Section A-Research paper
**ISSN 2063-5346**

# TO SATISFY OR NOT TO SATISFY THE CUSTOMER: A MACHINE LEARNING PERSPECTIVE

[1] Gözdegül Duran, [2] Hilal Onbaşı, [3] Süleyman Hilmi Beşli, [4] Nazım Taşkın, [5] Mesut Tartuk, [6] Furkan Taha Nurdağ
[1], [2], [3] Student [4] Lecturer, [5] R&D Manager, [6] Project Manager
[1], [2], [3], [4] Boğaziçi University, MIS Department, [5], [6] Comdata Türkiye, R&D Department
[1] gozdegulduran@gmail.com, [2] hilalonbasi@gmail.com, [3]s.hilmi.besli@gmail.com,
[4]nazim.taskin@boun.edu.tr , [5]mesut.tartuk@comdatagroup.com, [6]furkan.nurdag@comdatagroup.com

***Abstract*— CRM systems have been popular as they enable organizations manage their relationship with their customers and the relevant business processes more effectively. This study focuses on developing an efficient way to manage customers' needs and complaints using machine learning techniques. We adapted two topic modelling techniques, LDA and GSDMM, to find out the main themes mentioned by customers with negative sentiment. LDA and GSDMM were used in a complementary way considering major limitations and advantages of the both technique. Using LDA, four topics emerged from the data. GSDMM algorithm was used to analyze the data using the same number of topics. The results showed that the most commonly topic discussed by the customers were "unsubscribe" and "login" in our case. Organizations using the same method can gather further information from the analysis and also customers to find out a solution to their current problem. Combining machine learning techniques like in this study can help organizations to realize the problems and develop solutions simultaneously when these analysis are integrated into their system.***

**Index Terms— Customer satisfaction, call center, machine learning, topic modeling, LDA, GSDMM, Bert**

## I. INTRODUCTION

Communication with the customers has become one of the most important aspects for companies today. With the spread of the Internet and increased number of e-businesses, customer-seller relations took different dimensions and new ways of communication taking places every day [4]. Although call centers still have a very high portion of customer communication, new communication channels such as social media, AI customer representatives and chatbots are gaining importance [19].

In order to meet the pre- and post-sale needs of customers, companies have had to adapt the new ways of communication. It is important for companies to understand what their customers in these channels are talking about and what their requests are [10]. At this point, one of the biggest challenges is that, especially for the companies with a large number of customers, extensive requests come to and through communication channels by customers every day. This communication with customers creates a huge amount of data for companies that need to be analyzed.

Understanding the content of customer messages give companies opportunity to analyze the effects of their actions on their customers. One of the recent approaches to understand this effect is to check the sentiment of the messages from customers [3]. For example, by looking at the messages with positive sentiments, companies can identify what they are doing right and keep improving their service. At the same time, they can identify customer complaints from messages with negative sentiment and work on required solutions. In that sense, urgent problems can be prioritized and resolved in a short time. If there is a system error encountered by many customers, it can be detected and corrected. Problems that customers encounter during the purchasing process can be resolved faster. Facilitating this process with emerging and relevant technologies like machine learning can increase the efficiency of the work [12]. This study suggests that machine learning can be leveraged at determining the topics of customer messages and taking faster appropriate actions while considering the emergency of the topic using sentiment analysis.

Section A-Research paper
**ISSN 2063-5346**

This study aims to improve the effectiveness of call centers by developing a new approach that uses machine learning techniques. These techniques can be used to identify customer messages mainly in two categories, positive and negative messages and then identify emerging topics in each group. First, the polarity of the messages will be determined using sentiment analysis. By doing so we will cluster the messages according to whether the customer's request has positive, negative or neutral meaning. Then, we will separate the negative messages to prioritize. In order to understand the topics that the customer groups are complaining about, additional topic modelling algorithms like Latent Dirichlet allocation (LDA) and Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM) will be used to determine the topics of the negative messages to be handled by the call center employees.

The rest of the paper is organized as follows: First the relevant literature is explained. Following section discusses the research designed adapted for the study. This section is followed by results and discussion. Finally, the conclusion section concludes the paper.

