ISSN 1533-9211

DOI: 10.5281/zenodo.11257494

CLASSIFICATION OF PUBLIC OPINION REGARDING THE POLICIES OF THE CENTRAL GOVERNMENT VIA TWITTER DATA USING LATENT DIRICHLET ALLOCATION

ARIES DWI INDRIYANTI [1], RAHMAT GERNOWO [2] and EKO SEDIYONO [3]

[1,2,3] Diponegoro University, Semarang, Indonesia.
Email: [1]ariesdwi@students.undip.ac.id, [2]rahmatgernowo@lecturer.undip.ac.id, [3]eko@uksw.edu

Abstract

This study presents one popular topological modeling technique used in many research studies, namely Latent Dirichlet Allocation (LDA). In the context of the text analysis, LDA provides a strong framework for organizing documents according to the relevant topic sentences in the text. This method effectively identifies the relationship between the words in the document and transfers them to the relevant topik-topik that are extracted from the aforementioned content. The purpose of the LDA in this study is to reveal the topological structure that is hidden in the collection of text documents. By grouping documents according to related keywords or themes, LDA facilitates more in-depth understanding of the range of topics covered in the aforementioned text. The results of this LDA modeling can be used for more in-depth analysis, such as topological segmentation, document classification, or more canggih recommendation system development. In summary, the use of LDA in this study allows researchers to obtain more information on the structure and content of a collection of complex text documents. Through the identification of the pola-pola that are missing from the text data, this study can make a significant contribution to the understanding and retrieval of the information contained in the text.

Keywords: LDA, X, Government Policy.

INTRODUCTION

The primary goal of the Government of Essence is to increase the level of public awareness, whether it is through silent or noisy means. Because government policy is based on the assumption that the public is its intended audience, it can also be described as public policy. The policy of the government, which is also known as public policy, is shaped with caution and firmness, enhancing its impact on every individual in that country.

This is achieved via regulations that are interpreted in a systematic and comprehensive manner, while also requiring citations to be submitted in unison. The committee initiated this public policy phase in order to achieve mass population prosperity through effective and efficient regulations.

The Central Government regards public opinion as an essential component of the development strategy of the region. This policy is specifically designed to increase the level of social welfare in that region. Evaluation of policy is an essential tool to assess effectiveness and obtain more precise information about the public's response to implemented policies.

This is related to research by Berliana and Arsanti (2018), which suggests the importance of evaluating education in order to understand the dynamics of interactions between education and the general public in the Middle East[1].

There are several essential principles in European countries that serve as a guide for evaluating the performance of government employees. These tenets, which are referred to as good governance principles, serve as a guide for assessing public policy's effectiveness, accountability, and transparency.

In Belgium, these provisions are governed by the Administrative Rights Act (AROB) (Administrative Rechtspraak Overheids Beschikkingen). One of the most important aspects of AROB law is public opinion participation, which requires the government to consider and understand public opinion during the creation and implementation of public policy[2]. The efficiency of this approach in text classification and information retrieval has been shown in numerous research[3]. Latent Dirichlet Allocation (LDA) is a topological modeling technique that is widely used nowadays. For text analysis, a well-liked topological modeling approach is Latent Dirichlet Allocation (LDA). The way this method operates is that it creates topical topiks by clustering pages that contain either text or markup that appears frequently together. LDA's principal benefits are its scalability of data handling and ease of implementation[3]. The application of LDA facilitates more thorough mapping and analysis of pertinent policy documents in the context of East Javan government policy, helping stakeholders comprehend the substance of the policies, spot trends, and assess their efficacy. The research integrates Latent Dirichlet Allocation (LDA) and location mapping to uncover deep perceptions of social media data. This method generates patterns, relationships, and situations that are beneficial for users in making decisions. Mapping social media user spaces allows researchers and practitioners to understand the usage patterns and best applications of these communication services. The research conducted by Fernandez et al., (2022) uses sentiment analysis to examine public sentiment in two different public spaces in New York City: Bryant Park Park and Grand Central Station[4]. Data is collected from Twitter and analyzed using sentiment analysis tools. Results showed that public sentiment at Bryant Park was significantly more positive than at Grand Central Station. This difference may be due to several factors, such as the design of the space, facilities, and available activities.

