ICIC Express Letters
Part B: Applications
Volume **15**, Number **1**, January **2024**
 ISSN 2185-2766
pp. 51–60

MATERIALITY-BASED ONLINE COMPLAINT CLASSIFICATION: AN ANALYTICAL FRAMEWORK FOR EFFICIENT PUBLIC SERVICE USING TEXT MINING

Sehyoung Kim[1], Minju An[2], Hansol Lee[2] and Juyoung Kang[2,*]

[1]Department of Business Analytics
[2]Department of e-Business
School of Business
Ajou University
206 Worldcup-ro, Yeongtong-gu, Suwon 16499, Korea
{ sehyoung66; amj9898; lhs15 }@ajou.ac.kr; *Corresponding author: jykang@ajou.ac.kr

Received May 2023; accepted July 2023

Abstract. *This study presents a methodology for analyzing and processing online complaint data efficiently using big data analytics and text mining techniques. Inefficient complaint handling negatively impacts both complainants and government officials; however, existing studies have primarily focused on the complainant's perspective. For instance, the National Customer Satisfaction Index (NCSI) evaluates service quality in South Korea, although it lacks variables that consider the needs of government officials and employees who are responsible for handling complaints. This study aims to address this issue by clustering complaint data based on levels of dissatisfaction, specificity, and interest. The complaints are classified into three categories: high, medium, and low materiality. Subsequently, topic modeling techniques are employed to analyze the complaint topics based on their materiality levels. Finally, based on the findings of the analysis, an effective method for complaint handling is proposed. It is anticipated that by implementing the methods derived from this study to enhance complaint handling efficiency, both complainants and government officials will experience increased satisfaction. Additionally, using these methods in service complaint scenarios can alleviate stress among employees who are responsible for addressing such complaints and grievances.*
Keywords: Complaint, Text mining, Sentiment analysis, Clustering, Content analysis, Topic modeling

1. **Introduction.** Inefficient complaint handling negatively impacts both the complainant and the public servant who handles the complaint [1]. In addition, slow or unclear complaint handling can result in decreased satisfaction with the public administration [2]. To solve these problems, various studies have been conducted to increase the satisfaction of complainants. In particular, as the number of online complaints has increased, they have been utilized in various ways, such as utilizing them for usefulness and identifying improvements in products or services. Although the number of studies on resolving complaints and grievances has increased, most complaint studies have been limited to improving complainant satisfaction. There is no research on efficient complaint data management or the satisfaction of both complaint officials and complainants. Various problems, such as depression and health deterioration due to increased stress among complaint officials, have recently been reported in the media, emphasizing the importance of handling complaints more efficiently [3,4]. Therefore, we propose applying big data and text mining analysis techniques to complaint data and efficiently handling them to solve these problems. We clustered the complaint data based on the degree of dissatisfaction, specificity, and interest. Further, we categorized them into high, medium, and low based on their

DOI: 10.24507/icicelb.15.01.51

importance. Subsequently, we used the topic modeling technique to analyze which topics of complaints are related to each level of materiality. Using the analysis results, we propose a method for efficiently managing and handling complaints.

We expect that the results of this study will be applied in various fields. The National Customer Satisfaction Index (NCSI) is used to evaluate services in Korea. It includes perceived quality, customer expectations, perceived value, customer complaints, and customer loyalty, all of which are consumer-oriented variables [5]. In other words, employees working in service jobs or public officials handling complaints must render good service under all circumstances, which causes emotional exhaustion [6]. Excessive emotional exhaustion must be addressed because it increases employee stress and reduces work efficiency [7]. Consequently, this study proposes an effective method for processing complaint data using text mining methodology to solve this problem. Using the characteristics of the complaint data, such as length and the number of views, we clustered the data and derived the types of complaint problems that occur in each cluster. As a result of this study, the complaint data was categorized into three clusters, and the materiality level was added as high, medium, or low according to the characteristics and content of each cluster. Based on the results, we propose a method for efficiently handling complaints, and anticipate that the stress of employees who handle complaints and grievances can be reduced if methods to increase the efficiency of complaint handling are utilized in service complaints. We also believe this study can be used in various fields other than complaints in public administration.

