

## ANALYSIS ENGAGEMENT OF INSTAGRAM VISITORS AT UNIVERSITY OF MULTI DATA PALEMBANG BASED ON TOPIC USING LDA

Nathaniel Jerin Sisanto<sup>1\*</sup>, Muhammad Rizky Pribadi<sup>2</sup>, Rusbandi<sup>3</sup>

<sup>1,2,3</sup>Universitas Multi Data Palembang, Indonesia

[nathanieljerin07@mhs.mdp.ac.id](mailto:nathanieljerin07@mhs.mdp.ac.id)<sup>1\*</sup>

[rizky@mdp.ac.id](mailto:rizky@mdp.ac.id)<sup>2</sup>

[rusbandi@mdp.ac.id](mailto:rusbandi@mdp.ac.id)<sup>3</sup>

Received: 30-10- 2023

Revised: 01-11-2023

Approved: 02-11-2023

### ABSTRACT

*Social media has a big impact on everyday life, one of which is to communicate or to get information. Therefore, the development of social media applications makes people use social media applications to find information via the internet. The Instagram applications is one of the most popular social media because it has different topics based on post in the form of images or videos. Therefore, it is very difficult to identify a topic manually. One way to get implied information on social media is through topic modeling. This research was conducted to analyze the application of the LDA method to identify what topics are on Instagram at Multi Data Palembang University. The topics chosen in this study were obtained from LDA based on coherence values. This research uses 2 models, namely random forest and decision tree. Each model tested will produce different accuracy, precision, recall, and f1-score values. Tests were carried out on the LDA labeling dataset and manual labeling, the test results on the LDA labeling dataset were very good using the random forest model with an accuracy values of 78%, precision 80%, recall 66.66%, and f1-score 72.72%.*

### INTRODUCTION

Over time, today's technology users have had a big impact on everyday life, one of which is the development of information and communication technology. The growing development of information and communication technology makes people use social media to search for information via the internet with smartphones. Social media is a digital platform that offers every user the opportunity to engage in social activities. Several activities can be carried out on social media, such as communicating or interacting to provide information or content in the form of text, photos and videos. One of the digital platforms or social media that is widely used to receive information is Instagram [1].

Instagram can also be interpreted as an application that acts as a means of sharing photos and videos on social networks, allowing users to take photos and videos and add filters that give interesting effects to photos. Users can also provide negative and positive opinions or comments when discussing a topic that is currently trending [2]. Instagram is also often used for advertising media because instagram has features that support product marketing very well. The magnitude of this opportunity has fueled the rise of electronic accounts or online stores that sell various products on Instagram. Online stores are virtual stores where buyers and sellers cannot meet in person, so customer engagement is needed [3]

Customer Engagement is an emotional interaction between customers and companies formed by motivation, referrals and customer experiences through creative social media posts from brands that influence buyers [4] Therefore, user

engagement on social media is very influential, so that social media posts or accounts becomes popular and tends to attract more attention, if it generates higher engagement indicators for that post or account. So that the role of digital in higher education is very important to recruit prospective students and promote campus facilities or campus activities.

There are several universitas in Palembang, namely Sriwijaya University, Palembang PGRI University, Palembang Multi Data Palembang, Musi Charitas Catholic University, Tamansiswa University, Kader Bangsa University, Tridinanti University 3 Palembang, IGM University, Bina Darma University, Palembang Muhammadiyah University, and Islamic University Country [5]

There are many Universities in South Sumatra, the outhor chose Multi Data Palembang University to carry out the topic modeling method on Instagram Universitas Multi Data Palembang University. The researcher chose Multi Data Palembang University because Multi Data Palembang University was ranked second in the webimerics version of the university ranking in SOUTH SUMATRA July 2021 and the Multi Data Palembang University had 11.7 thousand followers [6], but likes and comments on Instagram posts of Multi Data Palembang University were very not in accordance with followers, so researchers choose Instagram social media at Multi Data Palembang University to do the modeling topic.

Topic modeling is a machine learning technique used to group documents into related topics or themes. The modeling topic includes many methods that can be used, such as Latent Semantic Analysis (LSA), Probabilistic Letent Semantic Analysis (PLSA), Latent Dirichlet Allocation (LDA). Latent Dirichlet Allocation (LDA) is an enhanced mixture model to capture word exchange and legacy document systems over PLSA and LSA. The most popular topic modeling method today is Latent Dirichlet Allocation (LDA). Latent Dirichlet Allocation (LDA) is a probabilistic algorithm that models each document as a combination of different topics and the distribution of words from each topic in the document. The Latent Dirichlet Allocation (LDA) algorithm tries to find the topics or themes that appear most frequently in documents, then determines how often certain words appear in those topics. The Latent Dirichlet Allocation (LDA) algorithm can be applied to various types of text data, such as news articles, business documents, and even text messages or social media. The Latent Dirichlet Allocation (LDA) algorithm is widely used in document grouping or classification applications, sentiment analysis, and natural language processing, so researchers chose the Latent Dirichlet Allocation (LDA) algorithm as the model for this research [7].

