JEBI

Journal of Economics and Business Issues

Application of Natural Language Processing to Extract Consumer Behaviors from Product Reviews

Poj Netsiri [1,*] and Marketa Lhotáková [2]

[1] Czech Technical University, Faculty of Information Technology, Department of Software Engineering, Prague, Czech Republic; netp03@vse.cz, Orcid: 000-0002-3309-2779
[2] Prague University of Economics and Business, Faculty of International Relations, Department of International Business, Prague, Czech Republic
* Correspondence: poj.netsiri@gmail.com

Abstract: Understanding consumer behavior is essential for effective marketing strategies. Market research often employs conventional qualitative analysis to uncover the motivations driving consumer purchases. However, a small sample size limits the ability to draw population-level conclusions, while analyzing large-scale data can be time-consuming. Moreover, qualitative research is prone to human error, subjectivity, and bias, leading to potentially misleading results. To tackle these issues, Natural Language Processing (NLP) is utilized to extract consumer behavior from large-scale product review datasets. The study focuses on product reviews for luxury cars, smart TVs, bread, and soaps to examine complex, dissonance-reducing, habitual, and variety-seeking purchasing behavior, respectively. The results, categorized using the Perceived Value Scale and sorted based on cluster size and ranking indicate that consumer purchase intentions are influenced by perceived economic (#3), emotional (#1), social (#4), and quality (#2) value, respectively in complex purchasing behavior. Dissonance-reducing behavior was found to be primarily driven by perceived quality value (#1), while habitual purchasing behavior was observed to be influenced by perceived economic (#3), emotional (#1), and quality (#2) value, respectively. Lastly, variety-seeking behavior was found to be guided by perceived emotional (#2) and quality (#1) value in purchase decisions, respectively. Therefore, NLP demonstrated an efficient, cost-effective, and unbiased approach to extracting consumer behaviors from large-scale product reviews.

Keywords: marketing research, consumer behavior, consumer perceived value, perceived value scale, purchase intention, qualitative analysis, artificial intelligence, natural language processing, topic modeling, Latent Dirichlet Allocation

Citation: Netsiri, P.; & Lhotakova, M. (**2023**) Application of Natural Language Processing to Extract Consumer Behaviors from Product Reviews. *Journal Of Economic and Business Issues, 3(2)*, 09-25.

Received: 07/06/2023
Accepted: 11/07/2023
Published: 31/07/2023

1. Introduction

1.1 Background

Marketing is a field that heavily relies on understanding consumer behaviors. Consumer behavior knowledge helps marketers understand the needs and wants of their target audience, enabling them to offer appropriate products to their customers [17]. Consumer behavior knowledge also helps marketers market and position their products successfully [9]. Recent research in this area has focused on understanding consumer purchase intention toward products. Previous studies have proposed four value dimensions that can influence consumer purchase intention, including perceived economic, emotional, social, and quality value [12,21].

Qualitative research is often used by marketers to understand why consumers behave and purchase products in certain ways. This type of research typically involves non-

numerical, contextualized, and unstructured data such as open-ended questionnaires. However, conventional qualitative research is criticized for its subjectivity, which can be partially attributed to the human analysis of data [1,2]. This method can also be time and cost-consuming, particularly when dealing with large-scale data. In addition, the poor quality of qualitative research can lead to inaccurate findings. Qualitative research with a small sample size may also be insufficient for population-level summaries. Additionally, there are limitations to qualitative research, including potential human error, self-selection bias, and poor questions or designs from researchers [1,2].

Natural language processing (NLP) is a subfield of Artificial Intelligence (AI) [18,20] that focuses on endowing computers with the ability to comprehend human language. This technology utilizes computational linguistics, statistical analysis, and machine learning algorithms to enable computers to process textual data and understand its meaning, intent, and sentiment. In recent years, NLP has been applied to various types of data sources, including open-ended feedback from customer satisfaction surveys and notes in electronic medical records (EMR) [1,2,8,11]. These studies have demonstrated the potential of NLP to capture the overall themes and analyze the content of large-scale unstructured data. Furthermore, NLP can significantly reduce the time and cost involved in analytical processes.

Thus, NLP has the potential to enhance conventional qualitative marketing research, as suggested by previous studies[2,6,10,16]. The current study collected large-scale product reviews for luxury cars [16], smart TVs, bread, and soaps from commercial websites in the U.S. These products were selected because they are associated with complex, dissonance-reducing, habitual, and variety-seeking purchasing behaviors, respectively. Applying NLP, consumer behavior-related terms were extracted from these large-scale datasets. Finally, these terms are categorized based on Perceived Value (PERVAL) and evaluated.

2. Literature review

2.1 NLP-enhanced Qualitative Research

Abram conducted a qualitative research study in 2018, employing a conventional method to unravel the significance of registered nurses' professional role in substance use disorder treatment [1]. The research shed light on the multifaceted responsibilities and contributions of these nurses in addressing substance abuse issues, revealing valuable insights into their crucial role in patient care. In a subsequent study conducted in 2020, Abram, et al. further explored the integration of Natural Language Processing (NLP) using the Latent Dirichlet Allocation (LDA) algorithm into the previous research [1,2]. The incorporation of NLP demonstrated notable advantages, including time and cost savings, highlighting the potential of this technology in enhancing qualitative research processes and analysis [2].

Furthermore, in 2019, Koleck, et al. utilized NLP techniques to process and analyze symptom information extracted from Electronic Health Records (EHR) [11]. The research showcased the potential of NLP in extracting valuable insights from unstructured clinical data, offering a promising approach to improving healthcare decision-making and patient outcomes. In 2022, Hagg, et al. conducted a study that investigated the growing utilization of LDA in the field of psychological science [8]. The research explored the applications of LDA in analyzing psychological data and its potential for advancing research in the field, providing valuable insights into human behavior and cognition.

The Latent Dirichlet Allocation (LDA) algorithm, proposed by Blei, et al. in 2003, is widely used for topic modeling of qualitative data [3]. This algorithm enables the extraction of topics from large-scale datasets, facilitating comprehensive analysis and understanding of textual information. Additionally, in 2014, Sievert, et al. introduced a method called "LDAvis" aimed at enhancing the visualization, interpretability, and comprehensibility of results obtained from LDA topic modeling [19]. The research focused on improving the effectiveness of visualizations, enabling researchers to gain deeper insights from

the extracted topics and better understand the underlying patterns and structures in their data.

2.2 PERVAL Model

On the other hand, Sweeney et al. developed the Perceived Value (PERVAL) model, which aims to evaluate customers' perceptions of the value of consumer durable goods at a brand level [22]. The term "consumer perceived value" refers to consumers' overall assessment of a product's utility based on their perceptions. The PERVAL model encompasses four dimensions: (1) emotional value, (2) social value, (3) economic value, and (4) quality/performance value [22]. Figure 1 presents the adopted PERVAL model in this study, which serves as a framework to evaluate the outcomes derived from NLP analysis.

