Topic-Based Segmentation in Email Marketing

Isti Surjandari[1, *], Harvey P. Gunawijaya[2,] Angella Natalia Ghea Puspita[1]
[1]Industrial Engineering Department, Faculty of Engineering, Universitas Indonesia, Indonesia
[2]Tokopedia, Jakarta, Indonesia

*Author to whom correspondence should be addressed:
E-mail: isti@ie.ui.ac.id

(Received October 21, 2023; Revised January 22, 2024; Accepted March 16, 2024).

Abstract: Email is one of the most common marketing channels businesses use to reach potential customers and advertise their products or services. To maintain the relevance of email messages and the performance of email marketing, personalization, or information from everyone is usually needed. This personalization is often derived from transactional data that reflects the behavior and interests of each customer. However, not all industries have enough transactional data for each of their customers. The challenge is how to use non-transactional data to discover customer interests and preferences. This study proposes to model customer preferences and build customer segmentation in email marketing based on topics and historical email interactions with customers. Biterm Topic Model (BTM) is used in this study for topic modeling because it is suitable for short texts such as email titles. Then segmentation is built based on the generated topics and customer email interaction history. Topic-based segmentation can increase the open rate by nearly two times on average. Marketers can use the findings of this study to develop new strategies and achieve higher email marketing performance.

Keywords: Biterm Topic Model, Customer Preferences, Email Marketing, Topic-based Segmentation

1. Introduction

The rapid development of technology disrupts almost every field of business. This technology has made many companies carry out digitalization processes. Digitalizing a company's business processes across all departments is called digital transformation. "Data" has emerged as the most significant digital asset of the companies as this process necessitates digitization [1)]. One of these fields of business that have digitalization is marketing.

Digital marketing methods such as email, social media, and push notifications have become critical in a brand's strategy to acquire and maintain customer loyalty. This digitization has proven that digital marketing activities are more effective and efficient in reaching the market than conventional marketing activities. Business enterprises are concentrating on improving the precision of marketing efforts in the context of digital technology to be more competitive and increase profit margins [2)].

Brands have become more creative in communicating with customers to build personal relationships. Various brands tailor their product and service recommendations based on customer activities and interests to provide more relevant suggestions to customers.

To better understand consumer behavior, personal relationships were presented as a metaphor for the links and interactions between customers and companies [3)]. According to [4)], providing recommendations using an interest profile can increase customer interaction on social media by up to 60%. The detected user behavior and purchasing patterns were used to produce recommendations; and this tactic boosts the efficacy and profitable economic position [5)].

Based on data from Statista, as many as 333.2 billion emails were sent daily in 2022 and are projected to increase to 392.5 billion daily emails by 2026 [6)]. Marketing using email (email marketing) is a profitable marketing channel, but it still needs more attention in the marketing literature [7)]. Recommendation and personalization of content in emails are important aspects of increasing interaction and performance of email marketing.

Many studies have successfully applied conventional topic models for email or topic modeling. For instance, a study shows that modeling with Latent Dirichlet Allocation (LDA) generates an 18% increase in clicks and a 10% increase in conversions [8)]. LDA can infer hidden implicit information from unstructured text data and explore potential topics contained therein [9)]. LDA has been used widely to explore topics in various fields, including opinion analysis from consumer reviews of hotels and ride-hailing service providers [10,11)].

Transaction data is widely used to build a more accurate customer personalization [12]. Customer behavior can be described by using many transaction data. The issue is that not all brands have the luxury of owning many transaction data.

Another data that a brand could utilize to create accurate personalization is email marketing data due to its purpose of maintaining customer retention [13]. Personalization is a fundamental idea in marketing, but it also has applications in other fields, including business management, computer science, decision science, information systems (IS), and psychology [14].

The problem with conducting analysis using transaction data is that the amount of data is large, and it is usually unstructured. For this reason, Machine Learning (ML) is used to assist in data processing to obtain hidden knowledge (including patterns) that can be used for personalization [15] and can automatically manage the trade-off between personalization and efficiency [16].

Marketers constantly monitor and improve the campaign's performance through email marketing. Most businesses that deal with business-to-business (B2B) and business-to-customer (B2C) transactions use email marketing extensively. According to a 2016 Ascend2 poll, 52% of B2B and B2C businesses use email technology as a component of their marketing strategy [17]. However, this email marketing ploy has grown widely employed in recent years, so recipients' behavior may alter as they grow accustomed to it [18].