Proof of the application of the Latent Dirichlet Allocation (LDA) Algorithm can be proven in research conducted by Akhsin Nurlayli and Morch. Ari Nasichuddin (2019), regarding Topic Modeling of Research by JPTEI UNY Lecturers on Google Scholar Using Latent Dirichlet Allocation. The research data was carried out using the scraping technique. Scraping results are stored in .csv format and the data obtained is 909 publication titles. The classification process is divided into 4 clusters or 4 topics and the algorithm used is Latent Dirichlet Allocation. Most research results are in the 1st cluster and the least in the 4th cluster [8].

Proof of the application of the Latent Dirichlet Allocation (LDA) Algorithm can be proven in research conducted by Alif Iffan Alfanzar, Khalid, and Indri Sudanawati Rozas regarding Thesis Topic Modeling Using the Latent Dirichlet Method. This research collected data using the scraping method. This research

conducted experiments on 5 different iteration tests, namely: 100, 500, 1000, and 5000. Each iteration test has a different number of topics, namely: 2, 3, 4, and 5. The test results using the Latent Dirichlet Allocation (LDA) method are available 584 thesis abstract data. The results of topic modeling were carried out in 5 iteration tests and each iteration test included a different number of topics, namely 2, 3, 4, and 5. The best topic cluster results were obtained with a number of topics of 3 [9].

Proof of the application of the Latent Dirichlet Allocation (LDA) Algorithm can again be proven in research conducted by Diandra Zakeshia Tiara Kannitha, Mustafid, and Puspita Kartikasari regarding Topic Modeling in Customer Complaints Using the Latent Dirichlet Allocation Algorithm in Twitter Social Media. This research collects data obtained from IndiHomeCare and FirstMediaCares Twitter. The data obtained was 5000 tweets, then duplicate data was removed so that the remaining 4828 tweets were for Indihome and 4973 tweets for FirstMediaCares. Test results using Latent Dirichlet Allocation (LDA) on the keyword FirstMediaCares produced a total of 10 topics while the keyword IndiHomeCare produced a total of 11 topics. Based on the trend results, the FirstMediaCares topic is the internet that shuts down while working, while IndiHomeCare is the internet that likes to disconnect and turn off. Based on the interpretation, the figure was 70% for FirstMediaCares and 81.81% for IndiHomeCare [10].

Furthermore, in previous research conducted by Bagus Wicaksono Arianto and Gangga Anuraga (2020), regarding Modeling Twitter User Topics regarding the "Ruangguru" application. This research data was conducted using Wordcloud tools. The algorithm used is Latent Dirichlet Allocation. The Latent Dirichlet Allocation (LDA) clustering method in the Ruangguru application groups Twitter data into 28 topics with the topic often discussed being Ruangguru discounts [11].

The difference between this research and previous research is that other studies label it manually, whereas I analyzed it based on the LDA labeler as a dataset and the dataset from LDA will be classified. Based on this description, the Latent Dirichlet Allocation (LDA) algorithm is very precise and suitable for topic modeling. In this research, we will use Instagram to view posts and compare topics discussed on Multi Data University Palembang's Instagram social media and the aim of this research is to find out trending topics that are currently being widely discussed on Multi Data University Palembang's Instagram.

## **METHODOLOGY**

In this implementation, there are several steps taken to get the best results. Figure 1 shows the research stages.

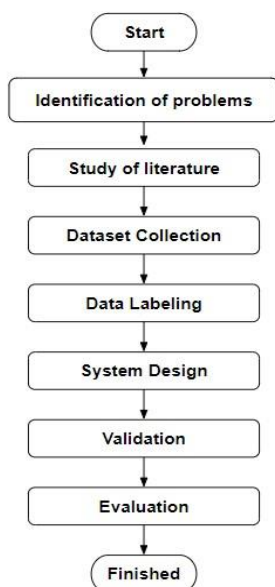


Figure 1. Research Stages

**A. Identification of problems**

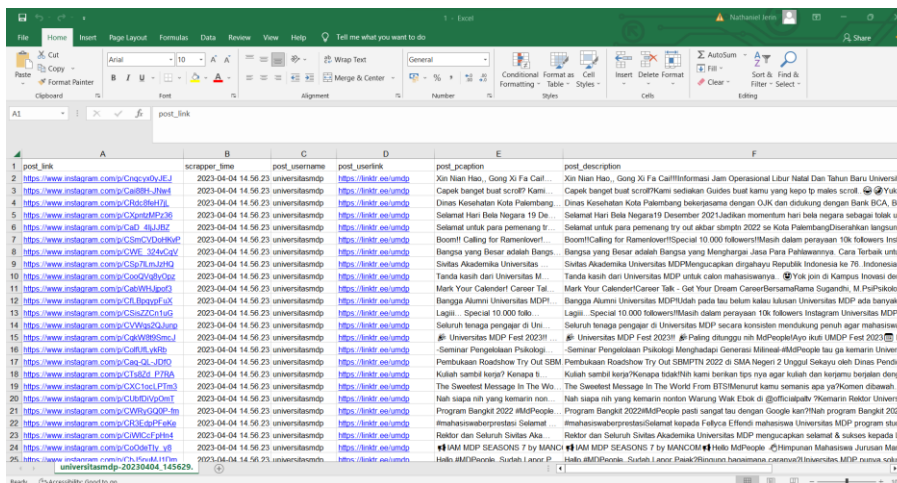
At this stage, problem identification will be carried out regarding topic modeling on the Palembang Multi Data University Instagram application using the Latent Dirichlet Allocation (LDA) Algorithm.

