A Novel Framework for Identifying Customers’ Unmet Needs on Online Social Media Using Context Tree

Taehoon Ko 1, Ilsun Rhiu 2, Myung Hwan Yun 3 and Sungzoon Cho 3,*

1 Department of Medical Informatics, The Catholic University of Korea, Seoul 06591, Korea; thko@catholic.ac.kr
2 Division of Future Convergence (HCI Science Major), Dongduk Women’s University, Seoul 02748, Korea; isrhiu@dongduk.ac.kr
3 Department of Industrial Engineering & Institute for Industrial System Innovation, Seoul National University, Seoul 08826, Korea; mhy@snu.ac.kr
* Correspondence: zoon@snu.ac.kr; Tel.: +82-2-880-6275

Received: 10 October 2020; Accepted: 25 November 2020; Published: 27 November 2020

Abstract: Customer needs and user contexts play an important role in generating ideas for new products or new functions. This study proposes a novel framework for identifying customers’ unmet needs on online social media using the Context Tree through the Hierarchical Search of Concept Spaces (HSCS) algorithm. The Context Tree represents the hierarchical structure of nodes associated with related keywords and corresponding concept spaces. Unlike other methods, the Context Tree focuses on finding the unmet needs of customers from online social media. The proposed framework is applied to extract customer needs for home appliances. Identified customer needs are used to make user scenarios, which are used to develop new functions of home appliances.

Keywords: Context Tree; HSCS algorithm; customer needs; user context; online social media

1. Introduction

In respect of user-centered design (UCD), customer needs contribute significantly to determining the initial direction of the new product development (NPD) process [1]. Customer needs refer to the various attributes of the potential product that the customer requires, also known as customer attributes or customer requirements. Considering customers’ unmet needs in the early stages of the NPD process, the design direction of new functions and specifications for the new product can be quickly set [2,3]. In addition, many studies have claimed that it is important to constantly grasp customers’ unmet needs and consider them in the full-cycle NPD process [4–7].

Companies attempt to identify customer needs through various channels, such as prototype, market research, focus group interview (FGI), and user observation [8–10]. These methods generate knowledge through question and answer between experimenter and user. However, these methods have several drawbacks. The answers of participants tend to be induced by the experimenter [11]. In addition, the survey method does not even provide hidden meaning to the basis of the results [12].

Recently, online social media has become an important area for investigating customer preferences [13]. Online social media is defined as services that enable users to establish online connections with their friends through the application and share information with them [14]. It has great potential to generate new knowledge and business value [15]. There is increasing use of social media data to derive customer needs. In marketing, some studies that attempt to extract insights for promotional strategies from social media are underway [16–18]. To derive perceptions of people, the
sentiment analysis methods can be applied to social media text [19]. These studies focus on calculating the number or ratio of keyword frequencies or latent topics in retrospectively collected data. However, while these studies are useful for finding customer context retrospectively, it is more important to find needs that do not satisfy customers in terms of NPD and UCD.

This study proposes a quantitative approach to identifying customers’ unmet needs in online social media, such as social network services (SNSs) or blog services. The proposed method is Context Tree, which is a tree structure expressing user context while searching for related keywords hierarchically. Each node of the Context Tree corresponds to a representative keyword and its related keywords. Various experiences of users with keywords are summarized in each node. On the business side, users’ summarized experiences can be used to derive the new features or specifications of new products. To create the Context Tree, it is necessary to collect a large amount of text data from online social media and pre-process the data using natural language processing (NLP) techniques. The branches of the Context Tree are expanded by the Hierarchical Search of Concept Spaces (HSCS) algorithm, which is proposed by this study. The HSCS algorithm’s purpose is to find unmet needs for the target product in online social media data, especially text data. This algorithm figures out hierarchically summarized user experiences for the target product \( p \) of the NPD process through related keywords to \( p \). Furthermore, it links user experiences, which are not relevant to \( p \) and are relevant to \( p \)'s-related keywords. The specific process of the HSCS algorithm is described in Section 3.2.

To verify the Context Tree generated by the HSCS algorithm, the proposed framework is used to discover ideas for implementing the NPD process for major home appliances. Target online social media data includes tweets on Twitter and blog posts on Korean web portals.

The remainder of this paper is organized as follows. In Section 2, several studies on market research, idea generation, and social media data are summarized. In Section 3, the Context Tree proposed in this study is introduced. Then, how the HSCS algorithm creates a Context Tree to find unmet needs is described in detail. Section 4 describes the online social media data and experimental settings used to show the effectiveness of the proposed method. Section 5 describes the result of creating the Context Tree from online social media data using the HSCS algorithm. For the five home appliances that are targeted, user contexts found in online social media and improvement scenarios for each product are created accordingly. In Section 6, a detailed discussion of the proposed method and experimental results are provided. Section 7 contains the conclusion of this study.

2. Literature Review

2.1. Methods for Identifying Customer Needs

This section describes the four most widely used methods of identifying customer needs, such as prototype, market research, FGI, and user observation [8–10]. Prototyping helps develop new products for selected subjects and obtain feedback. This method can be used when the product specification is clearly defined. Market research methods: market research refers to methods of delivering the questionnaire to many people and analyzing their responses. Structured questions about newly developed products or services are constructed, and individual opinions are recorded. It is important to configure the questions appropriately at an early stage to understand what potential users think about the product. FGI is an in-depth interview with a small group of potential users of products and services. The target size is small in comparison to market research; however, it is considered to obtain a more specific picture of what users think. Users’ observation methods help thoroughly understand users’ behaviors and analyze their problems and needs. The data analyzed by the experts include images or written texts of users’ daily records.

Despite being widely used, they have several drawbacks from the point of view of identifying customers’ unmet needs. The first weakness is that most of the methods require substantial monetary and time costs. It takes a lot of time and money to develop a prototype to obtain a user’s feedback. Market research and user observation often incur large costs due to the recruitment of participants and compensation for responses. The second weakness occurs in terms of sample size. Prototypes,
FGI, and user’s observation require substantial money and time to increase sample sizes for investigation. In market research, the more people who participate in the survey, the higher is the cost. The bigger problem, however, is that the specificity of survey results may deteriorate due to limitations on the number of questionnaires [20]. The third weakness is that people who participate in surveys already know the intentions of marketers. For example, if the company conducts a survey on a new product right after users experience it, some of them become aware of not only the intent of the questionnaire but also the responses the company wants [11]. This is likely to lead to biased results.

