Description	Item	Description
Screen size	Inches	Full/infinity screen	
Screen resolution	Number of pixels	Wireless charging	
Screen type	TFT, IPS, OLED	Security	Fingerprint, FaceID
AP speed	Application Processor, GHz	Network	Generation (5G)
AP count	Number of cores	Exterior design	Frame, back cover
Memory RAM	GB	Water/dust resistance	
Memory ROM	GB	Multi-cameras	
Rear camera	MP	Camera AF/zoom	Auto Focus
Front camera	MP	Power sharing	
Battery capacity	mAh	Sound UX	User eXperience, bluetooth
Battery life	Hrs	Mobile payment	Proximity payment
Price	USD		

the output value for subset z' and the output when attribute i is excluded from subset z', i.e., $f_x(z') - f_x(z' \setminus i)$. The model evaluates the difference in the output made by input i for all combinations of features.

$$\phi_i = \sum_{z' \subseteq x'} \frac{|z'|!(|M| - |z'| - 1)!}{|M|!} [f_x(z') - f_x(z' \setminus i)] \tag{6}$$

Recent studies employing NN models interpreted their models by SHAP (Deng et al., 2023; Joung & Kim, 2022). This study conducts SHAP on the trained NN model to estimate the effects of product features on customers' purchase decisions.

The obtained SHAP values are further analyzed to derive design guidance. In Fig. 6, the tables on the left represent an original format of the SHAP result. x_k^m is the spec value of feature k in alternative m, and ϕ_k^m represents the corresponding SHAP value for x_k^m. Since the NN model in this study is for classification with m classes, SHAP generates m datasets. Each dataset shows the impacts of input attributes on customers' purchase decisions for the mth alternative. The target classes (highlighted) are misaligned. Therefore, the result needs to be transformed so that the first column contains the values for the test design, i.e., the product tested for the customer's purchase. After concatenating all datasets, the method analyzes SHAP values of the test design by Eq. (7).

$$E_i^k = \frac{\sum_{n=1}^{N} SHAP_{in}^k}{N} \tag{7}$$

N is the number of choice sets with spec option i for feature k. Basically, Eq. (7) calculates the mean SHAP value for each spec option i of each product feature k. The magnitude and polarity of the resultant values give guidance for product design. For example, the spec values with positive mean SHAP are recommended options in product design because they have positive effects on customers' choices.

4. Case study

The proposed methodology was tested on a real-world dataset. The target product is a smartphone because (i) a large set of online reviews is available; (ii) consumers are familiar with smartphone features with a high penetration rate (86% globally) (Bankmycell, 2023).

4.1. Data

The review data was obtained from the top 100 best-selling cellphones on Amazon.com. The quality of the review dataset affects the reliability of the prediction result because the input to the prediction model is based on the review data, as explained in Section 3.2.1. In other words, noise in the dataset, such as fake reviews, produces incorrect customer choice sets, resulting in inaccurate design implications. To prevent such situations, this study filtered the collected reviews by verified purchases, an index that Amazon assigns to the reviewers with validated purchase records. After excluding unverified reviews and non-smartphone items, the data contained 44,691 reviews for 85 products, written from July 10, 2017 to March 24, 2022. The product description data was collected from GSMArena.com, a website focusing on mobile devices. It provides detailed information for released smartphones. After applying the method explained in Section 3.1.2, twelve existing features and eleven new features were determined, as shown in Table 2. Regarding these features, the spec values of each product were collected from GSMArena.com since it contains detailed spec information. More importantly, the website provides the 'battery life', which is difficult to collect because the data is based on simulations. The existing features have discrete spec values, and the new features have binary indicators. For example, the feature data of iPhone 11 is {6.1 inches, 828 × 1792, IPS, 2.7 GHz, Octa-core, 4 GB, 64 GB, 12 MP, 12 MP, 3110 mAh, 94 h, \$ 391} for the existing features and {1, 1, 1, 0, 1, 1, 1, 0, 1, 1} for the new features.

4.2. Choice sets

Customer choice sets were built based on the collected data and the sampling method in Fig. 3. Table 3 shows an example of constructed choice sets. This study assigned three alternatives to each choice set referring to the previous research (Suryadi & Kim, 2019; Wang & Chen,

Table 3
Choice sets.

Obs	Alt	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	...	f_{19}	f_{20}	f_{21}	f_{22}	f_{23}	Choice
1	1	5.5	1080	2	2.4	6	3	256	12	7	2691	...	1	0	1	1	1	0
	2	6.3	1080	3	2.8	8	8	256	12	10	3500	...	1	1	1	1	1	0
	3	6.1	1440	3	2.8	8	8	128	18	10	3400	...	1	1	0	1	1	1
2	1	5.8	720	2	1.6	8	2	32	8	5	3000	...	0	0	0	0	0	0
	2	6.1	1440	3	2.8	8	8	128	18	10	3400	...	1	1	0	1	1	1
	3	5.8	1125	3	2.4	6	3	64	12	7	2716	...	1	0	1	1	1	0

Table 4
SHAP dataset after modification.

Product features (Input values)							SHAP values						
f_1^T	f_2^T	f_3^T	...	$f_{21}^{C_2}$	$f_{22}^{C_2}$	$f_{23}^{C_2}$	f_1^T	f_2^T	f_3^T	...	$f_{21}^{C_2}$	$f_{22}^{C_2}$	$f_{23}^{C_2}$
0.54	0.63	0.50	...	0.10	0.00	0.00	0.128	0.242	0.003	...	0.012	−0.034	0.063
0.25	0.28	0.50	...	0.29	0.00	0.00	−0.012	0.021	0.009	...	−0.045	0.001	0.022
0.86	0.25	0.50	...	0.84	0.57	0.10	0.011	0.046	0.004	...	−0.022	−0.020	−0.008
1.00	1.00	1.00	...	0.10	0.81	0.44	−0.003	−0.043	−0.0365	...	0.092	0.585	0.133
0.75	1.00	1.00	...	0.10	0.66	0.00	0.023	0.017	−0.027	...	0.001	−0.363	0.059

2015), and each product has 23 features—12 ($f_1 - f_{12}$) existing and 11 ($f_{13} - f_{23}$) new features.

The values of the existing features were normalized, and the data of new features were modified by the NMF in Eq. (4). The time unit of one year was reasonable because major smartphone manufacturers hold global unpack showcases on a yearly basis. At this point, all feature data had the same scale.

