Article

User Need Prediction Based on a Small Amount of User-Generated Content—A Case Study of the Xiaomi SU7

Lingling Liu and Biao Ma *

School of Art and Design, Guilin University of Technology, Guilin 541006, China; 2007018@glut.edu.cn
* Correspondence: mabiao1023781430@gmail.com

Abstract: (1) Background: In the current competitive market environment, accurately forecasting user needs is crucial for business success. By analyzing user-generated content (UGC) on social network platforms, enterprises can mine potential user needs and discern shifts in these needs, thereby enabling more efficient and precise product design that aligns with user needs. For newly launched products with a limited presence in the market, the scarcity of UGC poses a challenge to businesses seeking to predict user needs from small datasets. (2) Methods: To address this challenge, this paper proposes a model using correlation analysis (CA) and linear regression (LR) combined with multidimensional gray prediction (a CA-LR-GM (1, N) model) to help enterprises use small sample data to predict user needs. Using the UGC of the Xiaomi SU7 as a case study, this paper demonstrates the prediction of user needs for the vehicle and refines the prediction outcomes through an optimization design informed by the principle of optimal key feature distribution. (3) Results: The findings validate the feasibility of the proposed theoretical framework, offering a technical solution for the identification and prediction of user need trends. (4) Conclusions: This research puts forward strategic recommendations for enterprises regarding the optimization of their products.

Keywords: UGC; CA-LR-GM (1, N) model; predict user needs; product optimization strategy

1. Introduction

With the rapid development of mobile Internet, big data, and the enhancement of consumer awareness, user needs exhibit characteristics of timeliness and complexity [1]. However, product development cycles are often lengthy and cannot keep pace with the dynamically changing user needs in real time, thus failing to make corresponding adjustments during the development phase [2]. Consequently, the capability to predict user needs comprehensively and accurately is crucial for a product to distinguish itself from numerous similar offerings and achieve development success. Xia et al. [3] have argued that product packaging design significantly influences consumer purchasing behavior. They proposed a deep learning-based approach to forecast user needs regarding packaging design. Concurrently, an increasing number of consumers voice their genuine opinions on product needs via e-commerce platforms and social networks, providing valuable market intelligence for enterprise growth. Scholars have started focusing on the analysis of online reviews, with Li et al. [4] leveraging semantic and sentiment similarity in online reviews to capture user needs. Lin et al. [5] introduced a competitive intelligence mining framework to extract user perceptions from user-generated content. These methodologies rely on substantial UGC data. However, for newly launched or niche products, which often lack extensive UGC, it remains challenging to predict user needs accurately using existing methods, especially when the goal is to determine the optimization direction for subsequent product iterations.

As a complex product, vehicles have a relatively long production cycle. It is necessary to obtain future user needs in the early stage of vehicle research and development and conduct product research and development based on them. In this way, the problem of companies spending a lot of time and resources on designing products that do not meet

user needs can be avoided. In the field of predicting the needs of vehicle users, existing research is still relatively lacking. Lash et al. [6] explored what factors influence users' intention to purchase electric vehicles and provided reasonable guidelines for enterprises to formulate marketing strategies. Zou et al. [7] studied users' preferences for shared electric vehicles by using the method of sentiment analysis based on big data, but did not predict future user needs. Moreover, as mentioned above, most of the existing research is based on a large amount of UGC in the vehicle industry as data for research. For vehicles that have just been launched and lack UGC, they also need to obtain user needs to optimize the next-generation products.

The multidimensional gray prediction model (GM (1, N)) is a method for constructing a mathematical model to forecast multivariate situations based on a small amount of incomplete information [8]. In this paper, GM (1, N) can be utilized to predict user needs based on a limited amount of UGC. However, GM (1, N) is applied with the assumption that there is no significant correlation between the factors. Yet, in reality, there is often a correlation between needs, and predicting a need in isolation may lead to a decrease in the accuracy of the prediction [9]. Correlation analysis is a statistical method for studying the relationship between two or more random variables in the same dataset [10]. Linear regression uses mathematical statistics in regression analysis to determine the interdependent quantitative relationship between two or more variables [11]. This paper will combine correlation analysis with linear regression to mine the connections between user needs, thereby enabling a more accurate analysis of user needs and improving the accuracy of need prediction. Consequently, the accuracy of need prediction can be enhanced.

In summary, while current research predominantly relies on extensive UGC data for need prediction, it often overlooks the necessity for predicting needs in the early stages of product launches. These studies tend to focus on the trend of individual needs without considering their interplay. The GM (1, N) model, capable of predicting from a limited dataset, combined with correlation and linear regression analyses, addresses this shortcoming by accounting for the mutual influences among needs, thereby refining the prediction indicators. Consequently, this paper introduces a novel approach to mining user needs from a modest amount of UGC and proposes a CA-LR-GM (1, N) model-based prediction method. This methodology aims to more precisely anticipate future user needs for products and to inform strategic product design planning within enterprises.

The main research questions of this paper are as follows:

(1) How to use a small amount of UGC to predict user needs.

(2) How to take into account correlations among user needs so as to improve the prediction accuracy when predicting user needs.

(3) How to guide product design to meet user needs after obtaining the predicted user needs.

This paper makes the following principal contributions:

(1) Recognizing the necessity of predicting user needs for products with a limited volume of UGC, we collect UGC for new products, employ a K-means topic model for clustering and analyzing comments, and calculate metrics of attention and satisfaction. Subsequently, we utilize the GM (1, N) model to forecast need indicators.

(2) Accounting for the interplay between various needs, this study employs correlation analysis and linear regression to dissect the influence dynamics among these needs. The methodology further involves the re-quantification of the attention and satisfaction metrics associated with the needs. This approach, by extension, enhances the predictive accuracy of user needs derived from a limited dataset of UGC.

(3) We introduce a strategy for product optimization grounded in the principle of optimal key feature distribution. This strategy mitigates the reliance on designers' subjective experience and clarifies the direction of product enhancement.

The remainder of this paper is structured as follows: Section 2 provides a literature review on user need prediction and UGC; Section 3 details our methodology; Section 4 presents a specific case study and the analysis of results; and Section 5 offers concluding remarks.

2. Literature Review

2.1. User Need Mining and Predicting

Many scholars have begun to explore how to mine valuable information from historical data and build practical and applicable need prediction models. The literature review on user need acquisition and prediction is summarized in Table 1.

Table 1. Literature review on acquiring and predicting methods of user needs (source: own elaboration).

Author	Reference	Year	Research Methods and Contents
Liu et al.	[12]	2018	Proposed a visually-aware temporal rating model with topics using review text to help mine visual dynamics and non-visual features for a rating prediction task.
Yusuf-Asaju et al.	[13]	2019	Employed machine learning algorithms for the prediction of customer satisfaction.
Lee et al.	[14]	2019	Proposed predictive global sensitivity analysis based on user-generated reviews for the prediction of the demand for hyperdifferentiated products.
Dou et al.	[15]	2019	Integrated a fuzzy Kano model and optimized gray model for the prediction of dynamic customer requirements and used House of Quality to calculate the optimal improvement plan.
Ali et al.	[16]	2020	Proposed an ontology-based method to extract the design features of products from online customer reviews.
Zhang et al.	[17]	2020	Used gray theory combined with Kansei engineering to mine the macro- and microscopic factors in product color design decision process based on the product color brand image.
Jiang et al.	[18]	2021	Used opinion mining and a fuzzy time-series method to predict the weights of customer needs.
Ostasz et al.	[19]	2022	Used the naïve Bayesian classifier for the prediction of the direction of product improvement.
Cheng et al.	[20]	2022	Proposed a model based on the quantitative Kano model and the customer satisfaction degree for the prediction of the customer satisfaction degree of the new component scheme.
Wang et al.	[21]	2023	Proposed an evolving demand satisfaction (EvoDESA) model to model a user's demand evolution for next-basket prediction.
Luo et al.	[22]	2023	Constructed an importance–satisfaction gap analysis (ISGA) model to obtain the changing trend of Chinese car users' needs from UGC.
Zhang et al.	[23]	2023	Used the gray–Markov model to predict the needs of automobile and mobile phone users from online reviews.
Fan et al.	[24]	2024	Studied the data of e-commerce product recommendations from the perspective of user preference and association rules.
Li et al.	[25]	2024	Developed a novel eye-tracking-based assessment tool to investigate user preference towards humanoid robot appearance design.
Luo et al.	[26]	2024	Used the chaos theory and cvxEDA algorithm to extract the features of the front face of automobiles and study the relationship between them and user perception.
Dinaryanti et al.	[27]	2024	Used topic modelling with the LDA algorithm and sentiment analysis with the naïve Bayes algorithm to analyze the factors that consumers consider when choosing smartphones.

It can be determined from the above literature review that the prerequisite for accurately predicting user needs is to accurately obtain their needs [21]. Research efforts in this domain typically fall into two broad categories. The first pertains to the realm of mobile application recommendations, where analysis of user behavior is utilized to discern needs and subsequently recommend content of potential interest to the user. For instance, Fan et al. [24] enhanced the precision of e-commerce recommendation systems by examining product recommendation data through the lens of user needs and association rules. The second category involves examining user needs for product features within the design sphere, aiming to refine product solutions to align with user needs [16]. This area, which is the focus of the present paper, has seen less research. Within this field, Li et al. [25] suggested employing eye-tracking tools to assess user needs regarding the

aesthetics of humanoid robots, while Luo et al. [26] explored the relationship between product personalization and user needs using questionnaires and assessments of physiological electrical signals. Although these methods leverage information collection tools to capture behavioral data and minimize subjective bias, they are often limited by small sample sizes and the susceptibility of users to experimental conditions when selecting needed elements. UGC, which comprises users' voluntary online expressions of their product needs, boasts a substantial user base and serves as a viable data source for user need analysis. Luo et al. [22] employed term frequency-inverse document frequency (TF-IDF) to evaluate text significance and uncover automotive user needs from online UGC. Li et al. [28] applied Word2Vec to transform text into word vectors for sentiment analysis and need determination. Dinaryanti et al. [27] conducted topic modelling and sentiment analysis on UGC to identify the factors that influence users' flagship smartphone choices. The existing body of research indicates that the study of user needs through UGC has reached a considerable level of maturity.

