Topic Modelling of Merdeka Belajar Kampus Merdeka Policy Using Latent Dirichlet Allocation

Sri Astuti Thamrin[1*], Nurul Rezki[2], and Siswanto Siswanto[3]
[1,3]Department of Statistics, Hasanuddin University, Makassar, Indonesia
[2]Data Analys, PT Bosowa Berlian Motor, Makassar, Indonesia
*Corresponding author: tuti@unhas.ac.id

Received: 12 June 2024
Revised: 3 September 2024
Accepted: 26 September 2024

ABSTRACT – **Topic modeling is the process of representing the topics discussed in text documents. In the current era of internet technology development, digital data is growing increasingly large, including tweet data from Twitter. This research aims to obtain topic modeling related to the Merdeka Belajar Kampus Merdeka policy on Twitter, which has been classified into positive and negative sentiments. The topic modeling method used is Latent Dirichlet Allocation (LDA). This method is for summarizing, clustering, connecting, or processing data from a list of topics. The data used in this research are tweets with the keyword "Kampus Merdeka" uploaded on Twitter. A total of 1579 tweets with these keywords were classified into 648 tweets and 931 tweets, respectively, with positive and negative sentiments. Each tweet with positive and negative sentiment produces 5 topics with parameter values α and β of 0.1. The coherence value in topic modeling for tweets with a positive sentiment (0.44) is more significant than for tweets with a negative sentiment (0.38) and represent for drawing conclusions about topics based on relationship between keywords in negative sentiment is more challenging compared to those in positive sentiment to the Merdeka Belajar Kampus Merdeka policy on Twitter.**

Keywords– **Topic Modeling, Latent Dirichlet Allocation, Coherence, Merdeka Belajar Campus Merdeka, Twitter.**

I. INTRODUCTION

The rapid development of internet technology at this time affects the increasing growth of the amount of digital data. Social media is one example of the largest internet data generator [1]. More than millions of people or organizations use social media as a form of their existence in cyberspace [2]. Establishing social relationships between users, sharing information and events will produce big data in real-time. As a social media, Twitter has become one of the ten most visited sites on the internet [3]. Twitter provides an application programming interface (API) that allows users to access and obtain information about tweets, user profiles, follower data, and other things. This makes Twitter a microblog that is in great demand by companies, organizations, and individuals in getting public opinion on a particular topic [4].

One of the issues that is currently being widely discussed on various social media and websites is the Merdeka Belajar Kampus Merdeka (MBKM) policy by the Ministry of Education and Culture, Research and Technology of the Republic of Indonesia. Based on an article released on January 29, 2020 by tirto.id, since its initial launch, this program has received many pros and cons from the public [5]. These reactions were conveyed through social media such as Twitter, Instagram, Facebook and so on.

Topic modeling is done to identify and map topics that appear in a particular opinion. Latent Dirichlet Allocation (LDA) is a popular topic modeling method used to summarize, cluster, connect or process data that produces a list of topics [6][7]. In a study conducted by Kherwa and Bansal [8] on topic modeling review, a comparison was made between Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA). The results showed that LDA produced higher coherence scores than LSA. Consequently, LDA is considered to be a highly effective tool for the statistical analysis of document collections [8]. Latent Dirichlet Allocation (LDA) in comment classification can significantly enhance the effectiveness of grouping comments into relevant categories or topics. This approach facilitates a clearer understanding of the underlying themes and subjects discussed by users in their comments [9]. The basic idea proposed by the LDA method is that each document is represented as a random mixture of topics, each topic has a character determined based on the distribution of words contained in it [10, 11]. Latent Dirichlet Allocation (LDA) requires defining the number of topics to be generated by the model [12]. Latent Dirichlet Allocation (LDA) allows the observation set to be explained by unobserved groups to explain the similarity of some parts of the data. The use of the LDA model makes it possible to obtain common topics from a document corpus [13, 14]. In 2021, a study on topic modelling using LDA on Twitter data with Indonesian keywords employed automated news analysis extracted from Twitter. The research successfully classified texts based on topics, which were then used to summarize, categorize, and process large datasets, generating a weighted list of topics for each document. This method can analyze vast volumes of documents, and the LDA approach ensures the accuracy of the generated topic model, both in terms of the topics and the words associated with them [15]. The purpose of this study is to conduct a modeling analysis related to the Merdeka Belajar Kampus Merdeka policy on Twitter. The LDA method is applied to group reviews given by the public related to the MBKM topic. Topic modeling identifies topics in reviews of the MBKM policy.