Good performance must be upheld and could be achieved through customer personalization [19]. Recommendation and personalization are not the only aspects of email marketing content. Most email campaigns, such as newsletters, ad hoc information, and specific offers, are sent only once, making email marketing less effective in giving recommendations.

Studies also show that many customers believe email marketing is often irrelevant [20], which implies that many emails were sent without affecting the customers, resulting in cost waste and poor customer interaction that would negatively impact domain reputation and future email deliverability. According to [21], email marketing's lack of success is caused by overly general content and the inability to segment customers. These two aspects could enhance email marketing effectiveness.

Email marketing content could be developed using a segmentation technique to target clients, mainly based on marketing strategy, to give a more personalized and relevant message, resulting in a better email marketing experience. Although running a successful email marketing campaign does not ensure increased sales, the data indicates that it plays a significant role in spurring expansion. Revenue growth is considerably more likely to occur for email marketers who are purposeful about optimizing their email campaigns [22]. Customer segmentation is necessary to achieve better email marketing performance and cost efficiency.

Email marketing usually does not produce maximum performance due to a lack of customer segmentation [21]. Doing an email blast to all customer emails is ineffective marketing because, on average, only 15.7% of emails are opened by customers in the hospitality industry [23]. [20] also found that email marketing is often irrelevant. This shows that a lot of money is wasted by sending emails to inappropriately targeted customers. Low customer interactions with email marketing will also worsen the company's domain reputation, impacting the ability to send subsequent emails (email deliverability).

Customer segmentation in marketing can help decision-makers make accurate decisions to improve product marketing and achieve optimal email marketing performance and cost efficiency. Several rule-based segmentations, such as demographics, RFM (Recency, Frequency, Monetary Value), and CLV (Customer Lifetime Value), may need to be revised for customer segmentation due to the complexity of human interactions in the digital environment.

In today's digital and big data era, Machine Learning (ML) is widely used to segment customers more efficiently to increase digital marketing growth [24]. The Biterm Topic Model (BTM) is a Machine Learning method used to process data in the form of text.

Machine Learning applications are widely used in various fields, including manufacturing, agriculture, services, and healthcare. The use of ML in Supply Chain Management (SCM) by companies in India has been studied by [2]. [25] used ML algorithms to automate the rice grain sorting process. Meanwhile, [26] uses the ML method to design the Dielectric Resonator Antenna (DRA). In the field of optimization, [27] integrates metaheuristic methods (i.e., swarm optimization) with ML to improve the performance of data processing machines (machine management).

To maintain the relevance of email messages and the performance of email marketing, personalization, or information from everyone is usually needed. This personalization is often derived from transactional data that reflects the behavior and interests of each customer.

However, not all industries have sufficient transactional data from each of their customers. One example is the hotel industry, which has fewer transaction frequency than other industrial sectors such as retail and e-commerce. The challenge is using non-transactional data to discover customer interests and preferences.

According to [28], transaction data is widely used to build more customer personalization. Personalization requires many transaction data that describes the individual activities of each customer, where this data is very minimal in the hotel industry in terms of frequency. However, in the hotel industry, they use email as a marketing communication channel with a lot of content and frequent interaction. Communication via email is used to maintain relationships and offer various promotions.

Email is a promising electronic customer marketing tool for the hotel industry. It enables operators to send promotional materials or communications directly to target customers, increasing customer engagement and business success [29,30].

The hotel sector sends more emails than any other sector, and a more significant proportion of these emails arrive in consumers' inboxes instead of being flagged as spam [30,31]. Recent studies indicate that the hotel sector had an email open rate of 40% [30,32] and an opt-in rate of 30% [30,33]; also, [30] found that the hotel sector was in the top five of the 13 industries they studied.

This study proposes to model customer preferences and build customer segmentation in email marketing based on topics and historical email interactions with customers. Biterm Topic Model (BTM) is used in this study for topic modeling because it is suitable for short texts such as email titles. Then, segmentation is built based on the generated topics and customer email interaction history.

Topic-based segmentation can offer more information about customer interests or preferences to improve relevance. These preferences are represented in the previously modeled email topics using BTM [34]. The performance measurement of the segmentation is measured not only by the email open rate but also by how well the segmentation excludes emails to customers who have the potential not to open their emails.

This study focuses on selecting the accurate target customers for each campaign rather than on email marketers' topic recommendations. The sent email would be more relevant using topic-based segmentation because it would only reach customers whose interests align with the topic.