Table 1 summarizes the characteristics of the four methods discussed above.

| Method               | Monetary Cost | Time Cost | Sample Size | Risk of Exposing the Intent of Questionnaire |
|----------------------|---------------|-----------|-------------|---------------------------------------------|
| Prototype            | high          | high      | small       | very high                                   |
| Market Research      | high          | high      | large       | high                                        |
| Focus Group Interview| (relatively) medium | high      | small       | high                                        |
| User’s Observation   | high          | high      | very small  | high                                        |

2.2. Idea Generation and Market Research

Girotra et al. defined the term ‘idea’ as several possible solutions to organizational problems [6]. This study examined the idea generation process through brainstorming in detail. Trott claimed that market research is more effective in the new product development (NPD) process when the difference in understanding of techniques between customers and the company is smaller [21]. Witell et al. showed that market research techniques could express manifest needs, though they have difficulty transferring latent needs in certain contexts [22]. Market research techniques, which concentrate on capturing customers’ previous experiences with a product or service, have been designed so that the participants respond to stimuli from the company [23].

Some studies showed the effect of crowdsourcing in terms of idea generation for a new product. Poetz and Schreier showed that both professionals and general users could provide ideas to solve an effective and relevant problem in the consumer goods market for baby products [24]. They evaluated all ideas derived by professionals and general users in terms of key quality dimensions, including novelty, customer benefit, and feasibility. Consequently, although they are blind to the source of ideas, users give ideas that are as effective as those from professionals. Bayus researched a case study with the Dell IdeaStorm community [25]. This study analyzed the pattern of crowdsourced and showed that initial generated ideas were the most effective in solving real business problems.

2.3. Studies in Various Business Fields Using Social Media Data

Schivinski and Dabrowski showed the effect of social media communication on consumer’s perceptions of brands [18]. This study compared the impact of firm-created and user-generated social media communication on brand equity, brand attitude, and purchase intention. This study claimed that user-generated social media communication had a positive influence on both brand equity and attitude, while firm-created communication affected only brand attitude. Kim and Ko proposed attributes of social media marketing activities and examined the relationships among those perceived activities, such as value equity, relation equity, brand equity, customer equity, and purchase intention, through a structural equation model [26]. They claimed that value and relationship equity had significant positive effects on purchase intention.

Some studies focused on extracting actionable topics, information, and knowledge from online social media data to solve business problems. Saha and Sindhwani proposed a framework for modeling the topic of evolution and discovering emerging themes in social media [27]. McAuley and Leskovec combined latent rating dimensions with latent review topic made by latent Dirichlet allocation (LDA) [28]. Tuarob and Tucker extracted information on several smartphones from Twitter
to predict product demand and longevity [29]. They analyzed tweets and figured out several strong, weak, and controversial features of each smartphone. Furthermore, they claimed a strong correlation between positive tweets and sales for each smartphone. Timoshenko and Hauser analyzed oral-care users’ reviews on Amazon, spanning the period from 1996 to 2014. This study used word embedding and convolutional neural network (CNN) to classify informative and uninformative sentences from users’ reviews. After that, non-repetitive sentences were identified through domain experts to derive the insights necessary for marketing actions [13]. Ko et al. used latent Dirichlet allocation (LDA) to extract topics from user reviews on Reddit.com [30]. This study constructed a KeyGraph based on the co-occurrences among the topics. Kühl et al. compared the insights on electric vehicles derived through literature and the insights derived by analyzing tweets on Twitter [31]. Jeong et al. applied the topic modeling method, especially latent Dirichlet allocation (LDA), to users’ reviews on Samsung Galaxy Note 5 to identify several topics and their sentiment weights [32]. Reisenbichler and Reutterer discussed how to use topic modeling methods in the marketing area [33].

Studies finding users’ experiences on specific products in online social media are published. Rhiu and Yun searched keywords in tweets and blog posts related to smartphones. They figured out that there were unsatisfactory experiences for using smartphones in terms of multi-functionality and connectivity [11]. Kim et al. analyzed user ratings and reviews on recliners on Amazon and found 15 clusters for affected users’ experience of recliners [34]. Joung et al. analyzed customer complaints about air-conditioners using feature selection and clustering to identify requirements at which customers want to obtain help from products [35]. He et al. conducted a case study to analyze tweets about laptop brands. Through sentiment analysis, this study quantitatively analyzed how positive and negative people had thoughts of each laptop brand [36].

The above-mentioned studies are summarized in Table 2.

Table 2. Related studies using online social media.

| Study                          | Purpose                                | Target Online Social Media or Data | Methods                                      |
|-------------------------------|----------------------------------------|-----------------------------------|----------------------------------------------|
| Schivinski and Dabrowski      | Figure out consumers perceptions of brands | Facebook                         | Structural equation model                     |
| (2016)                        |                                        |                                   |                                              |
| Kim and Ko (2012)             | Demonstrate the effectiveness of social media marketing | Questionnaires                    | Structural equation model                     |
| [26]                          |                                        |                                   |                                              |
| Saha and Sindhwani (2012)     | Capture the evolution and emergence of topics in social media | U.S. National Institute of Standards and Technology (NIST) Topic Detection and Tracking (TDT) corpus | Dynamic non-negative matrix factorization (NMF) |
| [27]                          |                                        |                                   |                                              |
| McAuley and Leskovic (2013)   | Find the relationship between hidden factors in ratings and hidden topic in reviews | Reviews on Amazon and Yelp         | Latent Dirichlet allocation (LDA), hidden factors as topics (HFT) |
| [28]                          |                                        |                                   |                                              |
| Tuarob and Tucker (2013)      | Obtain information on smartphones to predict product demand and longevity | Tweets on Twitter                 | Sentimental analysis                         |
| [29]                          |                                        |                                   |                                              |
| Timoshenko and Hauser (2019)  | Select informative sentences for identifying customer needs | Reviews on Amazon                 | Word embedding and convolutional neural network (CNN) |
| [13]                          |                                        |                                   |                                              |
| Ko et al. (2017)              | Extract topics for generating product opportunities | Reviews on Reddit                 | Latent Dirichlet allocation (LDA), Keygraph  |
| [30]                          |                                        |                                   |                                              |
Kühl et al. (2019) [31] Extract and categorize customer needs Tweets on Twitter Keyword searching

Jeong et al. (2019) [32] Extract topics and corresponding sentiments Reviews on Reddit Latent Dirichlet allocation (LDA), Sentiment analysis

Reisenbichler and Reutterer (2019) [33] Review paper about topic modeling in marketing - -