The correlation between urban metrics and sentiment analysis shows that shrinking cities generally have higher negative sentiment than developing cities. Research on the impact of location information on sentiment classification has the potential to produce better decision support tools[5]. The valuation of a business and its ranking of its geographic environment show a weak positive correlation. Combining geographic information such as business categories, their popularity, and the content of customer reviews can improve the accuracy of business ranking predictions[5].

There are two main approaches to predicting a tweet's location based on its content, which are based on using the tweet content alone or considering geographic context. The first approach groups the data based on the density of geotagged tweets in different regions. The second approach transforms the geographical space into a predefined grid of territories. Several studies have built language models for each area to predict location. The likelihood of a tweet being generated in an area is determined based on its relevance to geotagged tweets in that area. The closest distance is returned as the predicted location. The importance of geographic information in tweets and enabled apps is first described. The main approaches to the geo-location prediction section then describe the two main approaches used for geo-location prediction, based on tweet content or geographic context[6].

This study aims to evaluate the effectiveness of local government policies in East Java Province using Latent Dirichlet Allocation (LDA). LDAs are applied to classify government policies into specific categories, such as education, health, and infrastructure as well as map locations based on frequently emerging sentiments/opinions, and model key topics in policy. The results of the LDA classification will be integrated to produce an accurate and comprehensive information system on the effectiveness of East Java government policies

METHOD

The implementation of research procedures is carried out in order to achieve the goals to be achieved by carrying out several stages, namely preparation, design, programming, testing and improvement to draw conclusions from the results of the research conducted. The following are some of the procedures that will be carried out in this study as shown in Figure:

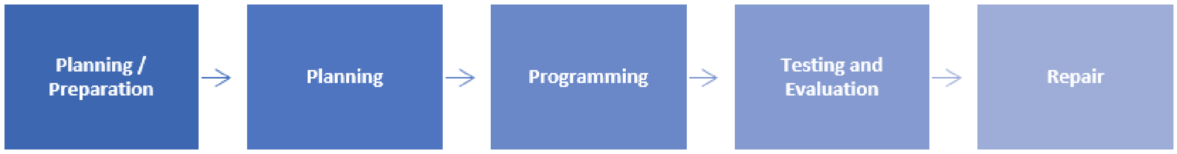


Figure 1: Research Procedure Flowchart

In the early stages focus on determining system requirements and workflow analysis. This stage includes determining the research topic, selecting appropriate classification methods for local government policy analysis of East Java Province, and collecting or crawling local government policy tweet data through the Twitter platform (X) or x. This stage is important to ensure that the data to be used in classification research is relevant and representative. The second stage is to design the architecture of the classification system with the method to be used. This includes selecting the features to be used in classification, dividing data into training data and test data, and setting the parameters to be used.

The third stage focuses on implementing a classification process using Python and Google Collaboratory to write, and run Python code. This stage includes preprocessing which includes text cleaning, tokenization, stopword removal, stemming process and non-standard word removal process. The next step is to group the corpus into topics using the Latent Dirichlet Allocation (LDA) technique. Once the corpus is grouped into several topics, the next step is to label each topic based on an ontology scheme that refers to the KBBI word field list using a hybrid measure. After the classification process has been programmed, the next stage is to test the classification model that has been made. The test was conducted using test data separate from the training data to measure the performance of the model in classifying local government policies of East Java Province. The last stage is carried out for defact improvements based on analyzing the results and evaluating the causes of poor model performance. After the testing

ISSN 1533-9211

DOI: 10.5281/zenodo.11257494

and improvement procedures were completed, an analysis was carried out on the results of the classification of local government policies in East Java Province using machine learning. The results of this analysis can then be used as a basis for compiling research reports and publication of research results in international journals as well as IPR submissions.