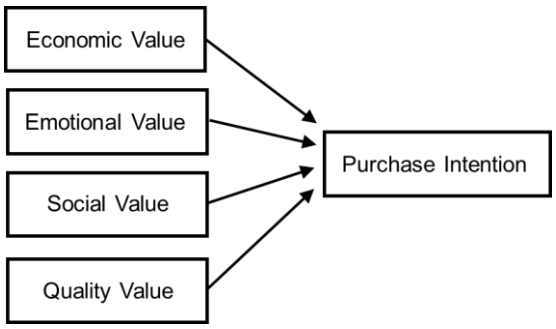

Figure 1. PERVAL Model for this Study

2.3 Research Gap

The rapid adoption of NLP and LDA topic modeling in various fields has indicated its potential application to business-related issues. Within the realm of marketing research, primary data, specifically product reviews obtained directly from customers, play a crucial role in understanding consumer purchasing behaviors. However, traditional qualitative research methods often prove to be time-consuming, particularly when dealing with large-scale datasets. Previous studies have demonstrated the cost-efficacy of Natural Language Processing (NLP) in extracting valuable information from textual data across various domains [2,8,11]. Hence, the current study aims to investigate the potential of NLP and LDA topic modeling in improving the efficiency and effectiveness of qualitative analysis in marketing research, specifically focusing on the extraction of meaningful insights from product reviews and understanding consumer behavior more comprehensively.

3. Research Methodology

3.1 Data Collection

This study involves the collection of large-scale raw data sets from product reviews. Four products are selected based on their distinct association with a specific consumer purchasing behavior [17], rather than their individual attributes or industry, as outlined below. Data sets are available at: https://github.com/pnetsiri/JEBI.

3.1.1 Complex Purchasing Behavior

For this purchasing behavior, product reviews of luxury cars were chosen as they are associated with complex purchasing behavior. Luxury car consumers tend to be highly

involved in the purchasing process and conduct extensive research before committing to a purchase. The data set was obtained from 809 product reviews of a Mercedes-Benz luxury car, collected from the website of CarMax, Inc. (www.carmax.com/reviews/mercedes-benz) in 2003. CarMax, Inc. is currently the largest retailer of pre-owned vehicles in the U.S., and its website offers large-scale product reviews of cars sorted by brand. The number of product reviews collected (809 records) is sufficient to make a population-level summary for this study, and the data set is appropriate as it represents the luxury car segment of the car market. A sample of the product and product review is shown in the previous publication [16].

3.1.2 Dissonance-reducing Purchasing Behavior

For the purpose of studying dissonance-reducing purchasing behavior, product reviews for smart TVs were chosen due to the high level of involvement consumers have in the purchase process, coupled with the difficulty in selecting a brand. The dataset used in this study consists of 1,937 product reviews of a Samsung AU8000 smart TV, collected from the Amazon.com Inc. website in 2023 (www.amazon.com/SAMSUNG-55-Inch-Crystal-AU8000-Built/dp/B08Z21BBWK/). As the current largest e-commerce company in the U.S., Amazon provides a vast array of product reviews of smart TV grouped by brand and model to the public. The large number of records (1,937 records) collected from this source is sufficient for a population-level summary of the study. Moreover, this dataset is relevant for this study because smart TV consumption could represent a segment of the TV market.

3.1.3 Habitual Purchasing Behavior

For extracting habitual purchasing behavior, product reviews for bread were selected because consumers of bread typically have minimal involvement with the brand and purchase their preferred bread. A raw data set consisting of 1,188 product reviews of Sola Sweet and Buttery Bread was collected from Amazon.com Inc.'s website in 2023 (www.amazon.com/Sola-Keto-Bread-Sweet-Buttery/dp/B082G8Z5JN/). Amazon.com Inc. is currently the largest e-commerce company in the U.S., offering large-scale product reviews of bread grouped by brand to the public. A large number of product reviews (1,188 records) obtained from this website is sufficient for making a population-level summary in this study. Moreover, this data set is relevant for the study because bread consumption could represent a segment of the bread market.

3.1.4 Variety-seeking Purchasing Behavior

The aim of this study was to extract variety-seeking purchasing behavior from product reviews of soaps, as consumers often buy new soaps not because they were dissatisfied with the previous one, but because they wanted to try something new such as a new scent. The raw data set included 904 product reviews of Dove Beauty Bar Skin Cleanser, collected from Amazon.com Inc. (www.amazon.com/Dove-Beauty-Bar-Shea-Butter/dp/B002TSA93Y/) in 2023, the current largest e-commerce company in the U.S. Their website provides large-scale product reviews of soaps grouped by brand to the public. The number of reviews (904 records) is considered sufficient to make a population-level summary for this study, and the data set is appropriate as soap consumption represents a segment of the soap market.

3.2 Pre and Post-processing

This study followed two stages of research methodology: (1) pre-processing, which involved preparing the data for analysis through standardization and formatting, and (2) post-processing, which entailed analyzing and visualizing the data [2,16]. The original data was collected in Excel format and then converted to JSON format. The study made

use of several primary packages and modules including NumPy (numpy.org) for scientific computing, Gensim library (pypi.org/project/gensim/) for topic modeling, NLTK library (www.nltk.org) for NLP, pyLDAvis library (pypi.org/project/pyLDAvis/) for interactive topic model visualization, and Jupyter Notebook (jupyter.org) for web-based interactive development. The Python program (Python.org) used in the study is available at: https://github.com/pnetsiri/JEBI.

During the first stage, standard data pre-processing methods were employed to prepare the data for NLP. This involved (1) standardizing text case to lower case, (2) removing stop words (such as "and", "or", "but", and "the"), and punctuation, and (3) stemming words (removing word endings to focus on the word root). Excluding bad data and words that do not contribute to the identification of relevant themes is also a standard process in many qualitative research approaches. A sample of the data after preprocessing is shown in the previous publication [16].

During the post-processing stage, the topic modeling [2,15,23] was performed using Latent Dirichlet Allocation (LDA) [2,3,5,13] from the Gensim library. The LDA output was used to develop topic themes and an overall theme for the data to conduct a qualitative assessment. Additionally, the pyLDAvis library was utilized to visualize the topics estimated by LDA [16,19]. This visualization tool provides both a global view of the topics and the top 10 terms associated with each topic [2,16]. The first topic (Topic 1), depicted as the largest red bubble, was selected for further analysis due to its largest size, suggesting its most significant influence from product reviews. The larger the bubble is, the more frequently the topic is found in the documents [19]. These top 10 terms from Topic 1 are categorized based on the PERVAL model. Since these terms might have multiple meanings, their categorization was manually determined by examing the entire context of the corresponding sample review. For example, if the context indicates an emotion, the term is categorized as an emotional value. Finally, each PERVAL value is sorted based on the cluster size and ranking of terms within each category. The cluster size refers to the number of terms within a category.