## II. LITERATURE REVIEW

Communication channels between customers and businesses have changed over time. The importance of written communication channels has increased as people prefer messaging more. According to the research conducted by Facebook IQ in 2018 [6], the importance of business messaging is increasing day by day. A recent study conducted by Facebook [6] across 15 different markets found that more than half of the people considered business messaging as the "modern way to communicate" (n.d.) in today's world. The participants also tended to consider such business messages as at least very good from effectiveness perspective compared to a conversation they would have over the phone with a company. Considering this perception about the messaging, customers would be more willing to communicate and shop with such businesses they find easier and effective to communicate. This also gives businesses a message that they should give more importance to their communication channels with their customers, especially the written ones, including social media. Furthermore, people and businesses exchanged over 20 billion messages on Messenger in a month in 2018. Statistics like Facebook conversation through chatbots growing more than 5 times in 2018 in USA [6] indicates that there is huge amount of data available for organizations to analyze.

According to Dixon, Freeman, and Toman [5] companies should focus on reducing the customer effort rather than trying to delight their customers with extras in order to increase customer loyalty. As explained by Dixon et al.[5] a Customer Contact Council conducted a study with a very large sample of about 75000 participants to examine the ways people interact with organizations through phone and other self-service channels provided by organizations. Study explaining the relationship between customer service and loyalty reported two major findings that can be important for organizations at strategy level. First, customer loyalty can be increased by providing ways to customers that will help them reduce their effort towards solving a problem rather than delighting the customers. One way to reduce customer effort could be improving the communication channels on which customers communicate. A system collecting and analyzing the messages properly and offering solutions to customers would help organizations to let their customer put less effort on the system and therefore improve their satisfaction and loyalty. Second, organizations can get benefits such as improved customer service, reduced service cost and even churn by addressing this finding [5].

Hendry et al. [11] examines the ways to improve the performance of chat bots and finds the intent of customer messages and how the customers should be responded effectively. Their study focuses on two tasks: first, to find new intents previously not known contents using the user messages through topic modelling; and second, organizing all the intents available by using topic modelling. In that sense, the authors compare LDA, Top2Vec and BERTopic approaches and adopt various metrics like "topic coherence, diversity, and quality" [11, p.1] to check the performance of LDA. Their results show that performances of BERTopic and Top2Vec algorithms were better compared to LDA model.

Literature provides a number of measures for evaluating performance of call centers such as number of calls answered within a specific time period, number of agents who are currently free, on call, or available and not available to take calls, etc. [9]. Also, for tracking quality of call center service, measures such as average time agents answer any call, quality of service provided to the clients and churn in the call center [7].

ISSN 2063-5346

As the employee churn is an important problem [21], in general and in call centers, organizations need to consider ways to reduce the churn. Reducing the churn and keeping the employees with tacit and explicit knowledge in the organization is part of any call center's success [21]. The automation that will come with using advanced algorithms and machine learning will reduce the unnecessary workload such as manual entries will improve the satisfaction of employees, leading to satisfaction of customers as well.

## III. RESEARCH DESIGN

### *A. Context: Call Center*

In this study, we analyzed transcribed Turkish chat data collected from a Turkish customer support center. Data consists of customer requests and request numbers. The data is used to understand and group the content of customer messages.

### *B. Data collection and preprocessing*

Raw data consisted of 502 user entered requests from clients of the support center. Collected data was preprocessed before conducting the analysis. First, we dropped the null values from the data set. Following this step, all characters in the remaining 497 entries were converted to lowercase. All the punctuation marks and stop words removed from the text. The final texts were tokenized before we use sentiment analysis and topic modelling algorithms [13]. Python was used for programming the topic modelling process. Furthermore, NLTK[1] was used for tokenization.

### *C. Sentiment Analysis*

Sentiment analysis is considered to be technique under the broad term of natural language processing (NLP). This technique is commonly used to determine if data, therefore participants' subjective state, has specifically positive, negative or a neutral sentiment. In business context, a common application of sentiment analysis is to check, monitor and understand the sentiment on their product, service or brand in general to be able to address and serve their customers' need better [15].

Sentiment analysis is used in this project to understand the polarity of customer's request and distinguish the messages especially with negative context. In this study, a BERT model BERTurk, which was developed for Turkish language, was used for sentiment analysis. BERT is a model trained with two techniques called Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). Unlike other models, BERTurk evaluates the sentence from both left-to-right and right-to-left, thus develops a better plan to extract the meaning and the relationships between the words more efficiently [20].

After the analysis, a new column called "Labels" was added to our data frame to show the results of the sentiment analysis. Since our primary goal is to identify and prioritize the customer complaints, we separated the negative labeled requests from positive ones.