RESEARCH MATERIALS AND TOOLS

To carry out implementation and testing on the system being developed in this study, the use of appropriate tools and materials is required. The material used in this study was tweet data taken from the Twitter platform or X.

Workflow

The data used in the study came from social media platforms such as Twitter or X. The next process involves a data pre-processing stage that includes text cleaning, tokenization, deletion of common words, a stemming process to reduce words to basic forms, as well as removal of non-standard words. After that, the corpus data is grouped into topics using the Latent Dirichlet Allocation (LDA) technique, and each topic is labeled based on an ontology scheme that refers to the KBBI word field list using the hybrid measure method.

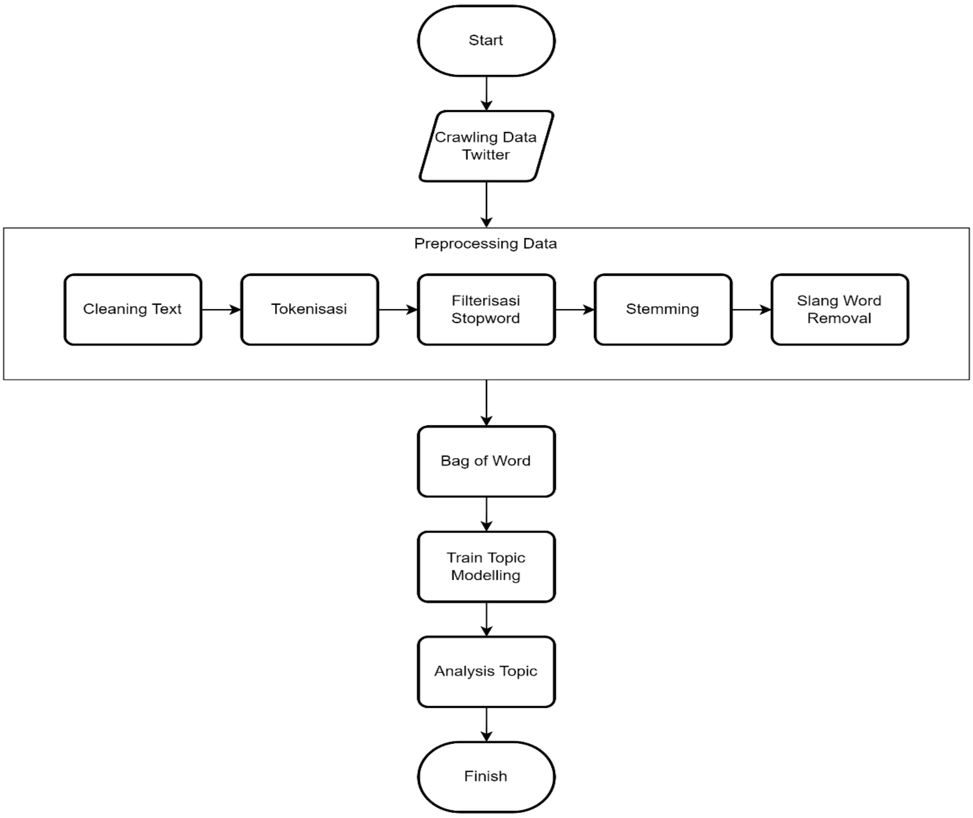


Figure 2: Workflow

DOI: 10.5281/zenodo.11257494

Data Collection/Crawling

The data used in the study was obtained through the crawling process of the social media platform Twitter (X). This data is obtained by entering specific parameters such as health, nutrition, and nutrition, as well as limiting geolocation with a radius range of between 20 to 30 kilometers from seven predetermined geolocation points, namely Surabaya, Malang, Bojonegoro, Madiun, Banyuwangi, Kediri, and Madura. Data capture is done using the Tweet Harvest library, which allows for easier data collection with search parameters such as language, tweet type, and date range. Using these parameters, the data obtained becomes more accurate and in accordance with the purpose of the study.