**B. Study of literature**

At this stage a literature study is carried out from several sources, for example journals and books that are still related to the research topic. It aims to obtain information and data that can be used to conduct this research.

**C. Dataset Collection**

Data collection on the Instagram site was carried out by a scraping process using Jupyter Notebook. Data scraping in this research was carried out on the Instagram site (<https://www.instagram.com/universitasmdu/>). The dataset was taken from 2021 to 2023. The data features used in the system were obtained from the Palembang Data University Instagram account and the data collected amounted to around 500 pieces of data. Figure 2 and Figure 3 are datasets obtained from scraping results.



	A	B	C	D	E	F
	post_link	scraper_time	post_username	post_userlink	post_pcaption	post_description
1	<a href="https://www.instagram.com/p/CqcyvUjEJ/">https://www.instagram.com/p/CqcyvUjEJ/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Xin Nian Hao, Gong Xi Fa Cai!	Xin Nian Hao, Gong Xi Fa Cai!#Informasi Jam Operasional Libur Natal Dan Tahun Baru Universitas
2	<a href="https://www.instagram.com/p/Ca8t9Uj4tHd/">https://www.instagram.com/p/Ca8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Capek banget buat scrrip? Kami	Capek banget buat scrrip?Kami sediakan Gudies buat kamu yang hepo tp males scrrip
3	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Dinas Kesehatan Kota Palembang	Dinas Kesehatan Kota Palembang bekerjasama dengan OJK dan didukung dengan Bank BCA, Bi
4	<a href="https://www.instagram.com/p/CXp8t9Uj4tHd/">https://www.instagram.com/p/CXp8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Selamat Hari Bela Negara 19 De	Selamat Hari Bela Negara 19 Desember 2021.Jadikan momentum hari bela negara sebagai titik ul
5	<a href="https://www.instagram.com/p/Ca8t9Uj4tHd/">https://www.instagram.com/p/Ca8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Selamat untuk para pemenang tr	Selamat untuk para pemenang try out sblar sblmptn 2022 se Kota Palembang!Disebabkan langsun
6	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Boomi Calling for Ramenlover!	BoomiCalling for Ramenlover!Special 10.000 followers!Malah dalam perayaan 10k followers Inst
7	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Bangka yang Besar adalah Banga	Bangka yang Besar adalah Banga yang Menghargai Jasa Para Pejuangny. Cara terbaik untu
8	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Svitas Akademika Universitas	Svitas Akademika Universitas MDPMengucapkan dipgahayu Republik Indonesia ke 76. Indonesia
9	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Tanda kasih dari Universitas M	Tanda kasih dari Universitas MDP untuk calon mahasiswaanya
10	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Mark Your Calendar! Career Tal	Mark Your Calendar!Career talk - Get Your Dream Career!Bertram Ramana Sugandi, M.PsPskolok
11	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Bangka Akemi Universitas MDP!	Bangka Akemi Universitas MDP!Malah pada tas belum kalau kelasau Universitas MDP ada kanyak!
12	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Lagii... Special 10.000 folo.	Lagii... Special 10.000 followers!Malah dalam perayaan 10k followers Instagram Universitas MDP!
13	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Seluruh tenaga pengajar di Uni	Seluruh tenaga pengajar di Universitas MDP secara konsisten mendukung penuh agar mahasiswa
14	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	# Universitas MDP Fest 2023!	# Universitas MDP Fest 2023! # Paling ditunggu nih MDPeople!ayo ikut UMDF Fest 2023!
15	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Seminar Pengabdian Psikologi	Seminar Pengabdian Psikologi Menghadapi Generasi Milenial!#People tag ga kemari! Urutan
16	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Pembukaan Roadshow Try Out SSM	Pembukaan Roadshow Try Out SSMPTN 2022 di SMA Negeri 2 Unggul Sekeloa oleh Dinas Pendid
17	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Kuliah sambil kerj? Kenapa t	Kuliah sambil kerj?Kenapa tidak?Nah kami berikan tips nya agar kuliah dan kerjmu berjalan deng
18	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	The Sweetest Message In The Wo	The Sweetest Message In The World From BTS!Menurut kamu semanis apa ya?Komen dibawah
19	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Nah siapa nih yang kemari mona	Nah siapa nih yang kemari mona Waturu Wati!Elok di @officialbts!#Komenin Raktor Univer
20	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Program Bangkit 2022 #MIDPeople	Program Bangkit 2022#MIDPeople pasti sangat tau dengan Google kan?#Nah program Bangkit 202
21	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	#mahasiswaberprestasi Selamat	#mahasiswaberprestasiSelamat kepada Falyca Effendi mahasiswa Universitas MDP program stud
22	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Rektor dan Seluruh Svitas Aka	Rektor dan Seluruh Svitas Akademika Universitas MDP mengucapkan selamat & sukses kepada l
23	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	#IAMI MDP SEASONS 7 by IANCI	#IAMI MDP SEASONS 7 by IANCI #IAMI MDP SEASONS 7 by IANCI #IAMI MDP SEASONS 7 by IANCI
24	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>	Halo #MIDPeople... Suciati Laver P	Halo #MIDPeople... Suciati Laver P... Suciati Laver P... Suciati Laver P... Suciati Laver P... Suciati Laver P...
25	<a href="https://www.instagram.com/p/C8S8t9Uj4tHd/">https://www.instagram.com/p/C8S8t9Uj4tHd/</a>	2023-04-04 14:56:23	universitasmdu	<a href="https://trkr.ee.usmdp">https://trkr.ee.usmdp</a>		