### *D. Topic Modeling*

Topic modelling is "a collection of algorithmic approaches that seek to find structural patterns within a collection of text documents, producing groupings of words that represent the core themes present across a corpus" [16, p.1250]. Therefore, topic modeling can be used to find out the main themes within a set of textual data. In this research, we used Latent Dirichlet Allocation (LDA) and Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM) topic modelling algorithms, the two very powerful and relevant, to understand the topics of our customer messages and cluster them accordingly. We aim to find the most accurate topic modelling algorithm to cluster our documents.

We used the well-known topic modelling algorithm Latent Dirichlet allocation (LDA) which is a "generative probabilistic model of a corpus" [2, p. 996]. With LDA words comprise documents and the documents are considered as set of randomly selected latent topics [2]. For LDA analysis, we used Gensim[2] in the Python environment.

One of the challenges with using LDA model is that LDA does not find the number of optimum clusters. Parameters should be tried manually to find the optimum number. LDA also does not label the clusters by itself; it shows associated words under each group and makes it possible to assign the topic labels to each cluster by looking at their contents [8].

One of the advantage of LDA is that it is possible to visualize the results of the model and see the relations between topics and associated words. LDAvis provides this visualization on the data

[1] NLTK from www.nltk.org

[2] Gensim from www.pypi.org/project/gensim/

interactively showing the cluster of topics over the Internet [18].

In addition to LDA, another popular algorithm for text-based analysis, GSDMM was used in the analysis. GSDMM, a useful technique with especially short text data, uses the Gibbs sampling algorithm for a Dirichlet Mixture Model [22]. The main difference between LDA and GSDMM in terms of how they work is that LDA assumes that several topics come together to form a document while in GSDMM only one topic forms a document [14]. Accordingly, the contribution of the topics within each document are different in these two algorithms. Considering the full contribution of a single topic in a document, GSDMM can offer better performance using short text documents like tweets [17] but LDA allows handling large text documents better. Another important advantage of GSDMM is that unlike LDA, GSDMM requires identifying only the upper limit for the topics to be set for the analysis. On the other hand, LDA requires the exact number of topics to set before the analysis [17]. Last but not the least, GSDMM does not offer a tool or an easier way to visualize the data that make the results easier to interpret unlike LDA.

Considering the aforementioned pros and cons of the two algorithms and the nature of data, short text in the form of Tweets, we use GSDMM for the data analysis.

## IV. RESULTS

After applying sentiment analysis, a word cloud was generated to visualize and understand the statements better. As negative sentiments may usually mean a potential problem a customer faced using the service by the company, handling these messages urgently can help organization to keep their customers satisfied. Therefore, our results includes only the texts with negative sentiments.


**Figure 1**: Word cloud generated by negative sentences

As shown in Figure 1, the negative entries are mainly about cancelling membership or subscription, forgetting password, login errors, etc. These common words identified in the word cloud can be used to get some insights about possible topics before applying topic modeling.

### A. LDA Results

Since LDA could not determine the optimum number of groups by itself, we tried to find the optimum number by giving the algorithm with different number parameters and determine the most meaningful cluster groups with highest precision. Since we are working with limited data, we started to experiment with small numbers. In our research, we tried LDA with the given number of topics which range from k=2 to k=8. The results showed a steady decrease after trying with seven topics. As can be seen in Figure 2, the optimum number of topics with the highest coherence score, 0,6583, for our data set is four.

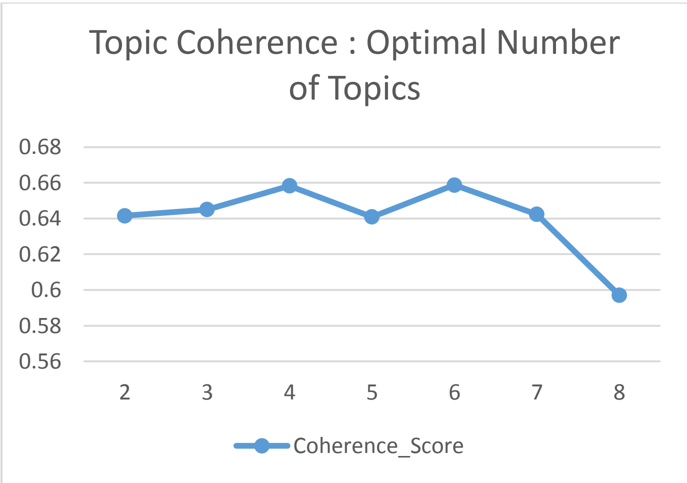


**Figure 2**: Coherence score by number of topics

By analyzing the terms under the topics carefully, four topics emerged and can be classified as "üyelik" (membership), "şifre" (password), "hesap" (account), "giriş" (login or access). The results can be found in Table 1.