4. Results

4.1 Complex Purchasing Behavior

This section focuses on examining the purchasing behavior of consumers by analyzing product reviews of luxury cars. For interpretation, Topic 1 in Figure 2 was selected (38.6% of tokens). The top 10 relevant terms associated with this topic were identified and ranked as: "car" (noun), "great" (adjective), "love" (noun), "good" (adjective), "ride" (verb), "smooth" (adjective), "comfortable" (adjective), "gas" (noun), "mercedes" (noun/brand), and "mileage" (noun). These terms indicate the underlying concept of consumer behavior in this context. Table 1 provides a summary of the terms, sample reviews, and sentiments related to this complex purchasing behavior, with all sample reviews expressing positive sentiments. Lastly, Table 2 provides a summary of the terms, and ranks (#) grouped by their value category. The category and sentiment for each term were manually identified by examing the entire context of the corresponding sample review in Table 1.

Table 1. Terms, Sample Reviews, and Sentiments Related to Complex Purchasing Behavior

Rank (#)	Term	Sample Review	Sentiment
1	car	This **car** is an amazing driving machine.	positive
2	great	It is a **great** car all around.	positive
3	love	I **love** this car and all of its features.	positive
4	good	This is a great car to drive and has enough power to make it feel **good**	positive
5	ride	It's a beautiful car, handling is great, **ride** is smooth and acceleration like no other.	positive
6	smooth	I found this beautiful car to be one of the **smooth**est most conservative machines on the road.	positive
7	comfortable	This is a very **comfortable** and great looking car.	positive
8	gas	Great **gas** mileage and you will just look good driving in comfort.	positive
9	mercedes	The **Mercedes** C300 is an affordable and classy automobile.	positive
10	mileage	Solid performance and good gas **mileage**	positive

Table 2. Categories, Terms, and Ranks Related to Complex Purchasing Behavior

Category and Rank (#)	Term and Rank (#)
Economic value (#3)	mercedes (#9), mileage (#10)
Emotional value (#1)	**car (#1), love (#3), good (#4), ride (#5), smooth (#6), comfortable (#7)**
Social value (#4)	mercedes (#9)
Quality value (#2)	great (#2), excellent (#6), gas (#8), mileage (#10)

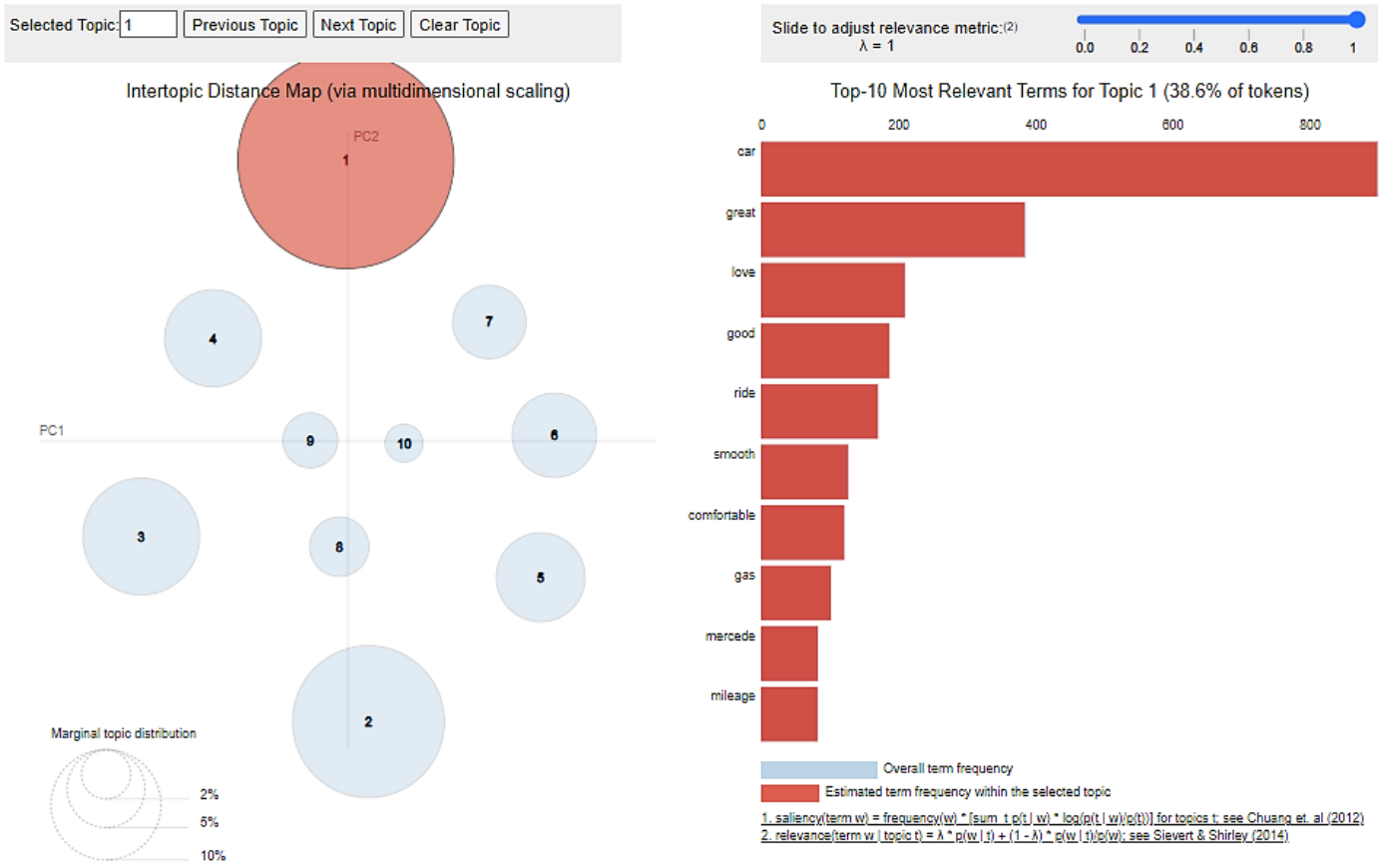

Figure 2. Topic 1 and Top 10 Terms Related to Complex Purchasing Behavior

4.2 Dissonance-reducing Purchasing Behavior

This section focuses on examining the purchasing behavior of consumers by analyzing product reviews of Smart TVs. For the purpose of interpretation, Topic 1 in Figure 3 was selected (37% of tokens). The top 10 relevant terms associated with this topic were identified and ranked as: "picture" (noun), "remote" (noun), "great" (adjective), "sound" (noun), "samsung" (noun), "quality" (adjective), "good" (adjective), "feature" (noun), "smart" (adjective), and "app" (noun). These terms indicate the underlying concept of consumer behavior in this context. Table 3 provides a summary of the terms, sample reviews, and sentiments related to this dissonance-reducing purchasing behavior. One sample review, marked by ** in Table 3 exhibited dissonance-reducing purchasing behavior of comparing brands. Two sample reviews show negative sentiment, and one shows neutral sentiment associated with poor product quality, while the rest express positive sentiment. Lastly, Table 4 provides a summary of the terms, and ranks (#) grouped by their value category. The category and sentiment for each term were manually identified by examing the entire context of the corresponding sample review in Table 3.