Figure 2. Sample Datasets

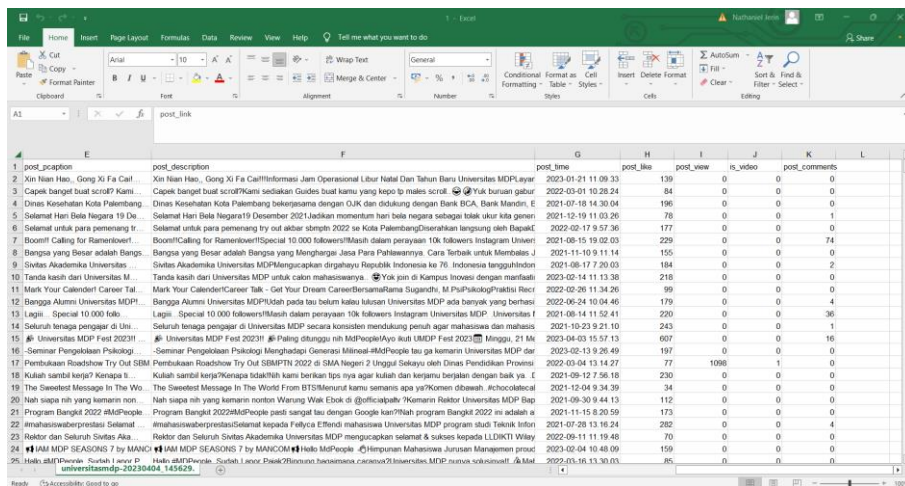


Figure 3. Sample Datasets

**D. Preprocessing Data**

In the Pre-processing Stage, the dataset obtained from the results of Multi Data Palembang University Instagram Scraping will carry out the Case Folding, Tokenize, Stopword and Stemming stages. Pre-processing aims to clean or remove text from data that is not needed so as to avoid problematic data and inconsistent data.

**E. Data Labeling**

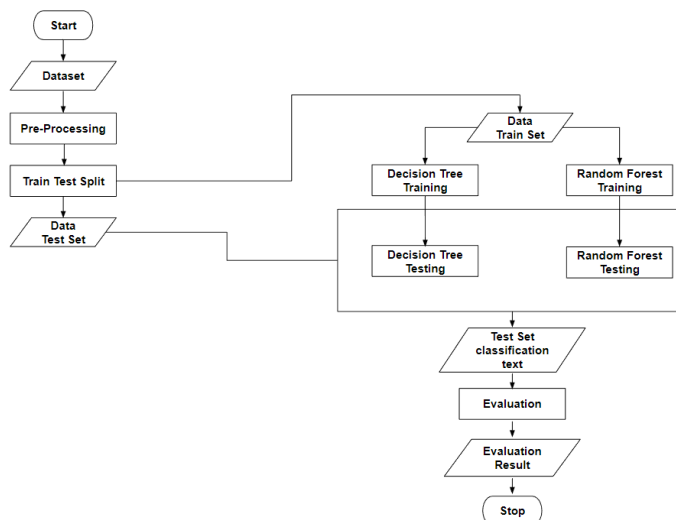
At the data labeling stage, the amount of data to be labeled is 500 data. Data labeling will be divided into 8 topic categories namely Achievement, Academic, Alumni, Tips, Graduation, Promotion, Event and Commemoration. At the data labeling stage it wa carried out by researchers and validated by the marketing department at Multi Data Palembang University. Data labeling is used to compare the results of data labeling performed by the LDA algorithm. Table 1 displays the labeling data carried out manually.

Tabel 1. Data Labeling

No	Topik	Jumlah
1	Akademik	25
2	Wisuda	3
3	Tips	19
4	Alumi	16
5	Peringatan	67
6	Event	190
7	Promosi	108
8	Prestasi	72
<b>Total</b>		<b>500</b>

**F. System Design**

The Decision Tree and Random Forest methods are used to predict the topics on Instagram posts, which are the number of topics on Instagram at Multi Data University Palembang. The overall system can be seen in Figure 4.



**Figure 4. System Design Chart**

### G. Validation

In this phase, model validation is carried out, targeted validation is the phase where the results obtained by the system are checked for correctness and run according to what is desired.