Rhiu and Yun (2018) [11] Explore users’ experience of smartphones Tweets on Twitter Keyword searching

Kim et al. (2019) [34] Extract affective user experience of recliners Reviews on Amazon Self-organizing map (SOM)

Joung et al. (2019) [35] Cluster customer complaints Customer complaints Hierarchical clustering

He et al. (2019) [36] Extract topics and corresponding sentiments Tweets on Twitter Sentiment analysis

3. Proposed Method

This section provides a specific description of the proposed methods for identifying customers’ unmet needs in online social media. The proposed framework consists of the following steps, illustrated in Figure 1. First, documents (or short paragraphs) written by the public should be collected from online social media and preprocessed to analyze the data. In this step, NLP techniques are used to normalize text data. After that, a keyword search is performed within the collected data using the HSCS algorithm. The HSCS algorithm designates a target product as a primal keyword and then searches not only the current user’s experience but also the basis of customers’ unmet needs by using related keywords. The basis for unmet demand can be used to design new features or specifications for target products in the future.
3.1. Gathering and Pre-Processing the Online Social Media Data

The first step is to determine the target online social media. Social media has highly diverse categories with numerous topic types. Social network services (SNSs), such as Twitter, Facebook, and LinkedIn, can enable gathering opinions from the unspecific majority, or comments by readers on various news sites and writing on community sites, such as Reddit, where various topics are discussed, can be good analysis targets. Although this study focuses on texts, services, such as Instagram, can also be considered if image data and corresponding hashtags are analyzed together. In addition, if customer’s needs to specific domains are required, companies can target comments on movie review sites or online shopping sites.

Once the target data is determined, it is necessary to search for a method to collect them. For online social media data collection, it is preferable to use application programming interfaces (APIs) of corresponding sites [37]. Note that APIs may not provide the data fields preferred by users or limited data quantities and collection time while using APIs. A strategic partnership may be necessary to utilize APIs or to share costs. In addition, companies may utilize dumped data disclosed for the purpose of research. Sites, such as Wikipedia, where various users participate, often provide updated dumped data, and Amazon.com discloses product review data for the purpose of research. A proactive method of gathering data is web page information through the web scrapping technique and extracting values that correspond to required data fields using HTML parsers.

After data gathering, preprocessing should be conducted. For text data, Natural Language Processing (NLP) tools in accordance with a language type should be used [38]. The important things in the proposed framework particularly are word extraction, word normalization, and out-of-vocabulary problem-solving. Word extraction extracts words used by the majority from free-formed texts. The word ‘normalization’ refers to a conversion from changed terms into basic type words,
which is significantly conducive to unifying information inside texts [39]. Tokenizing, part-of-speech tagging, stemming, and lemmatizing are necessary for word extraction and normalization, as a general procedure in NLP [40].

- **Tokenizing**: It is the extraction of linguistic units (e.g., words, phrases, or specific sequences of characters) used by people from unstructured texts. A white space and punctuations that determine the boundary of such units are often used in tokenizing, and variations, such as a change in character types, collocations, and idioms, are also considered in the tokenizing process.

- **Part-of-speech (POS) tagging**: POS tagging is the process of assigning part-of-speech tags, including noun, verb, adjective, adverb, etc., to words in a sentence or corpus. A POS tagger determines a POS tag for each word, considering a sentence structure and surrounding context. For English, the Penn Treebank tag set has been widely used. In this framework, it is necessary to derive nouns that refer to objects, adjectives and adverbs that reveal emotion, and verbs that express actions through POS tagging.

- **Stemming**: This is the process of transforming several words into a base root form. For example, ‘computer’, ‘computing’, ‘computed’, and ‘computation’ are transformed into a root form of ‘comput’. Stemming enables information collection from within documents containing the four existing words in an integrated manner.

- **Lemmatizing**: Lemmatizing also refers to the summarization of words utilized in various forms into a single representative lemma. Stemming considers only processing target words, whereas lemmatizing considers elements to identify the surrounding contexts, for example, part-of-speech (POS) tags. Thus, lemmatizing takes more time than stemming.

The most difficult aspect of analyzing social media data is processing slang or jargon [41]. Note that the importance of processing slang or jargon can differ with analysis topics. For example, if the purpose is to infer a user’s experience of a specific product, as shown in the case study later, processing slang or jargon is necessary in terms of emotional perspective. That is, a process to express specific emotions used by developing a separate dictionary without the need to change all slangs and jargons into semantically similar words. Accordingly, the number of terms to be processed becomes smaller, reducing the cost in terms of time taken for pre-processing.

3.2. Constructing the Context Tree Using Hierarchical Search Concept Space (HSCS) Algorithm

After the completion of data gathering and pre-processing, the **Context Tree**, a novel model that represents the user’s context and unmet needs in online social media, is ready to be created. The Context Tree refers to a data structure that extends branches by searching other detailed keywords related to a specific keyword iteratively. The **Hierarchical Search Concept Space (HSCS) algorithm** is also proposed as a method that creates the Context Tree. This algorithm has the property of simple ontology. A node in each tree refers to a corresponding keyword, and writings, including the corresponding keyword and parent keywords, belong to the node. In addition to this, a process to extract the user’s experience that can infer the unmet needs of users by controlling the inclusion and exclusion of words is included in the HSCS algorithm. The Context Tree enables intuitive and structural insight search for people who participate in the NPD process by summarizing documents that correspond to the keyword array effectively.

The detailed description of the HSCS algorithm for creating the Context Tree is as follows.

3.2.1. Determining the Primal Keyword

Before describing the process of constructing Context Trees, two important terms must be defined. First is a **concept space**, which is represented by some documents with an arbitrary keyword or keyword set. In this research, the target documents are the articles of the public on social media. Therefore, the concept space is a space that collects contexts that the public thinks about the corresponding keyword or keyword set. The other term is a **primal keyword**, which is a word that represents the target object. In the NPD process, the **primal keyword** refers to a term that is the target
product itself or closely related to the product. To acquire insights for new functions or specifications about specific products, for example, users’ experiences should be searched in social media. In this regard, the primal keyword becomes the product itself. The primal keyword serves as a seed query to construct the Context Tree.

Assume that the primal keyword is \( p \), the concept space of overall social media data is \( U \), and document, including the determined primal keyword \( p \), is a concept space \( P \). As the size of \( U \) is large in general, the relationship \(|U| \gg |P|\) is established. This progress so far is depicted in Figure 2.