Preprocessing Data

Tweet data obtained from crawling using tweetharvest needs to be processed before it can be used for topic modeling analysis. Most of Twitter(X)'s data is highly unstructured. There may be typos, slang usage, and grammatical errors. Then the cleanup step is applied to the crawled data, namely government policy tweet data to generate structured data.

There are several steps in preprocessing data including text cleaning, tokenizing, filtering stopword removal, stemming and Slang Word Removal:

1. Cleaning Text
2. Tokenizing
3. Stopword Removal
4. Stemming
5. Slang Word Removal (Non-Standard Word Removal)

Bag of Words

The bag of words is used to model each document by counting the number of occurrences of each word. Each document is represented by a bag of words model that ignores the order of words in the document, the syntactic structure of the document and sentences. The calculated value of the number of occurrences of each word is used in topic modeling.

Topic Modeling With Latent Dirichlet Allocation (LDA)

The topic modeling process aims to obtain the distribution of words that make up a topic and documents with a particular topic. Topic modeling has two stages that are carried out. The first stage is to conduct topic modeling based on adding and subtracting the number of topics. The second stage is to model the topic based on the number of iterations. The results of the two topic modeling were then analyzed by comparing the words of each cluster in the topic and seeing the visualization of the LDA modeling. The topic modeling process can be repeatable over a range of candidates, number of topics, and number of iterations specified. Latent Dirichlet Allocation (LDA) is one model of topic modeling. The LDA topic model is unsupervised machine learning. The model is useful in identifying hidden information in large sets of documents. This method can be solved using python, by first invoking the "LdaModel"

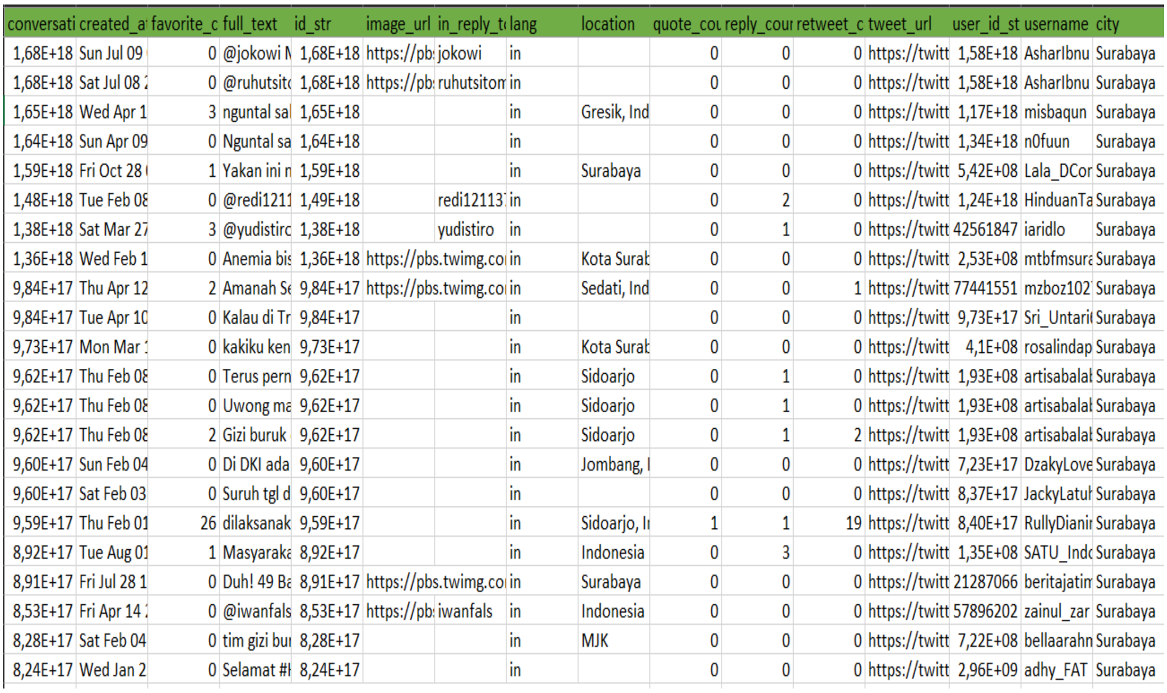
The Seybold Report
ISSN 1533-9211

DOI: 10.5281/zenodo.11257494

package in the gensim library. Package "LdaModel" to model the probability of occurrence of words in the document. Produce output data in the form of graphs that show the topic of the data studied.