II. MATERIALS AND METHODS

The data used in this study are primary data obtained from Twitter in the form of Indonesian-language tweets with the keyword "Kampus Merdeka" uploaded from January 20, 2020 to March 31, 2022. The data of 1579 tweets used are in text form. The data consists of 648 tweets with positive sentiment and 931 with negative sentiment. The data structure used in this study after preprocessing the tweet text data consists of predictor variables, namely the basic words of each tweet and response variables, namely the tweet sentiment classification (positive and negative).

Data Structure

The data structure used in this study after preprocessing the tweet text data consists of predictor variables, namely the basic words of each tweet and response variables, namely the tweet sentiment classification (positive and negative). The classification data in the support vector machine method is divided into training data and testing data with a ratio of 80:20 using 10-fold cross validation. Table 1 shows an example of the study data structure before preprocessing.

Table 1 Example of Study Data Structure

No	Tweet	Sentiment
1	@Batutuo terus berjuang wujudkan kampus merdeka ...disain program2nya agar dihasilkan SDM yg berkwalitas di Era 4.0	Positif
2	Program Kampus Merdeka ini keren banget sih, ngasih harapan buat perubahan sistem pendidikan Indonesia	Positif
3	Ngiri saya sama program kampus Merdeka	Positif
⋮	⋮	⋮
1579	labelnya kampus merdeka, tapi sesungguhnya sivitasnya terjajahnya sama berlapis-lapis regulasi	Negatif

Analysis Stages

The stages of analysis carried out in this study are as follows:

1. Crawling tweet data using snscrape stored in csv format.
2. Performing manual labeling using doccano which is an open-source tool for data annotation.
3. Preprocessing text data consisting of data cleansing, case folding, spelling normalization, stemming, stop word removal and tokenizing.
4. Performing LDA topic modeling on the results of positive and negative sentiment tweet data.
5. Evaluating LDA topic modeling using the Point-wise Mutual Information (PMI) Topic coherence value.

Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a statistical method used as a model for analyzing documents that represents documents through a mixture of topics, each generated based on specific probabilities. These topic probabilities reflect the clarity and relevance of the content within a document. The basic idea proposed by the LDA method is that each document is represented as a random mixture of topics. Then each topic has a character that is determined based on the distribution of words contained in it. Latent Dirichlet Allocation (LDA) requires defining the number of topics that will be generated by the model [12]. How LDA works is described using a graph called a probabilistic graphical model.

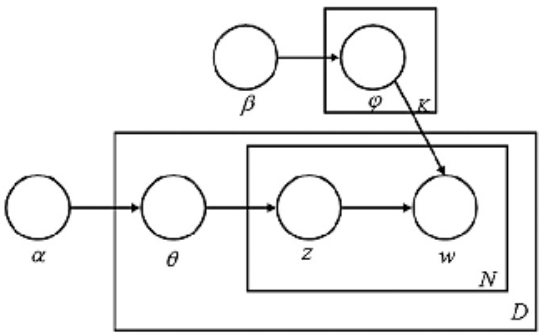

Figure 1. Probabilistic Graphical Model Work Process

In Figure 1, α and β are the topic distribution parameters of the document first or commonly called the Dirichlet prior parameters and the word distribution parameters of a topic. The values of both are positive real numbers that can be written as $0 \leq \alpha, \beta \leq 1$. The higher the value of α indicates that each document contains most of the topics and vice versa indicates that the document has the possibility of being represented by several topics. While the higher the value of β indicates that a topic contains a mixture of most of the words, conversely a topic only contains a mixture of several

words. The topic distribution of the document (α) results in the value of θ as a collection of topic mixtures in the form of a topic probability matrix for documents as in the following matrix:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\theta_{11}} & \cdots & \boldsymbol{\theta_{1K}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\theta_{D1}} & \cdots & \boldsymbol{\theta_{DK}} \end{bmatrix}.$$