Previous studies performed topic modeling based on patterns of co-occurring words at the document level. In contrast, in short texts such as email titles, these word patterns will be rare in each document. The novelty of this study is that it builds segmentation based on topics from marketing content in the form of short text, such as email titles, and then combines it with historical interactions from each customer.

This study can help email marketers determine the right target customers for each campaign that will be delivered so that the emails sent will be more relevant and will only reach the customer's email inbox if the customer's interests and preferences match the email marketing topic.

2. Methods

2.1. Email Topic modelling

Segmentation using transactional data is not the best option if the required data is insufficient. Another alternative for segmentation is to use data obtained through email marketing. The easiest method is to harness customer interaction data to measure every customer's responsiveness.

Email marketing data such as subject, body, or even the metadata are the only available customer interaction data that can be utilized. This study only used textual data such as email subject since many modern emails marketing uses image or animation as the body email.

Several other studies use email subject data to generate email subjects automatically [21] and provide suitable keyword recommendations for email subjects [35]. As for textual data in email subjects, it has never been explored for segmentation purposes.

[25] used email subject data to identify and model the topic for each email campaign. The resulting topics could also be combined with historical customer interaction data to reflect customer interests and preferences. This study assumed that historical customer interaction data on email could represent customer interest in the content.

Just like segmentation, where there is no standard method for finding the best number of groupings [36], there is no standard rule for determining the optimal number of topics. One quantitative approach that can be taken is to measure the coherence score.

Based on [37], the coherence score is proven to measure the quality of the topics produced. A coherence score is calculated from the most important keywords for each topic. The higher the coherence score, the more coherent the extracted topic is. In determining the number of topics, the number selected is usually determined when the coherence score begins to slope as the number of topics increases.

For this study, data were obtained from one of the companies in the hotel industry. The data was collected from the email marketing campaign period from July 2019 to December 2020, which totaled 177 campaigns.

The number of email addresses observed is 500 thousand customer emails, with around 30 million emails sent and 3 million emails opened. All this interaction data is attached to each campaign.

Biterm Topic Model (BTM) is used because it works better for short text than most traditional modeling techniques, such as Latent Dirichlet Allocation (LDA). Although LDA is commonly used for topic modeling techniques, it is less suitable for short text data [38].

BTM is a method used for studying topics by modeling terms on the corpus containing short text documents [39]. [40] used BTM on short review texts to analyze online textual content, and they found that BTM worked better than LDA.

What is meant by "term" is any pair of words that can be made without regard to the order of the document. For example, the line "Today's special offer" would become {today, special}, {special offer}, and {today, offer} via extraction using BTM.

BTM is more suitable for short texts because the data limitation is mitigated using terms, resulting in more information than individual words. BTM models the appearance of items throughout the document at once and learns from patterns found throughout the corpus. This learning process distinguishes BTM as a better solution to the data sparsity problem than conventional topic modeling methods [41].

Data sparsity will be problematic if traditional models (such as LDA) are applied directly to short texts. BTM creates terms that co-occur throughout a corpus to aid topic learning and is used to identify subjects in the case of short textual material [42].

The flexibility to model different subjects present in a text (which is a typical linguistic feature in short texts) cannot be accommodated by BTM because BTM was developed to explicitly model the production of word occurrence patterns (i.e., bitterms) found in a collection of short texts [43].

BTM also tries to exploit the occurrence pattern of strong global words to infer hidden subjects. Through the creation of co-occurrence patterns of words (biterms), the BTM model is used to extract latent topics from Short Texts (STs) [44].

BTM can be used for grouping documents into different topic categories. The concept is like text clustering, in which words that have semantically similar meanings are clustered together. The topics generated act as a key for topic-based segmentation and will be used to represent the campaign topic. Every customer email will have a set of topics assigned to them based on their historical interaction data.