Research in the domain of user need prediction is relatively scarce. Liu et al. [12] suggested predicting visual need components by examining fashion trends. Zhang et al. [17] employed a gray prediction model to forecast current product color trends. This paper also finds it pertinent to utilize gray prediction models for predicting user needs, with a critical consideration of the interrelations between various need indicators. Guo et al. [29] conducted a comprehensive correlation analysis of multiple perceptual aspects of color design. Li et al. [30] applied linear regression to correlate biometric data with user perceptual attributes, thereby informing bionic product design. Consequently, this study refines need indicators by integrating correlation and linear regression analyses, which account for the intricate interplay among these indicators. Furthermore, while the existing literature on user need prediction often discusses product optimization strategies at a macro level—Zhang et al. [23], for instance, outlined strategies for various aspects of automotive product optimization, including exterior and interior design—concrete implementation guidelines, particularly for aesthetic and interior elements, remain elusive. The absence of clear optimization directives in design practice is notably problematic. Gupta et al. [31] highlighted that a product's key features dictate the perceived intent towards the user, suggesting that targeting these features is essential for focused product appearance optimization.

2.2. UGC

The concept of UGC emerged from the Internet domain, enabling users to share their original content with others via online platforms [32]. As the Internet has evolved, it has accentuated the participatory nature of users, who now act as both users and producers of online content [33]. Businesses can glean insights into user needs from UGC, thereby informing and adjusting their design strategies [34]. Wang et al. [35] introduced a method for constructing a product usage context knowledge graph (PUCKG) utilizing UGC, which structures anecdotal cases to facilitate personalized product usage predictions, summaries, and reasoning. The methodology's effectiveness was assessed through a case study on robot vacuum cleaners, involving the analysis of over 2600 UGCs instances. Ng et al. [36] suggested employing fuzzy entity–relationship (ER) modelling to decipher user needs expressed in UGCs. Despite not necessitating the aggregation of all UGC, their case study amassed more than 2400 UGC samples—a figure unfeasible for fledgling products. Establishing a threshold for what constitutes a "small" volume of UGC is crucial. To ensure that study outcomes are significant, the quantity of UGC must be substantial. Given that the sales of different products can diverge markedly, affecting the volume of UGC, a statistical analysis of over 100 product types, each on the market for less than two months, was conducted. This analysis determined that a "small" quantity of UGC is represented by a count ranging from 200 to 500 items.

Regarding the analysis of UGC, the majority of current research focuses on analyzing users' emotions [37], where sentiment lexicons serve as tools for sentiment analysis based on semantic relationships [38]. Qi et al. [39] developed a sentiment lexicon to analyze

emotional needs from online comments, while Yu et al. [40] integrated various universal lexicons to investigate the emotional bias in a vast number of visitor comments. Beyond emotions, UGC also reflects the areas of product concern for users, making the analysis of these concern indicators an integral part of need analysis. Jing et al. [41] employed the TF-IDF model to extract design needs and applied word vector models along with clustering methods to categorize need features. Qin et al. [42] leveraged the Word2Vec model to discern user needs and preferences, achieving promising results. Similarly, Ma et al. [43] applied K-means clustering to segment users into three groups, predicting their needs for innovative cockpit features. In this study, we find it justifiable to adopt K-means clustering to categorize design features within UGC, thereby identifying key need elements.

3. Methodology

Extensive research has been dedicated to mining user needs and constructing predictive models based on the collected data [44]. However, many studies overlook scenarios with limited data and the interplay among various user needs, which hinders the accurate prediction of needs. This study addresses these gaps by analyzing and forecasting user needs through a four-stage process: mining and processing UGC, quantifying need indicators, correcting need indicators, and predicting need indicators. Initially, UGC is collected and processed to extract product features, which are then subjected to clustering analysis to infer user needs. Subsequently, the attention and satisfaction values of needs are calculated using the word frequency and sentiment scores associated with product features. Following this, correlation and linear regression analyses are employed to assess the relationships between need indicators, leading to their subsequent refinement. Ultimately, a GM (1, N) model is formulated for predicting needs, enabling a dynamic examination of shifting need trends. A detailed depiction of the research process is presented in Figure 1.
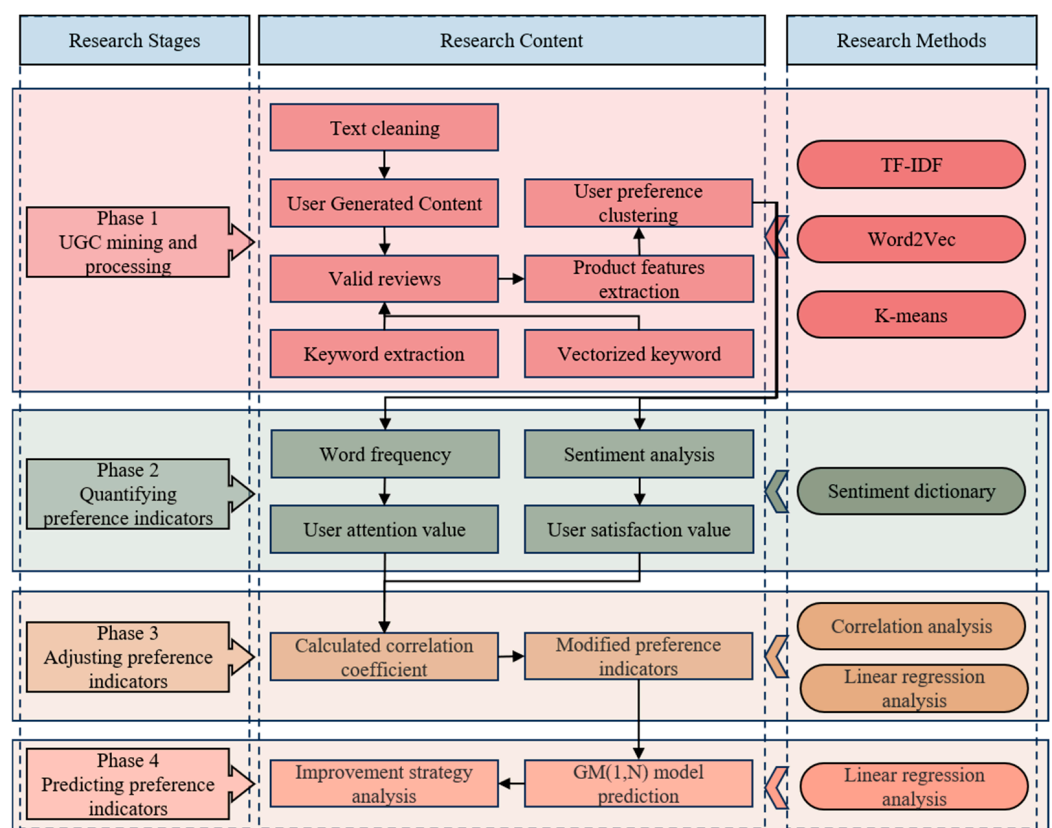

Figure 1. User need prediction flowchart (source: own elaboration).

3.1. UGC Mining and Processing

3.1.1. UGC Mining

UGC offers significant advantages over traditional methods of user need acquisition, such as the richness of data and the timeliness of access [45], making it the chosen data source for this study. However, the vastness and diversity of web users result in an uneven distribution of UGC, with a considerable portion being unrelated to product features, thereby diminishing the efficiency and accuracy of text mining. To enhance the precision of subsequent text analyses, this paper excludes four specific types of comments: (a) system default comments, which typically state that the user has not provided any evaluative content; (b) useless comments, which merely describe the user's purchase without revealing any needs; (c) duplicate comments, which are identical posts made by the same user within a short timeframe; and (d) irrelevant comments, which pertain to products unrelated to the one under consideration.

Following the data cleansing process, TF-IDF keyword extraction is applied to the dataset, a prevalent technique for keyword quantification that encompasses two components: term frequency (TF) and inverse document frequency (IDF) [46]. Term frequency indicates the occurrence of a word within a document, while inverse document frequency gauges the relative significance of a word across a document corpus. The significance of a word within a document set or an individual document is determined by the product of TF and IDF. As depicted in Equation (1), an elevated TF-IDF value corresponds to greater importance of the word within the document. Equation (1) is as follows:

$$\text{TF} - \text{IDF} = tf_{ij} \times idf_t = \frac{n_{ij}}{\sum_k n_{ij}} \times \log\left(\frac{|D|}{1 + |D_{ti}|}\right) \tag{1}$$

where n_{ij} denotes the number of occurrences of word i in document j, and $\sum_k n_{ij}$ denotes the summation of all occurrences of word i in the document. $|D|$ denotes the number of all texts in the text set; $|D_{ti}|$ is the number of texts containing word i in the text.

Word2Vec is a word vector training tool. It needs to be trained with large-scale text datasets, and then it can map and represent the unstructured text word information as a high-dimensional vector, which can further be used as the input of the model in natural language processing tasks. Compared with the traditional bag-of-words model, Word2Vec can better represent the semantic features of words and improve the sparsity of text features [47]. Therefore, in this paper, the preprocessed word texts can be converted into word vector representations by Word2Vec. Word2Vec includes two training modes, namely the continuous bag-of-words (CBOW) model and the skip-gram model [48]. The schematic diagram of the CBOW model is shown in Figure 2.

The continuous bag-of-words (CBOW) model is divided into three components: the input layer, the projection layer, and the output layer. The input layer comprises one-hot encoded vectors, which are binary representations of words using 0 s and 1 s. The projection layer's function is to transform the one-hot vectors from the input layer by multiplying them with a parameter matrix, thereby producing a lower-dimensional vector representation. Equation (2) delineates the computation for the projection layer's output, denoted as h, whereas Equation (3) details the input calculation for each node within the output layer. Equations (2) and (3) are as follows:
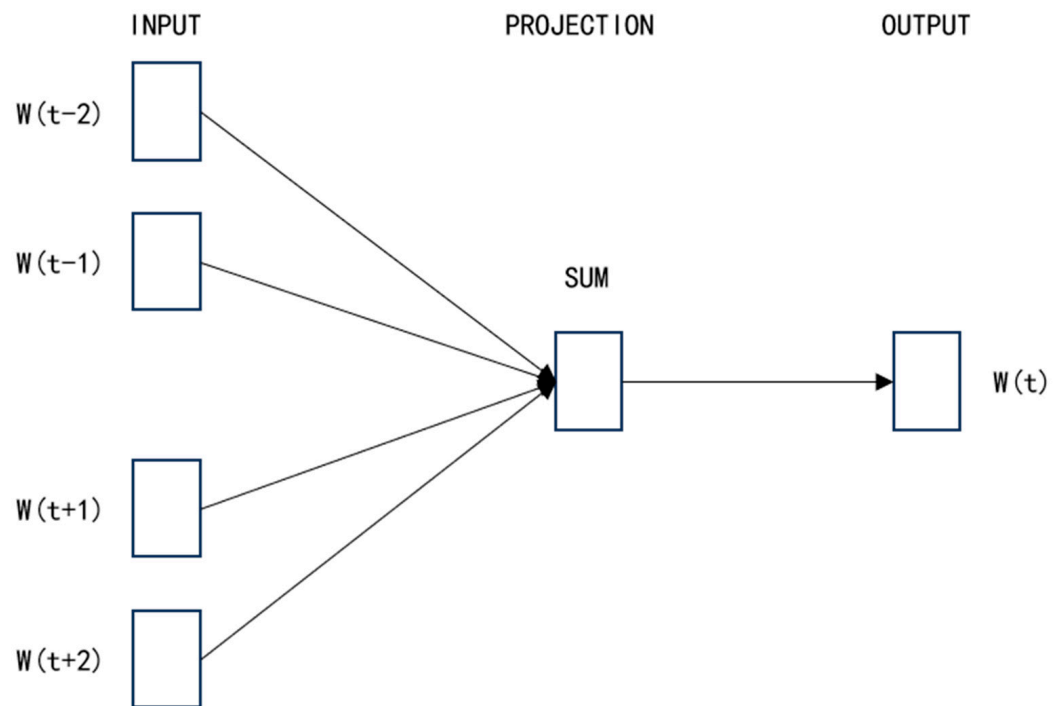
$$h = \frac{1}{C}WI\cdot\left(\sum_{i=1}^{c} x_t\right) \tag{2}$$

$$u_j = WO_j * h \tag{3}$$

where WI is the weight matrix and WO_j denotes the jth column of matrix WO.