Table 3. Terms, Sample Reviews, and Sentiments Related to Dissonance-reducing Purchasing Behavior

Rank (#)	Term	Sample Review	Sentiment
1	picture	Great **picture** quality, great sound without sound bar.	positive
2	remote	But the **remote** is not user friendly at all.	**negative**
3	great	**Great** colors and clarity for the picture.	
4	sound	I'm assuming the **sound** quality is just average due to fact that this tv is so thin.	**neutral**
5	samsung	I like the TV. I went from a Sony Bravia to this **Samsung**. I made a great choice.**	positive
6	quality	Good **quality** TV for the price, picture looks nice and sharp	positive
7	good	**Good** TV for the price.	positive
8	feature	First smart TV I have owned and love the **features**.	positive
9	smart	Crystal clear picture, perfect "**smart**" features.	positive
10	app	TV was slow and **app** integration was really poor.	**negative**

****Exhibition of dissonance-reducing purchasing behavior**

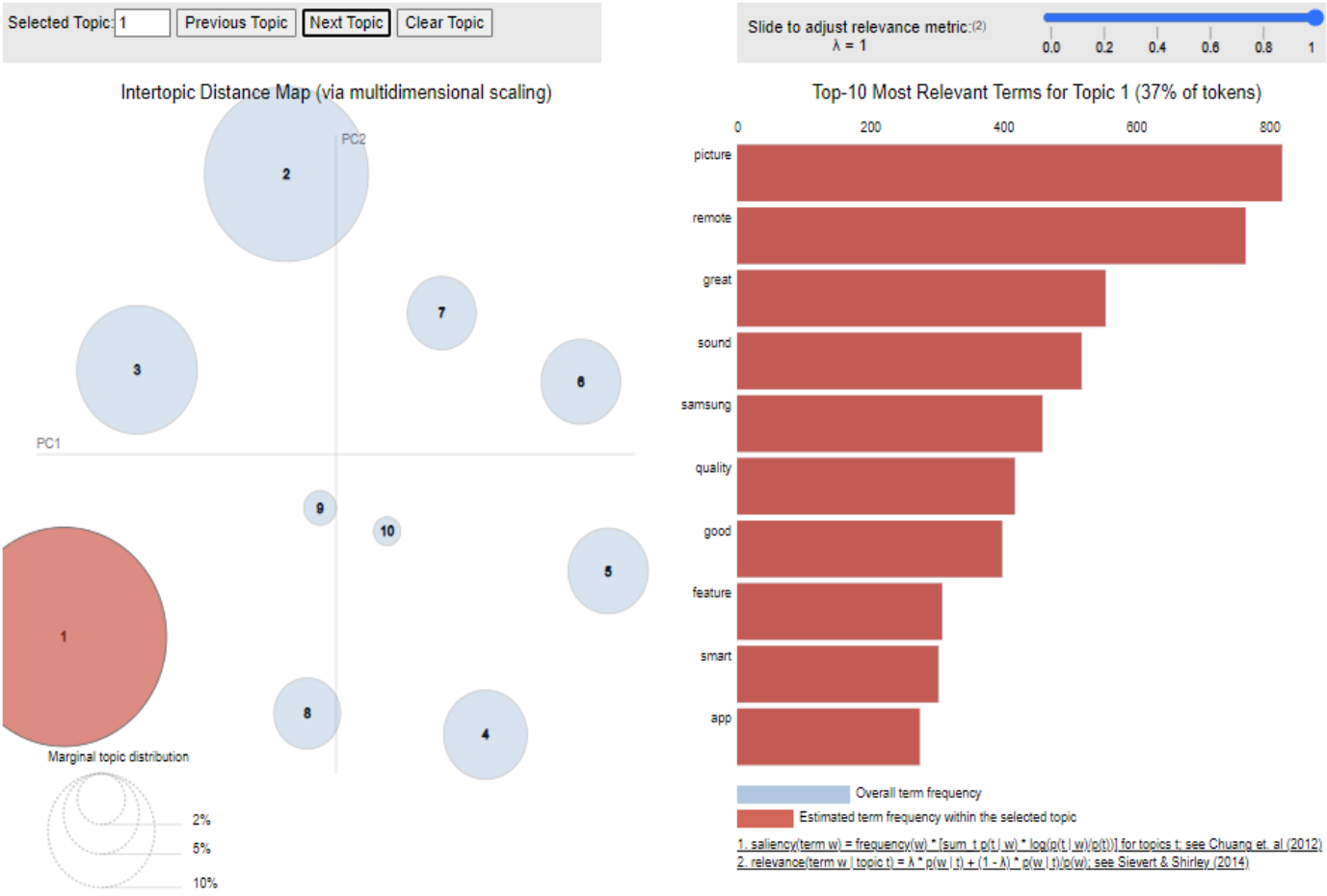

Figure 3. Topic 1 and Top 10 Terms Related to Dissonance-reducing Purchasing Behavior

Table 4. Categories, Terms, and Ranks Related to Dissonance-reducing Purchasing Behavior

Category and Rank (#)	Term and Rank (#)
Economic value	
Emotional value	
Social value	
Quality value (#1)	**picture (#1), remote (#2), great (#3), sound (#4), samsung (#5), quality (#6), good (#7), feature (#8), smart (#9), app (#10)**