BTM assumes that the corpus is a mixture of various topics and the bitterm are drawn from a topic z independently. The probability that a bitterm is taken from a topic is calculated from the probability that both words in the bitterm appear in a topic. The joint probability of $biterm\ b = (w_i, w_j)$ is depicted in Equation 1.

$$P(b) = \sum_z P(z)\, P(w_i|z)\, P(w_j|z) = \sum_z \theta_z \phi_{\mathrm{iIz}}\, \phi_{j\mathrm{Iz}} \quad (1)$$

Unlike conventional topic modelling, BTM does not model individual documents because all documents in the corpus are aggregated into bitterms. BTM cannot directly get the proportion of topics during the topic search process. The assumption in BTM is that the topic proportion of a document is the same as the topic proportion of the bitterms produced by that document. This assumption determines the topic for each document d and is depicted in Equation 2.

$$P\,(z|d) = \sum_b P(z|b)P(b|d) \quad (2)$$

$P(z|b)$ calculated using Bayes' theorem, where $P\,(z|d) = \theta_z$ and $P(w_i|z) = \phi_{(i|z)}$ as in Equation 3.

$$P(z|b) = \left(\frac{P(z)P(W_i|Z)P(W_j|Z)}{\sum zP(z)\, P(W_i|Z)P(W_j|Z)}\right) \quad (3)$$

While $P(b|d)$ is taken from the estimated empirical distribution of bitterms in documents described in Equation 4.

$$P(b|d) = \frac{n_d(b)}{\sum_b n_d(b)} \quad (4)$$

where $n_d(b)$ is the frequency of bitterm b in document d. $P(b|d)$ is usually uniformly distributed across all biterms in the document because the text is short.

2.2. Topic Based Segmentation.

The study aims to aid email marketers in identifying the ideal customers or groups most likely to be interested and engaged with email. Every email would be linked to a specific topic of interest depending on the customer's interaction. Only emails previously proven to indicate customer interest are included in a topic segment.

In this study, the performance of the original campaign will be compared to specific rule-based segmentations that email marketers commonly use and topic-based segmentation resulting from this study. Table 1 shows the five types of segmentation to be compared.

The segmentation performance measurements not only measure the email open rate when using each type of segmentation but also see how well the segmentation excludes customer emails that have the potential not to open the email. In other words, expected segmentation is segmentation that selects customers who are predicted to open the email and excludes customers who are predicted not to open the email.

Table 1. Types of Segmentation used for performance simulation.

TYPES OF SEGMENTATION	
Type	**Description**
Original	All original recipients of the email
A	Active customer email in the last year
B	Type A but filtered further using email topic preferences
C	Active customer email in the last three months
D	Type C but filtered further using email topic preferences

The metrics used for comparison are accuracy, precision, and recall. Equations 5, 6, and 7 show the accuracy, precision, and recall calculations, while Figure 1 shows the confusion matrix.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

		Predicted	
		Negative	Positive
Actual	Negative	TN True Negative	FN False Negative
	Positive	FP False Positive	TP True Positive

Fig. 1: Confusion Matrix

Figure 2 shows the flowchart of steps taken in this study.

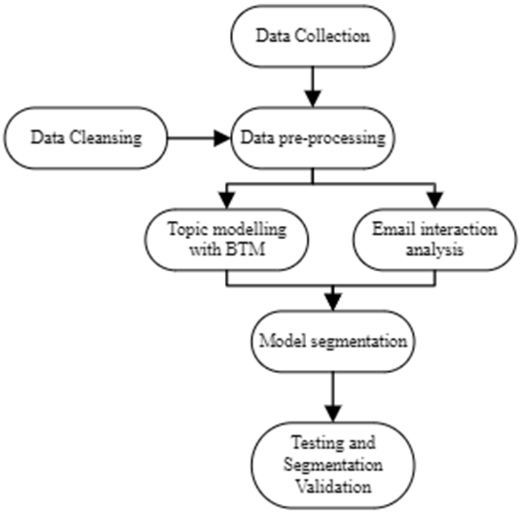

Fig. 2: Flow chart topic-based segmentation in email marketing

3. Results and Discussion

The simulation carried out in this study will compare the segmentation formed based on accuracy, precision, and recall. Accuracy shows how many correctly predicted users will open or not open an email. The higher the accuracy, the more customers are predicted to be correct whether they open or not the emails sent. High accuracy reflects the optimal volume of email sending, so that the remaining email sending quota can be used for other purposes.

Precision is the actual email open rate. Precision indicates how precisely segmentation can select customers who open emails compared to all customers in the segment. The higher the precision, the more customers in the segment are expected to open the email. In other words, the higher the precision, the higher the performance of the email open rate.

Recalls show how many customers in a segment opened the email. Sending emails to fewer customers means lowering the number of customers who can open the email. The higher the recall, the more customers open the emails in the segment.