This study is organized as follows: Chapter 2 describes the theoretical background of this study, and Chapter 3 introduces its methodology. Chapter 4 describes the results, whereas Chapter 5 concludes by discussing this study's contributions and limitations as well as the scope for future research.

2. **Literature Review.**

2.1. **Studies in complaint behavior.** Because consumer and user complaints reduce satisfaction with a service or product, research has been conducted to reduce complaints and increase satisfaction [8]. In the face of increasing complaints, Gao et al. [9] conducted a study to identify problems in complaint data and improve consumers' service satisfaction. Russell-Bennett et al. [10] studied the role and relationship of government third-party complaint agencies in improving consumer satisfaction. Tronvoll [11] investigated the impact of negative emotions on complaints and found that consumers complain when frustrated. Yilmaz et al. [12] examined the impact of customer response factors and organizational learning factors on the performance of gears as factors related to complaint management. The importance of studying complaints online has increased in recent years, particularly because anyone can view and verify complaints online [13]. Although surveys have limitations, such as low response rates and poor representation of consumer needs, they have traditionally been used to improve products or service [14]. Therefore, utilizing online complaint data can uncover consumer or user needs. Stevens et al. [15] proposed methods for effectively responding to online customer complaints. Dyussembayeva et al. [16] suggested conducting research to reduce and manage online complaints due to negative experiences, as these complaints discourage new users or consumers. These studies have focused on managing complaints or increasing complainant satisfaction. However, little research has been conducted to increase the satisfaction of both the complainant and the complaint official. We added a level of urgency to complaints in this study. We derived topics based on urgency to increase the satisfaction of complainants and complain officials' satisfaction by efficiently handling complaints. Unlike previous studies, which focused on the complainant only, this study is unique because it focuses on both the complainant and the complaint-handling official.

2.2. **Text mining approach in complaint data analysis.** Because online complaint data contains various consumers and users' opinions, including their complaints and needs, research has been conducted to analyze textual data on complaints and complaint reviews [17]. Much of this research has been conducted through topic modeling to discover key problematic situations in large amounts of complaint data. Topic modeling is a methodology for discovering meaningful topics in large volumes of documents and includes latent semantic analysis (LSA) and latent Dirichlet allocation (LDA) [18]. LSA is an analysis method that derives hidden meaning by reducing the dimensionality of the document term matrix (DTM) through singular value decomposition [19]. However, because LSA lacks information about the order or structure of words, LDA was developed [19,20]. LDA is the most popular topic modeling technique. It assumes that documents are composed of topics, which are represented by words [20]. LDA proceeds by estimating the probability that a word belongs to a particular topic and the probability of a particular topic in all documents containing the word as a combined probability. Shin [21] used topic modeling analysis to analyze the complaints of courier service workers and consumers. Zhang et al. [4] analyzed consumer complaints and perceptions in a sharing economy service using LDA topic modeling. Hu et al. [22] used structured topic modeling to analyze the complaints of consumers in hotels. In addition to topic modeling, studies have been conducted to analyze complaint data using various methods. Khedkar and Shinde [23] classified complaint data using machine learning-based models. Bastani et al. [24] studied the Consumer Financial Protection Bureau (CFPB) complaint data to analyze consumer complaints effectively. In this study, topic modeling for the type of complaint data was performed and applied to an expert system. However, this study only proposed an efficient system to provide information to experts and did not propose a method to process complaint data quickly. A recent study also proposed a complaint prediction model framework for effective government complaint management [25]. Previous research has attempted to automatically categorize complaints using predictive models [26]. However, this research is limited as it does not consider the content of the complaint data. Furthermore, this study is unique in that it divides the data according to the characteristics of the complaint data and proposes a method for effectively processing the complaint data based on the topic modeling results.

3. **Research Methodology.** This study proposes a method for efficiently analyzing and processing civil complaint data. Civil complaints are significant as they contain citizens' complaints and opinions. However, simple malicious complaints have a negative impact, such as stressing out civil servants and increasing citizen dissatisfaction if the processing of genuinely necessary complaints is delayed. Therefore, classifying civil complaint data by characteristics and suggesting solutions based on materiality through content analysis can increase the satisfaction of civil servants.