Finally, the output values are normalized using softmax as the activation function and the weights are updated using cross entropy, where cross entropy is shown in the following Equation (4):

$$L = \frac{1}{N}\sum_i L_i = -\frac{1}{N}\sum_i \sum_{C=1}^{M} y_{ic}\log(p_{ic}) \tag{4}$$

INPUT   PROJECTION   OUTPUT

W(t-2)

W(t-1)

SUM

W(t)

W(t+1)

W(t+2)

Figure 2. CBOW model structure diagram (source: own elaboration).

3.1.2. Product Features Extraction

The K-means algorithm, introduced by MacQueen in 1967 [49], is an unsupervised machine learning technique that does not necessitate manual labeling of the training data. It is renowned for its ability to cluster data samples based on their similarity, making it a prevalent method for processing unstructured data in information recommendation and pattern recognition tasks. The algorithm's core concept involves randomly selecting "m" points from the dataset as initial cluster centers. It then calculates the Euclidean or Manhattan distance between each of the remaining data points and these centers, assigning each point to the nearest cluster. This process is iteratively repeated until the cluster assignments stabilize, indicating that the algorithm has converged. The detailed procedure of the K-means algorithm is outlined below, assuming a dataset comprising n data samples denoted as $x_1, x_2, \cdots, x_n$.

Step 1: Determine the number of clusters k from the data;

Step 2: Calculate the Euclidean distance from each sample $x_i(i = 1, 2, \cdots, n)$ to each cluster center, that is, the centroid, and then assign the sample x_i to the nearest center based on the calculated distances. Equation (5) outlines the calculation method, as follows:

$$d(x_i, u_j) = \sqrt{\sum_{i=1}^{n}(x_i - u_j)^2} \tag{5}$$

Step 3: Recalculate the position of the centroid using the mean value method to obtain the updated centroid;

Step 4: If the centroid no longer changes, terminate the process. If changes persist, repeat Steps 2 and 3 until the centroid stabilizes.

3.1.3. User Need Clustering

The integration of K-means clustering results with the inherent characteristics of each keyword category allows for the assignment of descriptive attributes to each type of keyword, facilitating the classification of user needs, as depicted in Table 2. The collection of needs is represented by the set R, where each individual need is denoted as R_i. A product possesses n needs, which can be enumerated as $R = \{R_1, R_2, \cdots R_n\}$. Each need encompasses j keyword features, with j ranging from 1 to m. The keyword features associated with each need are indicated by $C_{i,j}$, and the comprehensive set of keyword features is expressed as $C = \{C_{1,1}, C_{1,2}, \cdots, C_{i,j}, \cdots, C_{n,m}\}$.

Table 2. User need clustering (source: own elaboration).

R_1	R_2	$\cdots$	R_n
$C_{1,1}$	$C_{2,1}$	$\cdots$	$C_{n,1}$
$C_{1,2}$	$C_{2,2}$	$\cdots$	$C_{n,2}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$C_{1,m}$	$C_{2,m}$	$\cdots$	$C_{n,m}$

3.2. Quantification of User Need Indicators

3.2.1. Calculation of User Attention

Utilizing the K-means clustering outcomes, keyword features $C_{i,j}$ are correlated with user needs R_i. The term frequency $f_{i,j}$ for each keyword feature is determined utilizing the TF-IDF metric. The attention value g_i^* for the i-th need is derived from the aggregate word frequency of all associated keyword features. To enhance data usability in subsequent analyses, this attention value is normalized and represented as g_i, with the calculation detailed in the following Equation (6):

$$g_i = \frac{g_i^*}{\sum_{i=1}^n g_i^*} = \frac{\sum_{j=1}^m f_{i,j}}{\sum_{i=1}^n \sum_{j=1}^m f_{i,j}}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m \tag{6}$$

3.2.2. Calculation of User Satisfaction

Sentiment analysis involves extracting and analyzing users' subjective emotional sentiments from UGC [50]. Utilizing a sentiment lexicon is a prevalent approach in sentiment analysis, aimed at assessing the emotional tone of online comments. This method involves establishing a set of rules for assigning sentiment scores, tailored to the specific context, as detailed below.

Sentiment dictionary construction. The data of BosonNLP Sentiment Dictionary comes from microblogs, forums, and other platforms [51], which are suitable for the analysis of product reviews. Therefore, this paper refers to the BosonNLP Sentiment Dictionary to construct the sentiment dictionary for product feature evaluation.

Emotion words. The utility value of sentiment polarity is represented by $SE_{i,j}$, as follows:

$$SE_{i,j} = \begin{cases} 1, & \text{positive sentiment words} \\ -1, & \text{negative sentiment words} \end{cases} \tag{7}$$

Negatives. Negatives can change the sentiment polarity of the sentence, and the change coefficient is expressed as λ_k, where k denotes the kth UGC, as follows::

$$\lambda_k = \begin{cases} 1, & \text{If the lexical category is negative and the previous word is detected as negative} \\ -0.5, & \text{If the lexical category is negative} \\ -1, & \text{If the lexical category is positive and the previous word is detected as negative} \end{cases} \tag{8}$$

Adverbs of degree. Degree adverbs lead to stronger emotions in user needs and double the weights. The score of degree adverbs in the kth UGC is denoted by μ_k, where k denotes the kth UGC.

$$\mu_k = \begin{cases} 2, & \text{If the lexical category is positive and the previous word detected is an adverb of degree} \\ -2, & \text{If the lexical category is negative and the previous word detected is an adverb of degree} \end{cases} \quad (9)$$

According to the above parameters, the satisfaction s_i^* of the ith user need can be calculated, in order to facilitate the use of the subsequent data, the appearance satisfaction index is normalized by removing the negative data, and the formula of s_i^* expressed by s_i is shown in the following Equation (10):

$$s_i = \frac{s_i^* - \min\{s_i^*\}}{\max\{s_i^*\} - \min\{s_i^*\}}, i = 1, 2, \cdots, n \quad (10)$$

3.3. Need Indicators Correction

3.3.1. Correlation Analysis Between User Need Indicators

Correlation analysis examines the nature of relationships between quantitative data, assessing both the presence and the strength of associations [52]. Such analysis aids in deciphering the interplay among variables, thereby enhancing the capacity to predict and elucidate phenomena. For instance, Rashid et al. [53] employed correlation analysis to discern functional needs among smartphone users. The correlation coefficient, specifically the Pearson's correlation coefficient, is utilized to ascertain the existence of relationships within the data. In this context, the set of correlation coefficients is denoted by r_o, where $r_o = \{r_1, r_2, \cdots r_{n^2}\}$, and each coefficient is derived according to the following formula:

$$r_o = \frac{\sum_{i=1}^{n} g_i s_i - \frac{\sum_{i=1}^{n} g_i \sum_{i=1}^{n} s_i}{n}}{\sqrt{\left(\sum_{i=1}^{n} g_i^2 - \frac{\left(\sum_{i=1}^{n} g_i\right)^2}{n}\right)\left(\sum_{i=1}^{n} s_i^2 - \frac{\left(\sum_{i=1}^{n} s_i\right)^2}{n}\right)}}, i = 1, 2, \cdots, n, o = 1, 2, \cdots, n^2 \quad (11)$$

3.3.2. Linear Regression Analysis Between User Need Indicators

The linear regression algorithm employs mathematical and statistical techniques to explore the relationships among multiple variables, with the primary objective of predicting unknown data points by establishing a linear fit between independent and dependent variables [54]. Owing to its intuitiveness and interpretability, linear regression has proven to be a robust tool across various applications, including data analysis, scientific research, and engineering. Luo et al. [55], for instance, utilized a statistical model to discern the relative importance of different designers' perceptual dimensions in shaping user needs for animal figures, thereby quantifying their individual contributions to overall need. In alignment with the linear regression methodology, the appearance need indices, which exhibit significant correlations as determined by the preceding correlation analysis, are sequentially input as dependent variables. The ensuing linear regression equation is formulated as detailed below in Equation (12):

$$C_i = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \cdots + \beta_j C_j, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m \quad (12)$$

where $\beta_0, \beta_1, \beta_2, \cdots, \beta_j$ are regression coefficients.

3.3.3. Revision of User Need Indicators

Need indicators are adjusted to reflect the mutual influence among different related need indicators. The self-influence coefficient for each need indicator is assigned a value of

1. Subsequently, the regression coefficients derived previously are normalized using the formula provided below:

$$\beta_i' = \frac{\beta_i}{\beta_0 + \beta_1 + \beta_2 + \cdots + \beta_i + \cdots + \beta_j + 1}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m \quad (13)$$

Based on the normalized regression coefficients, each need indicator was corrected with the following formula:

$$C_i' = \beta_0' + \beta_1' C_1 + \beta_2' C_2 + \cdots + \beta_i' C_i + \cdots + \beta_j' C_j, i = 1, 2, \cdots, n, j = 1, 2, \cdots, m \quad (14)$$

3.4. Need Indicator Forecasts

GM (1, N) represents an extension of the gray prediction model, particularly adept at addressing forecasting scenarios characterized by limited data and incomplete information yet influenced by multiple factors [56]. This model excels at extracting and analyzing incomplete data, thereby identifying underlying patterns and delivering predictions with a relatively high degree of accuracy in the face of complex and dynamic real-world challenges. The detailed methodology for prediction using GM (1, N) is as follows.

