Journal of Business Decision and Risk Sciences (JBDRS)

https://doi.org/

Journal Homepage: *www.JBDRSJournal.com*

ORIGINAL RESEARCH PAPER

Topic Modeling and Sentiment Analysis of Online Review for Persian E-commerce

Rouhollah Sohrabi [a, *]; Morteza Yousef Sanati [b]; Hamid Reza Abdolhosseiny [a]

[a] *Department of Management, Faculty of Economics and Social Sciences, Bu-Ali Sina University, Hamedan, Iran.*
[b] *Department of computer engineering, engineering faculty, Bu-Ali Sina University, Hamedan, Iran.*

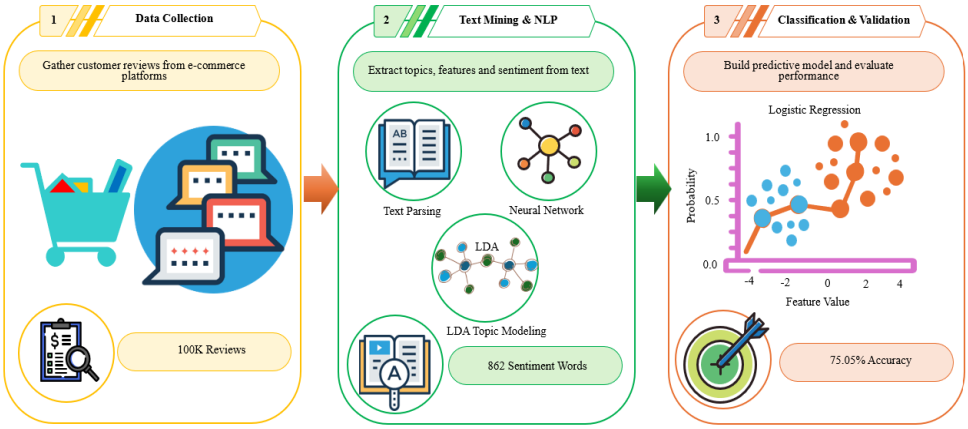
P A P E R I N F O

Paper history:
Received: 01/05/2026
Revised: 02/13/2026
Accepted in revised form: 03/01/2026
Published: 03/30/2026

***Keywords*:**
Data Mining;
Customer Satisfaction;
Latent Dirichlet Allocation;
Topic Extraction;
Big Data.

A B S T R A C T

Data mining methods can be used to analyze data and reveal unknown rules and links; Therefore, using the existing techniques and creating a relationship between the field of text mining and the field of management, we have analyzed the Persian text and extracted the dimensions of customer satisfaction and sentiment analysis in retail sites, which ultimately leads to 38 dimensions of satisfaction and creating a sentiment lexicon for Digikala website. In this research, the latent Dirichlet allocation method (LDA algorithm) has been used to extract the recurring topics discussed in the data set. We also made a sentiment lexicon in sentiment analysis, which produced a glossary containing 862 adjectives and frequent adverbs used in the comments. The data set of the DigikalaNext platform, which includes 100000 registered user comments provided to researchers, has been used. In this study, logistic regression was used to validate the steps, which proved the accuracy of 75.05%.



1. Introduction

Opinions are the center of almost all human activities and the primary influence of our behaviors. Our beliefs and perceptions of reality and our choices depend on how others see and evaluate the world. For this reason, when we need to make decisions, we often ask for the opinions of others [1]. Analysts estimate that more than 80 percent of today's data is stored unstructured (such as text, audio, video, video). Traditionally, qualitative data analysis methods such as manual coding have been used to analyze them. However, the size of the textual datasets available on the Internet makes manual analysis almost impossible [2]. For the first time in human history, large volumes of digitally interpreted data have been recorded for analysis [1]. In online shopping, customers cannot see the products up close, so the opinions of other users become more and more critical. These comments and ratings are vital for potential customers [3]. Online customer feedback has become an essential tool for companies and customers to receive and provide feedback on products and services .Online trading has grown exponentially over the past decade, accounting for 11% of all retail. One of the parameters of success in online trading is online reviews in which customers can determine their satisfaction with the purchase [5]. Given the significant role of e-commerce in the retail industry, Online commenting is becoming more and more important in purchasing decisions. Certain opinions affect sales in an e-commerce transaction [5,6]. The online review also reduces uncertainty and the cost of online commerce as long as quality information is embedded in consumer responsiveness [7,8]. Computer science, especially text mining, has come to the aid of management science and has provided a new method for researchers in this field [9]. Companies can use practical algorithms designed to extract effective customer satisfaction indicators and plan to improve these indicators to increase customer satisfaction.