4.3. Neural networks

An NN model was trained based on the choice sets. The feature data in a choice set was transformed into an array of size 69 (3×23) and used for NN input. The choice data in the last column was also converted to an array of size 3 and became the output of the NN model. The dimension of the hidden layer was set to be the same as the size of the input data referring to rule-of-thumb methods (Krishnan, 2021). Since the minimum sample size is 248,400 (50×(69+3)×69) (Alwosheel et al., 2018), this study constructed 268,146 choice sets with 6 observations for each reviewer. These choice sets were divided into two sets for training and testing the model. The division ratio is 8:2 because the model was evaluated by 5-fold validation. Hyperparameters for the NN model were determined by grid search, which tests all combinations of the candidate parameters. Those with the best performance were as follows: {Epochs: 200, Batch size: 20, Optimizer: Nadam, Learning rate: 0.001, Initializer: GloroNormal}. The activation function is ReLU for the hidden layer and softmax for the output layer. The model was implemented by the TensorFlow Keras in PYTHON.

The trained model was analyzed by SHAP. This study used Deep-Explainer of the SHAP package in PYTHON. After going through the modification process in Fig. 6, SHAP values were aligned, as shown in Table 4. Each column represents each feature of each alternative. Specifically, f_k^T indicates the kth feature of the testing alternative, and $f_k^{C_i}$ means the kth feature of competitors. Subsequently, SHAP values of the testing product went through further analysis for each product attribute. For example, the screen size has 14 spec values ranging from 4.0 inches to 6.8 inches. For each spec option, the mean SHAP value for the screen size (f_1^T) was calculated. And the same process was repeated for all 23 feature lists.

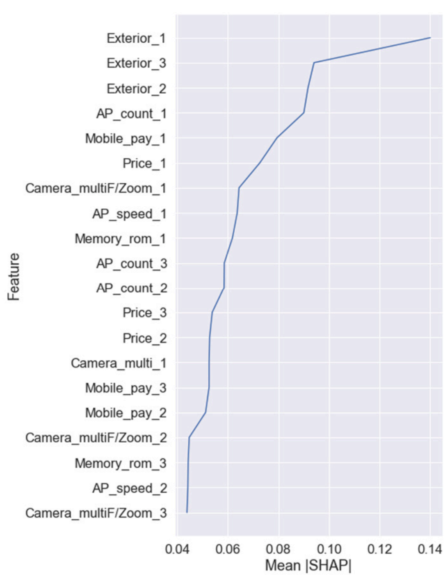
5. Result & validation

This section presents the results of the proposed methodology in two aspects: the effects of product features on customer purchases (5.1) and customer preference for feature specs (5.2, 5.3). The presented results are validated based on relevant research and customer review analysis. Also, the enhanced performance of the prediction model is evaluated (5.4).

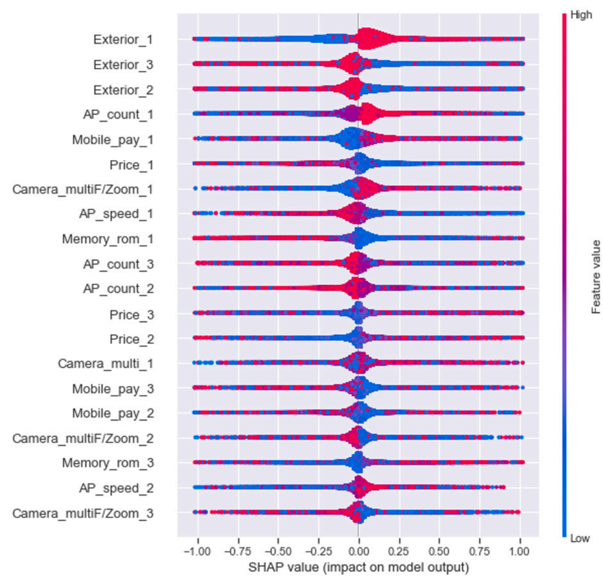
5.1. Effects of product features

The first result is about the effects of product features on customers' choices. Fig. 7(a) shows the magnitude of the impact of each input, i.e., each feature of each product option. The x-axis indicates the average of absolute SHAP values, and the y-axis lists the top 20 input factors. The exterior design of the target product, which is a new feature, has the largest effect on customer choices. Among the top five factors, four items belong to new features. The result implies that new features have higher impacts than existing features. To validate this implication, this study compares the mean of absolute SHAP values of two feature categories. The resultant mean values are 0.051 for existing features and 0.066 for new features. The two-sample t-test says that new features have larger impacts than existing features at the statistical significance of $\alpha = 0.0001$. This result matches the conclusion of relevant research by Zhou and Nakamoto (2007). The authors discovered that when consumers are familiar with a product category, they prefer a product with unique (new) features to one with enhanced (existing) features. Since customers are familiar with the smartphone, it is expected that new features have higher impacts than existing ones.

Fig. 7(b) shows a graphical representation of the SHAP result. The color bar on the right indicates feature spec values. The red color represents higher specs, and the blue color indicates lower specs. For example, the screen 6.7" is close to red and 5.5" is close to blue. The y-axis lists product features in the order of impact magnitude, and the x-axis shows the impact on the model output. In this study, the NN model predicts the customer's choice among different products, so the graph explains the relationship between feature specs and customer choices. For example, the fourth item in the y-axis is AP_count_1, which means the number of AP cores in the target product. The available spec options are quad, hexa, and octa-cores. If the product has octa-core (red), it positively affects the customer's purchase decision. When the product contains quad-core (blue), it has a negative impact on customer choices. This relationship implies that a higher spec is recommended for AP in the smartphone. Overall, Fig. 7 shows that the exterior designs of three alternatives are the top three significant features, which implies that the design has the highest influence on purchase decisions. Radford and Bloch (2011) suggested a similar implication in their research. They analyzed consumer responses to visual product newness using a survey and discovered that perceived newness is a key component of consumers' product evaluations.



(a) Impact magnitude



(b) SHAP graph

Fig. 7. SHAP analysis result.

Table 5
MNL result - existing features.

	Coefficient	P-Value
Screen size	0.3545	0.000
Screen resolution	0.0014	0.000
Screen type	−0.3488	0.000
AP speed	−3.2342	0.000
AP count	0.7684	0.000
Memory ram	0.3379	0.000
Memory rom	−0.0048	0.000
Camera rear	−0.0118	0.000
Camera front	0.0727	0.000
Battery capacity	−0.0009	0.000
Battery life	0.0439	0.000
Price	−0.0039	0.000

5.2. Preference for specs - existing features

The proposed methodology provides a novel design implication, a customer preference for feature specs. A multinomial logit model (MNL) was selected as a baseline model because MNL is a conventional approach for customer choice analysis. Sections 5.2 and 5.3 focus on comparing the results of MNL and the proposed method, while the theoretical background of MNL can be found in Section 6.1. Table 5 shows the result of MNL for existing features. The coefficient values indicate the customers' preferences for features. For example, 0.3545 for the screen size means that the screen size and customer utility for a product have a positive relationship. This implies smartphone users prefer larger screen sizes. On the other hand, −0.0039 for the price indicates that the price and customer utility have a negative relationship, suggesting that customers look for lower prices. Although this result shows the customer's tendency for product features (i.e., prefers higher specs or lower specs), it has a limitation in terms of industrial applications. In manufacturing industries, companies must decide the exact spec of each feature in the product design stage because they need to outsource product components and sign supply contracts. The new method suggested in this research solves this issue.