### 3.2.2. Finding Related Keywords that Appear with the Primal Keyword

This step extends the Context Tree by extracting related keywords \( w_1, w_2, w_3, \ldots, w_p \) from documents that correspond to \( P \). It is now necessary to evaluate the association between the primal keyword \( p \) and other words to extract related keywords. This study calculates co-occurrences of the primal keyword \( p \) and every word and selects words with high co-occurrence as related keywords. The document data space, where these two keywords are included, is called concept spaces. \( W_1, W_2, \) two concept spaces \( P \cap W_1, P \cap W_2 \) can be obtained when the Context Tree is extended once, as shown in Figure 3.
3.2.3. Finding Related Keywords Recursively

This step iterates the above process constantly. Let us assume that related keywords extracted by analyzing the document data, where previously extracted $w_i$ appears, are called $w_{ij}$, and the corresponding concept space is called $W_{ij}$. Here, child nodes of the Context Tree refer to $P \cap W_i \cap W_{ij}$ concept spaces. As this process is progressed once, the Context Tree is extended a step further. Furthermore, a narrower concept space can be generated as the extension is progressed further.

3.2.4. Discovering Contexts in Each Node of the Context Tree

Each node of the Context Tree has documents containing words, such as the target product $p$ and its related keyword $w_i$, etc. From these documents, contexts that correspond to each concept space can be searched and discovered. These contexts represent public awareness of the target products and their utility.

To prevent this problem, this study employed a method that identified the user’s context from the corresponding concept space based on the extracted related keywords. Extracting user contexts using related keywords has advantages of not only convenient expert interpretation but also quantitative evaluation through the frequency of keywords that respond to the corresponding context.

3.2.5. Discovering Contexts Used for Expanding the Concept Space of the Target Product

This step attempts to discover contexts that cannot be solved or correspond to the current target product. To discover the corresponding context, it is important to use documents whose related keyword $w_i$ of the primary keyword $p$ is present, but $p$ is not present. They belong to $P^c \cap W_i$ in the concept space. The result of expanding the Context Tree is illustrated in Figure 4.

![Figure 4. Discovering contexts used for expanding the concept space of the target product.](image)

If some parts can be solved by new functions or specifying the target product among the contexts, the concept space $P$ of the existing target product can be scalable to larger $P'$. Furthermore, $P' \cap P^c \cap W_i$ refers to potential areas where the target product can be used to satisfy customers’ unmet needs. They become a source for developing new functions or specifications of the target product. The following Figure 5 shows the detailed description.
Figure 5. Expanding the concept space of the target product. (a) The concept space $P$ is expanded to $P'$. (b) Shaded spaces can be regarded as evidence for the customer’s unmet needs and the new function of the target product.

The HSCS algorithm described so far is summarized in Table 3.

### Table 3. Hierarchical Search of Concept Spaces (HSCS) algorithm.

**Algorithm 1: Hierarchical Search of Concept Spaces (HSCS)**

1. Determine the primal keyword $p$.
2. Extract documents including $p$. Assign this corpus to concept space $P$.
3. Define the function named $\text{Pipeline}(\text{input\_space}=I, \text{number\_of\_keywords}=n)$:
   i) Preprocess the corpus in $I$. (tokenizing → POS tagging → stemming and lemmatizing → Processing slang or jargon)
   ii) Sort words by frequency in descending order and determine the top-words $w_i (i = 1, 2, ..., n)$.
   iii) Extract documents including $w_i$. Assign this corpus to each concept space $P \cap W_i$.
   iv) Return keyword set $\{w_i\}$ and concept space set $\{I \cap W_i\}$ for all $i = 1, 2, ..., n$.
4. Extract keyword set related to $P$ and corresponding concept space set:
   \[\{w_i\}, \{P \cap W_i\} = \text{Pipeline}(P, n_1) \text{ for all } i = 1, 2, ..., n_1\]
5. Extract keyword set related to each $P \cap W_i$ and corresponding concept space set:
   \[\{w_{ij}\}, \{P \cap W_i \cap W_j\} = \text{Pipeline}(P \cap W_i \cap W_j, n_2) \text{ for all } i = 1, 2, ..., n_1 \text{ and } j = 1, 2, ..., n_2\]
6. Find new concept space related to keywords $w_i$ and $w_{ij}$ (for all $i = 1, 2, ..., n_1$ and $j = 1, 2, ..., n_2$) but not related to $P$:
   - \[\{w_i\}, \{P_c \cap W_i\} = \text{Pipeline}(P_c, n_1) \text{ for all } i = 1, 2, ..., n_1\]
   - \[\{w_{ij}\}, \{P_c \cap W_i \cap W_j\} = \text{Pipeline}(P_c \cap W_i \cap W_j, n_2) \text{ for all } i = 1, 2, ..., n_1 \text{ and } j = 1, 2, ..., n_2\]

### 4. Experimental Setting

For solving a real business problem, the proposed method is used to identify customers’ unmet needs for home appliances. Target home appliances are refrigerator, washing machine, robot cleaner,
air conditioner, and oven. Based on identified customer needs and user contexts, several user scenarios are established for developing new functions of target products.

4.1. Background

Many companies have tried to develop new functions of the product based on exchanging information between smartphones and home appliances. In many cases, rather than considering customer-perceived values first, user scenarios based on available technologies and methodologies have been created. However, since these technology-specific scenarios are not from the user, the utility of the users is unknown. For example, if a company has a good voice recognition technology, product features can be developed to take full advantage of that technology. In this case, if the experience of customers who use the functions of the product is not sufficiently taken into consideration, the possibility of introducing a new product that is different from the user’s needs of the market increases.

To develop a function that fits the user’s experience, it is important to create a scenario that fully reflects the user’s experience and needs that occur during the actual user’s use of the target product. The efficient and effective NPD process is achieved by considering technology and user interface (UI), reflecting the user-friendly scenario in the early stage of product development. In the preceding example, it is difficult to make a smart product by speech recognition technology alone. If users are unspecified, it is necessary to concentrate on creating speech recognition technology robust to users’ accent and their surrounding noise. Besides, depending on whether it is important for the question and answer process to proceed smoothly and how it is important to form a bond with the user through dialogue with the virtual bots, the technology required to create an artificial intelligence (AI) model will change, and the model itself will be different.

Therefore, this study intends to develop a new function of smart home appliances from the viewpoint of user-centered design rather than technology-centered design. To achieve this goal, customers’ unmet needs (or user contexts) for five home appliances are derived from online social media using the Context Tree.