Data Collection Results

Making a classification system in this study uses hardware with the following specifications, namely a computer or laptop that has a minimum specification of core i3, 4 GB memory and 500 GB hard drive. In addition, in the process of program development, software such as Python Programming Language and Visual Studio Code are used for more efficient writing and organization of program code. The data used is qualitative data obtained from the results of Twitter Crawling Data on April 20, 2024 with the keywords used "Health", "Nutrition", and "Nutrition" and data taken based on cities that are carsidenan cities in East Java in the category of Indonesian-language tweets as many as 3707 tweets. It then removes duplicate data of tweets so that the data is reduced to 3687 tweets.



conversati	created_a	favorite_c	full_text	id_str	image_url	in_reply_t	lang	location	quote_cou	reply_cour	retweet_c	tweet_url	user_id_st	username	city
1,68E+18	Sun Jul 09	0	@jokowi M	1,68E+18	https://pb:	jokowi	in		0	0	0	https://twitt	1,58E+18	AsharIbnu	Surabaya
1,68E+18	Sat Jul 08 2	0	@ruhutsitc	1,68E+18	https://pb:	ruhutsiton	in		0	0	0	https://twitt	1,58E+18	AsharIbnu	Surabaya
1,65E+18	Wed Apr 1	3	nguntal sal	1,65E+18			in	Gresik, Ind	0	0	0	https://twitt	1,17E+18	misbaqun	Surabaya
1,64E+18	Sun Apr 09	0	Nguntal sa	1,64E+18			in		0	0	0	https://twitt	1,34E+18	n0fuun	Surabaya
1,59E+18	Fri Oct 28 (	1	Yakan ini n	1,59E+18			in	Surabaya	0	0	0	https://twitt	5,42E+08	Lala_DCor	Surabaya
1,48E+18	Tue Feb 08	0	@redi1211	1,49E+18		redi12113	in		0	2	0	https://twitt	1,24E+18	HinduanTa	Surabaya
1,38E+18	Sat Mar 27	3	@yudistirc	1,38E+18		yudistiro	in		0	1	0	https://twitt	42561847	iaridlo	Surabaya
1,36E+18	Wed Feb 1	0	Anemia bis	1,36E+18	https://pbs.twimg.co		in	Kota Surab	0	0	0	https://twitt	2,53E+08	mtbfmsura	Surabaya
9,84E+17	Thu Apr 12	2	Amanah Se	9,84E+17	https://pbs.twimg.co		in	Sedati, Ind	0	0	1	https://twitt	77441551	mzboz102	Surabaya
9,84E+17	Tue Apr 10	0	Kalau di Tr	9,84E+17			in		0	0	0	https://twitt	9,73E+17	Sri_Untari	Surabaya
9,73E+17	Mon Mar 1	0	kakiku ken	9,73E+17			in	Kota Surab	0	0	0	https://twitt	4,1E+08	rosalindap	Surabaya
9,62E+17	Thu Feb 08	0	Terus pern	9,62E+17			in	Sidoarjo	0	1	0	https://twitt	1,93E+08	artisabalab	Surabaya
9,62E+17	Thu Feb 08	0	Uwong ma	9,62E+17			in	Sidoarjo	0	1	0	https://twitt	1,93E+08	artisabalab	Surabaya
9,62E+17	Thu Feb 08	2	Gizi buruk	9,62E+17			in	Sidoarjo	0	1	2	https://twitt	1,93E+08	artisabalab	Surabaya
9,60E+17	Sun Feb 04	0	Di DKI ada	9,60E+17			in	Jombang, I	0	0	0	https://twitt	7,23E+17	DzakyLove	Surabaya
9,60E+17	Sat Feb 03	0	Suruh tgl d	9,60E+17			in		0	0	0	https://twitt	8,37E+17	JackyLatuh	Surabaya
9,59E+17	Thu Feb 01	26	dilaksanak	9,59E+17			in	Sidoarjo, I	1	1	19	https://twitt	8,40E+17	RullyDianir	Surabaya
8,92E+17	Tue Aug 01	1	Masyaraka	8,92E+17			in	Indonesia	0	3	0	https://twitt	1,35E+08	SATU_Indc	Surabaya
8,91E+17	Fri Jul 28 1	0	Duh! 49 Ba	8,91E+17	https://pbs.twimg.co		in	Surabaya	0	0	0	https://twitt	21287066	beritajatim	Surabaya
8,53E+17	Fri Apr 14 2	0	@iwanfals	8,53E+17	https://pb:	iwanfals	in	Indonesia	0	0	0	https://twitt	57896202	zainul_zar	Surabaya
8,28E+17	Sat Feb 04	0	tim gizi bu	8,28E+17			in	MJK	0	0	0	https://twitt	7,22E+08	bellaarahn	Surabaya
8,24E+17	Wed Jan 2	0	Selamat #I	8,24E+17			in		0	0	0	https://twitt	2,96E+09	adhy_FAT	Surabaya