Table 6 shows the result of the proposed method for existing features. The data is sorted by spec value in ascending order, and the highlighted numbers are positive SHAP values and corresponding spec values. Since the SHAP value represents the impact of the input data on the model output, the highlighted specs have positive effects on customers' purchase decisions. Therefore, the result shown in Table 6 provides recommended spec ranges for the existing features. Regarding the screen, recommend specs are the size 6.1” or ≥ 6.3ε, resolution ≤ 828, and type 2 (TFT). About the AP (Application Processor), the preferred spec is Octa-core with core speed ≤ 2.4 GHz. For the memory feature, RAM ≥ 6 GB and ROM ≤ 64 GB are recommended. The suggested spec range for the camera feature is {7, 11, 13-40} MP for the front camera and {12, 16, 18} MP for the rear one. Regarding the battery feature, the favored battery capacity is ≤ 2942 mAh, and the preferred battery life is between 81 h and 123 h. For the price, customers prefer a low tier in the range of ≤ $291.42. The price information was collected from Amazon, so it has a scale of two decimal places.

The result can be validated by interpreting reviews containing feature keywords (Park et al., 2023). For example, among 100 randomly selected reviews mentioning the screen size, 70% expressed their favor for larger screens. 24% were interested in the screen size but did not specify their preference for larger/smaller ones, and the remaining 6% were satisfied with smaller screens. This result supports customers' preferences for larger screens shown in Table 6. Regarding the battery, the preference for a smaller capacity was also explained by the review analysis. Most reviewers cared more about usage time than the capacity itself, which implies that customers are satisfied with the small battery capacity if it provides decent use time. The preference for longer battery life shown in Table 6 supports this implication.

5.3. Preference for specs - new features

The proposed method also gives a novel design implication for new features compared to MNL. Table 7 shows the conventional result from MNL. It provides the customer tendency for the new features in a similar way to the previous section. For example, 3.8384 for ‘exterior design’ indicates that customers prefer buying products with new looks. On the contrary, −0.4517 for ‘power share’ (a function that uses the smartphone battery to charge other mobile devices) shows that customers do not want this function. This result shows which new features are effective (preferred) and which are not effective (not welcomed). However, product designers need more in-depth explanations for the effectiveness of the new features, so they can adjust strategies and

Table 6
SHAP analysis result - existing features.

Screen size		AP speed		Camera front		Battery cap		Battery life		Price	
Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP
4.0	-0.014	1.3	0.179	2	-0.012	1300	0.073	59	-0.037	28.98	0.127
4.7	-0.055	1.4	0.140	5	-0.024	...	...	...	...	...	...
5.0	-0.019	1.6	0.151	7	0.003	2000	0.017	72	-0.006	278.95	0.010
5.5	-0.028	1.8	0.091	8	-0.008	2659	0.009	73	-0.002	279.00	0.005
5.8	-0.011	2.0	0.057	10	-0.005	2691	0.012	74	-0.004	284.00	0.005
6.0	0.000	2.1	0.030	11	0.001	2716	0.010	78	-0.006	286.99	0.002
6.1	0.001	2.2	0.024	12	-0.011	2942	0.004	79	-0.002	291.42	0.001
6.2	-0.002	2.3	0.022	13	0.023	3000	-0.003	81	0.001	293.00	-0.004
6.3	0.019	2.4	0.005	16	0.020	3046	-0.004	84	0.003	296.95	0.000
6.4	0.016	2.5	-0.013	20	0.043	3110	-0.004	86	0.004	304.94	-0.006
6.5	0.017	2.7	-0.035	25	0.005	3179	-0.003	87	0.002	318.00	-0.010
6.6	0.025	2.8	-0.048	32	0.026	3400	-0.004	...	...	329.00	-0.007
6.7	0.038	3.0	-0.040	40	0.003	...	...	123	0.026	...	...
6.8	0.039					5000	-0.020	144	-0.045	1399.99	-0.133

Screen resol		Screen type		AP count		Memory RAM		Memory ROM		Camera rear	
Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP
480	0.020	1	-0.018	4	-0.190	0.5	-0.030	16	0.069	5	-0.013
720	0.004	2	0.007	6	-0.043	1	-0.049	32	0.063	8	-0.012
750	0.031	3	-0.006	8	0.058	2	-0.029	64	0.042	12	0.012
828	0.015					3	-0.015	128	-0.005	13	-0.009
1080	-0.004					4	-0.002	256	-0.052	16	0.007
1125	-0.014					6	0.024	512	-0.168	18	0.007
1242	-0.017					8	0.038			25	0.000
1440	-0.022					12	0.026			48	-0.005
										50	-0.029
										108	-0.134

Table 7
MNL result - new features.

	Coefficient	P-Value
Network	−0.1798	0.001
Exterior design	3.8384	0.000
Resistance	2.6801	0.000
Full screen	−1.1122	0.000
Camera multi	0.7620	0.000
Camera AF/Zoom	−1.2288	0.000
Battery WC	−0.3258	0.004
Power share	−0.4517	0.000
Security	2.2539	0.000
Sound UX	−0.0719	0.317
Mobile pay	6.1376	0.000

develop new features that can attract customers. This research provides such evaluations for the new features.

Table 8 shows the result of the suggested method for the new features. As explained in Section 3.2.2, feature values range from 0.1 to 1. The value 0 indicates the absence of the feature. The positive SHAP values and corresponding spec values are highlighted. This study divides the new features into three categories based on the highlighted portions, as shown in Table 9. CAT 1 includes Exterior design, Sound UX, Mobile pay, Camera multi, Camera AF/Zoom, and Security. The common characteristic of these features is that the absence of the function negatively affects customer choices. Also, most features were accepted by customers early after their introduction to the market. CAT 2 contains Network and Power share. These features are distinct from others in that their existence negatively affects purchase decisions. The remaining features including Resistance, Full screen, and Battery wireless charging belong to CAT 3. In this category, both install/non-install have positive impacts on customer choices. It can be estimated that customers do not care much about features in CAT 3. This study further investigated the reasons behind different preferences based on relevant research and review data analysis.

In CAT 1, Exterior design has a positive influence regardless of spec levels. This is because the exterior design is the most significant feature, as shown in Section 5.1. Radford and Bloch (2011) presented that higher levels of visual newness engender more emotional responses from customers. Regarding Sound UX, Yoo and Ju (2018) surveyed customer opinions about Bluetooth earsets. They discovered that customers think using wireless earphones makes them look like intelligent and creative people. In addition, users prefer to carry lighter forms of accessories than to hang them. These reasons explain the positive influence of the new Sound UX. For Mobile pay, Liu and Mattila (2019) investigated customer evaluation on Apple Pay. The result showed that customers are satisfied with the mobile proximity payment because it gives them an elevated sense of coolness. Customers' preferences for this function can also be proven by the fact that US proximity mobile payment users keep increasing (Lebow, 2021). The analyses of these three features show that new features providing attractive images have positive effects on customers' purchase decisions.