### H. Evaluation

The final step is to evaluate the system that has been created. With the aim to minimize errors. At this stage, the calculation of the confusion matrix will be used to measure the quality of the clustering or grouping produced by the LDA topic modeling. In this stage, the confusion matrix is calculated with True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) data, using the confusion matrix formula to calculate Precision, Recall, F1-Score and Accuracy [12]:

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (2)$$

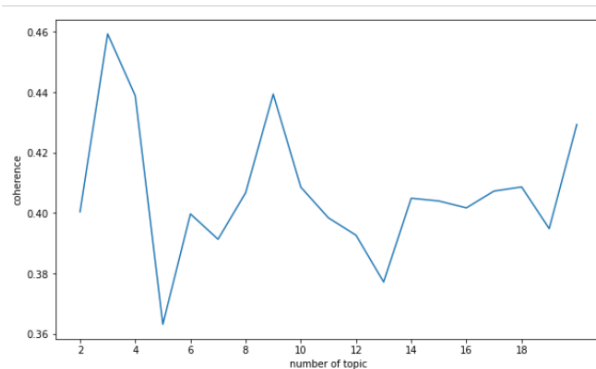
$$\text{F1-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (4)$$

## RESULT AND ANALYSIS

### A. LDA Topic Modeling

This stage uses the Latent Dirichlet Allocation method which functions to determine the topic words that represent each topic. The first step is to look at the coherence score, which is used to evaluate topic modeling, a good model produces topic with high topic coherence scores. Based on Figure 5, number 3 is the number of topics with the highest coherence value.



**Figure 5. Coherence Value**

The higher the coherence value, the better the interpretation results from the generated topic modeling. Based on the picture above Figure 5 the highest coherence value is topic 3, namely 0.46, the number of topics produced is then used as a reference, so that there are three best topics. Table 2 displays the results of the coherence values in the study.

**Tabel 2 Coherence Value**

<b>Topik</b>	<b>Nilai Koherensi</b>
2	0.4004023864073535
<b>3</b>	<b>0.4593123637768688</b>
4	0.43883807459081636
5	0.36314585792419274
6	0.3996940690970596
7	0.39130268921322936
8	0.40669563969414485
9	0.43934950078555307
10	0.40855590131307107
11	0.3983910904465393
12	0.39263883366145463
13	0.3771457546435214
14	0.4048858880106212
15	0.4039755147488607
16	0.40170122255123014
17	0.4072596580390873
18	0.4086295671300364
19	0.3948234511717473
20	0.42929531451422703

In selecting the number of topics, the Latent Dirichlet Allocation (LDA) model starts from 2, because 2 topics are the minimum number that can be selected. Then each word from the modeling topic has a weight. The weight value is the probability value of each word that appears in the topic. Table 3 displays the weight value of each word.



are three word cloud visualizations that show the topic modeling results. Wordcloud visualization is considered representative enough to display topic modeling results because it can determine the occurrence of common expressions based on the topic being modeled. Figure 7, 8, and 8 are Wordcloud Visualizations based on the topics formed.



Figure 7. Topic Wordcloud Visualization 1



Figure 8. Topic Wordcloud Visualization 2

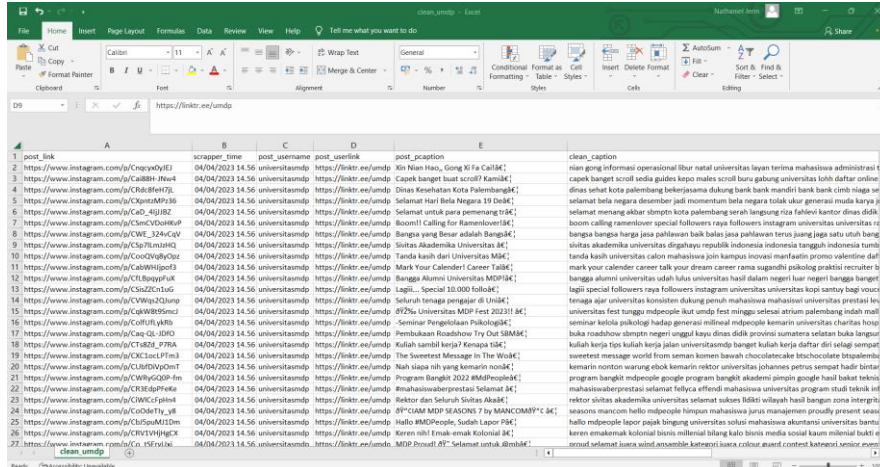


Figure 9. Topic Wordcloud Visualization 3

Figure 7, 8, and 9 show the terms that appear on the topic in the document, wordcloud visualization obtained from Tabel 3. The bigger the letters, the more words there are on the topic.