As shown in Figure 1, this study is conducted in three main stages. Step 1 involves data collection and preprocessing. We collect citizen complaints and perform preprocessing, such as date data processing and data type changes. After completing data collection and preprocessing, we create a dataset for analysis and proceed with EDA. Step 2 is to create a dataset and conduct EDA. Here, we conduct sentiment analysis and EDA. Sentiment analysis is performed on the complaint data to derive a sentiment score. In the case of complaint data, various opinions are included, and "sentiment" has a significant impact in addition to the content. Therefore, we added a sentiment score for full-scale data analysis in this study. Following the addition of the sentiment score, EDA is used to identify the characteristics of each variable. Finally, in Step 3, K-means clustering and topic modeling are used to cluster the complaint data according to its characteristics and analyze the contents. We cluster the complaint data using the degrees of dissatisfaction, concreteness, and interest and separate the data according to the clusters. We defined a

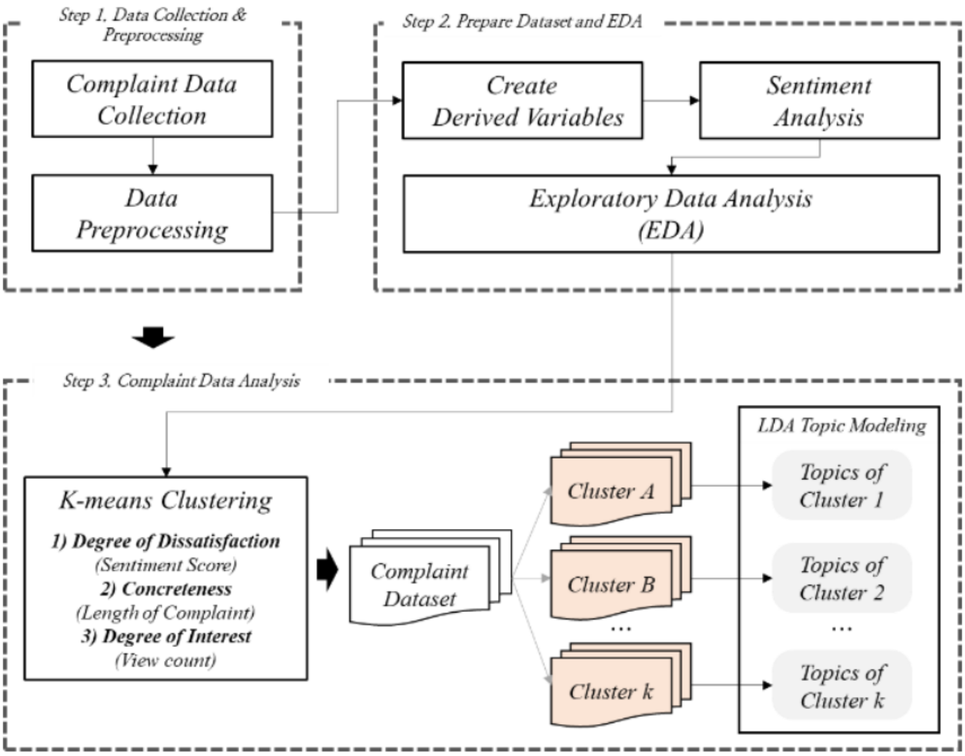

FIGURE 1. Research framework

"materiality" scale for each cluster's complaint data by analyzing the features of the degree of dissatisfaction, concreteness, and interest for these clusters. Next, for each cluster, we perform topic modeling to analyze which topics emerge. This allows us to analyze which complaints fall under which urgency level.