**Table 1**: Topic labels from negative sentences

| No | Topic Label | Selected Words |
|---|---|---|
| t1 | Üyelik / Membership | Üyelik/ membership, iptal/ cancellation, abonelik/ subscription, talep/ request, geri/ back, aylık/ montly, ödeme/ payment |
| t2 | Şifre / | Üyelik/ membership, |

| | | |
|---|---|---|
| | Password | unuttum/ forgot, şifre/ password |
| t3 | Hesap / Account | İptal/ cancellation, hesabım/ account, silmek/ delete |
| t4 | Giriş / Login | Yapamıyorum/ cannot do, giriş/ login or access, hata/ error, surekli/ always, alıyorum/ receiving |

To visualize the topics identified, LDAVis, using Latent Dirichlet Allocation was adapted. Figure 3 demonstrates topics within the whole dataset visualized by LDAVis. The areas of the corpus reflect the proportion of each topic within the corpus whereas λ determines the weight of a terms' probability under a specific topic in relation to its lift and its optimum value is 0.6 [18].

Overall, the result of LDA by analyzing negative sentences in the dataset gave some insights about business problems. The results shows that topic t1 and topic t2 are overlapping to some extent, which can be understood from both the terms they have in common and how close their bubbles are (see Figure 3). We come into consideration that most of the visitors who struggle with payment related issues are actually the ones who were facing problems with their subscription as usually they forget to cancel their subscription when the free trial ended and made a payment without realizing. Second topic, t2 includes the terms "forgot" and "password" indicating that the requests under the t2 are coming from the visitor who forgot their password and struggle with resetting it. Topic t3 consists of visitor requests about account deletion and cancellation issues.. By looking at our last topic, t4, we found out that there are visitors who had constant problems with their login. Receiving messages from a large number of visitors with the topic4 in a short time period may indicate that there may be a system error. ..

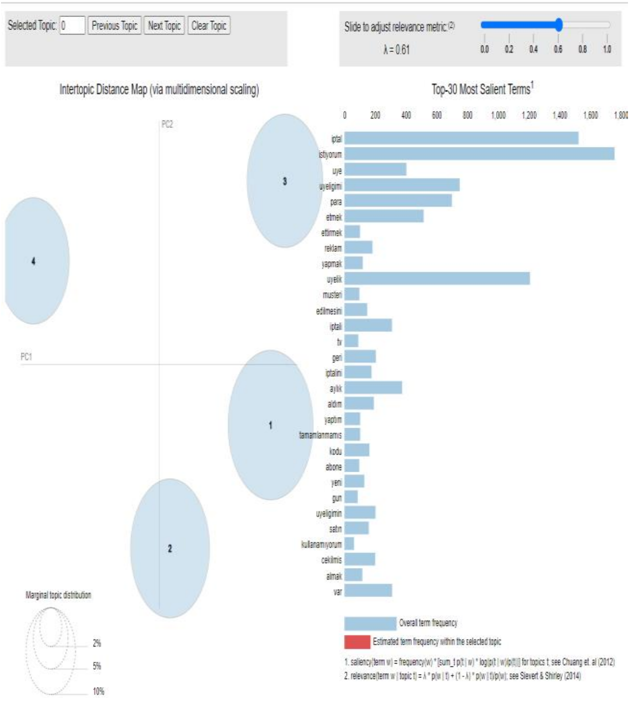


**Figure 3**: LDA topic modeling for dataset

*B. GSDMM RESULTS*

WITH analysis of GSDMM, four clusters of topic are populated. Coherence score of GSDMM algorithm resulted as 0.7781. Using the algorithm, four topics emerged from the data that can be classified as "üyelik" (membership), "giriş" (login or access), "hesap bilgisi" (account information), "ödeme" (payment). Table 2 shows the results of GSDMM algorithm.

**Table 2**: Topic labels from negative sentences

| No | Topic Label | Selected Words |
|---|---|---|
| t1 | Üyelik / Membership | Üyelik/ membership, iptal/ cancellation, üyeliğimi/ my membership, etmek/ to proceed |
| t2 | Giriş / Login or Access | Giriş/ Login or access, yapamıyorum/ cannot not do, sistemde/ system, sorun/ problem, error |
| t3 | Hesap bilgisi / Account information | Mail/ email, değiştirmek/ change, istiyorum/ request, bilgilerimi/ my information |
| t4 | Ödeme / Payment | Ödeme/ payment, iade/ return, aktif/ active, kart/ card, bilgi/ information |

The results of the GSDMM also support the findings of the results with LDA.