Step 1: Make an accumulation of the original numbers and set the original gray data:

$$X_i^0 = \left\{ X_i^0(1), X_i^0(2), \cdots, X_i^0(n) \right\} \quad (15)$$

Step 2: Perform one accumulation of Equation (15) to obtain the 1-AGO sequence, as follows:

$$X_i^1 = \left\{ X_i^1(1), X_i^1(2), \cdots, X_i^1(n) \right\} \quad (16)$$

where $X_i^1(k) = \sum_{n=1}^{k} X_i^0(m) i = 1, 2, \cdots, n; k = 1, 2, \cdots m$.

Step 3: Generate a sequence of immediate neighboring means of X_i^1, $Z_1^1(i)$.

$$Z_1^1(k) = \frac{1}{2} \left[X_1^1(k-1) + X_1^1(k) \right] k = 2, 3, \cdots, m \quad (17)$$

Step 4: Build the differential equation:

$$\frac{dX^1}{dt} + aX^1 = \sum_{i=1}^{N} b_i X_i^1 \quad (18)$$

where a is the development gray, b_i is the adjustment coefficient of X_i, and $\hat{a} = \left(B^T B \right)^{-1} B^T Y$ can be calculated by the least squares method.

Step 5: Next, compute the matrices B and Y.

$$B = \begin{bmatrix} -Z_1^1(2) & X_2^1(2) & \cdots & X_n^1(2) \\ -Z_1^1(3) & X_2^1(3) & \cdots & X_n^1(3) \\ \vdots & \vdots & \ddots & \vdots \\ -Z_1^1(n) & X_2^1(n) & \cdots & X_n^1(n) \end{bmatrix} \quad (19)$$

$$Y = \begin{bmatrix} X_1^0(2) \\ X_1^0(3) \\ \vdots \\ X_1^0(n) \end{bmatrix} \quad (20)$$

Step 6: The GM (1, N) prediction model is as follows:

$$\hat{X}_1^1(k+1) = \left[X_i^0 - \frac{1}{a}\sum_{i=2}^{n} b_{i-1}X_i^1(k+1)e^{-ak} + \frac{1}{a}\sum_{i=2}^{n} b_{i-1}X_i^1(k+1) \right] e^{-ak} \tag{21}$$

Step 7: The final cumulative reduction reduces the equation to the following Equation (22):

$$\hat{X}_1^0(k+1) = \hat{X}_i^1(k+1) - \hat{X}_i^1(k) \tag{22}$$

4. Case Studies

In recent years, the new energy vehicle (NEV) industry has experienced rapid growth, enhancing its competitive edge. Accurately predicting user needs is crucial for the success of NEV enterprises [57]. Xiaomi Group, an emerging player in the automotive sector, introduced the Xiaomi SU7, its first pure electric sedan, on 28 March 2024. This launch has garnered significant market attention. Given that the vehicle was launched less than six months prior to this study and has an adequate volume of UGC, this paper selects the Xiaomi SU7 as the subject of analysis to validate the proposed method's effectiveness. The methodology involves the following steps: First, the UGC related to the Xiaomi SU7 is crawled and processed using Python. Subsequently, product features are extracted and categorized using K-means clustering to deduce user needs. Following this, the identified need indicators are quantified, and metrics for user attention and satisfaction are calculated. The need indicators are then refined through correlation analysis and linear regression to enhance their accuracy. Subsequently, a GM (1, N) model is constructed for forecasting user needs, aiming to predict the trend in user interest for the Xiaomi SU7. Finally, the design optimization of the vehicle's taillight is undertaken to empirically test the precision of the prediction model presented in this study.

4.1. UGC Mining and Processing for the Xiaomi SU7

4.1.1. UGC Mining for the Xiaomi SU7

In this paper, the Python programming language in conjunction with the Selenium library was employed to harvest the user comments regarding the Xiaomi SU7 on the Autohome platform. Initially, the Chrome driver was configured, and the corresponding comment page was opened. A waiting time of 3 s was set to guarantee the successful loading of the page. Subsequently, a JavaScript script was executed multiple times to mimic the page scrolling action, thereby loading a greater quantity of comments. Subsequently, based on the specific identifiers of the comment elements within the page structure, both the comment content and the posting time were located and extracted. In total, 272 instances of UGC were gathered. After eliminating the UGC that bore no relevance to the Xiaomi SU7, 270 valid pieces of UGC were retained. Some of the UGC collected for the Xiaomi SU7 can be found in Supplementary Material S1. Python was then utilized to formulate regular expressions for the removal of emoticons. Subsequently, the Jieba tool was employed for word segmentation. Finally, stop words, such as "and", platform-specific terms, like "Autohome", and the product name "Xiaomi SU7" that held no pertinence to the product characteristics were expunged.

Following the text data cleaning, keyword extraction was performed using the TF-IDF method. The higher the TF-IDF value, the more significant the keyword, as nouns are ranked according to their TF-IDF values, as depicted in Figure 3. The figure reveals that nouns with higher TF-IDF values are predominantly associated with automotive attributes. Notably, certain terms, such as "appearance", "space", "styling", "interior", and others, represent product attribute features that garner significant user interest. The most salient keywords were selected to construct a network relationship diagram, as illustrated in Figure 4. This diagram demonstrates the interconnectedness of user needs associated with each keyword, underscoring the importance of considering these relationships when forecasting user needs.
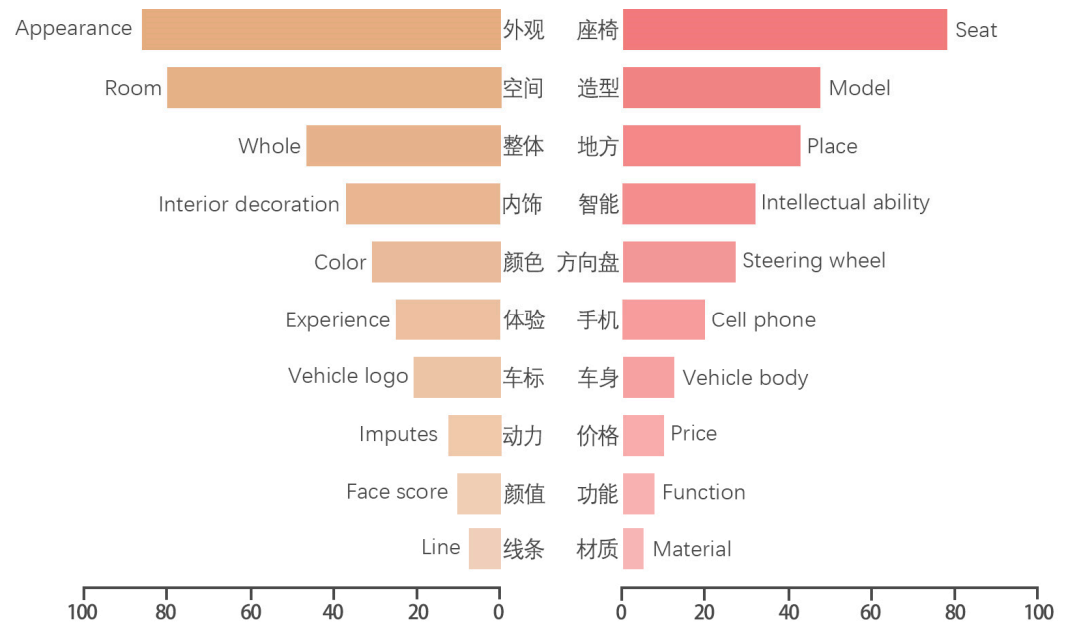
Figure 3. TF-IDF numerical ranking plot for nouns (source: own elaboration).
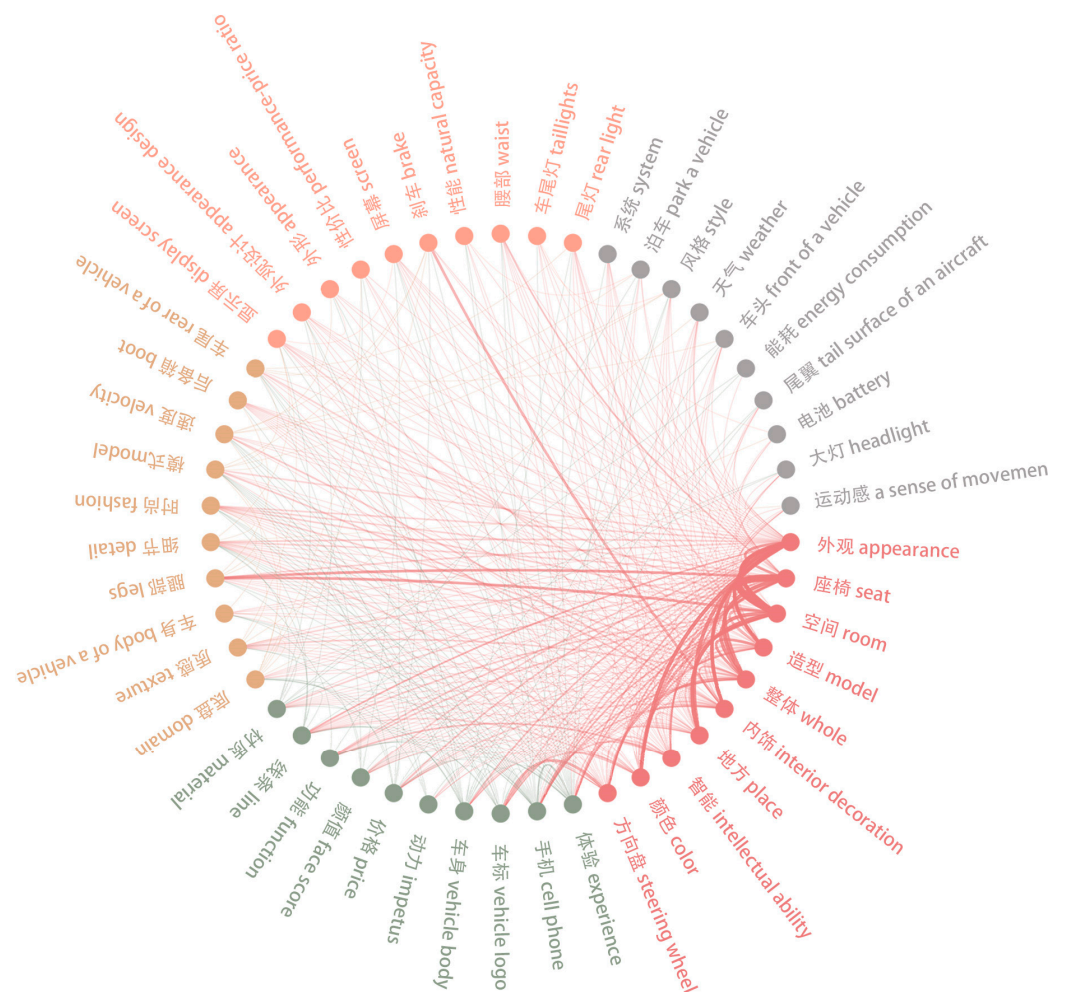

Figure 4. Network relationship diagram of keywords (source: own elaboration).