3.1. **Complaint data collection and preprocessing.** The data was collected from the "Saaeol e-complaint center in Gangseo-gu" using Selenium in python. Saaeol e-complaint is an online complaint service that enables citizens to use the Internet and mobile devices to file complaints and conduct inquiry consultations [27]. A total of 3,558 data points were collected, with the complaint data spanning the years 2008 to 2023. The variables of the collected data are the complaint's title, content, date, number of views, and whether it was answered shown in Table 1. We preprocessed the data after it was collected. First, we removed duplicate complaints. The criterion for duplicate complaints includes complaints with the same content that are posted because of late processing. Second, we determined how many characters were in the complaint. We calculated the number of characters and added it as a new variable because the more specific the complaint, the more characters it has.

TABLE 1. Characteristics of each complaint cluster

Attributes	**Description**
Title	Title of complaint
Content	Content of complaint
Date	Date of complaint
View_count	View count of complaint
Response_Info	Whether or not the complaint was answered

3.2. **Preparing the dataset and EDA.** To prepare the dataset for analysis, we conducted sentiment analysis and created derivative variables in this study. Firstly, we determined the degree of specificity in the complaint data by calculating the length of the complaint content. Because the length of the complaint varies, we included the length of the complaint data in our analysis. We also analyzed the complaint data to obtain a sentiment score. For this purpose, we utilized Naver Clova Sentiment. The sentiment score indicates the level of dissatisfaction expressed in the complaint. After completing the sentiment analysis, we conducted exploratory data analysis (EDA) to identify the characteristics of the variables to be analyzed. Figure 2 depicts the data used in this study, which ranges from 2008 to 2023.

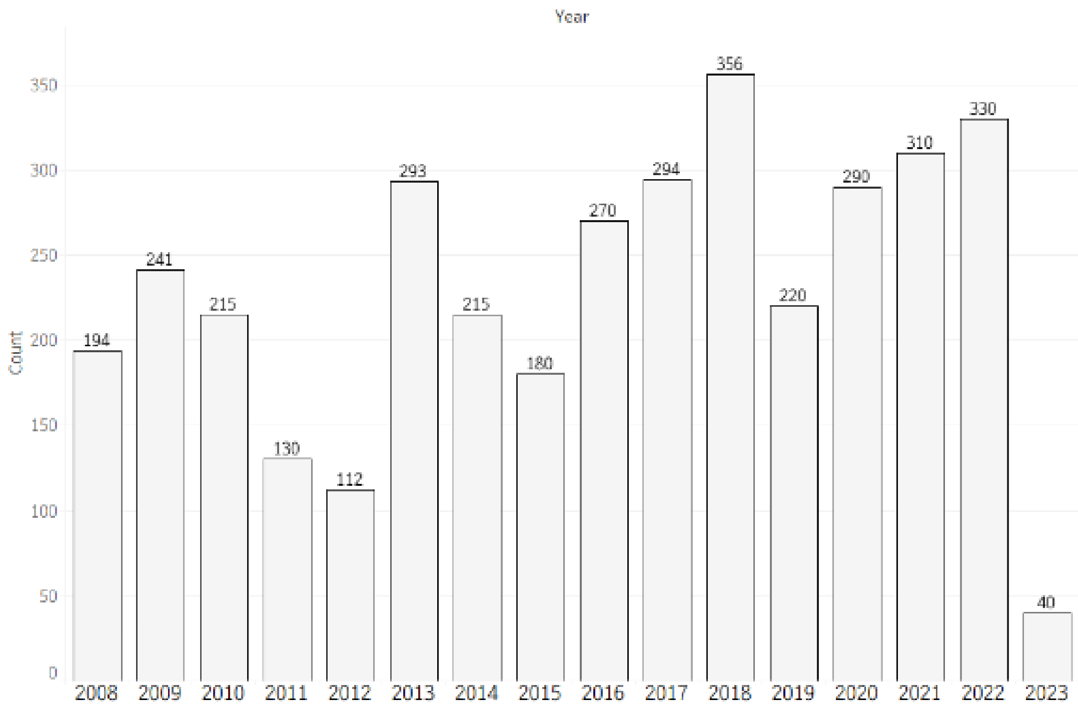

Figure 2. Number of complaints per year

The complaint data length indicates the number of texts in a complaint. A greater number of texts indicates greater specificity, whereas a lower number indicates a simpler complaint. This study used this length to assess the specificity of the complaint. Figure 3 represents complaints with higher concreteness on the right side of the "degree of concreteness" graph. The number of views on a complaint was used to indicate interest. More views signify a widespread issue that many people are aware of. This study used the number of views as a proxy for measuring the degree of interest in a complaint. The right side of Figure 3's "degree of interest" graph indicates data with a relatively high number of

