INTERNATIONAL JOURNAL OF eBUSINESS AND eGOVERNMENT STUDIES
Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108
Received: 27.12.2022 | **Accepted:** 28.04.2023 | **Published Online:** 25.04.2023

-RESEARCH ARTICLE-

QUANTIFYING CONSUMER PREFERENCES FOR GOODS AND SERVICES ACROSS CATEGORIES IN THE DIGITAL INFORMATION ERA

Oksana Yuldasheva

Department of Marketing, Saint Petersburg State University of Economics, Saint Petersburg, Russia
Email: uldasheva2006@yandex.ru
Orcid Id: https://orcid.org/0000-0002-5184-0929

Dmitriy Rodionov

Graduate School of Industrial Economics, Institute of Industrial Management, Economics, and Trade, Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia
Email: drodionov@spbstu.ru
Orcid Id: https://orcid.org/0000-0002-1254-0464

Olga Konnikova

Department of Marketing, St. Petersburg State University of Economics, Saint Petersburg, Russia
Email: olga.a.konnikova@gmail.com
Orcid Id: https://orcid.org/0000-0002-0978-0061

Evgenii Konnikov

Graduate School of Industrial Economics, Institute of Industrial Management, Economics, and Trade, Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia
Email: konnikov.evgeniy@gmail.com
Orcid Id: https://orcid.org/0000-0002-4685-8569

Citation (APA): Yuldasheva, O., Rodionov, D., Konnikova, O., Konnikov, E., Kryzhko, D., Pirogov, D. (2023). Quantifying Consumer Preferences For Goods And Services Across Categories In The Digital Information Era. *International Journal of eBusiness and eGovernment Studies, 15*(1), 157-173. doi: 10.34109/ijebeg.2023150108

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

Darya Kryzhko

Graduate School of Industrial Economics, Institute of
Industrial Management, Economics, and Trade,
Peter the Great St. Petersburg Polytechnic University,
Saint Petersburg, Russia
Email: kryzhko_da@spbstu.ru
Orcid Id: https://orcid.org/0000-0002-7006-6828

Dmitriy Pirogov

Department of Marketing, Saint Petersburg State University
of Economics, Saint Petersburg, Russia
Email: pirogovdm@gmail.com

─Abstract─

The current approach to demand analysis involves examining various product and service categories based on consumer requirements. To evaluate the significance level of these demand categories for the major consumer groups, an instrument that can objectively assess consumer preferences and requirements for these categories is required. This research seeks to create a tool that can be frequently updated using the Term Frequency-Inverse Document Frequency (TF-IDF) index. The aggregated consumer data was gathered from social networks, an excellent source for differentiating users based on their interests. Consumer properties were divided into three categories: those associated with the user's biology, society, and digital environment. The developed instrument quantifies the digital information environment and aggregates actual and potential consumers to determine the significance level of demand categories. By analyzing consumer characteristics across these three categories, the proposed tool can objectively and frequently assess the level of demand for various categories of products and services.

Keywords: Consumer preferences; demand analysis; Natural digital information; automated toolkit

1. INTRODUCTION

Several branches of the social and behavioral sciences investigate consumption patterns. Researchers in economics, marketing, psychology, and merchandising use findings from consumer behavior studies to better comprehend the factors influencing purchasing decisions. In consumer decision science, controlled field and laboratory studies on consumer preferences to reveal actualized preferences (Roma, Reed, DiGennaro Reed, & Hursh, 2017) and stated preference surveys (Ho, Mulley, &

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

Hensher, 2020) are typical methods for quantifying resource/product utility. Each strategy for consumer research has a lengthy history, but they all diverge significantly from the within-subject repeated-measures methodology utilized in behavior analysis (Montoya, 2019). In the current digital era, the proliferation of social networking sites has influenced the attention and habits of consumers. All forms of social media are now thoroughly integrated into individuals' daily lives. Even international consumers could not resist social media's allure. People throughout the globe are typically connected through technological means. Customers are attracted to these practices due to their similarities with the target market, such as language, ethnicity, sex, religion, and/or nationality.

Assessing the degree of importance of certain groups of products to potential consumers is a traditional and relatively well-developed component of the existing marketing framework (Dabbous & Barakat, 2020). Nonetheless, several fundamental limitations of this strategy should be noted. Traditional marketing tools for assessing the degree of importance of certain groups of products or services to the potential consumer may be biased (Chunaev, Nuzhdenko, & Bochenina, 2019; Ullah, Khan, Rahman, & Ullah, 2021) and are labor- and capital-intensive due to the requirement for consumer communication, which necessitates substantial human resources. To improve the representativeness of the evaluation results, it is necessary to compile a large amount of analytical data, which will undoubtedly increase the study's duration and expense (Ahmad, Turi, Al-Dala'ie, & Manan, 2022; Arias, Hashemi, Andersen, Træholt, & Romero, 2019).