* Corresponding Author: Rouhollah Sohrabi
E-mail address: r.sohrabi@basu.ac.ir
ISSN:
URL: https://www.jbdrsjournal.com/article_735679.html

Please cite this article as: R. Sohrabi, M. Y. Sanati & H. R. Abdolhosseiny, *Journal of Business Decision and Risk Sciences (JBDRS)*, 2026, *2*(1), 33-37. https://doi.org/

/© 2026 The Author(s). Published by Power of New Vision Movement Institute.

Topic modeling is a way to find the hidden structure of unlabeled data. In this method, documents are defined in the form of a combination of hidden topics, and each topic is a probability distribution over words. Most research in the field of Topic modeling includes "probabilistic latent semantic analysis" [8] and "Latent Dirichlet allocation" [9] The "Latent Dirichlet allocation" model is assumed to extract topics. When a user intends to write a text, he first thinks about it. Then, depending on each topic and the distribution of words in that topic, it selects a specific word to write. This is done repeatedly until the text is complete [11]. Marketing managers use online customer surveys to drive a company's sales and assess customers' feelings, brand, and absolute satisfaction [12]. One of the text analysis applications that data scientists have highly regarded is sentiment analysis. In general, sentiment analysis helps determine the emotions of a speaker, author, or emotional response to a document, event, or the overall polarity of a text. Usually, traditional methods (reading comments by humans) in extracting and summarizing comments will face many problems (Requires a lot of time and capital); Therefore, automated Sentiment analysis systems are needed [1]. The online retail industry operates in a highly competitive market in which achieving and maintaining a high level of customer satisfaction is considered a critical competitive advantage [13]. This research provides a new framework for measuring customer satisfaction in this industry. Using text-analysis methods, we examine online customer reviews to extract dimensions of satisfaction and sentiment analysis to guide online retail sites to improve competition. As a result, we extracted 38 dimensions of satisfaction and a sentiment lexicon specific to E-commerce.

In similar studies using Persian data, Samizade et al. [14] used machine learning methods to compare and classify Persian texts based on the emotions of active users in cyberspace. Shams et al. [15] conducted the first study on SA to target the Persian language. In that study, an unsupervised LDA-based approach was proposed. Their study can only be used to detect the polarity of the text and includes disadvantages such as not considering the problems of SA language in Persian. Shukri et al. [16], in a study, proposed a method for semantic clustering of texts based on Dirichlet latent allocation and genetic metaheuristic algorithm. Bagheri et al. [17] Published an article that addresses some of the challenges with Persian text Mining. In addition, different feature selection methods for detecting polarity were examined.

In a similar study using English data, Lucini et al. [18] published an article; This study provided a new framework for measuring customer satisfaction in the Airline industry. They reviewed online customer comments using analytics techniques To improve the competitiveness of airlines. They analyzed databases that Included more than 55,000 comments covering more than 400 airlines from 170 countries. Using an LDA model, 27 dimensions of satisfaction were identified, with 882 attributes. Guo et al. [19] analyzed a database of 266,544 online reviews for 25,670 hotels in 16 countries. In their study, they extracted the dimensions of customer satisfaction, the importance of different dimensions, and words related to dimensions through LDA. In a study by Xu et al. [20] in the hospitality industry, they examine customer feedback from Booking.com to obtain dimensions of customer satisfaction using latent semantic analysis. Medhat et al. [21], in a survey article, provide a comprehensive overview of the latest updates on sentiment analysis. This survey has reviewed and presented advances in recently proposed algorithms and various SA applications. The primary purpose of this study is to provide an almost complete picture of SA techniques and areas related to brief details.