4.2. Online Social Media Data and Processing Tool

Targeted online social media data consists of Korean tweets and blog posts on Naver.com, which is the biggest web portal in South Korea, containing several words representing previous target products. The tweets and blog posts used in the analysis were generated between May 2012 and April 2013. Because application programming interfaces (APIs) provided by Twitter and Korean web portals provide a limited number of data acquisitions, a web-based tool named ‘SOCIAL Metrics’ (http://socialmetrics.co.kr), which has already collected all tweets and blog posts through strategic alliances with online social media platforms, is used to prove the proposed method. SOCIAL Metrics was developed by DaumSoft, which is one of the leading social media [11]. SOCIAL Metrics parses all texts and converts them into words using a Korean natural language processing tool and normalized vocabulary developed in-house. Using this tool and database, the frequency of keywords and co-occurrence between keywords are calculated. Figure 6 illustrates the structure of SOCIAL Metrics. It derives related keywords with high co-occurrence by searching the words as queries.
5. Results

5.1. Applying the HSCS Algorithm to Online Social Media Data

The target products are five home appliances. This section describes an example that identifies customers’ unmet needs for ‘oven’, which is one of the target products, from online social media.

In this case, the primal keyword \( p \) is ‘oven’. Tweets and blog posts containing \( p \) are collected to make the concept space \( P \). This concept space is identical to the root node of the concept tree. As a result, 100,000 tweets and blog posts are found. After that, keywords related to \( p \) are determined. They are the top-\( k \) words that appear frequently in the concept space \( P \). Several concept spaces \( P \cap W_i \) (for \( i = 1, 2, ..., k \)) are branched out from the root node of the concept tree. The result is described in Table 4.

| Related Keywords | Term Frequency | Discovered User Contexts |
|------------------|----------------|----------------------------|
| Taste            | 66,455         | • Food made with the oven is usually delicious  
|                  |                | • New way of cooking by oven usually fails |
| Home             | 57,300         | • Do not know how to cook with an oven  
|                  |                | • Many people want to cook diverse cuisines, but they do not have any oven |
| Kids             | 36,078         | • Make a snack for the child using the oven  
|                  |                | • Oven should be kept out of reach of children |
| Snack            | 26,082         | • Want to make healthy snacks with the oven  
|                  |                | • Cooking snacks with the oven is sometimes tiresome |
| Mother           | 21,254         | • Although my mom has the oven, she doesn’t cook foods with the oven |
| Fail             | 2895           | • Usually fails to cook with the oven |
| Troublesome      | 269            | • More troublesome to cook with oven than microwave or gas stove |

Users’ contexts related to the primal keyword ‘oven’ and its related keywords are extracted by document clustering and frequently appeared words. For example, in the concept space with ‘oven’ and ‘fail’, there are many tweets that say that if you cook in the oven, you would fail. Besides, words like ‘usually’, ‘again’ are frequently appeared in that concept space. As a result, the user’s context named ‘usually fails to cook with oven’ is found, as shown in Figure 7.
Figure 7. Tweets and discovered context in the concept space with ‘oven’ and ‘fail’. It is found out that many people want to make healthy breads and snacks for their family. In the next step, tweets and blog posts, which do not contain the primal keyword ‘oven’ and contain keywords like ‘bread’, ‘snack’, ‘baking’ related to the primal keyword, are collected. By exploring the data, the following users’ contexts are discovered:

- Recipe is particularly important in baking.
- The main reason that baking fails is that many people make a mistake in the process of the recipe, such as not well controlled preheat temperature.
- I want to make snacks with healthy and good ingredients for my kids.
- I want to make healthy snacks for children, but I do not know what snacks to make.

The Context Tree of ‘oven’ is constructed based on these users’ contexts, illustrated in Figure 8.

5.2. Developing User Scenarios

After identifying customers’ needs from social media, it is necessary to create users’ scenarios in the same way as existing NPD processes. For this purpose, a task force team is made to create
scenarios based on the Context Tree. The team consists of product managers, engineers, market researchers, and (potential) customers. The team focuses on users’ pain points and corresponding solutions. Besides, the team does a lot of discussion about whether solutions are feasible and product specifications and price are appropriate for the customer.

As a result, five user scenarios for each target appliance are created. In this paper, one user scenario for a smart oven, named ‘Oven never fail to cook’, is described in detail. In the Context Tree of the oven described above, many customers want to make healthy foods, snacks, and breads with the oven. Besides, there is a strong sense that it is difficult to cook something with the oven. To satisfy these users’ contexts, the system where the oven manufacturer directly manages the recipes using its oven is proposed.

The scenario ‘Oven never fail to cook!’ is illustrated in Figure 9. The details of the scenario are as follows:

- Send recipes to oven online: Each oven has different product specifications, such as power consumption, space size, and temperature rise rate. For this reason, even with the same recipe, the result may vary depending on the oven. Thus, if the oven manufacturer manages detailed recipes that can make some foods using their own oven, consumers can receive a great deal of convenience. If the recipes are passed to the oven as eXtensible Markup Language (XML)—like data that is easy to parse, it will be possible for the oven to automatically adjust the temperature and heating time to the recipe.
- Automatic settings are made for cooking: The schedule of the main settings is automatically set. In addition, it transmits to the user a voice when the user inserts additional ingredients or subtracts the dish.
- Oven can interact with the user: There can be several interactions between the oven and its user. For example, if the food in the oven is overcooked and smoked, it will promptly notify the user to stop the operation or take out the food quickly. If the oven can detect a user’s mistake in terms of the recipe, it may be able to tell the user about the solution.

![Figure 9. An example scenario of oven: ‘Oven never fails to cook!’](image)

The users’ contexts targeted by this scenario is highlighted in the Context Tree, like in Figure 10.
In this manner, there are 18 user scenarios for satisfying customers’ needs and users’ contexts. Some of these scenarios are currently being implemented as new functions in home appliances. These scenarios are listed in Table 5.