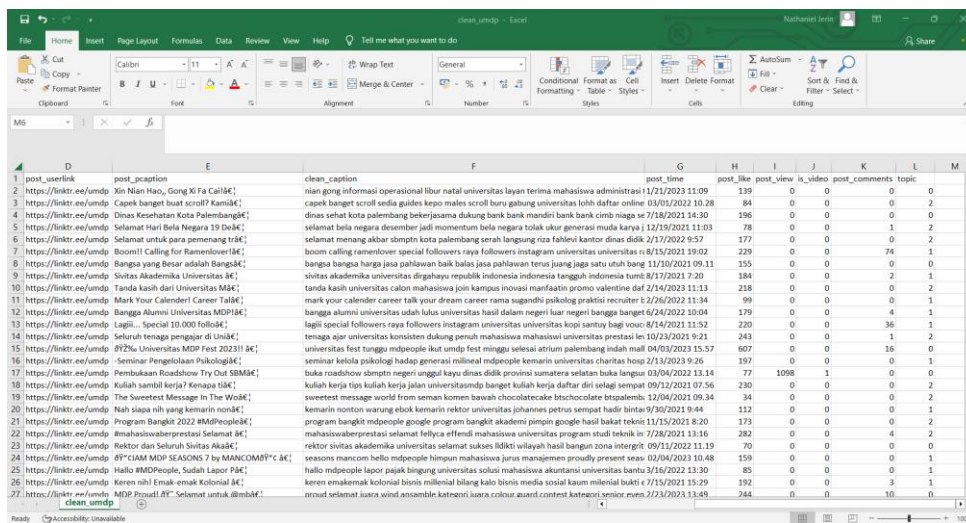
### C. Dataset berdasarkan label LDA

At the LDA data labeling stage, the number of data to be labeled is 499 data. Data labeling carried out by LDA is divided into 3 topic categories, namely topic 1 has the keywords competitions, webinars, and holidays, topic 2 has keywords list of lectures and scholarships, topic 3 has the keywords alumni achievement, academic, and performance. The 3 categories are obtained based on the coherence value. Figure 10 and Figure 11 are data labeling that has been carried out by LDA.



	A	B	C	D	E
1	post_link	scraper_time	post_username	post_userfick	post_caption
2	https://www.instagram.com/p/CqxydyEjE	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Xin Nian Hao, Gong Xi Fa Cai!
3	https://www.instagram.com/p/Ca8BhRhe4	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Capek banget buat scroll? Kamak!
4	https://www.instagram.com/p/C8kR6H7G	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Dinas Kesehatan Kota Palembang!
5	https://www.instagram.com/p/CXpMFPz36	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Selamat Hari Bela Negara 19 Desember!
6	https://www.instagram.com/p/Ca0_H4jBzR	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Selamat untuk para pemenang!
7	https://www.instagram.com/p/CWj2CvCvQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Boon! Calling for Ramenlover!
8	https://www.instagram.com/p/C4E_324vCvQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Bangsa yang Besar adalah Bangsa!
9	https://www.instagram.com/p/C5p7MmHqQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Sukses Akademika Universitas!
10	https://www.instagram.com/p/CvQd9HjDzQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Tanda kasih dari Universitas!
11	https://www.instagram.com/p/Ca6WHjg0f3	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Mark Your Calender! Career Talk!
12	https://www.instagram.com/p/CLB9p9yAa	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Bangga Alumni Universitas!
13	https://www.instagram.com/p/C5oZCvYvQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Lagi... Special 10.000 follow!
14	https://www.instagram.com/p/CV9Wz2QjU	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Sekolah tenaga pengajar di Unia!
15	https://www.instagram.com/p/CqkWB9Sjmc	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	872ha Universitas MDP Fest 2023!!!
16	https://www.instagram.com/p/Cvff7y4v8	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Seminar Psikologi Hadis Generasi!
17	https://www.instagram.com/p/Cac3G_0jD	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Pembukaan Roadshow Try Out SBM!
18	https://www.instagram.com/p/C1s8T_P7RA	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Kuliah sambil kerja? Kenapa tidak!
19	https://www.instagram.com/p/CXK3CvYvQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	The Sweetest Message In The World!
20	https://www.instagram.com/p/CvB9V9vQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Nah siapa nih yang kemarin nonde!
21	https://www.instagram.com/p/CvW9G2QjU	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Program Bangkit 2022 #MDPeople!
22	https://www.instagram.com/p/CvB9V9vQ	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Imahasiswa berprestasi Selamat!
23	https://www.instagram.com/p/CvW9G2QjU	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Rektor dan seluruh Siswa Akad!
24	https://www.instagram.com/p/CvW9G2QjU	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	87 CIAM MDP SEASONS 7 by MANCOMB!
25	https://www.instagram.com/p/CvW9G2QjU	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Hallo #MDPeople, Sudah Lapor PA!
26	https://www.instagram.com/p/CvW9G2QjU	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	Keren emakemak kolonial bisnis!
27	https://www.instagram.com/p/CvW9G2QjU	04/04/2023 14:56	universitasumd	https://linktr.ee/umdp	MDP Proud! #MDPeople - Selamat untuk #MDP!