$\boldsymbol{\theta_{DK}}$ shows the probability of the K[th] topic in the D-th document. From the collection of mixed topics (θ), each topic (z) can be separated from the mixture of topics. So that a new matrix is obtained containing the probability values of words against topics for each document as follows:

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{z_{11}} & \cdots & \boldsymbol{z_{1K}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{z_{N1}} & \cdots & \boldsymbol{z_{KN}} \end{bmatrix}.$$

The $\boldsymbol{Z_{KN}}$ value indicates the Kth topic in the N-th word. The probability of the topic obtained (z) and the distribution of words on the topic (β) produce the probability of words that appear as the final result of model formation (w), so that the results of this one-document model will produce words from the formed groups, these words can help in defining the categories of each group. So that the total probability based on the LDA model graph can be written mathematically as follows:

$$(\boldsymbol{w}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{d=1}^{D} \boldsymbol{P}(\boldsymbol{\theta_j}|\boldsymbol{\alpha}) \prod_{k=1}^{K} \boldsymbol{P}(\boldsymbol{\phi_k}|\boldsymbol{\beta}) \prod_{n=1}^{N} \boldsymbol{P}(\boldsymbol{z_{dn}}|\boldsymbol{\theta_j}) \boldsymbol{P}(\boldsymbol{w_{dn}}|\boldsymbol{\phi}, \boldsymbol{z_{dn}}),$$

with d and D indicating the document index, k and K indicating the topic index, n and N indicating the word index in the corpus and $\boldsymbol{\phi}$ is the corpus for the collection of documents that occurs due to the presence of β as in Figure 1 [16].

Evaluation of topics generated through LDA is done by looking at the topic coherence value that indicates the degree of cohesion among the words within a topic, derived from the analysis of semantic similarities and differences between the words in that topic [17], namely how easy the topic is to interpret using Point-wise Mutual Information (PMI) topic coherence as follows:

$$\boldsymbol{PMI(k)} = \sum_{j=2}^{N} \sum_{i=2}^{j-1} \boldsymbol{log} \left(\frac{\boldsymbol{P(w_i, w_j)}}{\boldsymbol{P(w_i)P(w_j)}} \right),$$

where N is the number of top words in topic k, $P(\boldsymbol{w_i}, \boldsymbol{w_j})$ is the probability of the i-th and j-th words appearing together in the document and $P(\boldsymbol{w_i})$ and $P(\boldsymbol{w_j})$ are the probability of the i-th and j-th words appearing in the document [18].

III. RESULTS AND DISCUSSION

Descriptive Analysis

Data were obtained from crawling results on Twitter with the keyword "kampus merdeka" uploaded on January 20[th], 2020 to March 30[th], 2022. The crawling results obtained were 1579 tweets in Indonesian. The results of the data collection obtained are shown in Table 2.