Topic-based segmentation consistently outperforms other methods in terms of accuracy. The accuracy achieved by adding topic information has doubled in segment A. The average accuracy attained with topic-based segmentation is 85%, while the average accuracy through conventional segmentation is only 60%. Higher accuracy indicates that customers in that segment are likelier to open and interact with the email.

Figure 3 shows a comparison of the accuracy of each segmentation group. The original sections contain the original campaign data, which is guaranteed to be 100% accurate.

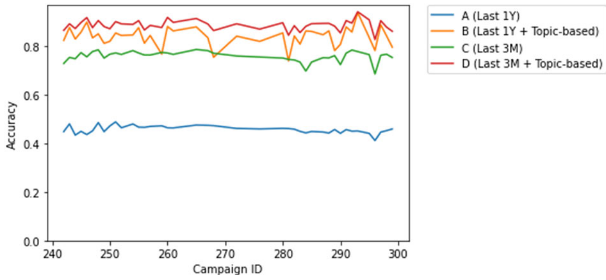

Fig. 3: Accuracy of each segmentation group

It has been demonstrated that adding information on topic interests and preferences can nearly double the performance of open email rates in terms of precision. An interesting finding revealed that active email user segmentation based on customers' preferred topic had a slightly lower email open rate in the last three months than active email user segmentation in the previous year. This result indicates that email marketing could attract fewer active users with a specific interest.

Figure 4 shows a comparison of precision for each segmentation. Topic-based segmentation reaches 55% in terms of recall. This score is lower than rule-based segmentation, which achieves over 90%. It is understandable, as more topic-based segmentation will result in fewer customer emails.

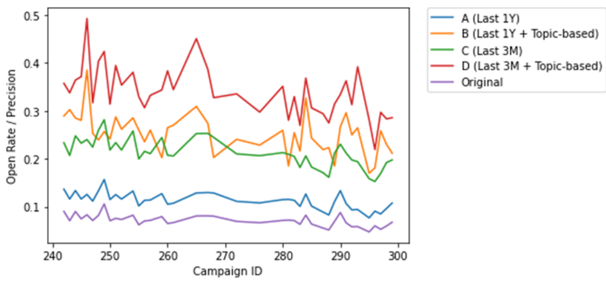

Fig. 4: Precision or open rate of each segmentation group

Figure 5 shows the recall comparison for each segmentation. Higher recall indicates that more customers who open the email are included in the segment.

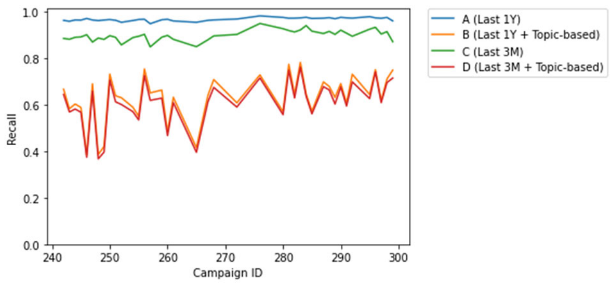

Fig. 5: Recall each segmentation group.

The original segment is the ground truth of campaign data to have a 100% perfect recall. Since the 40% decrease in recall from the rule-based segment is due to a reduction in the volume of emails sent, the average volume of emails sent for each segment must be monitored.

Figure 6 shows the average volume sent in each segment. Email marketing with primary rule-based segments has a substantially larger sending volume than topic-based segments. Topic-based segmentation has an average recall reduction of approximately 40%, with a nearly 70% decrease in volume sent from rule-based segments.

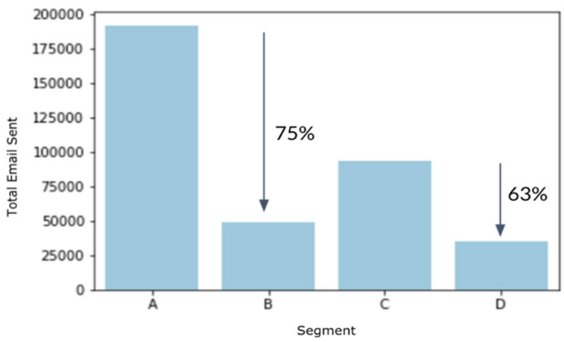

Fig. 6. Average volume sent for each segmentation group.

The proportion of the reduced volume sent is far more significant than the decrease in the recall, where the email open rate or precision attained is around two times higher.