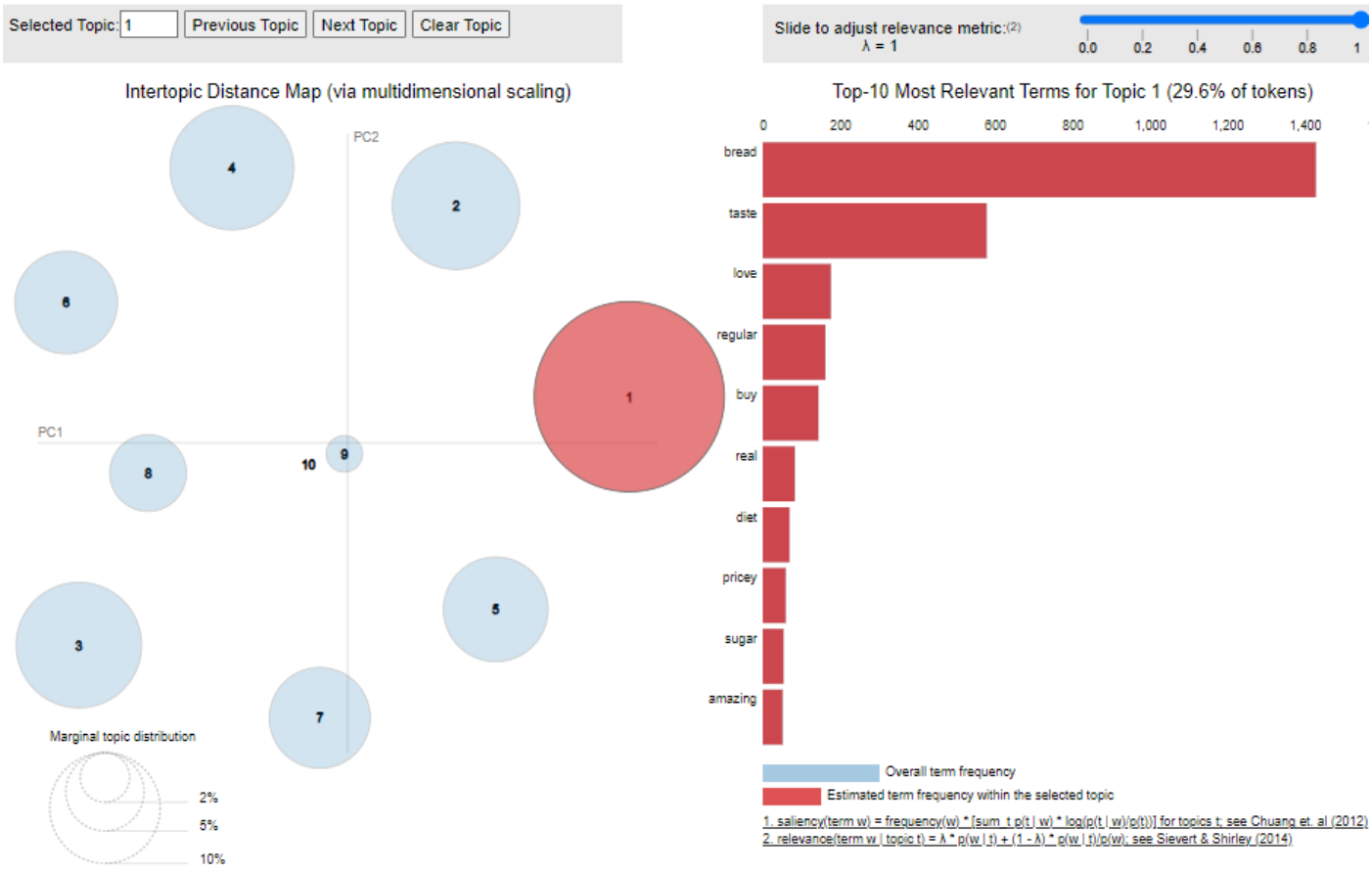

Figure 4. Topic 1 and Top 10 Terms Related to Habitual Purchasing Behavior

4.3 Habitual Purchasing Behavior

This section focuses on examining the purchasing behavior of consumers by analyzing product reviews of bread. For interpretation, Topic 1 in Figure 4 was selected (29.6% of tokens). The top 10 relevant terms associated with this topic were identified and ranked as: "bread" (noun), "taste" (noun), "love" (verb), "regular" (adjective), "buy" (verb), "real" (adjective), "diet" (noun), "pricey" (adjective), "sugar" (noun), and "amazing" (adjective). These terms indicate the underlying concept of consumer behavior in this context. Table 5 provides a summary of the terms, sample reviews, and sentiments related to this habitual purchasing behavior. Two sample reviews, marked by *** in Table 5 exhibited the habitual behavior of re-purchasing. One sample review expresses a negative sentiment associated with the high price of the product, while the remaining reviews show positive sentiment.

Lastly, Table 6 provides a summary of the terms, and their ranks (#) grouped by their value category. The category and sentiment for each term were manually identified by examing the entire context of the corresponding sample review in Table 5.

Table 5. Terms, Sample Reviews, and Sentiments Related to Habitual Purchasing Behavior

Rank (#)	Term	Sample Review	Sentiment
1	bread	This is the best low carb **bread** out there	positive
2	taste	It has a surprisingly good **taste**.	positive
3	love	**Love** this low carb bread!	positive
4	regular	This is the closest bread that I have found that tastes like **regular** bread	positive
5	buy	Will continue to **rebuy** this again.***	positive
6	real	Will buy again fresh and good taste like **real** bread.***	positive
7	diet	Excellent flavor and excellent for my **diet**!	positive
8	pricey	It's a little **pricey**, but I've found anything Low Carb costs more.	**negative**
9	sugar	This is a great bread for the keto diet and for diabetes since it has no **sugar**.	positive
10	amazing	**Amazing** flavor, wonderful texture.	positive

***** Exhibition of habitual purchasing behavior**

Table 6. Categories, Terms, and Ranks Related to Habitual Purchasing Behavior

Category and Rank (#)	Term and Rank (#)
Economic value (#3)	pricey (#8)
Emotional value (#1)	**taste (#2), love (#3), regular (#4), buy (#5), real (#6), amazing (#10)**
Social value	
Quality value (#2)	bread (#1), diet (#7), sugar (#9)

4.4 Variety-seeking Purchasing Behavior

This section focuses on examining the purchasing behavior of consumers by analyzing product reviews of soaps. For interpretation, Topic 1 in Figure 5 was selected (47.2% of tokens). The top 10 relevant terms associated with this topic were identified and ranked as: "soap" (noun), "skin" (noun), "dove" (noun/brand), "bar" (noun), "smell" (noun), "great" (adjective), "butter" (noun), "buy" (verb), "scent" (noun), and "sensitive" (adjective). These terms indicate the underlying concept of consumer behavior in this context. Table 7 provides a summary of the terms, sample reviews, and sentiments related to this variety-seeking purchasing behavior, with all sample reviews expressing positive sentiments.