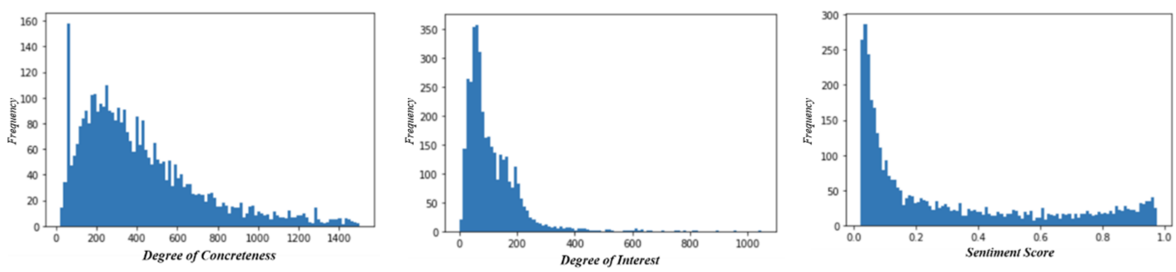

Figure 3. Degree of concreteness, degree of interest, and sentiment score

views, implying high interest. The sentiment score of a complaint was determined using sentiment analysis. Figure 3 depicts complaints with lower sentiment scores on the "sentiment score" graph as having high dissatisfaction. In order to evaluate dissatisfaction, this study used one minus the sentiment score as an indicator.

3.3. **K-means clustering.** Clustering was performed in this study using the "degree of concreteness", "degree of interest", and "degree of dissatisfaction" of the complaint data. The complaint data was clustered using K-means clustering, and the optimal number of clusters was determined using the elbow method. As a result of the analysis, the optimal number of clusters is three, as shown in Figure 4.

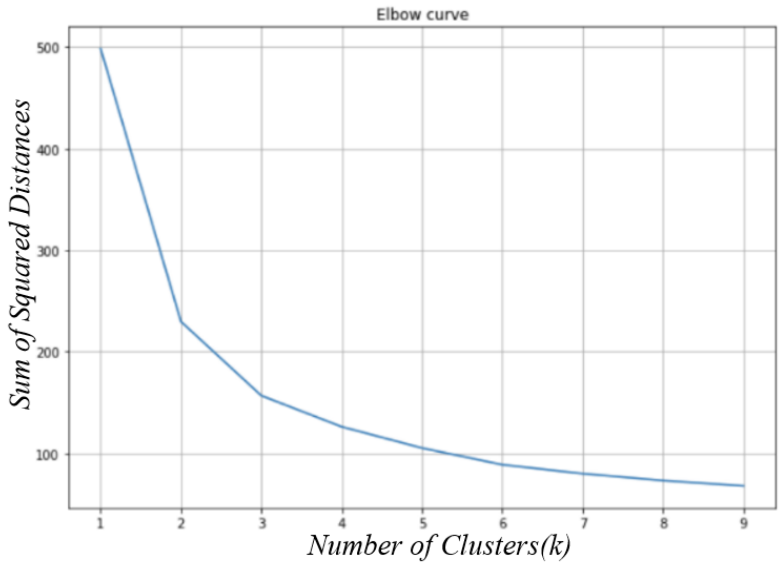

FIGURE 4. Optimal cluster identification

The characteristics of each cluster are shown in Table 2. As a result of clustering, the data per cluster is 1,992 in Cluster B, followed by 961 in Cluster A, and 605 in Cluster C. The cluster characteristics derived from the clustering result were used to derive the complaint handling materiality level. In this study, we propose clusters C > A > B in this order. Cluster C has the highest concreteness and interest, a high level of dissatisfaction, and is judged to have the highest materiality level. Cluster C has the lowest number of data (605); however, it is the most critical complaint cluster according to the characteristics of other variables. Cluster A, with a medium materiality level, has the second highest concreteness and interest and a low level of dissatisfaction. This cluster is characterized by the second-highest number of complaints and a low degree of dissatisfaction. Cluster B has the lowest materiality level and is characterized by the lowest concreteness, interest, and dissatisfaction. This cluster has the most complaints, at 1,992. Because simple complaints have the lowest materiality level and the highest number, techniques such as automation can reduce frustration for both the complainant and the official handling the complaint.