Figure 3: Data Crawling Screenshot

Data Preprocessing Results

To overcome the tendency of unstructured data, data preprocessing is carried out. The initial stage before data preprocessing is to take the tweet column only, which is then carried out cleaning text, tokenizing, stopword removal, stemming, and slang word removal in Indonesian. After the data goes through all the processes, the data that will be obtained is data that is clean and ready to be analyzed. Table 1 is an example of the results of the data preprocessing process.

ISSN 1533-9211

DOI: 10.5281/zenodo.11257494

Table 1: Processed Tweets

Tweet before	Tweet after
Kalau di trenggalek memang ada desa yang mengalami stunting dan gizi buruk mari kita cari solusinya. Jangan ditutup-tutupi #gusipulputimenang	Kalau trenggalek memang desa alami stunting gizi buruk cari solusi jangan ditutup tutupi
#jatim - bocah 8 tahun di jombang alami mikrosefalus dan gizi buruk tubuh tinggal tulang https://t.co/w4qbjlwgcb https://t.co/dub125yuwo	Bocah tahun jombang alami mikrosefalus gizi buruk tubuh tinggal tulang
Yakan ini memang diperuntukan buat balita yg masuk kategori bgm alias gizi buruk kan soalnya tanteku kader posyandu dan tiap mau ada pertemuan posyandu selalu bawa biskuit ini buat jaga^2 ada balita yg bgm	Yakan memang untuk buat balita yang masuk kategori bgm alias gizi buruk kan soal tante kader posyandu tiap mau temu posyandu selalu bawa biskuit buat jaga balita yang bgm

Then the results of tweets displayed visually can be seen in the word cloud visualization as shown in figure 4. In the tweets results column the words that often appear are healthy, nutrition, nutrition, good, less, no, and child. When viewed from the results of visualization, it can be concluded that public opinion in East Java is more directed to health related to the lack of nutrition and good nutrition for children. While in the following picture.