Table 7. Terms, Sample Reviews, and Sentiments Related to Variety-seeking Purchasing Behavior

Rank (#)	Term	Sample Review	Sentiment
1	soap	I really like using the **soap**.	positive
2	skin	This soap works the best for my **skin**, and for shaving	positive
3	dove	**Dove** leaves my face clean and feeling soft.	positive
4	bar	Does exactly what I would hope a **bar** of soap to do.	positive
5	smell	Good cleansing, pleasant **smell**, & long lasting.	positive
6	great	**Great** moisturizer for sensitive skin.	positive
7	butter	Love the Shea **butter** scent.	positive
8	buy	I'm a first-time **buy**er of Dove Shea Butter. I am impressed and really loving it.****	positive
9	scent	Leaves my skin soft, pleasant **scent**!	positive
10	sensitive	I have **sensitive** skin and this stuff doesn't bother at all... give it a shot!****	positive

****** Exhibition of variety-seeking purchasing behavior**

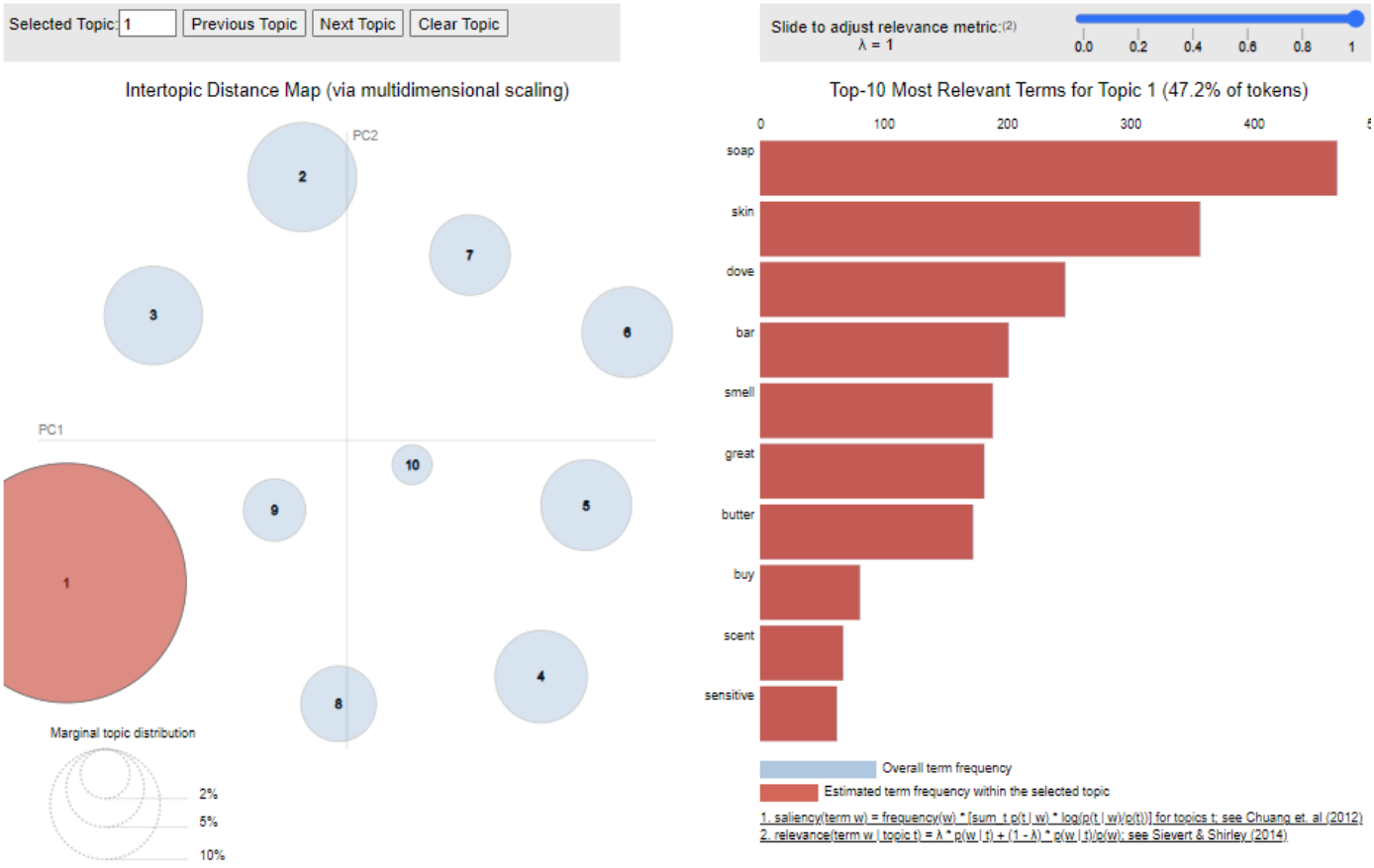

Figure 5. Topic 1 and Top 10 Terms Related to Variety-seeking Purchasing Behavior

Table 8. Categories, Terms, and Ranks Related to Variety-seeking Purchasing Behavior

Category and Rank (#)	Term and Rank (#)
Economic value	
Emotional value (#2)	dove (#3), smell (#5), butter (#7), buy (#8), scent (#9)
Social value	
Quality value (#1)	**soap (#1), skin (#2), bar (#4), great (#6), sensitive (#10)**

Two sample reviews, marked by **** in Table 7 exhibited variety-seeking purchasing behavior of trying something new. All sample reviews express positive sentiments. Lastly, Table 8 provides a summary of the terms, and their ranks (#) grouped by their value category. The category and sentiment were manually identified for each term by examing the entire context of the corresponding sample review in Table 7.

5. Discussion

In this study, the focus was on the analysis of the top 10 terms of the first topic extracted using NLP. These terms were categorized and explored to understand their potential influences within the PERVAL model for economic, emotional, social, and quality values. The additional analysis included ranking (#) these values based on the total cluster size and the rank of terms within each category.

5.1 Complex Purchasing Behavior

5.1.1 Economic Value (#3)

The ranking of this category in Table 2 suggests that economic value holds the third strongest influence on purchase intention. The influence was found positive, as suggested by its sentiments. Terms such as "mercedes" and "mileage" express the economic value associated with the luxury car, respectively. The third strongest influence suggested that the luxury car market is well-established and the price is no longer the primary driver for consumers when purchasing luxury cars. These results align with a previous study conducted by another research group [12].

5.1.2 Emotional Value (#1)

Drawing from the first ranking of this category in Table 2, it can be inferred that emotional value exerts the strongest influence on purchase intention. The influence was found positive, as suggested by its sentiments. Therefore, perceived emotional value strongly affects consumers' intentions to buy luxury cars. This result aligns with a previous study [12], which suggests that luxury car usage brings joy to consumers. Experiencing joy and pleasure plays a crucial role in driving a specific behavior. Terms such as "love", "good", "smooth", and "comfortable" reflect consumers' positive experiences with their luxury cars. Thus, a perceived emotional value significantly influences consumers' intentions to purchase luxury cars.

5.1.3 Social Value (#4)

Based on the fourth ranking of this category in Table 2, it is indicated that social value has the weakest influence on purchase intention. The influence could be positive, as suggested by its sentiment. Hence, consumers may partially believe that owning luxury cars can showcase their social status, supporting previous studies [12]. This finding weakly reinforces the understanding that individuals are concerned about their social identities and tend to choose luxury products to create and maintain a desirable self-image for the public. The term "mercedes" can signify consumers' social status associated with owning

these luxury cars. Consequently, perceived social value shows weakly influence on consumers' intentions to purchase luxury cars.