TABLE 2. Characteristics of each complaint cluster

Cluster	Count	Average concreteness	Average interest	Average dissatisfaction (1-sentiment score)
A	961	433.7	115.7	0.2
B	1,992	271.0	98.9	0.9
C	605	867.1	131.9	0.85

3.4. **Topic modeling.** After classifying the data into high, medium, and low materiality levels using the clustering results, we analyzed the topics of complaints according to each materiality level through topic modeling. Three rounds of topic modeling were conducted, with major topics derived for each cluster. For this study, we used latent Dirichlet allocation (LDA) topic modeling. In LDA topic modeling, the analyst must set the number of topics [24]. This study derived perplexity to derive the optimal number of topics as shown in Figure 5.

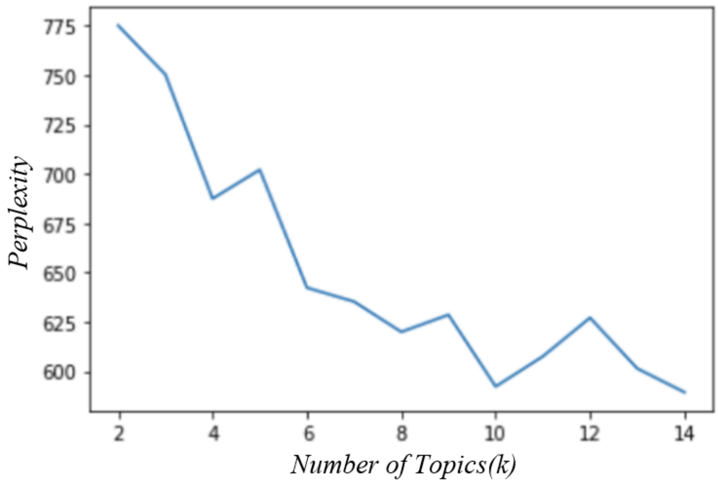

Figure 5. Optimal topic number identification

4. **Results.** Table 3 presents this study's high-materiality complaint topics, including facility use inquiries, additional sports facility requests, insect control complaints, parking problem resolution, and site development. These topics represent issues that require significant attention and have a high level of specificity and dissatisfaction. Collaboration between the responsible department and relevant parties, including higher-level departments, is crucial for providing specific solutions to these complaints.

Moreover, Table 3 identifies six major topics of medium materiality complaints, such as requests to enhance the traffic environment, resolve noise-related issues, address parking issues, modify bus routes, tobacco smoke complaints, and administrative service inquiries. These complaints have a relatively low level of dissatisfaction and can be efficiently addressed by resolving the underlying issues and providing the necessary information promptly.

Additionally, the table highlights seven topics of low materiality complaints, including the government's response to COVID-19, restaurant hygiene inquiries, the removal of illegally installed banners, minor illegal behaviors, illegal street vendors, grievances against administrative officials, and blocked drains. These complaints are typically straightforward and can be resolved quickly. Given their high dissatisfaction, low specificity, and low interest characteristics, automating the resolution process would be more efficient than relying on human agents.

5. **Conclusion.** This study proposes an effective method for handling online complaint data. From the complaint data, the "degree of concreteness", "degree of interest", and "degree of dissatisfaction" were derived. These variables were used to cluster, and the "materiality" variable was added based on cluster characteristics. Topics were identified based on their materiality level, and recommendations for effective handling were made. The proposed method aims to increase satisfaction for both complainants and complaint officials while improving the efficiency of complaint handling.