Two of the other features in CAT 1 are relevant to the camera. Peters and Allan (2018) researched smartphone camera usage. They mentioned that most personal photography (one trillion per year in 2015) was taken by phone cameras these days and found that people use smartphone cameras to capture and share moments in daily life. Therefore, it is natural that enhanced core functions (AF/Zoom) of the camera are welcomed by customers. The multiple cameras also provide improved reliability (facial recognition) and convenience (DXOMARK, 2019) (zoom, HDR, portrait modes, 3D, and low-light photo). The remaining CAT 1 feature, security, is the one that customers are concerned about. Ben-Asher et al. (2011) conducted a survey and discovered that smartphone users are concerned with security and data protection and that most users perceive the passcode as neither secure nor convenient. After the launch of biometric functions such as fingerprint scanning and face recognition, Baqeel and Saeed (2019) investigated the usability of face detection. The authors showed that most users are satisfied with Face ID and believe that the feature locks their smartphones safely and securely. This result implies that new features enhancing core usage or addressing user concerns have a positive impact on customers' purchase decisions.

CAT 2 contains the network feature, which means the advance in technology from 4G to 5G. In the network, the coverage and throughput

Table 8
SHAP analysis result - new features.

Exterior design		Sound UX		Mobile pay		Camera multi		Camera AF/Zoom		Security	
Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP
0.00	-0.189	0.00	-0.022	0.00	-0.054	0.00	-0.023	0.00	-0.101	0.00	-0.031
0.81	0.062	0.38	-0.006	0.10	-0.019	0.38	0.007	0.10	-0.051	0.10	-0.008
0.82	0.083	...	...	0.13	-0.011	0.41	0.007	...	...	0.13	0.005
0.83	0.087	0.49	-0.001	0.16	-0.003	0.44	0.011	0.35	0.000	...	...
0.84	0.079	0.52	0.001	0.19	0.006	0.47	0.010	0.38	0.003	0.35	0.001
0.85	0.081	0.54	0.003	0.23	0.014	0.49	0.011	0.41	0.010	...	...
0.86	0.079	0.57	0.005	0.26	0.020	0.52	0.012	0.44	0.009	0.70	-0.007
0.87	0.083	0.59	0.005	0.29	0.025	0.54	0.011	0.47	0.018	0.72	0.035
0.88	0.083	0.62	0.010	0.32	0.031	0.57	0.015	0.49	0.017	0.74	0.035
0.89	0.089	0.64	0.014	0.35	0.036	...	...	0.52	0.020	...	...
0.90	0.093	0.66	0.018	0.38	0.042	0.97	0.001	0.54	0.025	0.95	0.003
		0.68	0.018	0.41	0.046	0.98	0.032	...	...	0.96	0.005
		...	...	...	...	0.99	-0.003	0.99	0.042	...	...
		1.00	0.042	1.00	0.023	1.00	-0.016	1.00	0.046	1.00	-0.011

Network		Power share		Resistance		Full screen		Battery WC	
Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP
0.00	0.011	0.00	0.016	0.00	0.007	0.00	0.018	0.00	0.014
0.81	-0.053	0.81	-0.058	0.10	0.002	0.57	0.012	0.10	0.007
0.83	-0.058	0.83	-0.039	0.13	0.000	0.59	0.005	...	...
0.84	-0.051	0.84	-0.061	0.16	-0.001	0.62	0.006	0.35	0.005
0.85	-0.055	0.85	-0.062	0.19	-0.004	0.64	0.005	0.38	0.004
0.87	-0.059	0.87	-0.075	0.23	-0.006	0.66	-0.002	0.41	0.001
0.88	-0.062	0.88	-0.079	0.26	-0.008	0.68	-0.001	0.44	-0.006
...	...	...	...	...	...	...	...	...	...
0.96	-0.058	0.96	-0.058	0.81	-0.066	0.96	-0.006	0.95	-0.011
0.97	-0.060	0.97	-0.059	0.83	0.004	0.97	-0.006	0.96	-0.153
0.98	-0.067	0.98	-0.064	0.84	0.114	0.98	-0.006	0.97	-0.121
0.99	-0.068	0.99	-0.059	0.91	0.038	0.99	-0.005	0.99	-0.006
1.00	-0.072	1.00	-0.065	0.92	-0.016	1.00	-0.007	1.00	-0.145

Table 9
New feature categorization.

CAT	Label	Description
1	Positive	The absence of features negatively affect purchase decisions.
2	Negative	Customers are reluctant to features.
3	Neutral	Users do not care about features.

are significant factors that impact customer expectations. According to research by Fletcher (2021), the U.S. is top-ranked for 5G coverage but falls comparatively short and ranked last in throughput. The negative influence of the 5G network may be due to the issue in the essential function of smartphones—data communication. The other feature of CAT 2 is the power share function which uses a smartphone as a power source. Specifically, people can charge mobile accessories (smartwatches, earbuds) and other smartphones using their smartphones without charging cables. It is a new function providing novel usage but has no positive impact on customers' purchases. This study analyzed review data to discover the reasons behind it. Among 7179 reviews for the products with the power-sharing function, only 6 mentioned the feature. Table 10 shows the part of them. The customers said the feature is useful for charging their accessories but not for charging another smartphone. This is probably because battery life is a critical factor in usage (Xu et al., 2016), so customers may be reluctant to share their battery power with others. The CAT 2 shows new features harming essential functions of the product influence customer choices negatively.

Finally, CAT 3 includes Resistance (water/dust), Full screen, and Battery WC (wireless charging). Water resistance was one of the most wanted smartphone features (Tuarob & Tucker, 2015; Yu et al., 2019), and manufacturers implemented the relevant standards. However, many smartphone users complained about water damage to their IP-certified smartphones after falling into less than 1 m depth of water for a few seconds (Yu et al., 2019). Table 8 reflects this problem. In the early stage (spec value 0.83-0.91), the feature had a positive influence on sales because it was expected to satisfy user needs. But the positive impact disappeared shortly after since customers were disappointed with the function. Now people do not care much about water resistance, as indicated by the positive SHAP value for spec 0 (non-waterproof products). The other feature in CAT 3 is Full screen, the display covering the entire front. While the full screen was advertised as a key selling point, customers were not attracted by it in the early stage. This study analyzed the reviews and discovered that consumers have concerns about the camera hole in the middle of the screen, as shown in Table 10. Specifically, customers said *"Full screen video sounds good on paper, but there is a "hole" in the picture where the selfie camera is. It's annoying"*. These customers want to see videos and photos without any disturbances. They would prefer products with normal screens, which resulted in spec value 0 having a positive effect on customer choice. The last feature of CAT 3 is wireless charging for the battery. The word 'wireless' is reminiscent of freely moving while charging, but actual usage is different. First, users need to have their smartphones on top of the charging pad all the time. Second, it takes a longer time due to lower charging efficiency. As a result, customers are satisfied with wired charging as well as wireless charging, as shown in Table 8. The table also indicates that the feature was not popular for some time after the introduction. It is probably because customers need to purchase wireless charging pads, which are expensive. The results in CAT 3 imply that the discrepancy between what companies advertise and what customers experience makes customers indifferent to the new features, even the ones they requested.