## V. DISCUSSION

With the onset and rapid growth of industrialization, clear changes took place in people's purchasing habits. People started to pay attention not only to the product or service they receive, but also to their relationship with pre-sales and after-sales customer service. In this direction, companies started to establish customer service units and spend more wisely to bring them to better positions compared to their competitors.

One of the most effective and easy ways to manage customer service is to use customer relationship management (CRM) systems. Developing CRM systems with machine learning applications will both improve the customer experience and make it easier for employees to help customers.

Companies handling call centers or social media sites play an important role in customer satisfaction. These departments or sections of companies are the front face of their companies in the virtual world. In addition to usual work done by relevant departments in organizations, these sections can contribute to customer satisfaction by helping the company understand their customers of public opinion about the company better through call center data or posts and tweets on social media. While all data can be analyzed to find out what customers and public say about the company, urgency can be given to negatively loaded messages from their customers and even from larger society. This study aims to come up with a new approach that will serve companies to understand their customers or public better and urgently so that they can develop strategies to handle these cases and improve customer satisfaction. In this study, machine learning was applied to call center data.

Following the sentiment analysis that helped us finding out the comments with negative sentiment, we applied two algorithms of topic modeling, LDA and GSDDM techniques to explore the common themes emerged from the data. At the beginning of the topic modeling, the optimum number of groups was set to four, which had the highest coherence score. According to the analysis made in this part of the study, it is concluded that "receiving login error" is the subject that customers complain the most, while "membership cancellation" and "account deletion" are the most requested subjects.

In order to find out any possible missing themes, another technique that works well with short text, GSDMM was used. When the highest number of topics in GSDMM was determined as four in order to be compared with the LDA model, the most common complaint was "receiving login error", and the most requested services were "subscription cancellation" and "change account information".

When the consistency scores of the two different models are examined, it is seen that the score of the GSDMM is higher than the LDA. This finding contradicts with the findings of Batra and Pramod [1] where their results of the GSDMM were more interpretable. This could be because of the sample size or nature of the data. Since LDA finds more than one topic in a text, the "password" issue seems to be higher up. However, when we look at the GSDMM results, we see that the "login" problems are more prominent. The highest demand in both techniques is “unsubscribe”. The results clearly show that the highest demand in two different subject modeling applications is related to the subject of "subscription cancellation".

Further investigation on the reasons for coming up with these themes will help organizations to improve satisfaction of their customers or manage the social media better. As seen from the results, the important problem experienced by the customers, which is seen as common in the results of the two applications, is related to the "login". The company can determine what kind of problems customers encounter at the entrance and find improvements that can be made in software or design.

## VI. CONCLUSION, LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

In order to develop CRM systems with machine learning, it is necessary to develop modules that recognize words, work with larger data, and integrate the analyzes into the CRM system. This study aimed combining three algorithms, sentiment analysis, LDA and GSDMM, to understand customer needs or complaints. Main contribution of the study is integrating three analysis to help finding out more on customers. Results showed that combining these analysis help, indeed, finding out the main themes mentioned by customers or public using social media. The study is not without limitations. The major limitation of the study is about the size of the data. While the sample size

Section A-Research paper
**ISSN 2063-5346**

used for analysis is limited as only data from one company was used, it still does give an indication on what the problem was with the company and their relationship with customers. However, a further study can definitely replicate this techniques and find better results. This study used BERTurk structure and analysis was conducted using data in Turkish language. The dictionaries used by the analysis are limited compared to English language dictionaries. Studies extending the tokenization using Turkish words would also improve the accuracy of future studies using the same or similar algorithms. In addition, for further research, noise can be eliminated as much as possible, more Turkish stop words should be added manually, lemmatization of words may be implemented. These further steps will result in better topic segmentation which eventually will cause higher coherence scores. Further studies contributing to development of Turkish dictionaries would definitely be better for studies using text in this language.

In addition, topics that are generated can be used into live systems which will help agents to assign visitors entry to relevant help desk agents by quickly looking at the topic of the text data without reading it all. For further steps, an automation can be done which will automatically assign tasks by topics without any need of human intermediary which will save time and reduce costs for call center agents and increase customer satisfaction due to faster response rates.