4. Conclusion

In this study, topic modeling was proven able to produce topics that are unique from each other and build fairly accurate segmentation. The results of this study also show that lack of transactional data in certain industries does not prevent good personalization, where direct email marketing data can be used to find out customer preferences represented by topics.

Qualitatively, the topics generated from BTM are considered good by actual email marketers. Quantitatively, the resulting topics also produce good performance. The hit ratio for measuring the accuracy of these topics reached 69%, meaning most customers who opened the email had topic preferences that matched their historical interactions, which proves that customers have an interest and preference for certain topics.

Topic-based segmentation achieved a 25% greater accuracy on average, a 100% higher precision on average, and a 40% lower recall on average; this was achieved due to a 70% reduction in email volume. This segmentation can be further enhanced by using RF scores to prioritize subscribers more likely to open the email.

This study enables email marketers to understand their customers better and achieve a higher email marketing performance. The topics generated can also be used to develop future email marketing plans and strategies. Topic-based segmentation reduces email marketing costs and provides more relevance to customers, as proven by a higher level of interaction than rule-based segmentation.

Modeling topics in email marketing is a complex start problem requiring a large amount of historical campaign and interaction data. The average number of campaigns run on a large scale each month is between 5-7 campaigns. More accurate segmentation can be achieved by increasing the email marketing data available. The email content data can also be used to develop word recommendations for higher interaction.

This study only used data from one type of industry. A further study employing data from various industries could generate prospective topic preferences for cross-industry marketing.

Acknowledgements

The authors would like to express appreciation and gratitude to Universitas Indonesia for funding this study through PUTI Q2 Research Grants Universitas Indonesia [grant numbers NKB-1337/UN2.RST/HKP.05.00/2022].

References

1) R. Özsürünç, “The Role of Data Mining in Digital Transformation, ” *Multidimensional and Strategic Outlook in Digital Business Transformation*, 177-190 (2023). doi: 10.1007/978-3-031-23432-3.
2) R.H. Soto, F. Francis, M.A. Lopez, and J. P. Ramirez, “Implementation of Machine Learning in Supply Chain Management Process for Sustainable Development by Multiple Regression Analysis Approach (MRAA), ” *EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy*, 10(2), 1113-1119 (2023). doi: 10.2991/978-94-6463-124-1_58.
3) J. Story and J. Hess, “Segmenting customer-brand relations: beyond the personal relationship metaphor, ” *Journal of* Consumer *Marketing*, 23(7), 406-413 (2006). doi: 10.1108/07363760610712948.
4) Z. Zhao, Z . Cheng, L. Hong, and E.H. Chi, “Improving User Topic Interest Profiles by Behavior Factorization, ” *Proc. 24th International Conference on World Wide Web (WWW 2015),* 1406-1416 (2015). doi: 10.1145/2736277.2741656.
5) P. Rajapaksha and D. Asanka, “Recommendations to

Increase the Customer Interaction of E-commerce Applications with Web Usage Mining, ” *2023 International Conference on Contemporary Computing and Communications (InC4)*, 1-6 (2023). doi: 10.1109/InC457730.2023.10263131.