### Table 5. Scenarios for 5 home appliances.

| Product      | Scenarios                                                                 |
|--------------|---------------------------------------------------------------------------|
| Robot cleaner| Recognizing of obstacles                                                 |
|              | Recognizing and informing dirty place                                    |
|              | Customizing schedule of cleaning                                          |
|              | Recognizing and informing home condition (temperature, illumination, humidity, etc.) |
| Air conditioner | Recognizing the usual on/off pattern                                    |
|              | Informing the temperature around the air conditioner                     |
| Oven         | Informing the optimized recipe for the oven                              |
|              | (Subtitle: ‘Oven never fail to cook’)                                    |
|              | Informing the recipe according to the cooking level of a user            |
|              | Recommending the recipe according to context                             |
|              | Observing food through the camera in the oven                            |
| Refrigerator | Informing where to clean                                                  |
|              | Informing what kinds of food is insufficient                              |
|              | Managing the temperature of a refrigerator                               |
|              | Informing the effective method to store and manage food                  |
|              | Mini-refrigerator                                                        |
|              | Reducing the noise of the refrigerator                                   |
| Washing machine | Informing the best day to do laundry                                    |
|              | Informing where to dry laundry                                           |

### 6. Discussion

#### 6.1. Which is Right for Identifying Customer Needs: Tweets, Blog Posts, and Other Alternatives

The data used in this study are tweets and blog posts. As comparing the two, it is easier to identify the keyword-based user contexts about home appliances using tweets than blog posts. A single tweet is a short sentence within 140 characters (not strict criteria), while a blog post is long and
contains a lot of nested topics compared to tweets. For example, when the word ‘oven’ appears on a
tweet, the subject of the tweet is often highly related to the ‘oven’. On the other hand, even though
the word ‘oven’ is included in a blog post, the main topic of that blog post may be the daily record or
house interior, which are not significantly related to the oven.

Of course, the above example is not always true. For some subjects, there may be many posts
that deal with them in detail, and even the average length of those posts is long. It is a good idea to
check whether our target social media data is proper to extract users’ contexts through preliminary
analysis. However, if you do not have much time to collect and analyze all types of data, it is strongly
recommended to use short texts, such as tweets, product reviews, etc.

Except for tweets and blog posts, some websites and applications that deal with specific topics
in-depth are good for extracting users’ contexts and identifying customers’ needs. These types of
websites and applications include:

- Yelp: reviews on restaurants and dishes
- Amazon: reviews on many products (in specific categories related to target subjects)
- IMDB and Rotten Tomatoes: reviews on movies and TV shows
- Yahoo Answer: questions and corresponding answers about target subjects
- Reddit: Opinions on target subjects

6.2. Comparison between the Proposed Method and Conventional Methods for Identifying Customers’ Needs

In this section, the proposed method is compared with other methods in various aspects.

6.2.1. Sample Size and Sample Bias

In terms of sample size, the proposed framework is much stronger than conventional methods.
The number of social media users is huge, so it is easy to gather various opinions on social media.
With respect to sample bias, the proposed framework also has advantages and disadvantages. First,
because the number of social media users is huge, there is little bias towards the small sample size.
However, social media users may be biased toward age or sex. For example, it is not easy to get the
opinions of older people with online social media. This bias is difficult to control for researchers.
Therefore, it is a good idea to select proper subjects to utilize social media data.

6.2.2. The Direction of Discovering User Context

The proposed methodology starts from a broad concept space and reaches narrow concept
spaces by extending the Context Tree. It means that the proposed framework finds customer needs
in the direction of specific concept space in general concept space. On the other hand, conventional
methods try to derive several contexts from a small number of samples and generalize them.

It is important to consider this difference between the proposed framework and conventional
methods. If you need to track a user’s experience for a long time, it is not desirable to use social media
data. In this case, the Day Reconstruction Method (DRM) [42] is suitable for researchers to observe
users of the target product for a certain period. The use of social media data has the advantage of
being able to draw the user’s experience from many people. Besides, it does not convey the intent of
the researcher to people, so it is expected to obtain a more vivid, unfiltered experience.

6.2.3. Cost and Reusability

The proposed methodology requires a lot of costs. It requires an infrastructure to collect and
store social media data. In addition, some software programs are needed to develop, such as a natural
language processing toolkit for languages, which appear in the collected data, and a data exploration
tool for researchers. Therefore, the proposed framework can be more expensive than other methods.
However, once such a system is prepared, it is possible to easily extract the users’ contexts for any
subject. Although the initial investment cost is high, it has a great advantage in terms of reusability.

6.2.4. Easier to Implement than Machine Learning-Based Topic Modeling Methods
To find the context, machine learning-based topic model methods, including latent semantic analysis (LSA) and latent Dirichlet allocation (LDA), or clustering methods, including k-means clustering, may be employed. However, these mathematical and statistical approaches require extrapolations of analysts’ subjectivity to use latent dimensions or cluster centroid vectors in product development, and it is difficult to control good quality contexts via an algorithm. LSA and LDA extract latent vectors, also known as topic vectors, which are related to words. Each latent vector does not represent a single topic that humans can understand. After extracting the latent vector, the human must interpret and label it by looking at the related words and corresponding weights. Generally, this process is very subjective, and there is no guarantee that you can label with a clear topic. Therefore, deriving ideas for new products, which is this study’s goal, using machine learning-based topic model methods is likely to fail. Furthermore, LSA and LDA have several hyperparameters, which should be controlled for good performance. Especially, LDA is hard to determine the number of topics beforehand [33].

The clustering algorithm, one of the unsupervised machine learning methods, is also difficult to achieve our research purpose in the same context. By clustering documents, several cluster centroid vectors are derived, and they can be considered as the main topics. However, interpreting cluster centroid vectors is also not easy. Besides, it is difficult to produce good clustering results depending on the data.

To prevent this problem, this study employed a method that identified the user’s context from the corresponding concept space based on the extracted related keywords. Extracting users’ contexts using related keywords has advantages of not only convenient expert interpretation but also quantitative evaluation through the frequency of keywords that respond to the corresponding context. In addition, the proposed method is more suitable for NPD as it is more focused on the purpose of targeting the unmet needs of customers.

### 6.3. Criteria for Extracting Related Keywords

In the previous case study, term frequency is one of the criteria for extracting related keywords. In fact, three criteria for extracting keywords can be considered. The first criterion is the frequency of words. This is a way of preprocessing target texts and considering frequently appearing words as relevant keywords. The second one is the degree of diffusion. It means that many propagated articles are important. The third one is the rating of domain experts and researchers. They can select texts by their own knowledge. To prove out concepts, the article sets constructed by combinations of the above selection criteria are compared. Fifty tweets and 50 blog posts are included in each set. Fifteen non-experts and five experts are involved in the evaluation of these sets. They score how relevant corresponding articles are to the problem of identifying customer needs. A 10-point scale is used for scoring. The result of this experiment is shown in the following Table 6. Considering the frequency of words and rating of experts is the best scored.