In addition to the increased labor intensity, the inability to replicate the research process is determined by the requirement to create a novel study design to apply conventional marketing techniques to such activities. To determine a prospective customer's interest in different product or service categories, traditional marketing methods require respondents to comprehend the various needs that must be satisfied. This is essential, as the relative importance of various product and service categories constantly shifts in response to consumer demand, necessitating the periodic revision of assessment results (Sabbar, Kara, Said, Ahmed, & Asad, 2023). Before immediately implementing the capability to meet the corresponding requirements, it is necessary to develop the point of attraction of the demanding carrier in the case of evaluation to determine the need to discover a potential customer and develop practical methods for gathering information. Thus, the constraints mentioned above are significantly exacerbated. The restriction can be circumvented by assembling a functional set of communication resources. The assessment procedure requires more time and resources than usual due to the necessity of developing the required toolbox (Winstone, Mathlin, & Nash, 2019).

In addition, the significant diversity of consumers makes it impossible to employ a universal evaluation framework. As it requires communication between the respondent and the researcher, this limitation is systematically impossible with traditional marketing tools. Nonetheless, the emergence of social media has created new marketing opportunities (Kryvytska , Kovalenko, & Kovalchuk, 2021). In conclusion, it is important to observe the incorporation of the assessment subject into the respondent's needs reflection. The purpose of the present study is to develop a comparative toolkit by determining the need for a comparative base on the differentiation of consumers and groups of requirements.

2. LITERATURE REVIEW

2.1 Consumer Buying Behavior

According to Palalic, Ramadani, Mariam Gilani, Gërguri-Rashiti, and Dana (2021), "consumer buying behavior" refers to people's mental, emotional, and physical behaviors when selecting, purchasing, utilizing, and discarding products and services to satiate their needs and desires. It comprises purchasing and other consumption-related activities by parties involved in an exchange transaction. Economic factors such as the pattern of income and expenditure, the price of items, the price of complementary products, substitute goods, and elasticity of demand all impact consumer purchasing behavior (Qazzafi, 2019). According to Gu, Ślusarczyk, Hajizada, Kovalyova, and Sakhbieva (2021), psychological factors such as perception, attitudes, and learning influence it. Consumer behavior is influenced by several social and cultural factors, which affect an individual's decision to purchase a product and determine the type of product to buy (McClure & Seock, 2020).

2.2 Consumer Preferences

There is also a growing body of research on consumers' preferences for various product types (e.g., Dandage, Badia-Melis, and Ruiz-García (2017); Jin, Zhang, and Xu (2017); Liu, Li, Steele, and Fang (2018)). According to a number of studies, the amount of product information provided to customers significantly influences their preferences, and customers are willing to pay more for products that provide this level of information. On the other hand, the overwhelming majority of these studies merely consider product information as one of the features; in other words, they tend to ignore the heterogeneity of product information, which is of great interest to customers. In addition, to the best of our knowledge, there has not been a large-scale study of consumers' preferences for various categories of product information. Consumers use the country of origin to infer product quality based on past purchasing experiences (Claret et al., 2012; Eng, Ozdemir, & Michelson, 2016).

2.3 Consumer Preferences Evaluation

Thus, the task at hand is to develop tools that permit an objective assessment of the level of demand for specific categories of products and services without direct communication and affecting both actual and potential consumers, with a high update frequency. The solution to this issue is quantifying the digital information environment

(Campos-Castillo & Laestadius, 2020) by aggregating actual and potential consumers (Sarstedt et al., 2022; Ziyadin, Doszhan, Borodin, Omarova, & Ilyas, 2019). Social media is the primary aggregator of consumers in the digital environment, differentiating them within the categories of interest. When communicating and forming relationships within a social network, a person develops a set of personalized properties accessible to the constituents of the external environment, which is quantifiable (Mittal, Bhandari, & Chand, 2022). These characteristics can be subdivided conditionally into various categories.

First, the characteristics of the user as a subject of the biosphere are properties that characterize a person from a physical perspective, such as gender, age, build, and eye color. Depending on the social network, the sources of this information can include direct questionnaires/user profiles, photographic data (which can be analyzed using machine learning tools), or text analysis (Boursier, Gioia, & Griffiths, 2020; Rodionov, Gracheva, Konnikov, Konnikova, & Kryzhko, 2022). This information is unquestionably not essential for determining the importance of certain product or service categories to a prospective consumer. However, it allows for the categorization of consumers and, thus, the identification of the characteristics of the intended consumer.

The properties that characterize the user in the context of interaction (communication) are the characteristics of the user as an object of communication. Other described properties related to the user's communication objects and metaproperties of the communication environment (number of communication objects, frequency of their development, interaction type, etc.) may also be included (Ali, Ullah, Ahmad, Cheok, & Alenezi, 2023; Rodionov, Zaytsev, Konnikov, Dmitriev, & Dubolazova, 2021). The analysis of these properties is predicated primarily on the fundamental principle of the existence of social media: the possibility of establishing a digital communication circle. Practically every social network offers the opportunity to analyze these properties, but it is important to consider that social networks are often specialized. The objective of the communication process is determined by specialization. Chunaev et al. (2019) demonstrated that detecting communities not explicitly defined by social media is also possible.