Sun [22] and Xie et al. [23] have proposed a text-based clustering algorithm in this model to extract conceptual features and the model's dimensions. They have benefited from the Latent Dirichlet Allocation. In this model, each topic is a probabilistic distribution of keywords that probability can be considered the importance of that topic. Ma et al. [24] also proposed a clustering method based on topic extraction. This method is presented in three phases and uses the Latent Dirichlet Allocation model based on topic extraction and K-Means for clustering. He talked about movies on film review blogs. Jotheeswaran et al. [25], in a study, while expressing the critical and influential role of sentiment analysis in customer relationship management and online stores, examined the subject of sentiment analysis using the naive Bayes algorithm and used the decision tree approach.

In similar studies, the focus of the researchers is on improving the algorithm and optimizing it. The practical and prescriptive aspects have not received much attention, as in Persian research, this issue has been overused. A study conducted by Lucini et al. [18] has simultaneously implemented the LDA algorithm and sentiment analysis on a data set to provide solutions to improve service delivery, which is in dire need of the lack of such research in the Persian field. Following the example of this research, we decided to conduct a similar study in the area of Persian data, which may have been a step forward in this field.

2. Materials and Methods

2.1. Database

The dataset, made available by Digikala in 2020, is used in this study. The dataset includes 100000 comments registered by Digikala users about different products, published in a file in . xlsx format; in this file, various variables such as buyer ID, product ID, product title, buyers recommended the product, or not, comment text, strengths and weaknesses are given.

Table 1. Comment length statistics (number of words)

Min	Q1	Median	Q3	Max	Mean	Mode
1	9	19	36	1242	29.1	4

Source: Research Findings

2.2. Data preprocessing

According to Tirunillai et al. [26], text analysis using online customer reviews is complicated for three main reasons:

They do not have a standard structure, and most texts use words and grammar at an intermediate level.

Only a few words are informative about the product and its quality dimensions and require cleaning and selection steps.

The text must be transformed into a format that allows for numerical analysis.

The main steps needed to overcome those problems and prepare the ideas for statistical analysis are presented below.

Text preprocessing includes the following:

- Text normalization
- Convert text to sentences (Sentence Segmentation)
- Tokenizing
- Stop Word Removal
- Lemmatizing and Stemming
- Part-of-speech tagging

As in the study conducted by Lucini et al. [18] to identify and extract the dimensions of customer satisfaction in comments, only names were kept. On the other hand, only adjectives and adverbs were kept to calculate sentiment attributes. Finally, regression analysis used nouns, adjectives, and adverbs to validate the findings.

In data processing, the text of each comment is first normalized (the process of converting a token into its base form); In the normalization process, we want all texts to be standardized and integrated. For example: convert English numbers to Persian and delete extra spaces. After this step, the text of each comment is divided into its constituent sentences, and in the next step, each of these sentences becomes a list of its constituent words; then, the stop words in this list are removed, and in the next step, each of the remaining terms in the case Possibility becomes its primary root and in the last stage, using the tools available in the natural language processing in Persian, the role of each word in the sentence is identified and determined.Table 2 shows the effects and variations of each of these predicates on a few examples of raw comments in the data set.

Table 2. Examples of the preprocessed comments

Translation of Original comment	LDA preprocessed text	Sentiment analysis Preprocessed text
Really great. I am very satisfied with it		['really', 'great', 'very']
I have been buying this speaker for several years, and it does not matter; I have compared it with other brands such as Xiaomi, JBL, and other brands. A neck and neck are higher in all respects.	[years, word, Xiaomi, compare, neck, respect, doubt]	['really', 'not at all']
It has a pungent but fragrant smell. It lasts well and becomes more aromatic as it passes	[smell, longevity]	['pungent', 'fragrant', 'how much', 'more fragrant']

Source: Research Findings

2.3. Identification and extraction of customer satisfaction dimensions

To achieve this goal, we used the Latent Dirichlet Allocation (LDA), a Topic recognition method in machine learning and natural language processing. The LDA assumes that each document may be shown as a potential distribution of hidden topics (dimensions of customer satisfaction) and that the distribution of the topic in all documents has a common Dirichlet. Each topic in the LDA model is represented as a probabilistic distribution of words, and the word distribution on topics has a similar initial dichotomy. Everyone can potentially share topics, and each comment has its mix of topics. The LDA assumes the following production process for each document in a set [10,18].