4.1.2. Product Features Extraction and User Needs Clustering for the Xiaomi SU7

Upon observation and analysis of the extracted keywords, it was noted that some keywords did not pertain to product characteristics or user needs. To refine the keyword list for practical application, taking a TF-IDF value greater than 0.0015 as the threshold, the top 100 words associated with automotive attributes were selected and converted into word vectors using the Word2vec algorithm. These word vectors were subsequently input into the K-means clustering model for thematic categorization, as visualized in Figure 5. The K-means model sorted the diverse types of keywords into four distinct categories. By integrating the intrinsic characteristics of each keyword type with the contextual information from the automotive website, the keywords were assigned to various product attribute categories. This mapping was further aligned with the probability distribution of the keywords across themes, correlating with the user's need for specific product features. The product features $C_{i,j}$ were then mapped onto four user need categories of R_i, namely space (R_1), appearance (R_2), performance (R_3), and configuration (R_4), as detailed in Table 3.
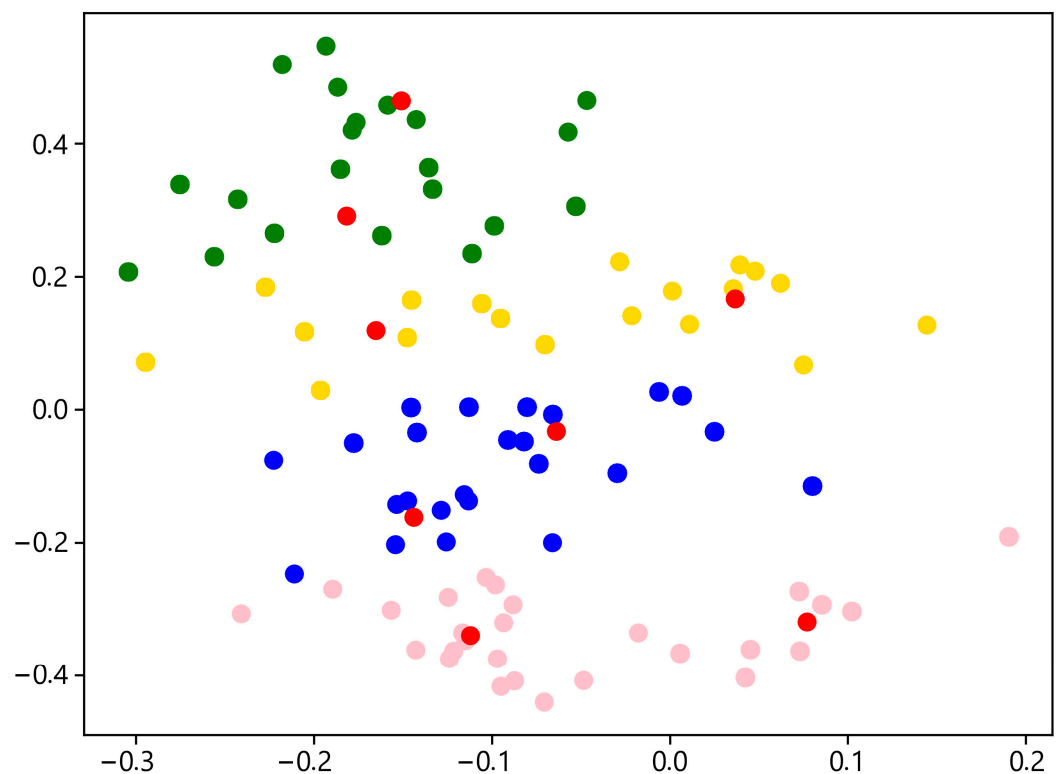

Figure 5. K-means clustering results for keywords (source: own elaboration).

Table 3. Keyword attribute clustering results (source: own elaboration).

User Need	Product Feature Keywords
Space (R_1)	Space($C_{1,1}$), seat($C_{1,2}$), back row ($C_{1,3}$), room($C_{1,4}$), leg room ($C_{1,5}$), front row($C_{1,6}$), trunk($C_{1,7}$), waist space($C_{1,8}$), roomy($C_{1,9}$), moderation($C_{1,10}$)
Appearance (R_2)	Design($C_{2,1}$), appearance($C_{2,2}$), trim($C_{2,3}$), automotive interior($C_{2,4}$), color($C_{2,5}$), whole($C_{2,6}$), model($C_{2,7}$), good-looking($C_{2,8}$), automotive logo($C_{2,9}$), material($C_{2,10}$), vehicle body($C_{2,11}$), face score($C_{2,12}$), line($C_{2,13}$), texture($C_{2,14}$), fashion($C_{2,15}$), rear($C_{2,16}$), stunning($C_{2,17}$), contour($C_{2,18}$), turnover rate($C_{2,19}$), appearance design($C_{2,20}$), taillight($C_{2,21}$), gray($C_{2,22}$), hidden($C_{2,23}$), smooth($C_{2,24}$), simple($C_{2,25}$), style($C_{2,26}$), atmospheric($C_{2,27}$), vehicle type($C_{2,28}$), aesthetic($C_{2,29}$), headstock($C_{2,30}$), headlamp($C_{2,31}$), luxurious($C_{2,32}$), empennage($C_{2,33}$), low key($C_{2,34}$), match($C_{2,35}$), element($C_{2,36}$), color matching($C_{2,37}$), headlight($C_{2,38}$), handsome($C_{2,39}$)

Table 3. *Cont.*

User Need	Product Feature Keywords
Performance (R_3)	Steer($C_{3,1}$), intelligent($C_{3,2}$), experience($C_{3,3}$), endurance($C_{3,4}$), control($C_{3,5}$), power($C_{3,6}$), function($C_{3,7}$), movement($C_{3,8}$), accelerate($C_{3,9}$), feel($C_{3,10}$), mode($C_{3,11}$), sports car($C_{3,12}$), speed($C_{3,13}$), subsidiary($C_{3,14}$), expedient($C_{3,15}$), command($C_{3,16}$), adapt($C_{3,17}$),property($C_{3,18}$), energy consumption($C_{3,19}$), system($C_{3,20}$), reaction($C_{3,21}$), sensitive($C_{3,22}$), soft ($C_{3,23}$), cozy($C_{3,24}$), ecology ($C_{3,25}$), comfy($C_{3,26}$), relaxed ($C_{3,27}$), science and technology ($C_{3,28}$), navigation ($C_{3,29}$), noise ($C_{3,30}$)
Configuration (R_4)	Steering wheel($C_{4,1}$), configuration($C_{4,2}$), shelter control panel ($C_{4,3}$), cell phone($C_{4,4}$), doorknob($C_{4,5}$), charging($C_{4,6}$), chassis($C_{4,7}$), turning engine($C_{4,8}$), detail($C_{4,9}$), Xiao ai($C_{4,10}$), voice($C_{4,11}$), display screen ($C_{4,12}$), screen ($C_{4,13}$), battery($C_{4,14}$), plastic($C_{4,15}$), wheelbase($C_{4,16}$), shock absorption ($C_{4,17}$), icebox ($C_{4,18}$), ventilate ($C_{4,19}$), key($C_{4,20}$), glass($C_{4,21}$)

4.2. Quantification of User Need Indicators for the Xiaomi SU7

4.2.1. Calculation of User Attention for the Xiaomi SU7

Since the Xiaomi SU7 was launched on 25 March, for the convenience of subsequent calculations and to ensure that the quantity of the collected UGC falls within the range of a small amount of UGC mentioned above, the time period for collecting UGC was set from 25 March to 25 August, lasting for a total of 150 days. Consequently, in this study, the UGC data are segmented into five distinct cycles, with each cycle representing a 30-day interval. The user attention value is derived using the formula presented in Equation (5), as delineated in Table 4. An analysis of the data presented in the table indicates that the highest level of user attention is directed towards "appearance", whereas "configuration" receives comparatively lower attention.

Table 4. User attention indicators in 5 periods (source: own elaboration).

Periods	Space (g_1)	Appearance (g_2)	Performance (g_3)	Configuration (g_4)
1	0.0484	0.0733	0.0535	0.0290
2	0.0839	0.1453	0.0363	0.0685
3	0.0486	0.0797	0.0267	0.0340
4	0.0623	0.0612	0.0380	0.0256
5	0.0574	0.0769	0.0361	0.0439

4.2.2. Calculation of User Satisfaction for the Xiaomi SU7

Sentiment analysis was conducted using Python3.12 to ascertain the level of user need satisfaction. The satisfaction values for user needs were computed in accordance with Equation (10), with the outcomes presented in Table 5. An examination of the data reveals that users exhibit the highest satisfaction with "performance", while satisfaction levels for other needs exhibit significant variability.

Table 5. User satisfaction indicators in 5 periods (source: own elaboration).

Periods	Space (s_1)	Appearance (s_2)	Performance (s_3)	Configuration (s_4)
1	0.4542	0.5079	0.3595	0.3984
2	0.6084	0.8428	1	0.5547
3	0.5938	0.3029	0.5128	0.1476
4	0.2695	0.5776	0.7772	0.3972
5	0.4844	0.5298	0.5770	0

4.3. Need Indicators Correction for the Xiaomi SU7

4.3.1. Correlation Analysis Between User Need Indicators for the Xiaomi SU7

In this study, SPSS 26.0 software was employed to calculate correlation coefficients using the Pearson, Spearman, and Kendall methods. The results obtained through Pearson's

method are presented in Tables 6 and 7. The correlation analysis indicates significant associations between the variables g_1, g_2, and g_4, as well as between s_2, s_3, and s_4.

Table 6. Results of the analysis related to the attention indicator (source: own elaboration).

	Average	Standard Deviation	g_1	g_2	g_3	g_4
g_1	0.031	0.015	1			
g_2	0.066	0.016	0.917 *	1		
g_3	0.029	0.007	0.551	0.331	1	
g_4	0.013	0.005	0.963 **	0.977 **	0.340	1

* $p < 0.05$, ** $p < 0.01$.

Table 7. Results of the analysis related to the satisfaction indicator (source: own elaboration).

	Average	Standard Deviation	s_1	s_2	s_3	s_4
s_1	0.031	0.015	1			
s_2	0.066	0.016	0.781	1		
s_3	0.029	0.007	0.591	0.935 *	1	
s_4	0.013	0.005	0.689	0.963 **	0.880 *	1

* $p < 0.05$, ** $p < 0.01$.