Lastly, the characteristics of the user as a subject of consumption and creation are properties that characterize the user's primary interests, making them crucial to the scope of this study. The specificity of the presentation of these properties is also multifaceted, dependent on the architecture of the social media, and typically presented in three ways: static presentation (information blocks in the user's questionnaire/profile), dynamic presentation (possibility of forming a multi-format information flow describing the user's interests), and integration presentation (possibility of developing content-thematic hubs). Almost every social network provides access to all three formats. The static and dynamic ones, however, assume that

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

the user has a realizable desire for a directed presentation of his interests in the external digital space, which is not given (Ahmad, Beddu, Itam, & Alanimi, 2019; Ketelsen, Janssen, & Hamm, 2020).

The viability of using social media as a source of quantifiable information about a prospective consumer's interests is determined by the possibility of forming content-thematic hubs. In addition, socio-thematic centers allow for the topographical delimitation of user groups and more accurate identification of potential consumers. This aspect has been the main research focus over the past decade, as evidenced by Capatina et al. (2020) and Zarrinkalam, Faralli, Piao, and Bagheri (2020) articles. Consequently, it is advisable to use information aggregated and systematized within the framework of social networks to determine the level of importance of certain groups of products or services to a potential consumer.

Moreover, this information can be separated into two primary streams. First, the flow that describes the potential consumers' interests. This flow is determined by the content-thematic nodes involved with the potential consumer. Second, the flow that describes the studied groups of goods or services is defined by content-thematic nodes devoted to the specific interests and requirements that each group of goods or services satisfies. Researchers have already developed methods for evaluating and forecasting demand (Xiong, Yuan, Zeng, & Zhou, 2022). However, this research aims to create a scalable method for frequent and largely automated assessments of the current demand for products and services requiring minimal skill and resources.

3. METHODOLOGY

The to-be-studied information channels are necessarily represented in their natural form and may include textual, photographic, and other types of information. Therefore, instruments are required to process and analyze natural information for quantitative purposes. The type of natural information primarily determines the distinctiveness of these instruments. For the development of practical analytical tools within the scope of this study, it is reasonable to limit natural information to its textual form. This limitation is predominantly due to this type of information's universal and pervasive nature. In addition, it should be noted that the tools required for the quantification and subsequent processing of photographic data are significantly more energy-intensive and require machine learning tools.

In contrast, natural textual information can be processed with analytical geometry's relatively less energy-intensive instruments. This is because textual information incorporates extremely natural information, such as an image, in part. Within the quantification framework, the above-described textual flows of natural information can be represented as vectors, where each coordinate represents a specific token (keyword), and its value indicates the token's importance within the flow. This significance can be

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

determined by occurrence frequency, but it increases the significance of low-content characters such as conjunctions, prepositions, and pronouns.

The Term Frequency-Inverse Document Frequency (TF-IDF) index can compensate for this effect by altering the frequency of occurrence according to the degree of uniqueness of the tokens. According to this index, the significance of a token is proportional to its frequency in a particular information stream and inversely proportional to its frequency in all analyzed information streams in Equation 1.

$$TF-IDF = \frac{n_t}{\sum_k n_k} * \log \frac{|D|}{|\{d_i \in D | t \in d_i\}|} \quad (1)$$

where:

- n_t is the frequency of occurrence of token *t* in a given stream;
- $\sum_k n_k$ is the total number of tokens in a given stream;
- $|D|$ is the number of information streams;
- $|\{d_i \in D | t \in d_i\}|$ is the number of streams in which token *t* occurs.

Using a vector where each coordinate is a token evaluated by the TF-IDF index, any textual information flow describing a single thematic component can be quantified universally. Evaluation of the degree of similarity between the quantitative flow describing the interests of potential consumers and the quantitative flow describing the studied groups of products or services will constitute the analytical core of the developed toolkit. Cosine similarity (a measure of similarity between two vectors of pre-Hilbert space that measures the cosine of the angle between them) can be used to evaluate this similarity:

$$\text{similarity}(A_i B_i) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \quad (2)$$

where:

- A is a vector quantifying the flow of information describing the interests of potential consumers;
- B is a vector quantifying the flow of information describing the groups of goods or services under consideration.