Figure 10. Sample Dataset based on LDA



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	post_userlink	post_caption	clean_caption	post_time	post_like	post_view	is_video	post_comments	topic				
2	https://linktr.ee/umdp	Xin Nian Hao, Gong Xi Fa Cai!	nian gong informasi operasional libur natal universitas layanan terima mahasiswa administrasi	11/21/2023 11:09	139	0	0	0	0	0	0	0	
3	https://linktr.ee/umdp	Capek banget buat scroll? Kamak!	capek banget scroll media guides kepo males scroll buru gabung universitas loh daftar online	03/03/2023 10:28	84	0	0	0	0	0	0	2	
4	https://linktr.ee/umdp	Dinas Kesehatan Kota Palembang!	dinas sehat kota Palembang bekerjasama dengan bank mandiri bank bank cimb bank bank bca	07/18/2021 14:30	196	0	0	0	0	0	0	0	
5	https://linktr.ee/umdp	Selamat Hari Bela Negara 19 Desember!	selamat bela negara desember jadi momentum bela negara tolah ukur generasi muda karya	12/19/2021 11:03	78	0	0	0	1	2	0	0	
6	https://linktr.ee/umdp	Selamat menang akbar sbmptn kota Palembang serah langsung riza fahevi kantor dinas didik	selamat menang akbar sbmptn kota Palembang serah langsung riza fahevi kantor dinas didik	2/17/2022 9:57	177	0	0	0	0	0	0	2	
7	https://linktr.ee/umdp	Boon! Calling for Ramenlover!	boon calling ramenlover special followers raya followers instagram universitas universitas	8/15/2021 19:02	229	0	0	0	0	0	0	74	1
8	https://linktr.ee/umdp	Bangsa yang Besar adalah Bangsa!	bangsa bangsa harga jasa pahlawan baik balas jasa pahlawan terus jagu jaga satu utuh bang	11/09/2021 09:11	255	0	0	0	0	0	0	0	0
9	https://linktr.ee/umdp	Sukses Akademika Universitas!	sukses akademika universitas dirgahayu republik indonesia indonesia tangguh indonesia tumb	8/17/2021 7:20	184	0	0	0	0	0	0	2	1
10	https://linktr.ee/umdp	Tanda kasih dari Universitas!	tanda kasih universitas calon mahasiswa join kampus inovasi manfaat promo valentine dal	2/14/2023 11:13	218	0	0	0	0	0	0	0	2
11	https://linktr.ee/umdp	Mark Your Calender! Career Talk!	mark your calender career talk your dream career rama sugandhi psikolog praktisi recruter	2/26/2022 11:34	99	0	0	0	0	0	0	0	1
12	https://linktr.ee/umdp	Bangga Alumni Universitas!	bangga alumni universitas udah lulus universitas hasil dalam negeri luar negeri bangga bang	6/24/2022 10:04	179	0	0	0	0	0	0	4	1
13	https://linktr.ee/umdp	Lagi... Special 10.000 follow!	lagi special followers raya followers instagram universitas universitas kopi santay bagi vouc	8/4/2021 11:52	220	0	0	0	0	0	0	36	1
14	https://linktr.ee/umdp	Sekolah tenaga pengajar di Unia!	tenaga ajar universitas konsisten dukung penuh mahasiswa mahasiswi universitas prestasi	10/23/2021 9:21	243	0	0	0	0	0	0	1	2
15	https://linktr.ee/umdp	872ha Universitas MDP Fest 2023!!!	universitas fest tunggu mdpeople ikut umdp fest minggu selesai atrium Palembang indah mall	04/03/2023 15:57	607	0	0	0	0	0	0	16	0
16	https://linktr.ee/umdp	Seminar Psikologi Hadis Generasi!	seminar kelika psikologi hadis generasi milenial mdpeople kemarin universitas charitas hosp	2/12/2023 9:20	197	0	0	0	0	0	0	0	1
17	https://linktr.ee/umdp	Pembukaan Roadshow Try Out SBM!	buka roadshow sbmptn negeri unggul kayu dinas didik provinsi Sumatera selatan buka bang	03/04/2022 13:14	77	1098	1	0	0	0	0	0	0
18	https://linktr.ee/umdp	Kuliah sambil kerja? Kenapa tidak!	kuliah kerja tips kuliah kerja jalan universitasmdp banget kuliah kerja daftar diri selaj	09/12/2021 07:56	230	0	0	0	0	0	0	0	2
19	https://linktr.ee/umdp	The Sweetest Message In The World!	sweetest message world from teman komen bawah chocolatecake btchocolate btspalemb	12/04/2021 09:34	34	0	0	0	0	0	0	0	2
20	https://linktr.ee/umdp	Nah siapa nih yang kemarin nonde!	kenarin nonon warung ebook kemarin rektor universitas johanes petrus sempat hadi bina	9/30/2021 9:41	112	0	0	0	0	0	0	0	1
21	https://linktr.ee/umdp	Program Bangkit 2022 #MDPeople!	program bangkit mdpeople google program bangkit akademi pemin google hasil bakat tekni	11/15/2021 8:20	173	0	0	0	0	0	0	0	2
22	https://linktr.ee/umdp	Imahasiswa berprestasi Selamat!	imahasiswa berprestasi selamat felicya effendi mahasiswa universitas program studi tekni	7/28/2021 13:16	282	0	0	0	0	0	0	4	2
23	https://linktr.ee/umdp	Rektor dan seluruh Siswa Akad!	rektor sukses akademika universitas selamat sukses lidki wilayah hasil bangun zona intergr	09/11/2022 11:19	70	0	0	0	0	0	0	0	0
24	https://linktr.ee/umdp	87 CIAM MDP SEASONS 7 by MANCOMB!	seasons mancom hello mdpeople himpun mahasiswa jurus manajemen proudly present seasi	03/04/2023 10:48	159	0	0	0	0	0	0	0	1
25	https://linktr.ee/umdp	Hallo #MDPeople, Sudah Lapor PA!	hallo mdpeople lapor pajak bingung universitas solusi mahasiswa akuntansi universitas bant	3/16/2022 13:30	85	0	0	0	0	0	0	0	1
26	https://linktr.ee/umdp	Keren emakemak kolonial bisnis!	keren emakemak kolonial bisnis milenial bilang kalo bisnis media sosial kaum milenial ba	7/1/2021 15:29	192	0	0	0	0	0	0	3	1
27	https://linktr.ee/umdp	MDP Proud! #MDPeople - Selamat untuk #MDP!	mdp selamat isara winid ansambel kategori isara colour award content kategori senior eum	7/23/2023 13:49	244	0	0	0	0	0	0	10	0

Figure 11. Sample Dataset based on LDA

### D. Test Results

At this stage the researcher will compare the results of the random forest and decision tree. Table 4 displays the results obtained from random forest and decision tree testing.