4.3.2. Linear Regression Analysis Between User Need Indicators for the Xiaomi SU7

Subsequently, linear regression analysis was conducted using SPSS 26.0 software to further analyze the correlated user need indicators. For instance, with g_1 as the independent variable and g_2 and g_4 as the dependent variables, the regression model yielded an R^2 value of 0.941, indicating a strong fit. The regression equation derived was $g_1 = 0.015 - 0.357g_2 + 1.783g_4$. Following the quantification method outlined in Equation (14), the re-quantified user need indicator $g'_1 = \beta'_0(g_1) + \beta'_2(g_2)g_2 + \beta'_4(g_4)g_4$ was obtained, and similar calculations were performed for other user need indicators. For g_3 and s_1, which lack correlated user need indicators, the initial values were retained. The corrected values of all user need indicators are summarized in Table 8.

Table 8. Modified user need indicators (source: own elaboration).

Periods	g_1	g_2	g_3	g_4	s_1	s_2	s_3	s_4
1	0.0281	0.0378	0.0535	0.0478	0.4542	0.4178	0.5161	0.5061
2	0.0591	0.0778	0.0363	0.1015	0.6084	0.6910	0.9781	0.8624
3	0.0327	0.0440	0.0267	0.0824	0.5938	0.804	0.3147	0.2975
4	0.0469	0.0679	0.0380	0.0734	0.2695	0.6362	0.6247	0.6348
5	0.0409	0.0871	0.0361	0.0785	0.4844	0.5944	0.5110	0.5827

4.4. Need Indicator Forecasts for the Xiaomi SU7

Utilizing the GM (1, N) model, predictions were formulated based on the outcomes of the correlation analysis. Specifically, the user need indicators g_1, g_2, and g_4 were forecast collectively as one group, while g_3 was predicted individually. Similarly, s_2, s_3, and s_4 were forecast as a separate group, with s_1 being predicted independently. Retrospective predictions spanning five periods were conducted, and the results of these forecasts are detailed in Table 9.

**Table 9.** Predicted results of user need indicators (source: own elaboration).

Periods	g_1	g_2	g_3	g_4	s_1	s_2	s_3	s_4
6	0.035	0.082	0.037	0.064	0.318	0.572	0.334	0.466
7	0.031	0.088	0.038	0.057	0.256	0.531	0.234	0.417
8	0.027	0.093	0.039	0.049	0.197	0.490	0.139	0.369
9	0.023	0.098	0.040	0.042	0.140	0.451	0.047	0.322
10	0.015	0.104	0.041	0.034	0.087	0.413	−0.040	0.276

In the realm of user need attention, g_2 exhibits a mean value of 0.0873 across the five cycles, signifying its status as the most favored need among users. Over the projected five cycles, g_2 demonstrates an upward trend, with the highest mean value reaching 0.0930. Conversely, s_2 exhibits a declining trend during the forecasted five cycles, suggesting a future increase in user focus on the appearance of the Xiaomi SU7, coupled with a potential decrease in satisfaction. This trend underscores the necessity for the company to prioritize user needs regarding the vehicle's aesthetics. Through the analysis of user comments regarding the appearance in UGC, it was found that the TF-IDF value of the term "taillight" is 0.0063, which is relatively high, and there are mostly negative comments about the design of the taillight. As an important part of the overall styling design of a vehicle, the design of the car's taillight can with the body lines and is crucial for enhancing the aesthetic appeal of the entire vehicle. Consequently, this study utilizes the optimized design of the taillight as a case study to validate the accuracy of the prediction model, as detailed in the subsequent section.

The aesthetic appeal of a product's appearance is determined by a multitude of key features, which are themselves influenced by critical points, lines, and angles. By fine-tuning these key feature indices, the product's appearance can be tailored to align with user needs [58]. This study involved collecting images of taillight designs from 30 high-selling cars, translating the taillight features into 2D curve representations, and consulting with several senior design professionals. Through this process, three types of constraints—points, angles, and distances—were identified as the primary determinants of taillight design. The taillight design of the Xiaomi SU7 serves as a case study to illustrate the positioning of these key features, as depicted in Figure 6, with detailed descriptions and calculations provided in Table 10. A panel of 10 transportation design experts, comprising 6 lecturers and 4 senior designers, was assembled to rate the taillight designs of the selected high-selling cars on a scale of 1 to 5. The scoring results are shown in Supplementary Material S2. To prevent bias towards a single brand's design language, the top five taillight designs from different manufacturers, based on average scores, were selected for further analysis. The key feature values of these five vehicles' taillights were calculated using the established formulae and compared with those of the Xiaomi SU7, with the results presented in Table 11. The data indicates that successful designs often share similarities in the values of their key features. The optimization of the Xiaomi SU7's taillight should aim to approximate these optimal key features, while also taking into account the practical challenges and costs associated with engineering implementation and corporate redesign efforts. Following the adjustment of the key features, the optimized taillight design was compared with the original, as illustrated in Figure 7. The key features that have been specifically adjusted are KI_5 and KI_6. KI_5 has been adjusted from the original 143.58° to 147.75°, and KI_6 has been changed from one feature index originally to two, that is, it has been adjusted from 43.94° to 48.25° and 120.25°.

Table 10. Taillight key characterization results (source: own elaboration).

| Feature Points	Connotation Explanation	Marking and Calculation
$P_8 - P_1 - P_7$	The ratio of the distance between the highest point of the spoiler and the height of the taillights to the height of the taillights.	$KI_1 = (y_8 - y_1)/(y_1 - y_7)$
$P_1 - P_7 - P_9$	The ratio of the height of the taillights to the distance between the bottom of the trunk and the taillights.	$KI_2 = (y_1 - y_7)/(y_7 - y_9)$
$P_3 - P_1 - P_7$	The ratio of the maximum height to the minimum height of the taillights.	$KI_3 = (y_3 - y_7)/(y_1 - y_7)$
$P_1 - P_2 - P_3$	The angle of the outer corner of the taillights.	$KI_4 = \pi - arctan[(y_3 - y_1)/x_2 - x_3]$
$P_2 - P_3 - P_4$	The inner corner angle of the taillights.	$KI_5 = arctan[(x_2 - x_3)/y_3 - y_2] + arctan[(x_3 - x_5)/(y_3 - y_5)]$
$P_4 - P_5 - P_6$	The edge corner angle of the taillights.	$KI_6 = arctan[(y_3 - y_5)/x_3 - x_5]$
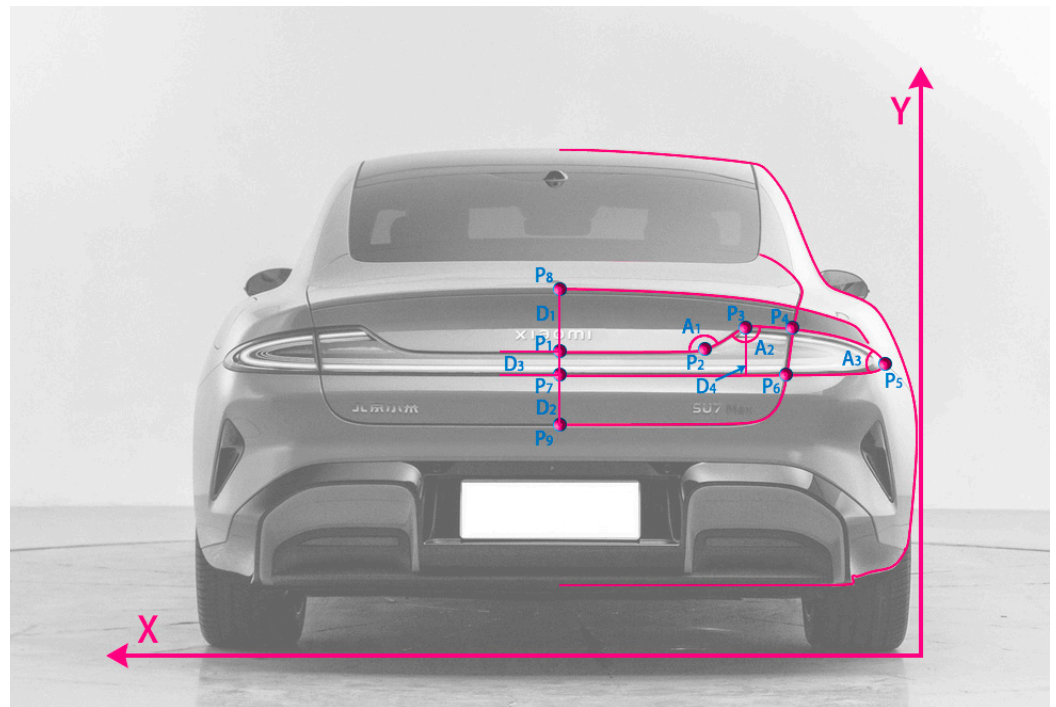
Figure 6. Illustration of key features of automotive taillights (source: own compilation based on an actual photo of the Xiaomi SU7 on Autohome).

Table 11. Calculation results of key features of tail lamps of various brands of automobiles (source: own elaboration).

	Qin PLUS	Model 3	Magotan	AION S	Hong Qi H5	Xiaomi SU7
KI_1	2.80	2.83	2.74	2.72	2.79	2.68
KI_2	0.42	0.44	0.42	0.46	0.43	0.48
KI_3	1.98	1.85	2.04	1.89	1.92	2.10
KI_4	144.25°	145.02°	144.28°	144.04°	143.56°	147.56°
KI_5	145.46°	147.25°	148.34°	145.86°	144.78°	143.58°
KI_6	45.96°	47.25°	46.87°	49.25°	48.36°	43.94°
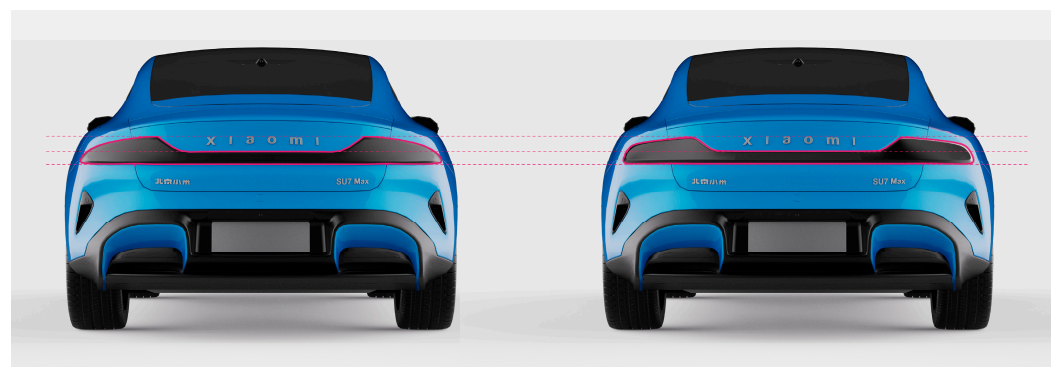
Figure 7. Comparative analysis of the taillight shape optimization for the Xiaomi SU7: pre- and post-modification effects (source: own compilation based on the 3D model of the Xiaomi SU7).