**Table 6.** Evaluation of criteria combinations for extracting keywords.

| Combinations of Criteria                  | Average Score |       |       |
|------------------------------------------|---------------|-------|-------|
|                                          | Non-experts   | Experts | Total |
| Frequency of words (F)                   | 5.27          | 5.6    | 5.3525|
| Degree of diffusion (D)                  | 5.07          | 3.6    | 4.7025|
| Rating of experts and researchers (R)    | 6.47          | 7.4    | 6.7025|
| F + D                                   | 5.6           | 6.2    | 5.75  |
| F + R                                   | 7             | 8.8    | 7.45  |
| D + R                                   | 6.4           | 6.8    | 6.5   |
| F + D + R                               | 6.27          | 7.8    | 6.6525|

### 7. Conclusions

In this paper, the Context Tree is proposed as a new approach to identifying significant customers’ unmet needs from online social media. The Context Tree consists of the primal keyword,
its related keywords, and concept spaces. Each node of the Context Tree has a representative keyword and a corresponding concept space. So, the Context Tree is useful for structurally exploring the users’ experience and context for the target product with respect to NPD. Besides, the proposed HSCS algorithm goes beyond simply searching for related keywords and identify the basis and evidence for customers’ unmet needs. Therefore, this basis can be used to derive ideas for designing new functions or specifications for the target product, which is foremost in the NPD process. With the case study, users’ contexts on five home appliances are identified using the Context Tree based on the HSCS algorithm. These contexts are used as sources to create users’ scenarios for new functions and specifications of the target appliances.

The main contribution of the Context Tree with HSCS algorithm is that the proposed method presented a way of utilizing online social media data, focusing on idea and concept screening stages in the early stages of the NPD process. According to [5], the idea screening stage and concept screening stage are critical to orienting the new product. The proposed method seeks not only the current user’s experience of using the target product but also the context in which the product is not being used to find the basis for unmet needs. As a result of a case study for home appliances, including the oven, users’ scenarios for the use of new products are successfully designed using evidence for customers’ unmet needs identified by the Context Tree. Besides, as mentioned in previous sessions, the Context Tree with HSCS algorithm is not parametric, so it can be applied to big data to easily derive results. In addition, the derived result is composed of structural keywords and context, which has the advantage of easy interpretation.

However, the proposed approach has some limitations because it still requires human effort to make ideas for new functions and specifications of a new product from the Context Tree. Besides, the process of extracting the users’ context may be subjective. Obviously, it would be necessary for the data analyst to participate in finding out which topics are contained in the corpus containing keywords. In this case, the keyword structure can help you. For example, suppose that the primal keyword is ‘oven’. After applying the pipeline to the corpus containing the primal keyword ‘oven’, it is found out that ‘snack’ is the most frequent word. By executing pipelines, suppose that the words ‘healthy’, ‘mom’, and ‘cookie’ come up frequently in the corpus containing ‘snack’. I can combine these keywords to extract the users’ context, like “Mom usually wants to make a healthy cookie using the oven for her children”. The process of extracting the users’ context may be subjective, but the materials, such as frequent keywords and the keyword structure used in the process, are objective through Big Data analysis. This is the biggest difference between conventional methods and the proposed method. However, it is important to reduce the user’s involvement in the above process. Reducing the user’s involvement will require further refinement of algorithms and analytical techniques.

One of the future works is to make users’ contexts intelligently. In this study, users’ contexts are described by keywords and not using intelligent algorithms, such as matrix decomposition algorithms and clustering algorithms. There is a possibility that a person’s subjectivity will intervene when that person views the keyword table and make contexts. Although the result of topic modeling methods like LSA, LDA, and clustering algorithms can be difficult to interpret, they have a strength in minimizing subjective intervention. Therefore, if a topic modeling algorithm with high interpretability is developed and applied to the Context Tree, it is possible to extract the user’s experience and context that can be more objectified.

Author Contributions: Conceptualization: M.H.Y. and S.C.; methodology, T.K. and I.R.; software, T.K.; validation, I.R.; formal analysis, T.K.; investigation, I.R.; resources, M.H.Y. and S.C.; data curation, T.K. and I.R.; writing—original draft preparation, T.K.; writing—review and editing, T.K. and I.R.; visualization, T.K.; supervision, S.C.; project administration, T.K.; funding acquisition, M.H.Y. and S.C. All authors have read and agreed to the published version of the manuscript.

Funding: The article publishing charge is funded by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, South Korea (grant number: HE20C0029).
Conflicts of Interest: The authors declare no conflict of interest.