Table 2. Study Data Structure

No	Publication date	Tweet	Username
1	2020-12-26 15:52:49+00:00	Paket kebijakan kampus merdeka dg fokus transformasi pendidikan tinggi dg 8 IKU akan sia sia jika pandemi ini akan berjalan lama. Seharusnya Mas Mentri fokus melakukan transformasi pendidikan ke era kebiasan baru. Sudah 2 semester pembelajaran pandemi, tapi tidak ada perubahan	Novensupra
2	2020-11-29 04:43:16+00:00	Kampus merdeka katanya, tapi mahasiswanya dijajah tugas terus, mana merdekanya?https://t.co/4MR1BB119v lewat @ChangeOrg_ID	DraftAnakUnpad
	2021-01-08 12:20:44+00:00	@gurulesmtk Kampus merdeka ribet wkwkw	Liberalnanggung
⋮	⋮	⋮	⋮
579	2021-08-19 03:07:30+00:00	Haduh enak banget yg belum menjadi mahasiswa akhir skrg, program kampus merdeka extremely incredible to develop our competence. udah dapet pengalaman berharga,	Smwrsunny

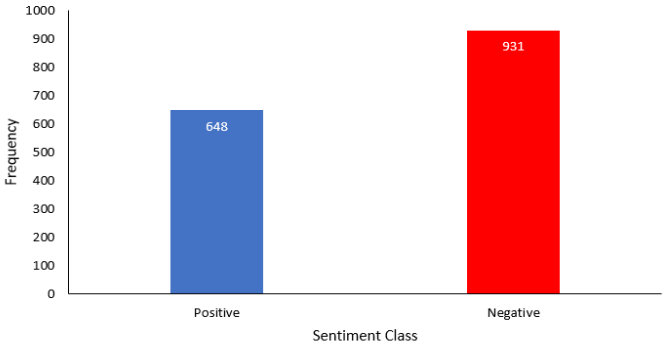

Figure 2. MBKM Policy Data Sentiment Class

Manual labeling was performed on the collected data. The data was labeled into two sentiment classes, namely positive and negative. Positive sentiment contains support for the MBKM policy, while negative sentiment contains rejection or distrust of the MBKM policy. Figure 2 shows the Bar Chart of the MBKM policy tweet sentiment class. Figure 2 shows that there is more negative sentiment data than positive sentiment data, which means that during the time span when the research data was collected, people uploaded more tweets on Twitter containing rejection or dissatisfaction than tweets containing support for the MBKM policy.

Data Preprocessing

Preprocessing is the process of preparing unstructured data into structured data so that it can be used for the next process [19]. Text data preprocessing is carried out on tweet data related to the collected MBKM policy. The text data preprocessing carried out consists of data cleansing, case folding, spelling normalization, stemming, stopword removal and tokenizing. Table 3 data structure before and after text data preprocessing.

Table 3. Data Structure Before and After Text Data Preprocessing

No	Data before preprocessing	Data after preprocessing
1	Paket kebijakan kampus merdeka dg fokus transformasi pendidikan tinggi dg 8 IKU akan sia sia jika pandemi ini akan berjalan lama. Seharusnya Mas Mentri fokus melakukan transformasi pendidikan ke era kebiasan baru. Sudah 2 semester pembelajaran pandemi, tapi tidak ada perubahan	[paket, bijak, kampus, merdeka, fokus, transformasi, didik, tinggi, sia, pandemi, jalan, lama, harus, menteri, fokus, lakukan, era, biasa, baru, semester, belajar, ubah]
2	Kampus merdeka katanya, tapi mahasiswanya dijajah tugas terus, mana merdekanya?https://t.co/4MR1BB119v lewat @ChangeOrg_ID	[kampus, merdeka, kata, mahasiswa, jajah, tugas]
3	@gurulesmtk Kampus merdeka ribet wkwkw	[kampus, merdeka, ribet, ketawa]
⋮	⋮	⋮
1579	Haduh enak banget yg belum menjadi mahasiswa akhir skrg, program kampus merdeka extremely incredible to develop our competence. udah dapet pengalaman berharga,	[aduh, enak, jadi, mahasiswa, akhir, program, kampus, merdeka, kembang, kompetensi, dapat, pengalaman, harga]

Topic Modelling

Latent Dirichlet Allocation (LDA) topic modeling is performed after the tweet data has gone through text data preprocessing. The trial-and-error method is used to determine the parameters α and β and the number of topics modeled. α and β are the topic distribution parameters of the document in advance or commonly called the Dirichlet prior parameters and the word distribution parameters of a topic [17]. The determination of the number of topics and the parameters α and β in the predicted tweets related to the positive sentiment MBKM policy are shown in Table 4.