1. Choose $N \sim \text{Poisson}(\xi)$
2. Choose $\theta \sim \text{Dirichlet}(\alpha)$
3. For each W_n in $W = \{W_1, W_2, \dots, W_N\}$:

a. Choose $Z_n \sim \text{Multinomial}(\theta)$

b. Choose W_n from $P(W_n|Z_n, \beta)$

The critical inferential problem that must be solved to use LDA is to calculate the posterior distribution of latent variables (Z|W and θ) of a document (equation):

$$P(\theta, Z|W, \alpha, \beta) = \frac{P(\theta, Z, W|\alpha, \beta)}{P(W|\alpha, \beta)}$$

However, the exact inference of this distribution is unsolvable due to the coupling between θ and β in the plural on latent topics; Thus, approximate algorithms such as inference of change (Airoldi et al. [27]; Blei et al.[10]; Hoffman et al. [9]) or the Markov Monte Carlo chain (Griffiths and Steyvers, [28]; Shivashankar et al., [29]; Tirunillai et al. [26]) are commonly used for it. In this study, we used the Batch Variational Bayes Inference (BVBI) algorithm (Hoffmann et al. [9]), which is available in the scikit-learn library (Hackeling [30] & Lucini et al. [21]).

2.4. Sentiment analysis

Like the method proposed in Duda et al., 2000 for identifying, extracting, and calculating the sentiments of adjectives commonly used by customers, the preprocessed data is arranged so that only attributes and adverbs remain in the comments. The sentiment strength is calculated by the probability of a class occurring (for example, positive or negative sentiments), given that the adjective is used in a comment. The higher the value of P (Class | adjective), the more informative the adjective is to classify a review as belonging to a specific class (i.e., the stronger the sentiment, negative or positive, associated with the adjective) [18].

$$P(Class|Adjective) = \frac{P(Adjective/Class)P(Class)}{P(Adjective)}$$

Some comments were categorized as presenting negative or positive emotions to perform this analysis. We considered negative comments to be those labeled "not recommend" that occurred in 16098 cases. Comments on the "recommend" tag, which occurred in 36,960 cases, were positive.

As a result, this method lists the most common and vital attributes that Digikala customers use when describing their emotions in comments and their sentiment scores [31].

To calculate the positive and negative probabilities for a particular word in the vocabulary, we use the following inputs:

- $freq_{pos}$ and $freq_{neg}$ The frequencies of that particular word in the class are positive or negative.
- N_{pos} and N_{neg} are the total number of positive and negative words for all documents, respectively.
- V The number of unique words in the whole document set for all classes, whether positive or negative.

We will use these to calculate the positive and negative probabilities of a particular word using this formula:

$$P(W_{pos}) = \frac{freq_{pos} + 1}{N_{pos} + V}$$

$$P(W_{neg}) = \frac{freq_{neg} + 1}{N_{neg} + V}$$

Log-likelihood:

To calculate the logical probability of the same word, we can implement the following equations:

$$loglikelihood = log\,(\frac{P(W_{pos})}{P(W_{neg})})$$

Words Take the probability of entering values from -inf to + inf [32].

Regression analysis:

We use a logistic regression classification to test and validate the dimensions of satisfaction and attributes extracted from online customer reviews [33] to predict product recommendations. In the regression, we used the entire preprocessed dataset: the independent variables were the significance of the satisfaction dimensions identified using the LDA and the sentiment score of each comment. The dependent variable was the recommendations given by the respondents about the goods. The dataset was divided into training and testing collections. Applying the best regression model to the whole data set also allows us to identify the most relevant dimensions for predicting product recommendations [18].

3. Results and Discussion

3.1. The optimal number of topics

One way to determine the optimal number of topics that can be extracted from a data set using the LDA algorithm is to use a coherence score; To do this, a program was written using the Python language that, by changing the number of topics in the range 2 to 98, on the preprocessed data, calculates the value of the coherence score each time and gives us an overview by plotting the changes. . Figure 1 shows the graph of changes in the coherence score in terms of the number of topics. Based on this figure, it can be concluded that the optimal number of topics extracted from the Digikala data set with a coherence score of 0.467 is 71.