References

1. Norman, D.A.; Draper, S.W. User Centered System Design: New Perspectives on Human-Computer Interaction, 1st ed.; Lawrence Erlbaum Associates Inc.: Hillsdale, NJ, USA, 1986.
2. Ulrich, K.T.; Eppinger, S.D.; Yang, M.S. Product Design and Development, 7th ed.; McGraw-Hill Education: New York, NY, USA, 2019.
3. Cooper, G.C. The Drivers of Success in New-Product Development. Ind. Mark. Manag. 2019, 76, 36–47.
4. Griffin, A.; Hauser, J.R. The voice of the customer. Mark. Sci. 1993, 12, 1–27.
5. Hart, S.; Hultink, E.J.; Tzokas, N.; Commandeur, H.R. Industrial Companies’ Evaluation Criteria in New Product Development Gates. J. Prod. Innov. Manag. 2003, 20, 22–36.
6. Girota, K.; Terwiesch, C.; Ulrich, K.T. Idea Generation and the Quality of the Best Idea. Manag. Sci. 2010, 56, 591–605.
7. Goffin, K.; Varnes, C.J.; Hoven, C.; Koners, U. Beyond the Voice of Customer: Ethnographic Market Research. Res. Technol. Manag. 2012, 55, 45–53.
8. Kinnear, T.C.; Taylor, J.R. Marketing Research: An Applied Approach; Subsequent Edition; McGraw-Hill Education: New York, NY, USA, 1996.
9. Vredenburg, K.; Mao, J.-Y.; Smith, P.W.; Carey, T. A Survey of User-centered Design Practice. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Minneapolis, MI, USA, 20–25 April 2002; Association of Computing Machinery: New York, NY, USA; pp. 471–478.
10. Sekaran, U.; Bougie, R. Research Methods for Business: A Skill Building Approach, 7th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2020.
11. Rhiu, I; Yun, M.H. Exploring User Experience of Smartphones in Social Media: A Mixed-Method Analysis. Hum. Comput. Interact. 2018, 34, 960–969.
12. Gable, G.G. Integrating Case Study and Survey Research Methods: An Example in Information Systems. Eur. J. Inf. Syst. 1994, 3, 112–126.
13. Timoshenko, A.; Hause, J.R. Identifying Customer Needs from User-Generated Content. Mark. Sci. 2019, 38, 1–20.
14. Noordhuis, P.; Heijkoop, M.; Lazovik, A. Mining Twitter in the Cloud: A Case Study. In Proceedings of the 2010 IEEE 3rd International Conference on Cloud Computing, Miami, FL, USA, 5–10 July 2010; IEEE: New York, NY, USA; pp. 107–114.
15. Dai, M.; He, W.; Tian, X.; Giraldi, A.; Gu, F. Working with Communities on Social Media: Varieties in the Use of Facebook and Twitter by Local Police. Online Inf. Rev. 2017, 41, 782–796.
16. Manigold, W.G.; Faulds, D.J. Social Media: The New Hybrid Element of the Promotion Mix. Bus. Horiz. 2009, 52, 357–365.
17. Baum, D.; Füller, J.; Pedit, T. Social Media Campaigns for New Product Introduction. In Proceedings of the 21st European Conference on Information Systems, Utrecht, The Netherlands, 6–8 July 2013; Association for Information Systems: Atlanta, GA, USA; pp. 1–12.
18. Schivinski, B.; Dabrowski, D. The Effect of Social Media Communication on Consumer Perceptions of Brands. J. Mark. Commun. 2016, 22, 189–214.
19. Tuaro, S.; Tucker, C.S. Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data. J. Comput. Inf. Sci. Eng. 2015, 15, 031003.
20. Adida, C.; Gottlieb, J.; Kramon, E.; McClendon, G. Response Bias in Survey Measures of Voter Behavior: Implications for Measurement and Inference. J. Exp. Political Sci. 2019, 6, 192–198.
21. Trott, P. The Role of Market Research in the Development of Discontinuous New Products. Eur. J. Innov. Manag. 2001, 4, 117–126.
22. Witell, L.; Kristensson, P.; Gustafsson, A.; Löfgren, M. Idea Generation: Customer Co-creation versus Traditional Market Search Techniques. J. Serv. Manag. 2011, 22, 140–159.
23. Verma, R.; Anderson, C.; Dixon, M.; Enz, C.A.; Thompson, G.; Victorino L. Key Elements in Service Innovation: Insights for the Hospitality Industry. In Proceedings of the Cornell Hospitality Roundtable, Ithaca, NY, USA, 1 November 2008; Cornell University, Ithaca, NY, USA; pp. 6-12.
24. Poetz, M.K.; Schreier, M. The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? J. Prod. Innov. Manag. 2012, 29, 245–256.
25. Bayus, B.L. Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community. *Manag. Sci.* 2013, 49, 226–244.

26. Kim, A.J.; Ko, E. Do Social Media Marketing Activities Enhance Customer Equity? An Empirical Study of Luxury Fashion Brand. *J. Bus. Res.* 2012, 65, 1480–1486.

27. Saha, A.; Sindhwani, V. Learning Evolving and Emerging Topics in Social Media: A Dynamic NMF Approach with Temporal Regularization. In Proceedings of the 5th ACM International Conference on Web Search and Data Mining, Seattle, WA, USA, 8–12 February 2012; Association for Computing Machinery: New York, NY, USA; pp. 693–702.

28. McAuley, J.; Leskovic, J. Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text. In Proceedings of the 7th ACM Conference on Recommender Systems, Hong Kong, China, 12–16 October 2013; Association for Computing Machinery: New York, NY, USA; pp. 165–172.

29. Tuarob, S.; Tucker, C.S. Fad or Here to Stay: Predicting Product Market Adoption and Longevity Using Large Scale, Social Media Data. In Proceedings of the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Portland, OR, USA, 4–7 August 2013; The American Society of Mechanical Engineers: New York, NY, USA; p. V02BT02A012.

30. Ko, N.; Jeong, B.; Choi, S.; Yoon, J. Identifying Product Opportunities Using Social Media Mining: Application of Topic Modeling and Chance Discovery Theory. *IEEE Access.* 2017, 6, 1680–1693.

31. Kühl, N.; Goutier, M.; Ensslen, A.; Jochem, P. Literature vs. Twitter: Empirical Insights on Customer Needs in E-mobility. *J. Clean. Prod.* 2019, 213, 508–520.

32. Jeong, B.; Yoon, J.; Lee, J-M. Social Media Mining for Product Planning: A Product Opportunity Mining Approach based on Topic Modeling and Sentiment Analysis. *Int. J. Inf. Manag.* 2019, 48, 280–290.

33. Reisenbichler, M.; Reutterer, T. Topic Modeling in Marketing: Recent Advances and Research Opportunities. *J. Bus. Econ.* 2019, 89, 327–356.

34. Kim, W.; Ko, T.; Rhiu, I.; Yun, M.H. Mining Affective Experience for a Kansei Design Study on a Recliner, *Appl. Ergon.* 2019, 74, 145–153.

35. Jung, J.; Jung, K.; Ko, S.; Kim, K. Customer Complaints Analysis Using Text Mining and Outcome-Driven Innovation Method for Market-Oriented Product Development. *Sustainability* 2019, 11, 40.

36. He, W.; Zhang, W.; Tian, X.; Tao, R.; Akula, V. Identifying Customer Knowledge on Social Media through Data Analytics. *J. Enterp. Inf. Manag.* 2019, 32, 152–169.

37. Lomborg, S.; Bechmann, A. Using APIs for Data Collection on Social Media. *Inf. Soc.* 2014, 30, 256–265.

38. Sun, S.; Luo, C.; Chen, J. A Review of Natural Language Processing Techniques for Opinion Mining Systems. *Inf. Fusion.* 2017, 36, 10–25.

39. Toman, M.; Jezek, K. Influence of Word Normalization on Text Classification. In Proceedings of the 1st International Conference on Multidisciplinary Information Sciences and Technologies, Merida, Spain, 25–28 October 2006; pp. 354–358.

40. Manning, C.; Schutze, H. *Foundations of Statistical Natural Language Process*, 1st ed.; The MIT Press: Cambridge, MA, USA, 1999; pp. 131–147.

41. Mostafa, M.M.; More Than Words: Social Network’s Text Mining for Consumer Brand Sentiments. *Expert Syst. Appl.* 2013, 40, 4241–4251.

42. Forlizzi, J. How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the Home. In Proceedings of the 2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Arlington, VA, USA, 9–11 March 2007; IEEE: New York, NY, USA; pp. 129–136.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.