Table 4. Parameter Determination in Positive Sentiment LDA Topic Modeling

α	β	Number of topics	Coherence	Increase in coherence value
0.1	0.1	5	0.44375	0.08826
	0.01	17	0.39846	0.03648
	0.001	8	0.39593	0.05577
0.01	0.1	5	0.42883	0.07303
	0.01	17	0.39737	0.05529
	0.001	8	0.40134	0.04467
0.001	0.1	5	0.42327	0.06747
	0.01	8	0.39987	0.04320
	0.001	17	0.38513	0.04880

Based on Table 4, the parameters α and β that show the highest coherence value are α and β of 0.1. Furthermore, Figure 3 shows the coherence value for the number of topics tested as many as 2 to 100 topics. The more topics modeled; the coherence value also tends to increase. However, the most significant increase in the coherence value occurs at the point of the number of topics 2 to 5 by 8.82% so that the number of topics modeled is 5 with parameters α and β of 0.1 which produces a coherence value of 0.44. The coherence value indicates the level of ease of interpretation of the modeled topic. The higher the coherence value, the easier each topic is to interpret or understand. Topic modeling on positive sentiment related to the MBKM policy produces a low coherence value so that the modeled topics are difficult to interpret or understand.

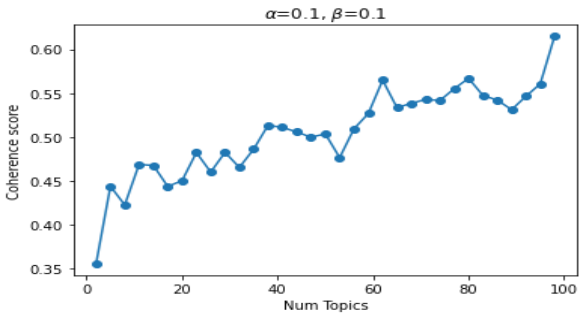

Figure 3. Coherence Value of Positive Sentiment Topic Modeling

The results of LDA topic modeling on positive sentiment related to the MBKM policy for 5 topics along with 5 keywords that contributed the most to the modeling of each topic are shown in Table 5.

Table 5. LDA Topic Modeling on Positive Sentiment of MBKM Policy

Topics	Key words				
	1	2	3	4	5
Topic 1	kampus	Merdeka	tertawa	Saku	Magang
Topic 2	merdeka	Kampus	dosen	Akademik	Lambat
Topic 3	merdeka	Kampus	mahasiswa	Uang	Ribet
Topic 4	anak	Mendikbud	Web	Stress	Ilmu
Topic 5	merdeka	Kampus	Mahasiswa	magang	Kuliah

The keywords are words that represent a topic with the first to fifth highest frequency of occurrence in all modeled documents or words that contribute the most to modeling a topic. In addition, the keywords represent the topic represented. If a topic modeling produces a high and easily interpreted coherence value, then the top keywords will be easily understood as related words and ultimately a topic can be concluded. However, LDA topic modeling on positive sentiment related to the MBKM policy produces a low coherence value so that concluding a topic based on the relationship of keywords is difficult. Table 4 shows that there are several words with a negative orientation that are keywords in the positive sentiment class topic modeling, namely "stress" and "complicated". This is influenced by the annotation error of the sentiment class.

The determination of the parameters α and β and the number of topics modeled were carried out through the trial-and-error method. The determination of the number of topics and the parameters α and β in the predicted tweets related to the positive sentiment MBKM policy are shown in Table 6.