5.1.4 Quality Value (#2)

According to the second ranking of this category in Table 2, the quality value was identified as the second strongest influence on purchase intention. The influence can be positive, as suggested by its sentiments. Luxury products are often associated with symbolic meanings such as exclusivity, high quality, aesthetics, prestige, and craftsmanship. This finding aligns with previous studies [12]. Terms such as "great", "excellent", "gas", and "mileage" convey the quality value of luxury cars. Therefore, the perceived quality value is found to influence consumers' intentions to purchase luxury cars.

In conclusion, for the complex purchasing behavior observed, consumers' perceived emotional, quality, economic, and social values can influence their purchase intentions regarding luxury cars, respectively. The influence was found positive, suggested by the sentiments. Furthermore, examining the attributes associated with each term can provide additional insights into their corresponding perceived value. Term "smooth", "love", and "comfortable" elicit specific feelings associated with the product. According to the findings, emotional value emerges as the most influential factor in purchase intentions toward luxury cars. Thus, it is observed that providing the highest quality product is not the most crucial aspect of marketing luxury cars. Instead, the development of positive emotions associated with luxury cars is observed to be of utmost importance. Additionally, the positive economic value is found to be of lesser importance while a positive social status linked to luxury cars is found to be of the least importance.

5.2 Dissonance-reducing Purchasing Behavior

5.2.1 Economic Value

No terms in this topic were found to associate with economic value. Consequently, the perceived economic value was observed to have no influence on consumers' purchase intentions regarding smart TVs.

5.2.2 Emotional Value

No terms in this topic were found to relate to emotional value. Therefore, the perceived emotional value was found to not influence consumers' purchase intentions toward smart TVs.

5.2.3 Social Value

No terms in this topic were found to pertain to social value. Hence, the perceived social value was found to not affect consumers' purchase intentions regarding smart TVs.

5.2.4 Quality Value (#1)

As the terms are exclusively found within this category in Table 4, it strongly suggests that quality value is the sole factor influencing purchase intention. Terms such as "picture", "remote", "great", "sound", "samsung", "quality", "good", "feature", "smart", and "app" represent the quality value of smart TVs, respectively. Therefore, perceived quality value is the sole factor that can influence consumers' intentions to purchase smart TVs. This finding aligns with a previous study [7], which highlights the significance of quality and features in attracting customers to TV brands. Furthermore, the influences were found positive, neutral, or negative, as suggested by their sentiments.

In conclusion, for the dissonance-reducing purchasing behavior observed, consumers' perceived quality value is the only factor that can influence their purchase intentions toward smart TVs. Furthermore, examining the attributes associated with each term can

provide additional insights into their corresponding perceived value. Terms such as "picture", "remote", "great", "sound", "samsung", "quality", "good", "feature", "smart", and "app" represent specific qualities associated with the product, respectively. According to the findings, quality value emerges as the primary influencer of purchase intentions toward smart TVs. Therefore, providing the best quality product is found to be the most important aspect of marketing them.

5.3 Habitual Purchasing Behavior

5.3.1 Economic Value (#3)

According to the third ranking of this category in Table 6, economic value exhibited the weakest influence on purchase intention. Therefore, perceived economic value only weakly affects consumers' intentions to purchase bread. The term "pricey" represents the economic value of the bread. Furthermore, the influence is found negative, as suggested by its sentiment.

5.3.2 Emotional Value (#1)

Based on the first ranking of this category in Table 6, it is indicated that emotional value holds the strongest influence on purchase intention. The influence was found positive, as suggested by its sentiments. Therefore, perceived emotional value strongly influences consumers' intentions to purchase bread. Terms such as "taste", "love", "regular", "buy", "real", and "amazing" express the emotional value associated with the bread, respectively. This finding aligns with a previous study that highlights issues such as staling and bad taste as significant factors contributing to bread wastage [4].

5.3.3 Social Value

No terms in this topic were found to relate to social value. Consequently, the perceived social value was found to have no influence on consumers' purchase intentions regarding bread.

5.3.4 Quality Value (#2)

According to the second ranking of this category in Table 6, the quality value was observed to have the second strongest influence on purchase intention. The influence was found positive, as suggested by its sentiments. Terms such as "bread", "diet", and "sugar" represent the quality value of the bread, respectively. Therefore, perceived quality value influences consumers' intentions to purchase bread. This result is consistent with a previous study [4], which emphasizes that quality is a primary consideration for consumers when choosing bread. Bread producers should prioritize product quality.

In conclusion, for the habitual purchasing behavior observed, consumers' perceived emotional, quality, and economic values can influence their purchase intentions regarding bread. Furthermore, examining the attributes associated with each term can provide additional insights into their corresponding perceived value. Terms such as "taste", "love", "regular", "buy", "real", and "amazing" evoke specific emotions associated with the product. According to the findings, emotional value emerges as the primary influencer of purchase intentions toward bread. Therefore, it is found that providing the best quality product may not be the most crucial aspect of marketing bread. Instead, developing positive emotions associated with bread is found to be of utmost importance. Furthermore, developing a positive economic value associated with bread is found to be of lesser importance.

5.4 Variety-seeking Purchasing Behavior

5.4.1 Economic Value

No terms in this topic were found to relate to economic value. Consequently, the perceived economic value was found to have no influence on consumers' purchase intentions regarding soaps.

5.4.2 Emotional Value (#2)

Based on the second ranking of this category in Table 8, it is suggested that emotional value holds the second strongest influence on purchase intention. The influence was found positive, as suggested by its sentiments. Terms such as "dove", "smell", "butter", "buy", and "scent" express the emotional value associated with soaps, respectively. Therefore, perceived emotional value influences consumers' intentions to purchase soaps. This finding aligns with a previous study [14], which highlights the strong relationship between soap brands and purchasing decisions.

5.4.3 Social Value

No terms in this topic were found to relate to social value. Consequently, the perceived social value was found to have no influence on consumers' purchase intentions regarding soaps.

5.4.4 Quality Value (#1)

Based on the first ranking of this category in Table 8, the quality value is found to exert the strongest influence on purchase intention. The influence was found positive, as suggested by its sentiments. Terms such as "soap", "skin", "bar", "great", and "sensitive" represent the quality value of the soap. Thus, perceived quality value strongly influences intentions to purchase soaps. This finding is consistent with a previous study [14].

In conclusion, for the variety-seeking purchasing behavior observed, consumers' perceived emotional and quality values can influence their purchase intentions regarding soaps. Additionally, examining the attributes associated with each term provided further insights into their corresponding perceived value. Terms such as "dove", "smell", "butter", "buy", and "scent" evoke specific emotions associated with the product, respectively. Similarly, terms such as "soap", "skin", "bar", "great", and "sensitive" represent specific qualities of the soap. According to the findings, quality value emerges as the primary influencer of purchase intentions toward soaps. Therefore, providing the best quality product is found to be the most important aspect of marketing them. Furthermore, developing a positive emotional value associated with soap is observed to be of secondary importance.