TABLE 3. Result of topic modeling

Materiality level	Topic	Keyword
High	Facility use inquiry	Welfare center, fire hall, open, use, facility, rental, weekend, Saturday
	Construction of sports facility	Tennis court, golf course, physical education, construction, building, permit, audit, facility, addition, basketball court, ...
	Bug complaint	Bugs, child, porch, neighborhood, happy, commute, space, greenery
	Request to solve the parking problem	Parking lot, public, construction, inconvenience, open, weekend, addition, construction
	Request for site development	Airport Boulevard, hope, development, residents, permission, country, request, thanks
Medium	Request for traffic improvement	Traffic light, traffic, sidewalk, pedestrian, complaint, grievance, crosswalk, signal
	Request to resolve noise	Noise, sound, damage, school, dormitory, morning, window, environment, resolution, office building
	Request for parking improvement	Parking lot, illegal, area, resident, bicycle, enforcement, space, sidewalk, Hwagok-dong
	Bus route request	Bus, route, line, change, bicycle, stop, village bus, increase
	Cigarette smell complaint	Cigarette, daycare center, playground, villa, house, cigarette smell
	Administrative service inquiry	Document, method, service, phone, application, health center, real estate, staff, reception
Low	Coronavirus response	Corona, quarantine, hospital, vaccine, mask, councilor, ignore, facility, enforcement
	Restaurant sanitation	Restaurant, sanitation, violation, complaint, eating out, audit, protest
	Banner removal	Banner, exposure, display, outdoor, citizen, protest, installation, removal
	Illegal behavior	Illegal parking, cigarette, smoker, illegal structure, enforcement, vehicle, unrest
	Street vendor complaint	Complaint of a street vendor, removal, accident, grievance, child, goods, elementary school, complaint
	Administrative complaint	Officer, inconvenience, grievance, government office, difficulty, complaint, stress
	Drainage complaint	Rain showers, land, repair, drainage, sidewalk, complex, passage, traffic

This study provides practical contributions by enabling efficient processing of complaint data and increasing satisfaction for complainants and officials. It offers a method to handle complaints efficiently while improving satisfaction by categorizing complaints based on their materiality level. Methodologically, text mining techniques were utilized to classify large amounts of complaint data and derive major topics. By categorizing frequently occurring complaints and proposing tailored handling methods, it improves efficiency over existing approaches. Additionally, this study has a social contribution by increasing satisfaction for both complainants and government officials who handle complaints. Unlike previous studies on complainant satisfaction, this study addresses the stress and depression experienced by the officials. Classifying complaints based on materiality and proposing corresponding solutions benefits both complainants and officials.

This study, however, has limitations. It only utilizes complaint data from Gangseo-gu, Seoul, which limits generalizability. Consequently, future studies should aim to collect and analyze data from various regions and apply the proposed framework to addressing this limitation. Furthermore, it is essential to validate the usefulness of this framework among practitioners, particularly civil servants who deal directly with complaints. Conducting research with their input and feedback would be valuable in further refining the proposed method.

Acknowledgment. This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2017-0-01637) supervised by the IITP (Institute for Information & Communications Technology Promotion).

REFERENCES

[1] S. S. Tax, S. W. Brown and M. Chandrashekaran, Customer evaluations of service complaint experiences: Implications for relationship marketing, *Journal of Marketing*, vol.62, no.2, pp.60-76, 1998.

[2] R. K. Al-Kiyumi, Z. N. Al-Hattali and E. R. Ahmed, Operational risk management and customer complaints in Omani banks, *Journal of Governance and Integrity*, vol.5, no.1, pp.200-210, 2021.

[3] X. Piao, J. Xie and S. Managi, Occupational stress: Evidence from industries affected by COVID-19 in Japan, *BMC Public Health*, vol.22, no.1, pp.1-12, 2022.

[4] N. Zhang et al., The impact of consumer perceived value on repeat purchase intention based on online reviews: By the method of text mining, *Data Science and Management*, vol.3, pp.22-32, 2021.

[5] C.-S. Lee, S.-H. Kim and J.-Y. Lim, The role of corporate image in the NCSI model: Experience from the KCSI study, *Journal of International Consumer Marketing*, vol.19, no.1, pp.7-34, 2006.