5.4. Model evaluation

The trained NN model is evaluated by prediction accuracy shown in Eq. (8), which is the ratio of correct predictions over all observations.

$$Accuracy = \frac{\#\ Correct\ Predictions}{\#\ Total\ Observations} \tag{8}$$

Table 10
Reviews with features.

Power sharing	*"I love the **power share** feature for my watch."* *"I love the fact I can **share my phone battery** with my galaxy buds case if it's running low."* *"The **wireless power share feature** is pretty nifty for charging a Galaxy watch or the Galaxy Buds, but it is too inefficient to use to charge another phone [...] works with the thinnest phone cases."*
Full Screen	*"The **camera being in the center of the screen** [...] That was my biggest worry before purchasing."* *"Full screen video sounds good on paper, but there is a "hole" in the picture where the selfie camera is. It's annoying [...] it shows up on full screen screenshots."*

Table 11
Prediction accuracy (5-fold validation).

	w/o NMF	w/ NMF
NN	73.90%	76.93%

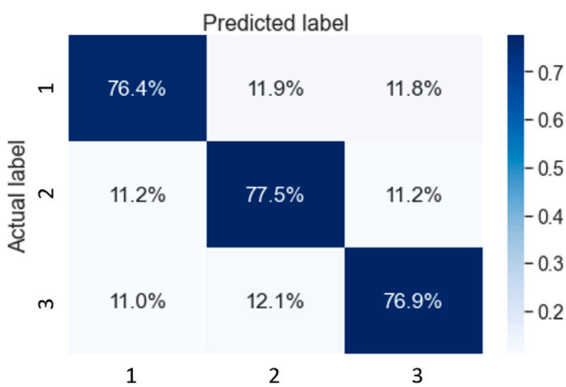


Fig. 8. Confusion matrix (5-fold validation).

Table 11 compares the accuracy of NN models with and without NMF. The first model predicts customers' choices based on the existing and new features, and the values of new features are binary indicators. The resultant accuracy is 73.90%. The second model is the one used in this study. It includes both feature categories and applies NMF for the new features. This model improves the result with 76.93% of accuracy.

The performance of the NN model is further analyzed by a confusion matrix in Fig. 8. There are different ways to represent the confusion matrix with four indices: True positive (TP), true negative (TN), false positive (FP), and false negative (FN). This study calculates the confusion matrix based on TP and FN, and Fig. 8 shows the result from 5-fold validation. The diagonal values show Recall ($\frac{TP}{TP+FN}$), and off-diagonal numbers denote wrong predictions in each class. Specifically, [row i, column j] with $i \neq j$ is the ratio of FN cases where customers are predicted to choose option j when they actually purchase option i. The confusion matrix shows that the NN model in this study provides balanced performance, i.e., the same level of accuracy for all classes.

6. Discussion

In this section, we highlight the methodological contribution of the research (6.1), followed by practical implications (6.2). The generality of the suggested method is validated by a new case study (6.3).

6.1. Methodological contribution

This study aims to provide a solution for design concept generation, i.e., how to configure new and existing features of a product at an engineering level. It requires proper spec ranges for parts and strategies for new features. Hence, the research focuses on two aspects: (i) interpretation of customer choices and (ii) evaluation of new features.

The conventional approach to interpreting customer choices is discrete choice analysis (DCA). The DCA is based on the principle of utility maximization, which means that a decision-maker chooses an alternative with the highest utility among available options (Ben-Akiva & Lermna, 1985). Regarding utility, Eq. (9) shows the most widely used random utility function, where U_{ni} represents the utility of customer n obtained by purchasing product i. ϵ is a random disturbance, and the deterministic part V_{ni} is a function of observable independent variables (Chen et al., 2012). For example, V_{ni} in Eq. (9) is the weighted sum of product features where x_{ik} represents the spec of feature k of product i, and β_{nk} is the importance that customer n has for feature k. Using this function, a multinomial logit (MNL) model yields the choice probability shown in Eq. (10). Pr_{ni} indicates the probability of customer n choosing product i, and J represents a set of available products.

$$U_{ni} = V_{ni} + \epsilon_{ni} = \sum_{k} \beta_{nk} x_{ik} + \epsilon_{ni} \tag{9}$$

$$Pr_{ni} = \frac{e^{V_{ni}}}{\sum_{j \in J} e^{U_{nj}}} \tag{10}$$

The conventional research analyzes customer choice sets shown in Table 3 using MNL and draws the relationship between product features and customer choices. Specifically, the resultant values for β_{nk} imply how much customers value feature k. However, it has limitations because the result cannot define the range of spec options preferred by customers. Let us assume that the β for the screen size is 0.3. The positive β suggests that the customer utility increases with a larger screen size, but the question is "What size is large enough?". Companies may need a lower bound for the screen size since the size is directly related to the cost of products. This research gives solutions to this limitation. The proposed method interprets customer choices by analyzing the marginal contribution of each spec value in each feature, thus providing spec ranges (options) that positively affect customer choices.

The second criterion for theoretical contributions is the evaluation of new features. As mentioned in Section 2, conventional methods for new feature evaluation are interviews and surveys. These traditional methods have limitations in that they require much time and cost. Also, the assessment depends on the subjective opinions of interviewees. This research provides an objective method to evaluate new features based on the newness merit function (NMF) and the marginal contribution of each input data (newness). The proposed method can classify new features into three categories by reflecting customer preferences.

6.2. Practical implications

In Section 5, this study presented the results of the proposed methodology and drew design implications such as feature importance and customer preference for feature specs. These implications can help companies set proper strategies for new product design.

Overall, new features have a higher impact than existing features in the familiar product category, which implies that companies should focus more on attractive new features than on improving existing ones. As mentioned in Section 5.2, this study classifies the new features into

Table 12
Design implications for new features (smartphone).