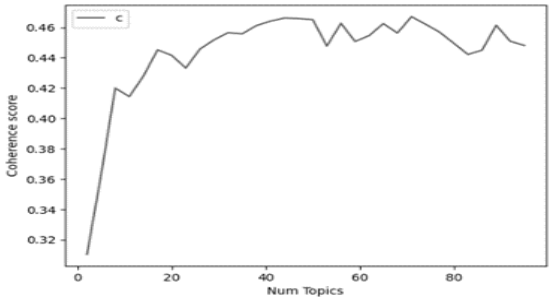


Figure 1. Coherence score changes by topic

Source: Research Findings

3.2. Detect and extract topics:

For a certain number of "t" subjects, we divided the data set into five parts (one for testing and four for training) and fitted an LDA model to training [18]. When the number of topics was set to 71, the entire dataset was used to conform to the LDA model to extract Latent topics in the text by this algorithm.

Table 3 shows one of the extracted topics, which includes ten words with the importance of each word in the topic.

Table 3. Example of attributes and importance in the topic

Attributes	Importance
Shoe	0.114
Color	0.086
Quality	0.037
Price	0.036
Size	0.036
Opinion	0.029
Picture	0.029
Order	0.019
Product material	0.018
Use	0.017

Source: Research Findings

3.3. customer satisfaction dimensions:

Dimension naming is done by two researchers separately, and the results are compared and discussed until an agreement is reached. Each researcher mentally tries to identify the logical connections between the essential words (for example, the top three words) for a topic and assign a name to it. The chosen name is mostly mentally tested through logical connection with other words in the list of top 10 words. The name chosen is considered permanent whenever a link is found. If a word is found that does not match the subject name, the naming process starts again and is repeated until the name logically connects to all the words. The last terms assigned to the topics are the dimensions of customer satisfaction expressed [18]. After completing the steps, 38 dimensions of satisfaction of the Digikala data set (some of the extracted items were also named because of their similarity to one dimension) were extracted as table 4:

Table 4. Dimensions of satisfaction of the Digikala data set

• Product color	• Durability of the product	• volume, and amount of space required to use the product
• Quality of construction and production	• Charger of the product	• Product style and appearance
• Item prices	• Product sound quality	• The amount of discount on the product
• Match the product specifications with the sent goods	• Speed of goods	• Type of expression in the text of the product description
• Match the product photo with the product itself	• Product power	• Non-allergenic product for skin and face in makeup
• Product ordering steps and postage time	• Accessories with goods	• Observance of ergonomic rules in the product
• Materials used in the manufacture of goods	• Product packaging	
• How to install and ease of use of the product	• Variety of product designs	
• Product function	• Product warranty	
• Product efficiency	Return services	
• The memory cards in electrical storage devices	• The amount of energy consumption of the product	

- Type of brand
- Battery life in electrical goods
- Give it as a gift

Product taste

- Aroma, smell, and taste of the product
- Support and after-sales service
- The product is real or fake
- Quality of wires and cables with electrical appliances
- The quality of the camcorder in the mobile phone
- Display on phone, tablet, and laptop

Source: Research Findings

3.4. Sentiment analysis

To extract and identify the sentiment score of the words in the data set, reviews that recommend purchasing the product to other users are considered positive reviews, and reviews that are not recommended to buy the product are considered negative reviews; thus, 36960 positive reviews and 16098 Negative reviews finally remains from the data set. We analyze the new data set after proper preprocessing of sentiment analysis. Finally, 862 words with calculated sentiment scores were obtained, and Table 5 lists 20 positive and negative words with the highest sentiment score.

Table 5. 20 words with the most negative and positive sentiment score

Attribute	Sentiment score	Attribute	Sentiment score
Lie	-9.904	Funny	10.834
Haram	-9.904	Powerful	10.739
Middle	-9.722	Has space	10.473
Awful	-9.722	Powerful	10.344
True	-9.722	Separately	10.344
Unconsciously	-9.722	Dark	10.344
Gratuitous	-9.722	Useful	10.282
Movement	-9.616	Unique	10.282
Possible	-9.616	Valuable	10.228
Problem	-9.616	Good-looking	10.11
Unconsciously	-9.498	Beautiful color	10.11
Enjoy	-9.498	Useful	10.046
Inefficient	-9.498	Computer	10.046
Current	-9.365	Excellent	10.046
So much	-9.211	Exclusive	9.977
Instantaneous	-9.211	Juicy	9.977
That's bullshit	-9.211	Equipped	9.903
Bad Taste	-9.211	Beautiful	9.823
Dryer	-9.211	Accessibility	9.736
The original	-9.028	European	9.736

Source: Research Findings

After determining the sentiment lexicon for Digikala and Persian online retailers, it is time to calculate the sentiment score of the initial data set. To do this, divide each comment into its constituent words, and the words that have a semantic load are added to their scores, and the sum of these scores creates the sentiment score of that comment.