This measure is chosen because it considers only non-zero dimensions, which is particularly effective for unloaded vectors, which are always textual information quantifiers. This value ranges from 0 to 1, as the TF-IDF index cannot be negative.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

It is necessary to consider the peculiarities of this indicator's dispersion. Because the text description implies a substantial amount of technical and low-content vocabulary, many elements will be extremely similar when forming a multidimensional vector (from 1000 to 5000 coordinates), resulting in a low variance of possible cosine similarity values. Given that, this indicator is relative and subject to comparison within need categories and consumer groups, increasing the range of potential values is prudent. Logit transformation can be utilized to implement this mathematical procedure. This mathematical procedure entails the transformation from a linear function, which describes the derived cosine similarity parameters, to a sigmoid function:

$$logit(-\text{similarity}(A_i B_i)) = \frac{1}{1 + e^{-12 * \frac{\text{similarity}(A_i B_i) - \text{similarity}_{min}}{\text{similarity}_{max} - \text{similarity}_{min}} + 6}} \tag{3}$$

where:

- $\text{similarity}(A_i B_i)$ is the cosine similarity of vectors *A* and *B*;
- similarity_{min} is the minimum value of the cosine similarity measure in the entire array of information flow vector comparisons;
- similarity_{max} maximum is the maximum value of the cosine similarity measure in the entire array of information flow vector comparisons.

The logit transformation enables us to visually compare the vector vectors of information flows and determine whether certain categories of consumer requirements are more or less significant. It is also feasible to differentiate the comparison threshold value. Utilizing a set of statistically universal measures, such as the arithmetic mean and root mean square deviation (RMSD), can enhance the analytical significance of the conclusions. Consequently, it is possible to establish three fundamental levels of the need's significance and binarize based on the rule of the first sigma:

1. Extreme demand - determined by exceeding the logit-transformed value of the cosine similarity of the sum of the mean and RMSD:

$$logit(\text{similarity}) > \frac{\sum_{i=1}^{n} logit(\text{similarity})}{n} + \sqrt{\frac{\sum_{i=1}^{n} \left(\text{similarity}_i - \overline{\text{similarity}}\right)^2}{n}} \tag{4.1}$$

2. Standard demand - lies in the range between the mean value and the sum of the mean value and RMSD:

$$\frac{\sum_{i=1}^{n} logit(\text{similarity})}{n} + \sqrt{\frac{\sum_{i=1}^{n} \left(\text{similarity}_i - \overline{\text{similarity}}\right)^2}{n}} > logit(\text{similarity}) > \frac{\sum_{i=1}^{n} logit(\text{similarity})}{n} \tag{4.2}$$

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

3. Baseline demand - lies in the range between the mean value and the difference between the mean value and the RMS:

$$\frac{\sum_{i=1}^{n} logit(\text{similarity})}{n} > logit(\text{similarity}) > \frac{\sum_{i=1}^{n} logit(\text{similarity})}{n} - \sqrt{\frac{\sum_{i=1}^{n}\left(\text{similarity}_i - \overline{\text{similarity}}\right)^2}{n}} \qquad (4.3)$$

Thus, according to the results of quantification, indicators are formed, the sum of which allows to assess the importance of the need on a four-level scale, where:

- 0 is the complete absence of demand for the product or service in the analyzed consumer group;
- 1 is minimal demand for the corresponding product or service in the analyzed consumer group;
- 2 is the average demand for the corresponding product or service in the analyzed consumer group;
- 3 is extraordinary demand for the related product or service in the analyzed consumer group.

As mentioned above, the proposed toolkit is comparative, which determines the need to form a comparative base on the differentiation of consumers and groups of needs. The multi-level nature of the developed toolkit requires its algorithmic character and automation.

4. RESULTS AND DISCUSSION

The proposed toolkit for comparative assessment of the significance of need categories for consumer groups is based on natural digital information describing consumers and need categories. Thus, the first step in applying the developed toolkit is to seek and systematize natural digital information sources. It was decided to analyze the information environment of the Russian Federation to conduct this study. This is primarily attributable to the availability of automated information aggregation in social media and the widespread use of the Internet. The VK social network is the most effective information aggregator within the Russian Federation's information ecosystem. In this instance, the relevant communities (groups) are the content and thematic concentrations of consumer information and information on categories of requirements. These organizations can be separated into distinct categories.

First is consumer information concentration. Note that these nodes do not contain information describing explicitly targeted consumers. However, they include unique identifiers of consumers (domains and IDs) that enable references to the consumers' pages, which have the necessary descriptive natural information. Additionally, it should

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

be noted that these centers can be divided into two fundamental subgroups. i) Information aggregators regarding current consumers. For this function, it is recommended to utilize communities specifically devoted to the subject of analysis. ii) Information aggregators for prospective consumers. According to the territorial characteristics of the subject of analysis, these centers can be defined. For instance, if we are discussing an offline retail facility, it is recommended to use it as the hub of information about potential consumers' communities of nearby residential complexes, nearby competitors and social facilities (schools, kindergartens, etc.), nearby business centers and/or other enterprises (factories, plants, etc.), and, if it is necessary to achieve the maximum analytical coverage - communities of districts of the settlement or communities of other mu

Second, information aggregators for requirements. First, these concentrators should be differentiated based on the requirements category. The specificity of the definition of effective information concentrators is considerably less formal than that of consumer information concentrators and is primarily determined by the requirements category. For example, for goods about the beauty industry, the concentrators can be communities devoted to the corresponding product brands, communities of interest, concentrating both specialists in the field and people interested in it, and communities of beauty salons, hairdressing salons, and other categories of businesses offering related services. For the categories of entertainment industry services, however, the hubs may be communities devoted to the respective forms of entertainment and centers for providing these services. Consequently, selecting suitable nodes is not technical and necessitates a multi-level approach, requiring multiple approvals.