## REFERENCES

[1] R. Batra, and D. Pramod, "Exploring Customer comments using Latent Dirichlet Allocation," IEEE Pune Section International Conference (PuneCon), pp. 1-6, Dec. 2021.

[2] D. M. Blei, A. Y. Ng and, M. I. Jordan, "Latent dirichlet allocation," Journal of Machine Learning Research, vol. 3, pp. 993-1022, Jan. 2003.

[3] W. Chamlertwat, P. Bhattarakosol, T. Rungkasiri, and C. Haruechaiyasak, "Discovering Consumer Insight from Twitter via Sentiment Analysis," J. Univers. Comput. Sci., vol. 18, no. 8, pp. 973-992, 2012.

[4] A. Constantinescu-Dobra, and M. A. Coțiu, "Communication Channels," Communication Management. IntechOpen, 2021.

[5] M. Dixon, K. Freeman, and N. Toman. "Stop Trying to Delight Your Customer," Harvard Business Review, vol. 88, no. 7/8, pp. 116-122, 2010.

[6] Facebook, "Why messaging businesses is the new normal," 2018, Retrieved on June 9, 2022, from https://www.facebook.com/business/news/insights/why-messaging-businesses-is-the-new-normal#Business-messaging-has-gone-globa

[7] R. A. Feinberg, I.-S. Kim, and L. Hokama, "Operational Determinants of Caller Satisfaction in the Call Center," International Journal of Service Industry Management, vol. 11, no. 2, pp. 131–141, 2000.

[8] J. Gan, and Y. Qi, "Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example," Entropy, vol. 23, no. 10, pp. 1301, 2021.

[9] A. Gilmore, and L. Moreland, "Call centres: how can service quality be managed?," Irish Marketing Review, vol. 13, no. 1, pp. 3, 2000.

[10] G. A. Gorry, and R. A. Westbrook, "Can you hear me now? Learning from customer stories," Business Horizons, vol. 54, no. 6, pp. 575-584, 2011.

[11] D. Hendry, F. Darari, R. Nurfadillah, G. Khanna, M. Sun, P. C. Condylis, and N. Taufik, "Topic modeling for customer service chats," International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 1-6, Oct. 2021.

[12] S. Kumar, and M. Zymbler, "A machine learning approach to analyze customer satisfaction from airline tweets" Journal of Big Data, vol. 6, no. 1, pp. 1-16, 2019.

[13] S. Liu, and I. Lee "A hybrid sentiment analysis framework for large email data," 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), pp. 324-330, 2015.

[14] J. Mazarura, and A. De Waal "A comparison of the performance of latent Dirichlet allocation and the Dirichlet multinomial mixture model on short text," Pattern Recognition Association of South Africa and Robotics and Mechatronics

Section A-Research paper
**ISSN 2063-5346**

International Conference (PRASA-RobMech), pp. 1-6, Nov. 2016.
[15] MonkeyLearn, "What Is Sentiment Analysis?" 2022, Retrieved on November 17th, 2022, from https://monkeylearn.com/sentiment-analysis/
[16] M. J. Mortenson, and R. Vidgen "A computational literature review of the technology acceptance model," International Journal of Information Management, vol. 36, no. 6, pp. 1248-1259, 2016.
[17] R. Pelgrim, "Short-Text Topic Modelling: LDA vs GSDMM. Towards Data Science," 2021, Retrieved on November 18th, 2022. from https://towardsdatascience.com/short-text-topic-modelling-lda-vs-gsdmm-20f1db742e14
[18] C. Sievert, and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," Proceedings of the workshop on Interactive Language Learning, Visualization, and Interfaces, pp. 63-70, June, 2014.
[19] Statista, "Communication channels customers prefer to use to resolve customer service issues in the United States in 2020," 2020, Retrieved on June 9, 2022, from https://www.statista.com/statistics/818566/preferred-channels-customer-service-issues-united-states/
[20] K.T. Ucar, "BERT Modeli ile Türkçe Metinlerde Sınıflandırma Yapmak," 2020, Retrieved on June 9, 2022, from https://medium.com/@toprakucar/bert-modeli-ile-t%C3%BCrk%C3%A7e-metinlerde-s%C4%B1n%C4%B1fland%C4%B1rma-yapmak-260f15a65611
[21] M. Upal, "Telecommunication service gap: Call center service quality perception and satisfaction," Communications of the IBIMA, vol. 3, pp. 18-27, 2008.
[22] J. Yin, and J. Wang, "A dirichlet multinomial mixture model-based approach for short text clustering," Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 233-242, August 2014.