Data interpretation in the conventional method is typically based on the initial design of the study, whereas the NLP method relies on the topics selected by the researcher [2,8,11]. One notable advantage of the NLP method is the absence of subjectivity, bias, and errors introduced by humans. Additionally, the conventional method is time-consuming and expensive as it relies on manual data processing by humans. In contrast, the NLP method offers faster and more cost-effective analysis since it automates various tasks. Overall, qualitative analysis utilizing NLP provides several advantages in comparison to conventional methods. These advantages include increased efficiency and cost-effectiveness, along with the elimination of bias, error, and subjectivity. However, it's important to note that the NLP method requires careful preprocessing, setup, and adjustments to ensure accurate and meaningful results.

5.5 Future Research

Netsiri et al. [16] have identified several potential applications arising from their study. One such application is in the field of marketing, where the identified terms could be used to guide the creation of strategies. Product designers could also leverage these

terms, rankings, and categories to optimize their designs to attract customers. Manufacturers, on the other hand, could use this information to improve the quality and economic value of their products.

Further research should explore the possibility of using this method to extract various other types of information that may be useful in marketing and related fields. For instance, by analyzing customer feedback, topic modeling could help identify significant terms and provide insights into topics that could positively influence the customers.

6. Conclusions

The purchasing decisions of consumers are intricate and have multiple dimensions. In this study, NLP was applied to extract different consumer-perceived values related to purchasing intentions toward luxury cars, smart TVs, bread, and soaps, respectively. These products represent different consumer behaviors including complex, dissonance-reducing, habitual, and variety-seeking purchasing, respectively. The data was collected from product reviews posted on commercial websites and analyzed using NLP and LDA topic modeling.

The study found that for complex purchasing behavior, consumers' perceived emotional, quality, economic, and social value were found to influence their intentions to purchase, respectively. For dissonance-reducing purchasing behavior, the solely perceived quality value was found to influence purchase intentions. For habitual purchasing behavior, the perceived economic, emotional, and quality value was found to influence purchase intentions, respectively. Finally, for variety-seeking purchasing behavior, the perceived emotional and quality value was observed to influence purchase intentions, respectively.

The terms identified in the study not only indicate perceived values toward purchase intention but also provide insights into their corresponding perceived values [16]. For example, the word "smooth" can indicate a joyful and pleasant feeling associated with driving a luxury car. The rank and cluster size associated with each value category could also assist marketers to prioritize and optimize their marketing strategy. Thus, NLP presents a promising technique to perform qualitative analysis for efficient, cost-effective, and unbiased extraction of meaningful information for marketing research.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: Data sets and Python code are available to download at: https://github.com/pnetsiri/JEBI.

References

1. Abram, M.D. (2018). The role of the registered nurse working in substance use disorder treatment: a hermeneutic study. *Issues in Mental Health Nursing*. 39(6):490-498.
2. Abram, M.D., Mancini, K.T. and Parker, R.D. (2020). Method to integrate Natural language processing into quantitative research. *International Journal of Qualitative Methods.* 19:1-6.
3. Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 3:993-1022.
4. Demirtas, B., Kaya, A. and Dagistan, E. (2018). Consumers' bread consumption habits and waste status: Hatay/Turkey example. *Turkish Journal of Agriculture – Food Science and Technology*. 6(11):1653-1661.
5. Dutta, B. (2022). What is Latent Dirichlet Allocation (LDA) in NLP? Retrieved from www.analyticssteps.com/blogs/what-latent-dirichlet-allocation-lda-nlp.
6. Gkikas, D.C. and Theodoridis (2022). AI in consumer behavior. *Advances in Artificial Intelligence-based Technologies*. Springer. 22:147-176.
7. Gnanavel, T. and Sivaskthi, D. (2020). A study on consumer satisfaction toward Samsung LED television with special reference to Coimbatore city. *EPRA International Journal of Multidisciplinary Research*. 6(11):339-343.

8. Hagg, L.J., Merkouris, S.S., O'Dea, G.A., Francis, L.M., Greenwood, C.J., Fuller-Tyszkiewicz, M., Westrupp E.M. and Macdonald, J.A., Youssef, G.J. (2022). Examining analytic practices in Latent Dirichlet Allocation within psychological science: Scoping review. *Journal of Medical Internet Research*. 24(11).
9. Hampasagar, A. (2021). Consumer behavior and its importance in marketing. Retrieve from www.linkedin.com/pulse/consumer-behaviour-its-importance-marketing-anant-hampasagar/.
10. Jarek, K. and Mazurek, G. (2019). Marketing and artificial intelligence. *Central European Business Review*. 8(2):46-55.
11. Koleck, T.A., Dreisbach, C., Bourne, P.E. and Bakken, S. (2019). Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *Journal of American Medical Informatics Association*. 26(4):364-379.
12. Lou, X., Chi, T., Janke, J. and Desh, G. (2022). How do perceived value and risk affect purchase intention toward preowned luxury goods? An empirical study of U.S. consumer. *Sustainability*. 14, 11730:1-16.
13. Ma, X. (2019). Dimensionality-Reduction with Latent Dirichlet Allocation. Retrieved from towarddatascience.com/dimensionality reduction with latent dirichlet allocation-8d73c586738c.
14. Mahajan, R. and Behare, N. (2022). Comparative study to understand factors affecting consumer purchase behavior of different types of soaps and to check soap preference in current pandemic conditions with respective Pune. *Journal of Pharmaceutical Negative Results*. 13(9):7004-7013.
15. Monkeylearn (2023). Topic modeling: An introduction. Retrieved from monkeylearn.com/blog/introduction-to-topic-modeling/.
16. Netsiri, P. and Lhotakova, M. (2023). Application of natural language processing to enhance qualitative research used for marketing. *Proceedings of 23rd International Joint Conference: Central and Eastern Europe in the changing business environment*. 171-180.
17. Radu, V. (2022). Consumer behavior in marketing-patters, types, segmentation. Retrieved from https://www.omniconvert.com/blog/consumer-behavior-in-marketing-patterns-types-segmentation/.
18. Rampton, J. (2022). 3 examples of AI in marketing. Retrieved from blogs.oracle.com/marketingcloud/post/ai-in-marketing
19. Sievert, C. and Shirley, K.E. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceeding of the workshop on interactive language learning, visualization, and interface*. 63-70.
20. Stedman, C. and Hughes, A. (2023). Data mining. Retrieved from www.techtarget.com/searchbusinessanalytics/definition/data-mining.
21. Stolz, K. (2022). Why do(n't) we buy preowned luxury products? *Sustainability*. 14, 8686:1-24.
22. Sweeney, J.C. and Soutar, G.N. (2001). Consumer perceived value: The development of a multiple item scale. *Journal of Retailing*. 77:203–220.
23. Textrics (2021). Topic modeling: How can you use it for analyzing unstructured data? Retrieved from textrics.medium.com/topic-modelling-how-can-you-use-it-for-analysing-unstructured-data-28045dd502cc.