CAT	Label	Implications
1	Positive	An effective strategy is to develop new features providing differentiated emotional experiences. Upgrading core usage or solving user concerns has positive effects on customers' purchases.
2	Negative	Features that undermine critical functions of the product are not welcomed among consumers.
3	Neutral	New features must guarantee a certain level of quality in usage.

three categories, shown in Table 12. CAT 1 represents new features with positive effects. Analyses in Section 5.2 show that customers highly prefer new features providing positive emotional experiences, e.g., visual newness and cool lifestyles. As pointed out by Yoo and Ju (2018), *"analyzing emotional experience factors and making them reflect specifically on products is important in achieving differentiated competitiveness"*. Therefore, an effective strategy for the new product design is to develop new features providing differentiated emotional experiences. For example, a device with a novel exterior form would be a good strategy for new smartphones. In recent years, foldable phones have provided the highest level of visual newness, and they get a lot of attention from customers. Global foldable shipments grew 64% YoY to 2.5 million units in Q1 2023 (Counterpoint, 2023). CAT 1 also suggests that new features that upgrade core usage or solve user concerns have positive effects on customers' purchase decisions. This result implies that companies can attract customers by identifying core usage of products and improving it or by discovering user concerns and providing solutions to them. CAT 2 is labeled as Negative because customers are reluctant to the features in this category. It shows features that harm critical functions of the product are not welcomed by consumers. Therefore, new features related to essential product functionality should be carefully planned and designed. CAT 3 is named Neutral since users are not sensitive to the features in this category. The relevant research shows that features in CAT 3 failed to meet customer expectations. When there exists a discrepancy between what companies advertise and what people expect, the feature does not get much attention, even if it is something customers have requested. Almsalam (2014) pointed out that customer satisfaction positively relates to customer expectations and perceived service quality. Therefore, newly introduced features must guarantee a certain level of quality in usage.

For existing features, this study provides recommended spec ranges for product design. Companies can narrow down component candidates for new products, thus making part sourcing more efficient. For example, let us assume that a smartphone manufacturer wants to enhance its screen feature with a minimum budget. Then the company can choose 6.1'', the smallest screen size with positive impacts on customer purchase (Table 6). Also, the suggested spec ranges give guidelines for design problems. In practices where product design is formulated as an engineering design optimization (EDO) problem (Gowharji & Whitefoot, 2021), the guidance can be reflected as upper/lower bounds for spec values. Therefore, companies can obtain design solutions with the spec configuration preferred by customers.

6.3. Application domains

The proposed method has a few requirements for its application: (i) The target product needs to have a feature architecture that can be easily classified into new and existing features; (ii) There should exist a broad set of competing products in the market; (iii) A large number of reviews for the target products should be available. Despite these restrictions, we can ensure a certain level of generalizability thanks to the era of digital communications. For example, Amazon.com provides customer reviews for over 20 million products, part of which will satisfy the above requirements.

This section presents a new case study for a different product category with a relatively simple feature structure than smartphones to demonstrate the general applicability of the proposed methodology. The target product was a coffee machine, and 44,275 reviews were collected from the top 13 items on Amazon.com. The review dates ranged from April 23, 2002 to July 26, 2022. The feature list contained 5 existing features (price, capacity, height, width, and depth) and 6 new features (programmable on, auto pause, brew options, fast brew, cup size options, and auto-off). The data went through the proposed methodology and resulted in the NN model with 76.49% accuracy. The result of the SHAP analysis is shown in Table 13. It provided spec guidance for the existing features as in the smartphone products. Regarding new features, the method divided them into the same three categories—CAT 1 (positive), CAT 2 (negative), and CAT 3 (neutral). One can draw design implications for new features in a similar way to the previous case study. For example, the auto pause is a function that stops brewing when users hold the container in the middle of brewing. Customers like this function probably because they want to save time in the busy morning. On the other hand, users do not like the fast brew function. The function may have some drawbacks, such as decreasing the flavor of the coffee. Moreover, this function can be complemented by 'programmable on' which enables users to set up the time for automatic brewing. This new case study shows that the suggested methodology can be applied to different product domains.

7. Conclusion & future works

As user-generated data has become a popular source for consumer product research, there exist various studies to extract design implications from online user data. However, the resultant implications are more proper for the enhancement of existing features rather than the strategy for new features, while companies need strategies for both feature categories. This study bridged this gap by proposing an explainable neural network-based approach to extract comprehensive design strategies from online user-generated data. Specifically, the proposed methodology

- Distinguishes existing/new features and quantifies the time-varying value of new features.
- Trains a neural network model that predicts customer choices and interprets the model using SHAP.
- Draws comprehensive design implications by analyzing the SHAP result.

The methodology was tested on the products of different domains—smartphones and coffee makers. The result provided spec guidance for the existing features, i.e., spec ranges favored by customers. Regarding new features, the result defined three categories based on customer responses and drew innovation strategies by interpreting preference patterns in each category. The obtained design implications were validated by their consistency with relevant market research.

While the result of this study provides comprehensive design strategies, it has some limitations to be addressed. First, this study analyzed the whole customer base as a single group. In reality, customers have different interests and preferences for product features. Therefore, some people may not agree with the interpretations presented in Section 5. In future research, customer segmentation will be considered, and different prediction models will be trained for each segment. It will solve the issue by reflecting different customer preferences in the result.

Table 13
SHAP analysis result - case 2.

Price ($)		Capacity (lb)		Height (inch)		Width (inch)		Depth (inch)	
Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP
24.99	0.034	12	-0.052	7.3	0.058	4.3	0.059	6.3	0.043
34.99	0.013	14	-0.096	9.6	0.038	4.5	0.057	7.4	0.018
39.99	0.017	25	-0.100	10.1	0.045	4.7	0.070	8.0	0.052
49.43	-0.010	40	-0.013	10.2	0.038	4.8	0.060	8.1	0.028
51.25	-0.022	...	...	...	...	...	...	...	...
...	...	48	-0.022	12.2	0.007	10.0	0.001	10.7	0.002
159.95	-0.012	50	0.006	12.5	-0.008	10.1	-0.010	10.8	-0.011
159.99	-0.016	...	...	...	...	...	...	...	...
169.99	0.012	66	0.018	14.0	-0.008	12.0	-0.071	14.1	-0.014
189.99	-0.001	70	0.039	14.6	-0.002	12.2	-0.099	15.2	-0.020
199.99	0.020	75	0.022	15.0	0.009	12.8	-0.051	15.4	-0.028
229.95	0.037	96	0.110	16.4	-0.011	15.3	-0.147	16.0	0.012