Aggregating concentrator categories does not necessitate automation, but the ultimate form of aggregation can be universalized. The corresponding natural digital information aggregation stage follows the complete form for fundamental information aggregation. The consumer information aggregation and needs information aggregation processes can be implemented concurrently for optimization. Consider first the process of consumer information aggregation. This procedure can be divided into two sub-procedures.

First, the VK API and the aggregation of current and potential user IDs enable the transmission of queries to a community based solely on its ID. To accomplish the objectives of the proposed toolkit, however, only users who have publicly disclosed their birth year (for age determination) and gender on their page are of analytical interest. Therefore, the initially extracted array must be filtered, and only the identifiers of individuals granted access to this information must be stored. The array of generated user IDs is then sorted according to the specified categories. "Teens (under 18)", "Young (18-25) women", "Young (18-25) men", "Adults (26-45) women", "Adults (26-45) men", "Mature (46+) women", and "Mature (46+) men" were identified as user categories for this study.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

Second, the aggregation of natural digital information characterizes the set of current and potential consumer interests. Categorically distributed arrays of identifiers, which result from the previous stage's implementation, permit targeted access to the information aggregated on user pages. As previously mentioned, it is advisable to use the description of content-thematic centers whose members are users as information describing their interests. For this purpose, the VK API contains a method that enables the extraction of all available community information. This stage consists solely of technical tasks, allowing it to be automated. Based on the results of this stage, consolidated data frames describing the set of current and prospective consumer interests for each of the audiences identified in the first stage are created.

Third, the aggregation of the descriptions of need categories. This stage is implemented concurrently with the preceding stages, and the resulting form is similar. Similar to the array described in the preceding section, this text array is aggregated. The tools used in this procedure are identical to those used in stage 2. Additionally, the VK API method contains a method for extracting the information flow of the content theme concentrator at once. As a consequence of the first three stages, a single consolidated data frame is created, the structure of which is differentiated by consumer categories and types, as well as the description of consumer need categories;

Standardization of aggregated text arrays is the fourth objective. According to the results of constructing an aggregated data frame, the analyst has a substantial quantity of quantifiable natural digital data. Fifth is the vectorization of standardized text arrays. The principles of analytic geometry define the quantification procedure described previously and imply the vector representation of standard text arrays. The vector coordinates correspond to the TF-IDF index values for the selected array of tokens. In this instance, comparing interests necessitates comparing relatively heterogeneous text arrays and forming a large collection of standard tickets serving as vector coordinates.

Sixth is the constructed vectors' cosine kernel (cosine similarity) is determined. The set of vectors obtained in the preceding step, describing both categories of requirements and the set of interests of target consumer groups, is subjected to a quantitative comparison for comparative similarity, for which the metric of kernel cosines or cosine similarity can be used. Notably, the TF-IDF index determines the coordinates of the derived vectors, which depend on the total size of the "bag of words" and the total number of texts analyzed. The resulting vectors must be normalized because these parameters differ considerably between the arrays describing the categories of needs and those describing the categories of consumer interests. A standard algorithm can be used for this purpose, which entails subtracting the array's minimum value from each coordinate and dividing the result by the array's range (the difference between its maximum and minimum values);

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

Seventh, logit transformation of cosine similarity matrices of consumer need categories and interest vectors. The cosine similarity matrices obtained in the preceding step permit us to compare the level of interest of consumer groups in the target requirements categories. However, the proposed method of quantifying natural numerical information tends to center the values of the comparison parameter, preventing obvious differentiation between the objects being compared. The logit transformation, which involves transitioning from a linear to a sigmoid function, can mitigate this issue.

Last, binarization of logit-transformed cosine similarity matrices of vectors of requirements categories and vectors of consumer groups' interests. This is the final stage of the devised algorithm and involves dividing the obtained logit-transformed cosine similarity matrices by the significance threshold. In this instance, the significance threshold is exclusively determined statistically. It entails the designation of three levels: extreme (above the upper limit of the first sigma), average, and fundamental (above the lower limit of the first sigma). Due to the logit transformation performed at the previous stage, which increased the variance of the array of resulting values, this statistical method is applicable.

According to the implementation results of the proposed algorithm, the researcher has at least six matrices categorized by consumer types (actual and prospective) and demand levels (extraordinary, average, and minimal). The matrices obtained permit the formation of a multidimensional set of comparative inferences.