3.5. Regression analysis

The importance of satisfaction and sentiment scores assigned to online customer reviews was tested and validated using a logistic regression classifier to predict Digikala product recommendations. The average accuracy of the experimental set was 75.05%. Table 6 calculates the P-value for all variables and states that, as can be seen, the P-value for topics 3, 5, 7, 8, 10, 36, 39, 40, 44, 47, 53, 55, 56, 65, 66, 68, 69 and 70 are more significant than 0.05, and this shows that the coefficients of these issues are not substantial in the final result with 95% confidence.

Table 6. Dimension p-value in logistic regression

Dimensions	P_value	Dimensions	P_value	Dimensions	P_value
Topic 1	0	Topic 25	0	Topic 49	0
Topic 2	0	Topic 26	0	Topic 50	0
Topic 3	0.686	Topic 27	0	Topic 51	0
Topic 4	0	Topic 28	0.001	Topic 52	0.007
Topic 5	0.287	Topic 29	0.001	Topic 53	0.212
Topic 6	0	Topic 30	0	Topic 54	0
Topic 7	0.083	Topic 31	0	Topic 55	0.137
Topic 8	0.423	Topic 32	0.015	Topic 56	0.105
Topic 9	0	Topic 33	0.003	Topic 57	0.032
Topic 10	0.891	Topic 34	0.037	Topic 58	0
Topic 11	0	Topic 35	0	Topic 59	0.004
Topic 12	0.043	Topic 36	0.236	Topic 60	0
Topic 13	0	Topic 37	0	Topic 61	0.011
Topic 14	0	Topic 38	0	Topic 62	0
Topic 15	0	Topic 39	0.524	Topic 63	0
Topic 16	0	Topic 40	0.152	Topic 64	0.049
Topic 17	0	Topic 41	0	Topic 65	0.053
Topic 18	0	Topic 42	0	Topic 66	0.115
Topic 19	0.001	Topic 43	0.01	Topic 67	0
Topic 20	0.005	Topic 44	0.398	Topic 68	0.639
Topic 21	0.001	Topic 45	0.006	Topic 69	0.186
Topic 22	0	Topic 46	0	Topic 70	0.817
Topic 23	0.018	Topic 47	0.162	Topic 71	0
Topic 24	0.008	Topic 48	0	Sentiment score	0

Source: Research Findings

Identifying the dimensions of customer satisfaction is crucial to precisely determine how users evaluate Digikala's site's services. One of the essential issues that sellers should pay attention to is the accuracy of the declared items in the description of their product and its compliance with the shipped goods, for example, the items " Product color," " Match the product specifications with the sent goods" and " Match the product photo with the product itself" show the carelessness of the sellers of Digikala site when introducing and sending their products, the sellers can get more satisfaction from their customers due to these cases and higher accuracy. Online buyers are very punctual and accurate. If the seller sells poor quality products, buyers with negative ratings and comments on that product will prevent the seller from selling and making more profit, such as " Quality of construction and production," " Materials used in the manufacture of goods," "Product function," "Type of brand," "Durability of the product," "The product is real or fake," "Quality of wires and cables with electrical appliances" and "Observance of ergonomic rules in the product" Extracted from the recorded comments of buyers who indicate this. In today's world, customers are always looking for more services. To attract buyers' satisfaction, this point should be taken into account and not neglected. Commodity "among the extracted dimensions indicates this issue, so sellers in today's market will be successful if there is no shortage in this area. The case of "product ordering process and post time" has been one of the most critical issues concern of buyers. It is high and causes his dissatisfaction, and for this reason, the possibility of returning the goods will increase. Hence, the speed of sending the goods is essential, which should be considered by the sellers and Digikala.