Table 6. Parameter Determination in Negative Sentiment LDA Topic Modeling

α	β	Number of Topics	Coherence	Increase in coherence Value
0.1	0.1	5	0.38062	0.09092
	0.01	14	0.42701	0.05481
	0.001	5	0.38657	0.05156
0.01	0.1	5	0.37304	0.09661
	0.01	14	0.39859	0.05671
	0.001	5	0.36818	0.05929
0.001	0.1	5	0.37326	0.08644
	0.01	14	0.39683	0.05359
	0.001	14	0.39678	0.05141

Based on Table 6, the parameters α and β that show the highest coherence value are α and β of 0.1. Furthermore, Figure 4 shows the coherence value for the number of topics tested as many as 2 to 100 topics. The more topics are modeled, the coherence value also tends to increase. However, the most significant increase in the coherence value occurs at the point of the number of topics 2 to 5 by 8.82% so that the number of topics modeled is 5 with parameters α and β of 0.1 which produces a coherence value of 0.44. Topic modeling on negative sentiment related to the MBKM policy produces a lower coherence value than topic modeling on positive sentiment so that the topics modeled are also more difficult to interpret or understand compared to topics modeled on positive sentiment.

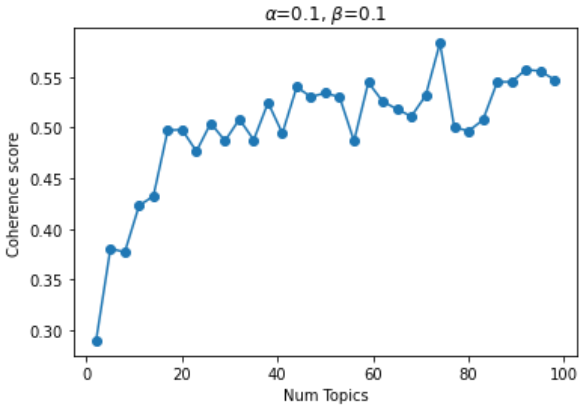

Figure 4. Coherence Value of Negative Sentiment Topic Modeling

The results of LDA topic modeling on positive sentiment related to the MBKM policy for 5 topics along with 5 keywords that contributed the most to the modeling of each topic are shown in Table 7.

Table 7. LDA Topic Modeling on Negative Sentiment of MBKM Policy

Topics	Key Words				
	1	2	3	4	5
Topic 1	merdeka	Kampus	mahasiswa	Bijak	sengsara
Topic 2	merdeka	Kampus	mahasiswa	Magang	tertawa
Topic 3	bayar	Mas	susah	menteri	penuh
Topic 4	merdeka	Kampus	banget	Magang	Keren
Topic 5	kampus	Merdeka	uang	Magang	Kuliah

Latent Dirichlet Allocation (LDA) topic modeling on negative sentiment related to MBKM policy produces lower coherence values compared to positive sentiment topic modeling so that concluding topics based on the relationship of keywords to negative sentiment is more difficult. Table 7 shows that there are words with a positive orientation that are keywords in the negative sentiment class topic modeling, namely "cool". This is influenced by sentiment class annotation errors.

IV. CONCLUSIONS AND SUGGESTIONS

Topic modeling on public reviews related to the MBKM policy by the Ministry of Education and Culture, Research and Technology using the Latent Dirichlet Allocation (LDA) method produced five topics. The number of topics determined has produced a collection of words that form a good topic, and have their respective percentages. Latent

Dirichlet Allocation (LDA) topic modeling on positive and negative sentiments related to the MBKM policy on Twitter, each with five topics and parameters α and β of 0.1 with a coherence value of 0.44 in positive sentiment topic modeling and 0.38 in negative sentiment topic modeling. Consequently, drawing conclusions about topics based on relationship between keywords in negative sentiment is more challenging compared to those in positive sentiment. The information generated from the five topics generated in this study can be used as evaluation material to improve services related to MBKM. In the future study, it is recommended to estimate the parameters α and β in LDA topic modeling, utilize alternative evaluation methods, such as employing both coherence and prevalence value and perform a comparison between LDA and LSA topic modelling.