One of the neighborhood retail centers in the Vyborg district of St. Petersburg, Russia, served as the test subject. Communities of the shopping center in the VK social network were used to aggregate the actual audience. In contrast, communities of residential areas in the Vyborg district were used to aggregate the prospective audience. As the categories of demand, the following products and services offered in the analyzed shopping center were selected: i) supermarket, ii) clothing and shoes, and iii) leather and accessories. iv) domestic goods, v) home appliances, vi) cosmetics, health, and beautification aids, vii) infant goods, viii) sports and recreation goods, ix) jewelry, x) HoReCa (catering services), xi) Impulse Services, xii) adult entertainment, xiii) children's amusement, xiv) sports and fitness, xv) beauty services, and xvi) others.

According to the analysis, we obtained a system of matrices, each of which has analytical value, but the data matrix of audience distribution is the most analytically robust. This matrix comprises the sum of matrices indicating both actual and potential audiences' demand significance levels (extreme, standard, and basic). The value of the resulting indicator is determined on a scale from 0 to 6 points, where 0 represents the category of needs with the least importance, and 6 represents the category of needs with the most importance. The survey results are displayed in Table 1.

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

Table 1. Data Matrix of Shopping Center Audience Distribution

Categories	Teens (Under 18)	Young women (18-25)	Young men (18-25)	Adult women (26-45)	Adult men (26-45)	Mature women (46+)	Mature men (46+)
Supermarket	3	4	4	4	2	2	3
Clothing and shoes	4	4	5	6	4	3	3
Leather and accessories	3	4	4	4	2	2	3
Housewares	6	6	6	6	6	5	6
Household appliances	6	6	6	6	6	6	6
Cosmetics, health & beauty products	1	0	2	2	0	1	1
Baby products	4	4	5	6	3	3	3
Sports and recreation products	3	2	4	4	2	2	2
Jewelry	1	0	1	2	0	1	0
Other	6	4	6	6	4	5	4
HoReCa (catering services)	0	0	0	0	0	0	0
Impulse Services	4	2	4	4	2	3	3
Entertainment for adults	6	5	6	6	5	5	5
Entertainment for children	2	0	1	2	0	1	1
Sports and Fitness	4	3	4	5	3	3	3
Beauty services	0	0	0	0	0	0	0

As seen in the matrix, all possible variations of the final indicator are presented. The least contentious regarding their significance are: i) household products (score approaches six). This may be primarily attributable to the area's relatively dense population growth. High construction rates determine the demand for household products. ii) Household appliances (6 points). The primary factors for this are the same as those for the demand for household goods. iii) Entertainment for adults (score exceeds 5). This indicates the public's social activity, which may result from their capacity to pay. iv) HoReCa (0 points). This demonstrates the constraints of the proposed toolkit. This category of requirements is not representative of society due to its inherent specificity. Therefore, it is necessary to establish interest types that indirectly indicate this need or omit them from the proposed methodology. v) Charm (score of 0). This category's conclusions are identical to those regarding the HoReCa category. vi) Cosmetics, cosmetics, and health products (average score of 1, maximum score of 2). In this instance, both flawed methodology and gender specificity are

Vol: 15 **No:** 1 **Year:** 2023 **ISSN:** 2146-0744 (Online) (**pp.** 157-173) **Doi:** 10.34109/ijebeg.2023150108

applicable. vii) Jewelry (the average score is one, with a maximum score of two). This category's conclusions are identical to those for "Cosmetics, health, and beauty products." viii) Entertainment for minors (the average score is one, with a maximum of two). This fact suggests that either consumer do not have children (which may also be related to the density of development in the area) or the methodology in this instance needs improvement.

5. CONCLUSION AND RECOMMENDATIONS

Based on the quantification of natural digital information, we proposed in this study an automated tool for assessing the degree of importance of certain groups of products or services to the potential consumer. The proposed mechanism is useful primarily because of its relative objectivity and minimal labor intensity. However, it is necessary to consider its adaptation to distinct demand categories and the potential distortions of consumer information in social media. Consequently, it is most efficient to use this tool in conjunction with conventional marketing research tools and within the context of continuously analyzing shifting consumer preferences. Utilize sophisticated data analytics techniques to extract insightful information from social network data. This can provide a greater comprehension of consumer preferences and requirements, thereby improving the precision of demand assessments. Integrate user-feedback mechanisms into the tool to continuously enhance performance and meet changing consumer needs. This feedback cycle guarantees the tool's continued relevance and responsiveness to shifting consumer behaviors.

6. LIMITATIONS AND FUTURE STUDIES

Recognize the limitations of exclusively relying on social network data, as it may not accurately represent the entire consumer population. Consider augmenting the data with additional sources for a more thorough analysis. Consider potential biases in social network data, such as user demographics and preferences, which can affect the evaluation outcome. Ensure precise compliance with privacy regulations to safeguard user information. Conduct longitudinal studies to validate the precision and efficacy of the devised instrument over an extended period. This will provide invaluable insight into its predictability and dependability. Encourage sociology, marketing, and computer science researchers to collaborate to enhance the tool's multidimensional analysis and address complex consumer dynamics. Investigate real-time consumer preferences and requirements monitoring so businesses and policymakers can make timely decisions. Utilize dynamic data streams and emerging technologies to gather consumer insights. Create an instrument that is scalable and adaptable across industries and regions. This will enable widespread applicability and market-specific customization. By implementing these recommendations, addressing these limitations, and investigating future research avenues, the developed tool can contribute to a more accurate and up-to-date consumer demand assessment, which will benefit various market stakeholders.