[6] C. K. Lam, X. Huang and O. Janssen, Contextualizing emotional exhaustion and positive emotional display: The signaling effects of supervisors' emotional exhaustion and service climate, *Journal of Applied Psychology*, vol.95, no.2, pp.368-376, 2010.

[7] L. Ramarajan, S. G. Barsade and O. R. Burack, The influence of organizational respect on emotional exhaustion in the human services, *The Journal of Positive Psychology*, vol.3, no.1, pp.4-18, 2008.

[8] K. Gelbrich and H. Roschk, A meta-analysis of organizational complaint handling and customer responses, *Journal of Service Research*, vol.14, no.1, pp.24-43, 2011.

[9] J. Gao et al., Research on user complaint problem location and complaint early warning stragegy based on big data analysis, *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (Trustcom)*, pp.1504-1509, 2022.

[10] R. Russell-Bennett, C. E. Härtel and J. Drennan, Consumer satisfaction and redress with a government third-party complaints agency, *The Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, vol.23, pp.65-82, 2010.

[11] B. Tronvoll, Negative emotions and their effect on customer complaint behaviour, *Journal of Service Management*, 2011.

[12] C. Yilmaz, K. Varnali and B. T. Kasnakoglu, How do firms benefit from customer complaints?, *Journal of Business Research*, vol.69, no.2, pp.944-955, 2016.

[13] Y. Cho et al., An analysis of online customer complaints: Implications for web complaint management, *Proc. of the 35th Annual Hawaii International Conference on System Sciences*, 2002.

[14] B. Yang et al., Exploiting user experience from online customer reviews for product design, *International Journal of Information Management*, vol.46, pp.173-186, 2019.

[15] J. L. Stevens et al., Timeliness, transparency, and trust: A framework for managing online customer complaints, *Business Horizons*, vol.61, no.3, pp.375-384, 2018.

[16] S. Dyussembayeva et al., It makes me feel vulnerable! The impact of public self-disclosure on online complaint behavior, *International Journal of Hospitality Management*, vol.88, DOI: 10.1016/j.ijhm.2020.102512, 2020.

[17] A. P. Kirilenko, S. O. Stepchenkova and X. Dai, Automated topic modeling of tourist reviews: Does the Anna Karenina Principle apply?, *Tourism Management*, vol.83, DOI: 10.1016/j.tourman.2020.104241, 2021.

[18] I. Vayansky and S. A. Kumar, A review of topic modeling methods, *Information Systems*, vol.94, DOI: 10.1016/j.is.2020.101582, 2020.

[19] T. K. Landauer, P. W. Foltz and D. Laham, An introduction to latent semantic analysis, *Discourse Processes*, vol.25, nos.2-3, pp.259-284, 1998.

[20] D. M. Blei, A. Y. Ng and M. I. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research*, vol.3, no.1, pp.993-1022, 2003.

[21] J. G. Shin, Analysis regarding complaints of courier consumers and workers in the parcel delivery service by using topic model, *Journal of Convergence for Information Technology*, vol.10, no.2, pp.39-48, 2020.

[22] N. Hu et al., What do hotel customers complain about? Text analysis using structural topic model, *Tourism Management*, vol.72, pp.417-426, 2019.

[23] S. Khedkar and S. Shinde, Deep learning and ensemble approach for praise or complaint classification, *Procedia Computer Science*, vol.167, pp.449-458, 2020.

[24] K. Bastani, H. Namavari and J. Shaffer, Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints, *Expert Systems with Applications*, vol.127, pp.256-271, 2019.

[25] S. Chen, Y. Zhang, B. Song, X. Du and M. Guizani, An intelligent government complaint prediction approach, *Big Data Research*, vol.30, 100336, 2022.

[26] B. D. Yelkenci, G.Özdağoğlu and B. İlter, Online complaint handling: A text analytics-based classification framework, *Marketing Intelligence & Planning*, 2023.

[27] C. Y. Lee and T. Kim, A study of the standard structure for the social disaster and safety incidents data, *Journal of the Society of Disaster Information*, vol.17, no.4, pp.817-828, 2021.