5. Discussion

5.1. Model Validation

The prediction of user needs in this study employs a systematic approach. Initially, the word frequency and vector representation within UGC are determined through the application of TF-IDF technology and Word2vec. This is followed by the utilization of

K-means clustering, which leverages the word vectors to extract and categorize product attribute features from the UGC, thereby identifying user needs. Subsequently, the attention value associated with user needs for the product is calculated based on word frequency. The satisfaction value is then ascertained using an emotion dictionary. To address the interrelationships among need indicators, correlation analysis is conducted to reveal the mutual influence between needs. This analysis is complemented by linear regression to refine the need indicator values. Finally, the developmental trends in both attention and satisfaction levels of user needs are forecast using the multidimensional gray prediction model GM (1, N).

In this study, the Markov prediction model and the long short-term memory network (LSTM) are selected for comparative analysis. To evaluate the performance of the proposed model against these established models, the initial five cycles of data are utilized as the sample dataset. Subsequently, predictions for the five cycles are made and compared against actual data. The accuracy of these predictions is assessed using the mean absolute percentage error (MAPE), the root mean squared error (RMSE), and the mean absolute error (MAE). Lower values of MAPE, RMSE, and MAE indicate a smaller deviation between predicted and actual values, thereby suggesting higher model accuracy. The evaluation results of the model prediction accuracy are shown in Supplementary Material S3. The CA-LR-GM (1, N) model demonstrates the lowest average MAE values of 0.0032 for user attention and 0.1731 for user satisfaction, respectively, among the three models. This indicates that the CA-LR-GM (1, N) model outperforms the others in terms of prediction accuracy, making it suitable for practical application. Additionally, a questionnaire-based study was conducted to evaluate the optimized taillight and overall vehicle design of the Xiaomi SU7. A total of 50 participants, comprising 5 professional automotive designers, 5 university faculty members specializing in transportation design, 10 design students, and 20 general users, were selected to assess the aesthetic appeal. Evaluations were based on a five-point Likert scale, with the mean value considered as the final assessment. The final evaluation results are shown in Supplementary Material S4. The original and optimized taillight designs received final ratings of 4.02 and 4.24, respectively, while the original and optimized overall vehicle designs were rated at 4.35 and 4.42, respectively. The results indicate an improvement in the scoring following the design optimization.

5.2. Analysis of Improvement Strategy of Product Design

The CA-LR-GM (1, N) model's predictions of user need attention and satisfaction values serve as a basis for validating the model's efficacy, using the Xiaomi SU7 taillight styling optimization as a case study. The following insights inform the strategic enhancement of user needs for the Xiaomi SU7. Notably, the need indicators g_2 and g_3 show an increasing trend and exhibit an upward trend, with g_2 showing a more pronounced increase, while g_1 and g_4 indicate a decline. Conversely, satisfaction across all four user needs is observed decrease. This trend may stem from the initial influx of positive feedback for newly launched products, which tends to diminish as users encounter and report more defects over time. Among these factors, the satisfaction for s_3 experiences the most significant decline. In conclusion, it is imperative for enterprises to implement proactive measures to preempt potential issues. Enhanced research and development focus should be directed towards the appearance need, which records the highest increase in user attention, and the performance need, which witnesses the steepest decline in satisfaction. Furthermore, targeted optimization efforts should be informed by the pain points frequently highlighted by users in UGC to address specific areas of concern.

6. Conclusions

This study proposes a novel approach to predicting user needs by leveraging limited UGC, with particular emphasis on the Xiaomi SU7. Through the integration of multiple advanced techniques, namely TF-IDF, Word2Vec, K-means clustering, correlation analysis, linear regression, and GM (1, N), we have meticulously constructed a comprehensive

framework for analyzing and forecasting user needs. The amalgamation of user attention and satisfaction metrics has been empirically validated to be highly effective in capturing the intricate nature of user needs. Our research findings unequivocally demonstrate that the CA-LR-GM (1, N) model can accurately prognosticate the trends of these metrics, thereby furnishing invaluable insights for optimizing product design. The triumphant application of this model in the refinement of the Xiaomi SU7's tail lamp design robustly attests to its practical utility. The research outcomes are encapsulated as follows:

(1) We have successfully showcased a method for predicting user needs with a relatively scant amount of UGC, and the prediction accuracy outperforms that of other models. By deftly applying a suite of techniques encompassing data cleansing, feature extraction, and cutting-edge predictive modelling, we have convincingly demonstrated that meaningful insights can be gleaned even from restricted data sources.

(2) Via correlation analysis and linear regression, we have adroitly incorporated the interrelationships among user needs into the prediction model. This has not only enhanced the precision of our predictions but also enabled a more nuanced comprehension of how diverse user needs interact and exert influence on one another.

(3) Based on the prediction outcomes of our experiment, the optimized design of the Xiaomi SU7's tail lamp serves as a tangible exemplar of how user need data can be transmuted into actionable design enhancements. This firmly validates the practical applicability of our research in guiding product design decisions.

Nevertheless, this paper is not without limitations that warrant improvement, as follows:

(1) Our research is centered around the newly introduced electric vehicle, the Xiaomi SU7, and the application of this method holds profound significance for automotive optimization design. However, it must be conceded that our research purview is circumscribed. Notably, a more extensive comparison with automobiles from other countries would augment the generality of our conclusions. Future research could expand the research scope to encompass a more heterogeneous array of vehicles and markets, further validating and refining the proposed methodology.

(2) The results of this study are predicated solely on UGC. However, other data sources, such as transaction data and battery performance indicators, may also mirror the future requirements of users. Transaction data can offer valuable perspectives on consumer purchasing behavior and long-term needs, while battery performance metrics are pivotal for apprehending the long-term viability and environmental ramifications of electric vehicles. By assimilating these supplementary data sources, future research could attain a more holistic understanding of user needs and behaviors within the context of electric vehicles.

(3) Seasonal fluctuations and promotional activities may precipitate oscillations in user needs, and our current model falls short of fully capturing these vicissitudes. Additionally, the rapid evolution of technology and consumer trends poses formidable challenges to the long-term prediction accuracy. Future research ought to contemplate incorporating dynamic factors and real-time data to augment the adaptability of the prediction model.

7. Future Research Avenues

To surmount these limitations and build upon the current research, several prospective research directions are posited.

(1) Firstly, augmenting the dataset to span a broader spectrum of products and user cohorts would enhance the universality of the research findings. This could entail collecting UGC from multiple platforms and regions, as well as integrating data from disparate sources, such as customer surveys and social media.

(2) Secondly, it is imperative to enhance the prediction model by incorporating state-of-the-art machine learning algorithms and techniques. This could encompass exploring deep learning architectures for more accurate feature extraction and prediction, as well as devising hybrid models that amalgamate the strengths of diverse algorithms.

(3) Finally, it is crucial to consider the impact of exogenous factors, such as technological advancements and regulatory modifications, on user needs. Future research should probe into how these factors interact with user needs and formulate corresponding product design and marketing adjustment strategies.

In summation, this study lays a solid foundation for future research in the domain of user need prediction and product design optimization. By addressing the limitations and exploring novel research avenues, we aspire to contribute to the advancement of the automotive industry.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/wevj15120584/s1, Supplementary Material S1. Some of the crawled UGC of the Xiaomi SU7. Supplementary Material S2. 30 Coupe taillight styling evaluation sheet. Supplementary Material S3. Comparison results of model prediction accuracy. Supplementary Material S4. Comparison results of the Xiaomi SU7 taillight styling score before and after optimization.

References

1. Li, X. Dynamic acquisition method of a user's implicit information demand based on association rule mining. *Int. J. Auton. Adapt. Commun. Syst.* **2022**, *15*, 361–378. [CrossRef]
2. Sun, H.; Guo, W.; Shao, H.Y.; Rong, B. Dynamical mining of ever-changing user requirements: A product design and improvement perspective. *Adv. Eng. Inform.* **2020**, *46*, 11. [CrossRef]
3. Xia, B.H.; Sakamoto, H.; Wang, X.T.; Yamasaki, T. Packaging Design Analysis by Predicting User Preference and Semantic Attribute. *ITE Trans. Media Technol. Appl.* **2022**, *10*, 120–129. [CrossRef]
4. Li, Q.X.; Yang, Y.; Li, C.J.; Zhao, G. Energy vehicle user demand mining method based on fusion of online reviews and complaint information. *Energy Rep.* **2023**, *9*, 3120–3130. [CrossRef]
5. Lin, J.; Jiang, X.Y.; Li, Q.; Wang, C. A competitive intelligence acquisition framework for mining user perception from user generated content. *Appl. Soft Comput.* **2023**, *147*, 14. [CrossRef]
6. Lashari, Z.A.; Ko, J.; Jang, J. Consumers' Intention to Purchase Electric Vehicles: Influences of User Attitude and Perception. *Sustainability* **2021**, *13*, 6778. [CrossRef]
7. Zou, P.Y.; Zhang, B.; Yi, Y.; Wang, Z.H. How does travel satisfaction affect preference for shared electric vehicles? An empirical study using large-scale monitoring data and online text mining. *Transp. Policy* **2024**, *146*, 59–71. [CrossRef]
8. Tien, T.L. A research on the grey prediction model GM(1,n). *Appl. Math. Comput.* **2012**, *218*, 4903–4916. [CrossRef]
9. Yang, L.; Liu, Y.Z.; Jiang, Y.C.; Wu, L.; Sun, J.S. Predicting personalized grouping and consumption: A collaborative evolution model. *Knowl.-Based Syst.* **2021**, *228*, 20. [CrossRef]
10. Zhou, Y.; Zhang, Q.; Singh, V.P.; Xiao, M.Z. General correlation analysis: A new algorithm and application. Stoch. Environ. *Res. Risk Assess.* **2015**, *29*, 665–677. [CrossRef]
11. Höskuldsson, A. Common framework for linear regression. *Chemom. Intell. Lab. Syst.* **2015**, *146*, 250–262. [CrossRef]
12. Liu, Y.N.; Shen, Y.M. Personal Tastes vs. Fashion Trends: Predicting Ratings Based on Visual Appearances and Reviews. *IEEE Access.* **2018**, *6*, 16655–16664. [CrossRef]
13. Yusuf-Asaju, A.W.; Dahalin, Z.B.; Ta'a, A. *Towards Real-Time Customer Satisfaction Prediction Model for Mobile Internet Networks*; Recent Trends in Data Science and Soft Computing; Springer Publishing: New York, NY, USA, 2019; pp. 95–104.
14. Lee, C.; Xu, X.; Lin, C.-C. Using Online User-Generated Reviews to Predict Offline Box-Office Sales and Online DVD Store Sales in the O2O Era. *J. Theor. Appl. Electron. Commer. Res.* **2019**, *14*, 68–83. [CrossRef]

15. Dou, R.; Li, W.; Nan, G. An integrated approach for dynamic customer requirement identification for product development. *Enterp. Inf. Syst.* **2019**, *13*, 448–466. [CrossRef]

16. Ali, M.M.; Doumbouya, M.B.; Louge, T.; Rai, R.; Karray, M.H. Ontology-based approach to extract product's design features from online customers' reviews. *Comput. Ind.* **2020**, *116*, 103175. [CrossRef]

17. Zhang, X.; Yang, M.; Su, J.; Yang, W.; Qiu, K. Research on product color design decision driven by brand image. *Color Res. Appl.* **2020**, *45*, 1202–1216. [CrossRef]

18. Jiang, H.; Sabetzadeh, F.; Kwong, C.K. Dynamic analysis of customer needs using opinion mining and fuzzy time series approaches. In Proceedings of the 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Luxembourg, 11–14 July 2021; pp. 1–6.