**Tabel 4. Comparison of Random Forest and Decision Tree**

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
Random Forest	74%	68.75%	70.96%	69.84%
Decision Tree	57%	61.53%	43.24%	50.79%

In Table 4 the test was carried out using a confusion matrix through a random forest library and a decision tree, from 499 test data. The results obtained from the tests carried out on the random forest model yielded 74% accuracy, 68.75% precision, 70.96% recall, and 69.84% f1-score, while the decision tree model produced 57% accuracy, 61.53% precision, 43.24% recall, and f1-score 50.79%. So in terms of accuracy the random forest model is better than the decision tree model.

At this stage the researcher will compare the manual labeling dataset and the LDA labeling dataset using the random forest and decision tree models. Table 5 and Table 6 display the results obtained from testing based on manual labeling and LDA labeling.

**Tabel 5. Manual Labeling Results**

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
Random Forest	79%	77.35%	66.32%	66.66%
Decision Tree	56%	66.66%	33.33%	50%

**Tabel 6. LDA Labeling Results**

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
Random Forest	78%	80%	66.66%	72.72%
Decision Tree	52%	44.68%	65.62%	53.16%

In Table 5 and Table 6 the tests were carried out using a confusion matrix through a random forest library and a decision tree. The results obtained from the manual labeling tests carried out on the random forest model yielded 79% accuracy, 77.35% precision, 66.32% recall, and 66.66% f1-score, while the manual labeling tests performed on the decision tree model produced 56% accuracy, 66.66% precision, 33.33% recall, and 50% f1-score, then results obtained from the LDA labeling tests carried out on the random forest model yielded 78% accuracy, 80% precision, 66.66% recall, and 72.72% f1-score, while the decision tree model resulted in 52% accuracy, 44.68% precision, 65.62% recall, and 53.16% f1-score. So in terms of f1-score, the random forest model and decision tree model labeled with LDA are better than those with manual labels. Here it can also be seen from the results of manual labeling and LDA labeling that there is an improvement, which means that LDA has a good

comparison compared to manual.

## CONCLUSION

Results comparison of testing models with the LDA labeling dataset and manual labeling that the random forest model gets good results on LDA modeling with an accuracy of 78%, precision 80%, recall 66.66%, and f1-score 72.72%, while the LDA labeling test is carried out on the decision tree model to get unfavorable results with an accuracy value of 52%, precision 44.68%, recall 65.62%, and f1-score 53.16%. For further research to get improvements you must look for the appropriate classification.

## REFERENCES

- [1] Umam, "Definition of social media, history, functions, types, benefits and development," <https://www.gramedia.com/literasi/pengertian-media-sosial/>.
- [2] Rudi Dian Arifin, "Understanding Instagram along with its history, function, purpose, benefits, etc.," <https://dianisa.com/pengertian-instagram/>.
- [3] S. Rohadian and M. T. Amir, "Efforts to build customer engagement through Instagram social media," *Journal of Entrepreneurship, Management and Industry (JEMI)*, vol. 2, no. 4, pp. 179–186, 2019.
- [4] R. M. H. Ashari and O. F. Sitorus, "The influence of content marketing on customer engagement with memorable coffee," *Jurnal EMT KITA*, vol. 7, no. 1, pp. 38–46, 2023.
- [5] Raihannisa Fitriah, "11 Universities in South Sumatra, number 9 view of the Ampera bridge," <https://sumsel.inews.id/berita/universitas-di-sumatera-selatan/all>.
- [6] Universitas MDP, "Instagram," <https://www.instagram.com/universitasmdp/>.
- [7] M. R. Pribadi, E. P. Widiyanto, D. Hermanto, D. I. Ricoida, and D. Pibriana, "Analysis of marketplace social media user engagement by topic," in *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, 2022, pp. 35–39.
- [8] A. Nurlyayli and M. A. Nasichuddin, "Jptei UNY lecturer's research modeling topic on Google Scholar using latent dirichlet allocation," *Elinvo (Electronics, Informatics, and Vocational Education)*, vol. 4, no. 2, pp. 154–161, 2019.
- [9] A. I. Alfanzar, K. Khalid, and I. S. Rozas, "Thesis topic modeling uses the latent dirichlet allocation method," *JSil (Jurnal Sistem Informasi)*, vol. 7, no. 1, pp. 7–13, 2020.
- [10] D. Z. T. Kannitha, M. Mustafid, and P. Kartikasari, "Topic modeling on customer complaints using the latent Dirichlet allocation algorithm in Twitter social media," *Jurnal Gaussian*, vol. 11, no. 2, pp. 266–277, 2022.
- [11] B. W. Arianto and G. Anuraga, "Topic modeling of Twitter users regarding the 'ruangguru' application," *Jurnal ILMU DASAR*, vol. 21, no. 2, pp. 149–154, 2020.
- [12] N. L. P. M. Putu and A. Z. Amrullah, "Sentiment analysis and modeling of Lombok tourism topics using the naive Bayes algorithm and latent Dirichlet allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 123–131, 2021.