7. FUNDING

The Ministry of Science and Higher Education of the Russian Federation funded the research under the strategic academic leadership program "Priority 2030" (Agreement 075-15-2023-380 dated 20.02.2023).

REFERENCES

Ahmad, M., Beddu, S., Itam, Z. b., & Alanimi, F. B. I. (2019). State of the art compendium of macro and micro energies. *Advances in Science and Technology. Research Journal*, *13*(1), 88-109. doi: https://doi.org/10.12913/22998624/103425

Ahmad, M., Turi, J. A., Al-Dala'ie, R. N. S., & Manan, A. (2022). Potential use of recycled materials on rooftops to improve thermal comfort in sustainable building construction projects. *Frontiers in Built Environment*, *8*, 1014473. doi: https://doi.org/10.3389/fbuil.2022.1014473

Ali, M., Ullah, S., Ahmad, M. S., Cheok, M. Y., & Alenezi, H. (2023). Assessing the impact of green consumption behavior and green purchase intention among Millennials Toward Sustainable Environment. *Environmental Science and Pollution Research, 30*(9), 23335-23347. doi: https://doi.org/10.1007/s11356-022-23811-1

Arias, N. B., Hashemi, S., Andersen, P. B., Træholt, C., & Romero, R. (2019). Distribution system services provided by electric vehicles: Recent status, challenges, and future prospects. *IEEE Transactions on Intelligent Transportation Systems, 20*(12), 4277-4296. doi: https://doi.org/10.1109/TITS.2018.2889439

Boursier, V., Gioia, F., & Griffiths, M. D. (2020). Do selfie-expectancies and social appearance anxiety predict adolescents' problematic social media use? *Computers in Human Behavior*, *110*, 106395. doi: https://doi.org/10.1016/j.chb.2020.106395

Campos-Castillo, C., & Laestadius, L. I. (2020). Racial and ethnic digital divides in posting COVID-19 content on social media among US adults: secondary survey analysis. *Journal of medical Internet research*, *22*(7), e20472. doi: https://doi.org/10.2196/20472

Capatina, A., Kachour, M., Lichy, J., Micu, A., Micu, A.-E., & Codignola, F. (2020). Matching the future capabilities of an artificial intelligence-based software for social media marketing with potential users' expectations. *Technological Forecasting and Social Change*, *151*(C), 119794. doi: https://doi.org/10.1016/j.techfore.2019.119794

Chunaev, P., Nuzhdenko, I., & Bochenina, K. (2019). Community detection in attributed social networks: a unified weight-based model and its regimes. In *2019 International Conference on Data Mining Workshops (ICDMW)* (pp. 455-464). IEEE. doi: https://doi.org/10.1109/ICDMW.2019.00072

Claret, A., Guerrero, L., Aguirre, E., Rincón, L., Hernández, M. D., Martínez, I., . . . Rodríguez-Rodríguez, C. (2012). Consumer preferences for sea fish using conjoint analysis: Exploratory study of the importance of country of origin, obtaining method, storage conditions and purchasing price. *Food Quality and Preference*, *26*(2), 259-266. doi: https://doi.org/10.1016/j.foodqual.2012.05.006

Dabbous, A., & Barakat, K. A. (2020). Bridging the online offline gap: Assessing the impact of brands' social network content quality on brand awareness and purchase intention. *Journal of retailing and consumer services, 53*(C), 101966. doi: https://doi.org/10.1016/j.jretconser.2019.101966

Dandage, K., Badia-Melis, R., & Ruiz-García, L. (2017). Indian perspective in food traceability: A review. *Food Control, 71*, 217-227. doi: https://doi.org/10.1016/j.foodcont.2016.07.005

Eng, T.-Y., Ozdemir, S., & Michelson, G. (2016). Brand origin and country of production congruity: Evidence from the UK and China. *Journal of Business Research, 69*(12), 5703-5711. doi: https://doi.org/10.1016/j.jbusres.2016.01.045

Gu, S., Ślusarczyk, B., Hajizada, S., Kovalyova, I., & Sakhbieva, A. (2021). Impact of the covid-19 pandemic on online consumer purchasing behavior. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(6), 2263-2281. doi: https://doi.org/10.3390/jtaer16060125

Ho, C. Q., Mulley, C., & Hensher, D. A. (2020). Public preferences for mobility as a service: Insights from stated preference surveys. *Transportation Research Part A: Policy and Practice, 131*, 70-90. doi: https://doi.org/10.1016/j.tra.2019.09.031

Jin, S., Zhang, Y., & Xu, Y. (2017). Amount of information and the willingness of consumers to pay for food traceability in China. *Food Control, 77*, 163-170. doi: https://doi.org/10.1016/j.foodcont.2017.02.012