6) L. Ceci, “Number of sent and received e-mails per day worlwide from 2017 to 2026,” https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/ (accessed January 21, 2024).
7) X. Zhang, V. Kumar, and K. Cosguner, “Dynamically managing a profitable email marketing program,” *Journal of Marketing Research*, 54, 851-866 (2017). doi:10.1509/jmr.16.0210.
8) A. Noulas and M.S. Einarsen, “User Engagement through Topic Modelling in Travel, ” *Proc. 2nd Workshop on User Engagement Optimization*, 2-7 (2014). http://www.ueo-workshop.com/wp-content/uploads/2014/04/noulasKDD2014.pdf (accessed 2020).
9) D. Yu, A. Fang, and Z. Xu, “Topic research in fuzzy domain: Based on LDA Topic Modelling, ” *Information Sciences*, 648, 119600 (2023). doi: 10.1016/j.ins.2023.119600.
10) R. Annisa, I. Surjandari, and Zulkarnain, “Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation,” *Procedia Computer Sciences*, 161, 739-746 (2019). doi: 10.1016/j.procs.2019.11.178.
11) I. Surjandari, R.A. Wayasti, Zulkarnain, E. Laoh, A.M Masbar Rus, I. Prawiradinata, “Mining Public Opinion on Ride-Hailing Service Providers using Aspect-Based Sentiment Analysis,” *International Journal of Technology*, 10(4): 818-828 (2019). doi: 10.14716/ijtech.v10i4.2860.
12) Z.L. Bawn and R.P.D. Nath, , “A Conceptual Model for Effective Email Marketing,” *Proc. 17th International Conference on Computer and Information Technology (ICCIT 2014),* 250-256 (2014). doi: 10.1109/ICCITECHN.2014.7073103.
13) L. Chittenden and R. Rettie, , “An Evaluation of Email Marketing and Factors Affecting Response,” *Journal of Targeting, Measurement and Analysis for Marketing,* 11(3), 203-217 (2003). doi: 10.1057/palgrave.jt.5740078.
14) S. Chandra, S. Verma, W.M. Lim, and S. Kumar, “Personalization in personalized marketing: Trends and ways forward,” *Psychol Mark*, 39, 1529-1562 (2022). doi: 10.1002/mar.21670.
15) E. Egorov, F. Nikitin, V. Alekseev, A. Goncharov, K. Voronstov, “Topic Modelling for Extracting Behavioral Patterns From Transaction Data,” *Proc. IEEE 2019 International Conference on Artificial Intelligence: Application and Innovation (IC-AIAI)*, (2019). doi: 10.1109/IC-AIAI48757.2019.00015.
16) S. Tofangchi, A. Hanelt, D. Marz, and L.M. Kolbe, “Handling the Efficiency-Personalization Trade-Off in Service Robotics: A Machine-Learning Approach,” *Journal of Management Information Systems*, 38(1), 246-276 (2021). doi: 10.1080/07421222.2021.1870391.
17) A. Zhang, V. Kumar, and K. Cosguner, “Dynamically Managing a Profitable Email Marketing Program,” *Journal of Marketing Research*, 54(6), 851-866 (2017). doi: 10.1509/jmr.
18) L. Defau and A. Zauner, “Personalized subjects lines in email marketing,” *Marketing Letters* (2023). doi: 10.1007/s11002-023-09701-7.
19) "6 Ways to Improve Your Email Campaigns By Next Month," https://www.campaignmonitor.com/blog/email-marketing/6-ways-to-improve-your-email-campaigns-by-next-month/ (accessed Jnauary 21, 2024).
20) M. Hartemo, “Email Marketing in The Era of The Empowered Consumer,” *Journal of Research in Interactive Marketing,* 10(3), 212 - 230 (2016). doi: 10.1108/JRIM-06-2015-0040.
21) G. Singh, H. Singh, and S. Shriwastav, “Improving Email Marketing Campaign Success Rate Using Personalization,” *Advance in Analytics and Application*, 77-83 (2019). doi: 10.1007/978-981-13-1208-3_8.
22) “Validity and Demand Metric Report Proves Email Optimization Key For Revenue Growth: 80% of marketers report email marketing effectiveness is holding steady or increasing, according to 2019 State of Email Marketing Report,” *NASDAQ OMX's News Release Distribution Channel*, (2019). https://www.proquest.com/docview/2296047736/fulltext/EB5A1B34D6C74EC6PQ/1?accountid=17242 (accessed 2023).
23) “Ultimate Email Marketing Benchmarks for 2020: By Industry and Day,” https://www.campaignmonitor.com/resources/guides/email-marketing- benchmarks/ (accessed January 21, 2024).
24) C. Wang, “Efficient customer Segmentation in digital marketing using deep learning with Swarm intelligence approach,” *Information Processing and Management*, 59, 103085 (2022). doi: 10.1016/j.ipm.2022.103085.
25) K. Sujatha, N.P.G. Bhavani, V. George, T.K. Reddy, N. Kanya, and A. Ganesan, “Innovation in Agriculture Industry by Automated Sorting of Rice Grains,” *EVERGREEN Joint Journal of Novel Carbon Resource Science & Green Asia Energy*, 10(01). 283-288 (2023). doi: 10.5109/6781076.
26) P. Ranjan, S. Yadav, H. Gupta, and A. Bage, “Design and Development of Machine Learning Assisted Cylindrical Dielectric Resonator Antenna,” *EVERGREEN Joint Journal of Novel Carbon Resource Science & Green Asia Energy*, 10(01) 308-316 (2023). doi: 10.5109/6781085.