19. Ostasz, G.; Siwiec, D.; Pacana, A. Universal Model to Predict Expected Direction of Products Quality Improvement. *Energies* **2022**, *15*, 1751. [CrossRef]

20. Cheng, F.; Yu, S.; Chu, J.; Fan, J.; Hu, Y. Customer satisfaction-oriented product configuration approach based on online product reviews. *Multimed. Tools Appl.* **2022**, *81*, 4413–4433. [CrossRef]

21. Wang, S.; Wang, Y.; Hu, L.; Zhang, X.; Zhang, Q.; Sheng, Q.Z.; Orgun, M.A.; Cao, L.; Lian, D. Modeling User Demand Evolution for Next-Basket Prediction. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 11585–11598. [CrossRef]

22. Luo, H.; Song, W.; Zhou, W.; Lin, X.; Yu, S. An Analysis Framework to Reveal Automobile Users' Preferences from Online User-Generated Content. *Sustainability* **2023**, *15*, 13336. [CrossRef]

23. Zhang, N.; Qin, L.; Yu, P.; Gao, W.; Li, Y. Grey-Markov model of user demands prediction based on online reviews. *J. Eng. Des.* **2023**, *34*, 487–521. [CrossRef]

24. Fan, Z. E-Commerce Data Mining Analysis based on User Preferences and Assiciation Rules. *Scalable Comput. Pract. Exp.* **2024**, *25*, 1765–1772. [CrossRef]

25. Li, F.; Chen, C.-H.; Liu, Y.; Chang, D.; Cui, J.; Sourina, O. Autoencoder-enabled eye-tracking data analytics for objective assessment of user preference in humanoid robot appearance design. *Expert Syst. Appl.* **2024**, *249*, 123857. [CrossRef]

26. Luo, S.; Shan, P.; Bian, Z.; Lin, H.; Zhang, Y.; Cui, Z.; Shen, C.; Wang, L. Effects of product personalisation degree on user perception in car front design. *J. Eng. Des.* **2024**, *35*, 944–971. [CrossRef]

27. Dinaryanti, R.F.; Indrawati. Analysis Of Consumer Preferences In Choosing Smartphone Through User Comments On Youtube. *Qual. Access Success* **2024**, *25*, 323–330. [CrossRef]

28. Li, S.G.; Zhang, Y.Q.; Li, Y.M.; Yu, Z.X. The user preference identification for product improvement based on online comment patch. *Electron. Commer. Res.* **2021**, *21*, 423–444. [CrossRef]

29. Guo, F.; Li, F.X.; Nagamachi, M.; Hu, M.C.; Li, M.M. Research on color optimization of tricolor product considering color harmony and users' emotion. *Color Res. Appl.* **2020**, *45*, 156–171. [CrossRef]

30. Li, X.R.; Hou, X.G.; Yang, M.; Zhang, L.; Guo, H.Y.; Wang, L.Y.; Li, X.Y. A method of constructing an inspiration library driven by user-perceived preference evaluation data for biologically inspired design. *Adv. Eng. Inform.* **2022**, *52*, 15. [CrossRef]

31. Gupta, R.K.; Gurumoorthy, B. Feature-based ontological framework for semantic interoperability in product development. *Adv. Eng. Inform.* **2021**, *48*, 23. [CrossRef]

32. Naab, T.K.; Sehl, A. Studies of user-generated content: A systematic review. *Journalism* **2017**, *18*, 1256–1273. [CrossRef]

33. dos Santos, M.L.B. The "so-called" UGC: An updated definition of user-generated content in the age of social media. *Online Inf. Rev.* **2022**, *46*, 95–113. [CrossRef]

34. Lee, J.Y.H.; Yang, C.S.; Chen, S.Y. Understanding Customer Opinions From Online Discussion Forums: A Design Science Framework. *Eng. Manag. J.* **2017**, *29*, 235–243. [CrossRef]

35. Wang, X.Z.; Liu, A.; Kara, S. Constructing Product Usage Context Knowledge Graph Using User-Generated Content for User-Driven Customization. *J. Mech. Des.* **2023**, *145*, 14. [CrossRef]

36. Ng, C.Y.; Law, K.M.Y. Investigating consumer preferences on product designs by analyzing opinions from social networks using evidential reasoning. Comput. *Ind. Eng.* **2020**, *139*, 11. [CrossRef]

37. Chan, K.Y.; Kwong, C.K.; Kremer, G.E. Predicting customer satisfaction based on online reviews and hybrid ensemble genetic programming algorithms. *Eng. Appl. Artif. Intell.* **2020**, *95*, 13. [CrossRef]

38. Yan, M.; Lou, X.R.; Chan, C.A.; Wang, Y.; Jiang, W. A semantic and emotion-based dual latent variable generation model for a dialogue system. *CAAI T. Intell. Technol.* **2023**, *8*, 319–330. [CrossRef]

39. Qi, J.Y.; Zhang, Z.P.; Jeon, S.M.; Zhou, Y.Q. Mining customer requirements from online reviews: A product improvement perspective. *Inf. Manag.* **2016**, *53*, 951–963. [CrossRef]

40. Yu, Y.Y.; Chen, J.Q.; Mehraliyev, F.; Hu, S.K.; Wang, S.B.; Liu, J. Exploring the diversity of emotion in hospitality and tourism from big data: A novel sentiment dictionary. *Int. J. Contemp. Hosp. Manag.* **2024**, *36*, 4237–4257. [CrossRef]

41. Jing, L.T.; Yang, J.W.; Ma, J.F.; Jing, X.; Li, J.Q.; Jiang, S.F. An integrated implicit user preference mining approach for uncertain conceptual design decision-making: A pipeline inspection trolley design case study. *Knowl. Based Syst.* **2023**, *270*, 27. [CrossRef]

42. Qin, J.W.; Jiang, Y.P. Recommender resources based on acquiring user's requirement and exploring user's preference with Word2Vec model in web service. *Int. J. Internet Protoc. Technol.* **2019**, *12*, 144–152. [CrossRef]

43. Ma, J.; Gong, Y.Q.; Xu, W.X. Predicting User Preference for Innovative Features in Intelligent Connected Vehicles from a Cultural Perspective. *World Electr. Veh. J.* **2024**, *15*, 130. [CrossRef]

44. Yakubu, H.; Kwong, C.K. Forecasting the importance of product attributes using online customer reviews and Google Trends. *Technol. Forecast. Soc. Change* **2021**, *171*, 13. [CrossRef]

45.   Nasrabadi, M.A.; Beauregard, Y.; Ekhlassi, A. The implication of user-generated content in new product development process: A systematic literature review and future research agenda. *Technol. Forecast. Soc. Change* **2024**, *206*, 19. [CrossRef]

46.   Kim, D.; Seo, D.; Cho, S.; Kang, P. Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Inf. Sci.* **2019**, *477*, 15–29. [CrossRef]

47.   Church, K.W. Emerging Trends Word2Vec. *Nat. Lang. Eng.* **2017**, *23*, 155–162. [CrossRef]

48.   Lee, C.; Jeon, D.; Ahn, J.M.; Kwon, O. Navigating a product landscape for technology opportunity analysis: A word2vec approach using an integrated patent-product database. *Technovation* **2020**, *96–97*, 102140. [CrossRef]

49.   Sinaga, K.P.; Yang, M.S. Unsupervised K-Means Clustering Algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [CrossRef]

50.   Yang, L.; Li, Y.; Wang, J.; Sherratt, R.S. Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access* **2020**, *8*, 23522–23530. [CrossRef]

51.   Wang, Z.Q.; Qin, Y.T. The Impact of Shanghai Epidemic, China, 2022 on Public Psychology: A Sentiment Analysis of Microblog Users by Data Mining. *Sustainability* **2022**, *14*, 9649. [CrossRef]

52.   Chen, X.H.; Chen, S.C.; Xue, H. Large correlation analysis. *Appl. Math. Comput.* **2011**, *217*, 9041–9052. [CrossRef]

53.   Rashid, A.; Zeb, M.A.; Rashid, A.; Anwar, S.; Joaquim, F.; Halim, Z. Conceptualization of smartphone usage and feature preferences among various demographics. *Clust. Comput.* **2020**, *23*, 1855–1873. [CrossRef]

54.   Ari, B.; Güvenir, H.A. Clustered linear regression. *Knowl. Based Syst.* **2002**, *15*, 169–175. [CrossRef]

55.   Luo, S.J.; Zhang, Y.F.; Zhang, J.; Xu, J.H. A User Biology Preference Prediction Model Based on the Perceptual Evaluations of Designers for Biologically Inspired Design. *Symmetry* **2020**, *12*, 1860. [CrossRef]

56.   Zeng, B.; Duan, H.M.; Zhou, Y.F. A new multivariable grey prediction model with structure compatibility. *Appl. Math. Model.* **2019**, *75*, 385–397. [CrossRef]

57.   Cheng, Y.; Li, Y.P.; Zhang, N.; Chen, L.J.; Cao, J. A knowledge graph-enabled multi-domain mapping approach supporting product rapid design: A case study of new energy vehicles. *Adv. Eng. Inform.* **2024**, *62*, 19. [CrossRef]

58.   Hou, X.G.; Gou, B.C.; Chen, D.K.; Chu, J.J.; Ding, N.; Ma, L. A method to assist designers in optimizing the exterior styling of vehicles based on key features. *Expert Syst. Appl.* **2024**, *254*, 14. [CrossRef]