After creating the appropriate sentiment lexicon, we calculated the sentiment score of each comment in the data set. This score indicates the positive or negative degree of customers' feelings towards that product and whether or not other users purchase it. It can be used in the regression algorithm to predict the buyer's opinion about the product recommendation. In this way, it can help decide to invest in different parts of the site and advance the company's goals [34]. After completing the written programs, it is now possible to give each Digikala user's comments registered in cyberspace to the program and determine the amount of coverage of each topic in that comment along with the score of emotions and product recommendation or not. The importance and application of this program become apparent when we intend to evaluate and categorize these results among the millions of registered user comments at the cyberspace level.

Research Limitations:

1. Lack of suitable and optimal data processing libraries for Persian data [35]
2. Unavailability of data for different categories of Digikala products
3. Use of variables with a p-value greater than 0.05 in logistic regression

Suggestions for future research:

1. Using other subject modeling algorithms and comparing their performance with each other
2. Work on improving the performance of Persian language processing libraries and optimizing their performance to mechanize future research and projects further
3. Using translated emoticons dictionaries and comparing their performance with the results of this research

4. Using other forecasting algorithms and comparing their performance
5. Use of data sets more focused on specific areas of products and their analysis
6. Use larger data sets and optimize results
7. Integrating the steps performed in this research and building its application
8. Testing and implementing the steps of this research on stream data

4. Conclusions

In this paper, we present a novel text mining-based approach to exploring dimensions of E-commerce customer satisfaction from the analysis of Online Customer Reviews (OCRs). To analyze the comments of users of the online retail site Digikala, a very reliable method has been used to identify the dimensions of their satisfaction. LDA is one of the main methods of extracting latent topics from unstructured texts. Also, in this research, a simple Bayesian classifier has been used to analyze the sentiment scores of words. Finally, a logistic regression classification is used to predict the retail recommendations of Digikala. The percentage of prediction obtained in the logistic regression stage was 75.05, which is close to the cases operated in English articles.

Our paper explores an existent gap in the literature regarding retrieval and information mining from OCRs in the Persian E-commerce industry. Our method allows extracting dimensions of satisfaction from raw text, and It is an innovative way of measuring customer satisfaction suitable for future academic and applied studies.

Author Contributions

Authors contributed equally to the conception, preparation, and writing of this manuscript.

Conflict of Interest

R. Sohrabi, in the capacity of Editorial Board, had no involvement in the selection of reviewers or the management of the peer-review process, and the other authors, while observing publication ethics in referencing, declare the absence of any conflict of interest.

The authors, in the capacity of Editoral, had no involvement in the selection of reviewers or the management of the peer-review process, and the other authors, while observing publication ethics in referencing, declare the absence of any conflict of interest.

Acknowledgement

The authors gratefully acknowledges the support provided by the Bu-Ali Sina University in conducting this research.

AI Used Statment

The authors retain full responsibility for the accuracy, originality, integrity, legality, and ethical compliance of all text, data, tables, figures, analyses, references, and conclusions, and have verified all AI-assisted outputs. Any fabrication, falsification, manipulation, plagiarism, copyright infringement, or other misuse of AI remains solely the responsibility of the authors and does not transfer or diminish their accountability for the published work.