REFERENCES

[1] V. Dhawan and N. Zanini, "Big Data and Social Media," *Research Matters: A Cambridge Assessment Publication*, vol. 18, pp. 36–41, 2014.

[2] U. Sivarajah, Z. Irani, S. Gupta, and K. Mahroof, "Role of big data and social media analytics for business to business sustainability: A participatory web context," *Industrial Marketing Management*, vol. 86, pp. 163–179, Apr. 2020, doi: 10.1016/j.indmarman.2019.04.005.

[3] A. M. Zuhdi, E. Utami, and S. Raharjo, "Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN," *Jurnal Informa: Jurnal Penelitian dan Pengabdian Masyarakat*, vol. 5, no. 2, pp. 2442–7942, 2019.

[4] T. Kurniawan, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naive Bayes Classifier dan Support Vector Machine," Institut Teknologi Sepuluh November, Surabaya, 2017.

[5] H. Prabowo, "Pro dan Kontra atas Kebijakan 'Kampus Merdeka' Nadiem."

[6] P. Madzík, L. Falát, and D. Zimon, "Supply chain research overview from the early eighties to Covid era–Big data approach based on Latent Dirichlet Allocation," *Comput Ind Eng*, 2023.

[7] P. Madzik, L. Falat, L. Jum'a, M. Vrábliková, and D. Zimon, "Human-centricity in Industry 5.0–revealing of hidden research topics by unsupervised topic modeling using Latent Dirichlet Allocation," *European Journal of Innovation Management*, 2024.

[8] P. Kherwa and P. Bansal, "Topic Modeling: A Comprehensive Review", *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 24, pp. 1–16, 2020.

[9] D.K. Bustami and S. Noviaristanti, "Service Quality Analysis of Tokopedia Application Using Text Mining Method", *International Journal of Management, Finance and Accounting*, vol. 3, no. 1, pp. 1–21, 2022.

[10] S. Zhou, P. Kan, Q. Huang, and J. Silbernagel, "A guided latent Dirichlet allocation approach to investigate real-time latent topics of Twitter data during Hurricane Laura," *J Inf Sci*, vol. 49, no. 2, pp. 465–479, 2023.

[11] P. Madzík, L. Falát, and D. Zimon, "Supply chain research overview from the early eighties to Covid era–Big data approach based on Latent Dirichlet Allocation," *Comput Ind Eng*, 2023.

[12] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Indonesian of Health Information Management Journal (INOHIM)*, vol. 8, no. 2, pp. 100–109, 2020.

[13] I. M. K. B. Putra and R. P. Kusumawardani, "Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA)," *Jurnal Teknik ITS*, vol. 6, no. 2, pp. 311–316, 2017.

[14] D. Marutho and N. A. Setiyanto, "Comprehensive Exploration of Machine and Deep Learning Classification Methods for Aspect-Based Sentiment Analysis with Latent Dirichlet Allocation Topic Modeling," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 1, 2024.

[15] K.H. Musliadi, H. Zainuddin and Y. Wabula, "Twitter Social Media Conversion Topic Trending Analysis Using Latent Dirichlet Allocation Algorithm", *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 4, no. 1, pp. 390–399, 2022.

[16] F. Gurcan, O. Ozyurt, and N. E. Cagitay, "Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation," *The International Review of Research in Open and Distributed Learning*, vol. 22, no. 2, pp. 1–18, Jan. 2021, doi: 10.19173/irrodl.v22i2.5358.

[17] J. Stolee, "An Evaluation of Topic Modelling Techniques for Twitter", *Research Paper*, pp. 1-11, 2016

[18] L. Yao *et al.*, "Incorporating Knowledge Graph Embeddings into Topic Modeling," 2017. [Online]. Available: www.aaai.org

[19] J. Ipmawati, Kusrini, and E. T. Luthfi, "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen," *Indonesian Journal on Networking and Security*, vol. 6, no. 1, pp. 28–36, 2017.