Programmable on		Auto pause		Brew options		Fast brew		Cup options		Auto off	
Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP	Spec	SHAP
0.00	-0.026	0.00	-0.021	0.00	-0.013	0.00	0.108	0.00	0.025	0.00	0.038
0.10	0.015	0.10	0.014	0.10	-0.002	0.10	-0.194	0.10	0.006	0.10	-0.004
0.13	0.002	0.13	0.004	0.13	-0.003	0.13	-0.156	0.13	-0.002	0.13	-0.002
...	...	...	...	...	...	0.16	-0.199	0.16	0.006	0.19	-0.007
0.35	0.053	0.35	0.046	0.32	0.021	0.19	-0.139	...	...	...	...
0.38	0.070	0.38	0.005	0.35	0.022	0.23	-0.204	0.84	0.193	0.84	1.396
0.47	0.262	0.47	0.099	0.38	0.013	...	...	0.86	-0.049	0.91	-0.191
0.49	0.095	0.49	0.017	0.41	0.017	0.96	-0.304	0.87	0.039	0.92	-0.003
0.52	2.759	0.52	1.769	0.44	0.014	0.97	-0.332	...	...	0.93	-0.054
...	...	...	...	...	...	0.98	-0.363	0.98	-0.061	0.94	0.022
0.99	0.191	0.99	0.114	0.99	0.026	0.99	-0.334	0.99	-0.013	0.99	-0.131
1.00	0.495	1.00	0.243	1.00	0.073	1.00	-0.310	1.00	-0.041	1.00	-0.249

Second, interactions among features were not discussed. For example, the bigger screen size would require a larger battery capacity because the screen size is directly related to power consumption. Also, there may exist uncovered relationships between existing and new features. Future research will address this limitation by investigating the SHAP result from the NN model. Finally, we will enhance the generality of the proposed method by testing datasets of more diverse product categories. Also, the result with different NN structures (Table A.14) will be further investigated to improve the performance of the method.

CRediT authorship contribution statement

Seyoung Park: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Validation. **Harrison Kim:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix. Different neural network structures

Table A.14 shows the results of different NN structures. Each model was implemented by a keras library in PYTHON (keras.applications - DenseNet121/ ResNet50, keras.layers - GRU/ LSTM). The prediction accuracy was measured by 5-fold validation.

Table A.14
Prediction accuracy.

Category	Model	Accuracy
FFNN	SNN (H = 1)	76.93%
	DNN (H = 5)	76.00%
	DNN (H = 10)	73.95%
CNN	DenseNet121	76.90%
	ResNet50	76.82%
RNN	GRU	58.66%
	LSTM	58.52%

References

Almsalam, S. (2014). The effects of customer expectation and perceived service quality on customer satisfaction. *International Journal of Business and Management Invention*, *3*(8), 79–84.

Alwosheel, A., Cranenburgh, S., & Chorus, C. G. (2018). Is your dataset big enough? sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, *28*, 167–182.

Bankmycell (2023). How many smartphones are in the world? https://www.bankmycell.com/blog/how-many-phones-are-in-the-world.

Baqeel, H., & Saeed, S. (2019). Face detection authentication on smartphones: End users usability assessment experiences. In *2019 International conference on computer and information sciences* (pp. 1–6). IEEE.

Ben-Akiva, M., & Lermna, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. The MIT Press.

Ben-Asher, N., Kirschnick, N., Sieger, H., Meyer, J., Ben-Oved, A., & Möller, S. (2011). On the need for different security methods on mobile phones. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services* (pp. 465–473).

Bi, J.-W., Liu, Y., Fan, Z.-P., & Cambria, E. (2019). Modelling customer satisfaction from online reviews using ensemble neural network and effect-based kano model. *International Journal of Production Research*, *57*(22), 7068–7088.

Bigorra, A. M., Isaksson, O., & Karlberg, M. (2019). Aspect-based kano categorization. *International Journal of Information Management*, *46*, 163–172.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Brace, I. (2018). *Questionnaire design: how to plan, structure and write survey material for effective market research*. Kogan Page Publishers.

Çalı, S., & Baykasoğlu, A. (2022). A Bayesian based approach for analyzing customer's online sales data to identify weights of product attributes. *Expert Systems with Applications*, *210*, Article 118440.

Casakin, H., & Goldschmidt, G. (1999). Expertise and the use of visual analogy: implications for design education. *Design studies*, *20*(2), 153–175.

Chen, W., Hoyle, C., & Wassenaar, H. J. (2012). *Decision-based design*. Springer.

Christensen, K., Nørskov, S., Frederiksen, L., & Scholderer, J. (2017). In search of new product ideas: Identifying ideas in online communities by machine learning and text mining. *Creativity and Innovation Management*, *26*(1), 17–30.

Chung, W., & Tseng, T.-L. B. (2012). Discovering business intelligence from online product reviews: A rule-induction framework. *Expert Systems with Applications*, *39*(15), 11870–11879.

Counterpoint (2023). Global foldable smartphone market continues to expand, underpinned by China. https://www.counterpointresearch.com/global-foldable-smartphone-market-continues-expand-underpinned-china/.

Daly, S. R., Yilmaz, S., Christian, J. L., Seifert, C. M., & Gonzalez, R. (2012). *Design heuristics in engineering concept generation*. Wiley Periodicals, Inc..

Darko, A. P., & Liang, D. (2022). Modeling customer satisfaction through online reviews: A FlowSort group decision model under probabilistic linguistic settings. *Expert Systems with Applications*, *195*, Article 116649.

Deng, S., Zhu, Y., Duan, S., Yu, Y., Fu, Z., Liu, J., Yang, X., & Liu, Z. (2023). High-frequency forecasting of the crude oil futures price with multiple timeframe predictions fusion. *Expert Systems with Applications*, Article 119580.

Du, Y., Liu, D., Morente-Molinera, J. A., & Herrera-Viedma, E. (2022). A data-driven method for user satisfaction evaluation of smart and connected products. *Expert Systems with Applications*, *210*, Article 118392.

DXOMARK (2019). Multi-camera smartphones: Benefits and challenges. https://www.dxomark.com/multi-camera-smartphones.

Fabijan, A., Olsson, H. H., & Bosch, J. (2016). Time to say'good bye': Feature lifecycle. In *2016 42th Euromicro conference on software engineering and advanced applications* (pp. 9–16). IEEE.

Fletcher, B. (2021). U.S. scores top rank for 5G coverage, falls short on speed. https://www.fiercewireless.com/5g.

Gemser, G., & Leenders, M. A. (2001). How integrating industrial design in the product development process impacts on company performance. *Journal of Product Innovation Management: an International Publication of the Product Development & Management Association*, *18*(1), 28–38.

Goldberg, D. M., & Abrahams, A. S. (2022). Sourcing product innovation intelligence from online reviews. *Decision Support Systems*, *157*, Article 113751.

Gowharji, W., & Whitefoot, K. S. (2021). Influence of omitted variables in consumer choice models on engineering design optimization solutions. *Journal of Mechanical Design*, *143*(12), 121–703.

Han, Y., & Moghaddam, M. (2021). Eliciting attribute-level user needs from online reviews with deep language models and information extraction. *Journal of Mechanical Design*, *143*(6).