References

[1] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions (1st ed.)*. Cambridge University Press, 2015.
[2] S. Debortoli, O. Müller, I. Junglas & J. vom Brocke. *Communications of the Association for Information Systems*, 2016. **39(1)**, 110–135. https://doi.org/10.17705/1cais.03907.
[3] D. H. Park, J. Lee, & I. Han, *International Journal of Electronic Commerce*, 2007, **11(4)**, 125–148. https://doi.org/10.2753/jec1086-4415110405.
[4] P. Chaovalit & L. Zhou, *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 2005. https://doi.org/10.1109/hicss.2005.445.
[5] K. Soen, B. Yin, *Customer Behaviour Analysis of E-commerce What information can we get from customers' reviews through big data analysis*, Master of Science Thesis, KTH Industrial Engineering and Management Industrial Management, Stockholm, 2019.
[6] S. A. Razavi, S. Pirani, *Business Review*, 2019, **3(4)**, 28-36. https://doi.org/10.22050/pbr.2019.119151
[7] N. Hu, L. Liu, & J. Zhang, *SSRN Electronic Journal*, 2008, **9(3)**, 201-214.
[8] M. D. Hoffman, D. M. Blei & F. Bach, Online Learning for Latent Dirichlet Allocation. Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 2010.
[9] S. A. Razavi, A. K. Manshad, L. Kashani, *Energy Strategy Reviews*, 2026, **65**, 1022261-30. https://doi.org/10.1016/j.esr.2026.102226
[10] D. M. Blei, A. Y. Ng & M. I. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research*, 2003, **3(1)**, 993-1022.
[11] M. Shokohian, A. Asemi, A. Shabani & M. Cheshmeh Sarabi, *Journal of Health Information Management*, 2019. **16(4)**, 1-7.
[12] L. M. Willemsen, P. C. Neijens, F. Bronner & J. A. de Ridder. *Journal of Computer-Mediated Communication*, 2011, **17(1)**, 19–38.
[13] S. A. Razavi, M. Nasr Esfahani, P*olitical Strategic Studies*, 2020, **9(35)**, 101-134. https://doi.org/10.22054/qpss.2020.42777.2320
[14] R. Samizade & S. M. Abad, *Journal of Information Technology Management*, 2018, **10(1)**, 101-122.
[15] M. Shams, A. Shakery, H. Faili, *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing*, Shiraz, 2012.
[16] S. Shukri, B. Masoumi, Fourth International Conference on Science and Technology Research, Tehran, 2016.
[17] A. Bagheri, M. Saraee, F. de Jong, *21st Iranian Conference on Electrical Engineering (ICEE)*, Mashhad, 2013.
[18] F. R. Lucini, L. M. Tonetto, F. S. Fogliatto, M. J. Anzanello, *Journal of Air Transport Management*, 2020, **83**, 101760.
[19] Y. Guo, S. J. Barnes, Q. Jia, *Tourism Management*, 2017, **59**, 467–483.
[20] H. Xu, F. Zhang, W. Wang, *Knowledge-Based Systems*, 2015, **76**, 166–175.
[21] W. Medhat, A. Hassan, H. Korashy, *Ain Shams Engineering Journal*, 2014, **5(4)**, 1093–1113.
[22] X. Sun, 1*0th International Conference on Semantics, Knowledge, and Grids*, Beijing, 2014.
[23] P. Xie, E.P. Xing, ArXiv, 2013, abs/1309.6874.
[24] Y. Ma, Y. Wang, B. Jin, *Expert Systems with Applications*, 2014, **41(18)**, 8203–8210.
[25] J. Jotheeswaran, D. Y. Kumaraswamy, *International Journal of Engineering and Computer Science*, 2013, **2(10)**, 2893-2898.
[26] S. Tirunillai, G. J. Tellis, *Journal of Marketing Research*, 2014, **51(4)**, 463–479.
[27] E. M. Airoldi, D. M. Blei, S. E. Fienberg, E. P. Xing, *Journal of Machine Learning Research*, 2008, **9(65)**, 1981–2014.
[28] T. L. Griffiths, M. Steyvers, *Proceedings of the National Academy of Sciences*, 2004, **101(1)**, 5228–5235.
[29] S. Shivashankar, S. Srivathsan, B. Ravindran, A. V. Tendulkar, *Bioinformatics*, 2011, **27(13)**, 61–68.
[30] G. Hackeling, *Packt Publishing*, 2014.
[31] H. Mofid, M. R. Khosravi-Nikou, S. A. Razavi, M. Mohammadian, *Applied Thermal Engineering*, 2024, **245**, 1228101-17. https://doi.org/10.1016/j.applthermaleng.2024.122810
[32] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, Wiley-Interscience, 2000.
[33] Python-Bloggers, How to get started with Advanced Analysis with Python for Copilot in Excel. (2021). https://python-bloggers.com/2024/10/how-to-get-started-with-advanced-analysis-with-python-for-copilot-in-excel/.
[34] S. A. Razavi, H. Motavali, Journal of Industrial Engineering and *Management Studies*, 2023, **10(2)**, 106-115. https://doi.org/10.22116/jiems.2023.399150.1516
[35] S. A. Razavi, A. Tarnian, *Journal of Environmental Economics and Chemical Processes*, 2026, **3(1)**, 29–33. https://doi.org/10.22034/jeecp.2024.735000
[36] X. Zhang, D. R. Jeske, J. Li, V. Wong, *Computational Statistics & Data Analysis*, 2016, **94**, 238–249.