27) J. Grace L.K., A. Maneengam, P. Kumar, J. Alanya-Beltran, "Design and Implementation of Machine Learning Modelling through Adaptive Hybrid Swarm Optimization Techniques for Machine Management," *EVERGREEN Joint Journal of Novel Carbon Resource Science & Green Asia Strategy*, 10(2), 1120-1126 (2023). doi: 10.5109/6792811.
28) Z. Lian and R. Nath, " A Conceptual Model for effective email marketing," *Proc. 17th International Conference on Computer and Information Technology (ICCIT 2014)*, (2014). doi: 10.1109/ICCITechn.2014.7073103.
29) A. Marinova, J. Murphy, and B.L. Massey, "Permission e-mail marketing: as a means of targeted promotion," *Cornell Hotel and Restaurant Administrative Quarterly,* 43(1), 58-67 (2002). doi: 10.1016/S0010-8804(02)80009-X.
30) K. Yang, J.H. Min and K. Garza-Baker, "Post-stay email marketing implications for the hotel industry: Role of email features, attittude, revisit intention, annd leisure involvement level," *Journal of Vacation Marketing*, 25(4), 405-417 (2019). doi: 10.1177/1356766718814081.
31) A.S.Abraham, T. Chaudhary, and J.K. Deane, "A multi-industry, longitudinal analysis of the email marketing habits of the largest United States franchise chains," *Journal of Direct, Data and Digital Marketing Practice*, 11, 187-197 (2010). doi: 10.1057/dddmp.2009.31.
32) E. Brudner, "Email open rates by industry: see how you stack up," https://blog.hubspot.com/sales/average-emailopen-rate-benchmark (accessed 2018).
33) D. Chaffey, "Email marketing engagement and response statistics 2018," https://www.smartinsights.com/email-marketing/email-communications-strategy/statistics-sources-for-email-marketing/ (accessed 2018).
34) H. Gunawijaya dan I. Surjandari, "Modeling User Email Preferences Using Biterm Topic Model," *ACM 4th Asia Pacific Conference on Research in Industrial and System Engineering (APCORISE 2021)*, 348-352 (2021). doi: 10.1145/3468013.3468369.
35) A. Conceição and J. Gama, "Main Factors Driving the Open Rate of Email Marketing Campaigns," *Discovery Science*, 145-154 (2019). doi: 10.1007/978-3-030-33778-0_12.
36) I. Surjandari, R. Rindrasari, and A. Dhini, "Evaluation of Efficiency in Logistic Company: An analysis of Last-Mile Delivery," *EVERGREEN Joint Journal of Novel Carbon Resource Science & Green Asia Strategy*, 10(2), 649-657 (2023). doi: 10.5109/6792811.
37) D. Mimno, H.M. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," *ACM 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 262–272 (2011).
38) I. Vayansky and S. Kumar, ," A review of topic modeling methods," *Information Systems*, 94, 1-14 (2020). doi: 10.1016/j.is.2020.101582.
39) X. Yan, J. Guo, Y. Lan, and X. Cheng, "A Biterm Topic Model for Short Texts," *22nd International Conference on World Wide Web (WWW 2013)*, 1445-1456 (2013). doi: 10.1145/2488388.2488514.
40) Q. Ma, H. Du, and Z. Wang, " A Novel Perspective to Mining Online Hotel Reviews Based on Biterm Topic Model," *2021 IEEE 6th International Conference on Big Data Analysis (ICBDA)*, (2021). doi: 10.1109/ICBDA51983.2021.9403060.
41) J. Schneider and M. Vlanchos, "Topic Modeling Based on Keywords and Context," *SIAM International Conference on Data Mining (SDM 2018),* 369-377 (2018). doi: 10.1137/1.9781611975321.42.
42) H. Alhazmi, "Detection of Students' Problems in Distance Education Using Topic Modelling and Machine Learning," *Future Internet*, 14, 170 (2022). doi: 10.3390/fi14060170.
43) K. Zhang, Y. Zhou, Z. Chen, Y. Liu, Z. Tang, L. Yin, and J. Chen, "Incorporating Biterm Correlation Knowledge into Topic Modelling for Short Texts," *The Computer Journal*, 65(3), 537-553 (2022). doi: 10.1093/comjnl/bxaa079.
44) B.A.H. Murshed, J. Abawajy, S. Mallappa, M.A.N. Saif, S.M. Al-Ghuribi, and F.A. Ghanem, "Enhancing Big Social Media Data Quality for Use in Short-Text Topic Modelling," *IEEE Acces*, 10, 105328-105351 (2022). doi: 10.1109/ACCESS.2022.3211396.