Homburg, C., Schwemmle, M., & Kuehnl, C. (2015). New product design: Concept, measurement, and consequences. *Journal of Marketing*, *79*(3), 41–56.

Hou, T., Yannou, B., Leroy, Y., & Poirson, E. (2019). Mining customer product reviews for product development: A summarization process. *Expert Systems with Applications*, *132*, 141–150.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177).

IDEO (2002). IDEO method cards. http://www.gillianhayes.com/Inf231F12/wp-content/uploads/2012/10/IDEOMethodCards.pdf.

Jiao, Y., & Qu, Q.-X. (2019). A proposal for kansei knowledge extraction method based on natural language processing technology and online product reviews. *Computers in Industry*, *108*, 1–11.

Joung, J., & Kim, H. M. (2022). Explainable neural network-based approach to kano categorisation of product features from online reviews. *International Journal of Production Research*, *60*(23), 7053–7073.

Joung, J., & Kim, H. (2023). Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews. *International Journal of Information Management*, *70*, Article 102641.

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

Kim, J., Park, S., & Kim, H. M. (2022). Analysis of dynamic changes in customer sentiment on product features after the outbreak of COVID-19 based on online reviews. *Journal of Mechanical Design*, *144*(2).

Krishnan, S. (2021). How to determine the number of layers and neurons in the hidden layer? https://medium.com/geekculture/introduction-to-neural-network-2f8b8221fbd3.

Lebow, S. (2021). Proximity mobile pay is on the rise worldwide. https://www.insiderintelligence.com/content/proximity-mobile-pay-on-rise-worldwide.

Lee, D., Derrible, S., & Pereira, F. C. (2018). Comparison of four types of artificial neural network and a multinomial logit model for travel mode choice modeling. *Transporation Research Record*, *2672*, 101–112.

Lee, S., Kyung, G., Yi, J., Choi, D., Park, S., Choi, B., & Lee, S. (2019). Determining ergonomic smartphone forms with high grip comfort and attractive design. *Human Factors*, *61*(1), 90–104.

Lipizzi, C., Iandoli, L., & Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter streams. *International Journal of Information Management*, *35*(4), 490–503.

Liu, S. Q., & Mattila, A. S. (2019). Apple pay: Coolness and embarrassment in the service encounter. *International Journal of Hospitality Management*, *78*, 268–275.

Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774.

Osborn, A. F. (1953). *Applied imagination*. Scribner's.

Park, S., Joung, J., & Kim, H. (2023). Spec guidance for engineering design based on data mining and neural networks. *Computers in Industry*, *144*, Article 103790.

Park, S., & Kim, H. M. (2022a). Finding social networks among online reviewers for customer segmentation. *Journal of Mechanical Design*, *144*(12), Article 121703.

Park, S., & Kim, H. M. (2022b). Phrase embedding and clustering for sub-feature extraction from online data. *Journal of Mechanical Design*, *144*(5).

Peters, C., & Allan, S. (2018). Everyday imagery: Users' reflections on smartphone cameras and communication. *Convergence*, *24*(4), 357–373.

Radford, S. K., & Bloch, P. H. (2011). Linking innovation to design: Consumer responses to visual product newness. *Journal of Product Innovation Management*, *28*(s1), 208–220.

Rathore, A. K., & Ilavarasan, P. V. (2020). Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products. *International Journal of Information Management*, *50*, 111–127.

Roth, A. E. (1988). *The shapley value* (pp. 31–40). Cambridge University Press.

Ryan, M., Kolstad, J. R., Rockers, P. C., & Dolea, C. (2012). *How to conduct a discrete choice experiment for health workforce recruitment and retention in remote and rural areas: a user guide with case studies*: *Technical report*, The World Bank.

similarweb (2023). Top websites ranking for e commerce and shopping in the world. https://www.similarweb.com/top-websites/category/e-commerce-and-shopping/.

Suryadi, D., & Kim, H. (2018). A systematic methodology based on word embedding for identifying the relation between online customer reviews and sales rank. *Journal of Mechanical Design*, *140*(12), Article 121403.

Suryadi, D., & Kim, H. M. (2019). A data-driven methodology to construct customer choice sets using online data and customer reviews. *Journal of Mechanical Design*, *141*.

Talke, K., Salomo, S., Wieringa, J. E., & Lutz, A. (2009). What about design newness? Investigating the relevance of a neglected dimension of product innovativeness. *Journal of product innovation management*, *26*(6), 601–615.

Thölke, J. M., Hultinka, E. J., & Robbenb, H. S. (2001). Launching new product features: a multiple case examination. *Journal of Product Innovation Management: An International Publication of the Product Development & Management Association*, *18*(1), 3–14.

Tuarob, S., & Tucker, C. S. (2015). Automated discovery of lead users and latent product features by mining large scale social media networks. *Journal of Mechanical Design*, *137*(7).

Wang, M., & Chen, W. (2015). A data-driven network analysis approach to predicting customer choice sets for choice modeling in engineering design. *Journal of Mechanical Design*, *137*.

Xu, W., Liang, W., Peng, J., Liu, Y., & Wang, Y. (2016). Maximizing charging satisfaction of smartphone users via wireless energy transfer. *IEEE Transactions on Mobile Computing*, *16*(4), 990–1004.

Yin, S., Wang, Y., & Shafiee, S. (2023). Ranking products through online reviews considering the mass assignment of features based on BERT and q-rung orthopair fuzzy set theory. *Expert Systems with Applications*, *213*, Article 119142.

Yoo, H. S., & Ju, D. Y. (2018). Analysis of the emotional user experience elements of wireless earphones. In *International conference on applied human factors and ergonomics* (pp. 247–255). Springer.

Yu, Q., Xiong, R., Li, C., & Pecht, M. G. (2019). Water-resistant smartphone technologies. *IEEE Access*, *7*, 42757–42773.

Zhang, M., Fan, B., Zhang, N., Wang, W., & Fan, W. (2021). Mining product innovation ideas from online reviews. *Information Processing & Management*, *58*(1), Article 102389.

Zhou, K. Z., & Nakamoto, K. (2007). How do enhanced and unique features affect new product preference? The moderating role of product familiarity. *Journal of the Academy of Marketing Science*, *35*(1), 53–62.

Seyoung Park is currently a postdoctoral researcher in the department of industrial and enterprise systems engineering at the University of Illinois Urbana-Champaign (UIUC). She received Ph.D. in industrial engineering at UIUC. Her main research interests include data mining, machine learning, and data-driven product design.

E-mail: seyoung7@illinois.edu

Harrison Kim is currently a professor in the department of industrial and enterprise systems engineering at the University of Illinois Urbana-Champaign. He received Ph.D. in mechanical engineering at the University of Michigan. His research interests include engineering systems design and optimization, green, sustainable product design, manufacturing and remanufacturing, product design analytics, and trend-mining design.

E-mail: hmkim@illinois.edu