Ketelsen, M., Janssen, M., & Hamm, U. (2020). Consumers' response to environmentally-friendly food packaging-A systematic review. *Journal of Cleaner Production, 254*, 120123. doi: https://doi.org/10.1016/j.jclepro.2020.120123

Kryvytska , O. R., Kovalenko, L. V., & Kovalchuk, V. M. (2021). Methodical Approach to Assessing the Demand for Higher Education in Ukraine. *The Problems of Economy, 4*(50), 42–49. doi: https://doi.org/10.32983/2222-0712-2021-4-42-49

Liu, C., Li, J., Steele, W., & Fang, X. (2018). A study on Chinese consumer preferences for food traceability information using best-worst scaling. *PloS one, 13*(11), e0206793. doi: https://doi.org/10.1371/journal.pone.0206793

McClure, C., & Seock, Y.-K. (2020). The role of involvement: Investigating the effect of brand's social media pages on consumer purchase intention. *Journal of retailing and consumer services, 53*, 101975. doi: https://doi.org/10.1016/j.jretconser.2019.101975

Mittal, A., Bhandari, H., & Chand, P. K. (2022). Anticipated positive evaluation of social media posts: social return, revisit intention, recommend intention and mediating role of memorable tourism experience. *International Journal of Culture, Tourism and Hospitality Research, 16*(1), 193-206. doi: https://doi.org/10.1108/IJCTHR-12-2020-0287

Montoya, A. K. (2019). Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models. *Behavior research methods, 51*, 61-82. doi: https://doi.org/10.3758/s13428-018-1088-6

Palalic, R., Ramadani, V., Mariam Gilani, S., Gërguri-Rashiti, S., & Dana, L. P. (2021). Social media and consumer buying behavior decision: what entrepreneurs should know? *Management Decision, 59*(6), 1249-1270. doi: https://doi.org/10.1108/MD-10-2019-1461

Qazzafi, S. (2019). Consumer buying decision process toward products. *International Journal of Scientific Research and Engineering Development, 2*(5), 130-134. Retrieved from https://www.researchgate.net/publication/336047692

Rodionov, D., Gracheva, A., Konnikov, E., Konnikova, O., & Kryzhko, D. (2022). Analyzing The Systemic Impact of Information Technology Development Dynamics on Labor Market Transformation. *International Journal of Technology (IJTech), 13*(7), 1548-1557. doi: https://doi.org/10.14716/ijtech.v13i7.6204

Rodionov, D., Zaytsev, A., Konnikov, E., Dmitriev, N., & Dubolazova, Y. (2021). Modeling changes in the enterprise information capital in the digital economy. *Journal of Open Innovation: Technology, Market, and Complexity, 7*(3), 166. doi: https://doi.org/10.3390/joitmc7030166

Roma, P. G., Reed, D. D., DiGennaro Reed, F. D., & Hursh, S. R. (2017). Progress of and prospects for hypothetical purchase task questionnaires in consumer behavior analysis and public policy. *The Behavior Analyst, 40*, 329-342. doi: https://doi.org/10.1007/s40614-017-0100-2

Sabbar, S. D., Kara, M. H., Said, S., Ahmed, S., & Asad, M. (2023). Awareness of Halal Branding and Marketing: Consumer Perception in Makassar, Indonesia. *Journal of Advances in Humanities Research, 2*(2), 98-124. doi: https://doi.org/10.56868/jadhur.v2i2.128

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L., & Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing, 39*(5), 1035-1064. doi: https://doi.org/10.1002/mar.21640

Ullah, S., Khan, U., Rahman, K. U., & Ullah, A. (2021). Problems and benefits of the China-Pakistan Economic Corridor (CPEC) for local people in Pakistan: a critical review. *Asian Perspective, 45*(4), 861-876. doi: https://doi.org/10.1353/apr.2021.0036

Winstone, N. E., Mathlin, G., & Nash, R. A. (2019). Building feedback literacy: Students' perceptions of the Developing Engagement with Feedback Toolkit. *Frontiers in Education, 4*, 39. doi: https://doi.org/10.3389/feduc.2019.00039

Xiong, Y., Yuan, H., Zeng, W., & Zhou, J. (2022). Research on the Establishment and Application of Demand Value Analysis Model for Automobile Products. In *Proceedings of China SAE Congress 2021: Selected Papers* (pp. 1209-1222). Springer. doi: https://doi.org/10.1007/978-981-19-3842-9_94

Zarrinkalam, F., Faralli, S., Piao, G., & Bagheri, E. (2020). Extracting, mining and predicting users' interests from social media. *Foundations and Trends® in Information Retrieval, 14*(5), 445-617. doi: http://dx.doi.org/10.1561/1500000078

Ziyadin, S., Doszhan, R., Borodin, A., Omarova, A., & Ilyas, A. (2019). The role of social media marketing in consumer behaviour. *E3S Web of Conferences, 135*, 04022. doi: https://doi.org/10.1051/e3sconf/201913504022