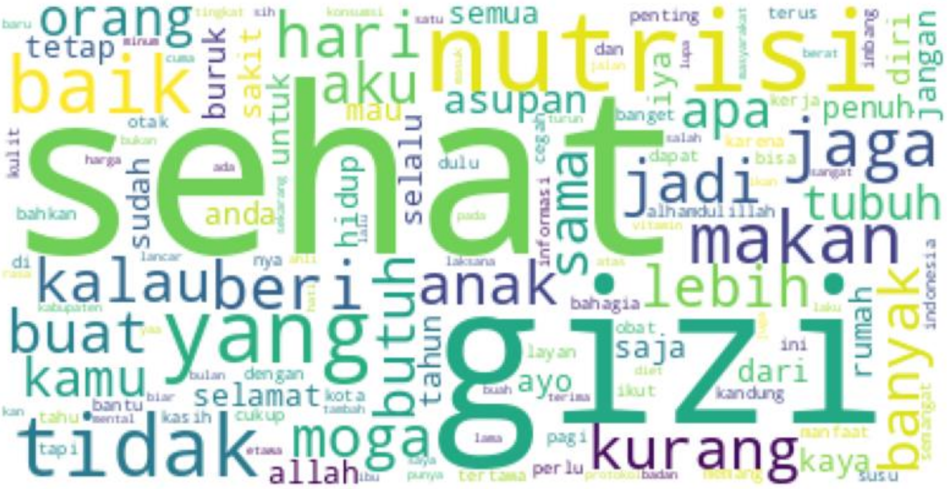


Figure 4: Word cloud view of tweets

Currently, most studies use the obfuscation index to select the optimal number of topics in the LDA model. However, Michael R and Both A et al. have proven that the coherence index is the most consistent measure of human interpretation[7]. Therefore, the topic coherence score is selected as a quantitative index of the optimal number of topics in the model. In general, the higher the topic coherence score, the better the quality of the model. It can be seen from the results in Figure 5 that abscissa is the number of topics, and ordinate is the coherence of topics corresponding to various topics.

The higher the coherence, the more diluted the model. Model interpretation is also acceptable, when the number of topics is found to be 9, then the topic coherence score is highest at 0.410638. In terms of coherence score, the optimal number of topics in the LDA model is 5. To

DOI: 10.5281/zenodo.11257494

prevent too few topics from causing the information in the topic to become too abstract and difficult to interpret, we tested several topics with higher coherence scores, and after many experiments, when the number was 5 then the model was the best, so the number of topics in the model was finally set to 5.

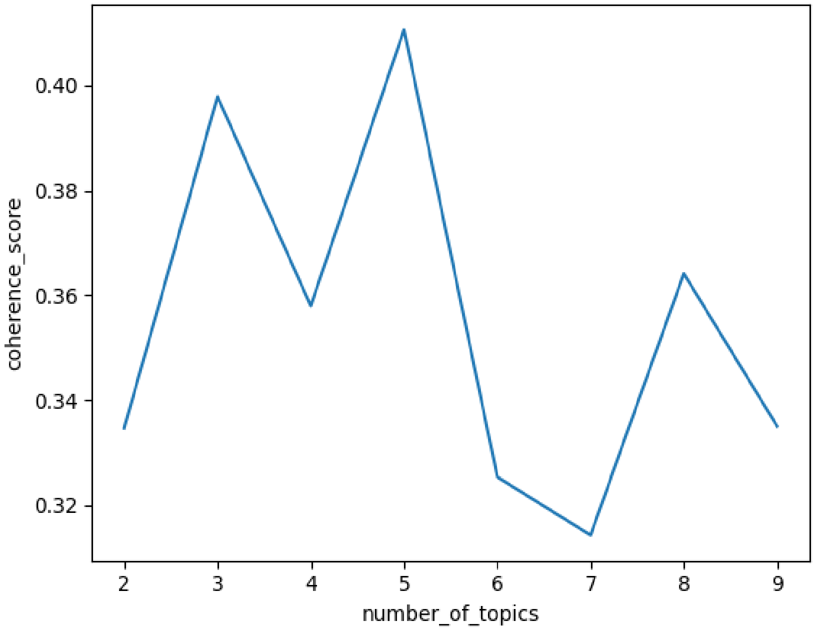


Figure 5: The coherence score of different number topics

Pre-processing data is further processed using the Latent Dirichlet Allocation (LDA) method. The information contained in the data will be interpreted in the form of a collection of main topics. The first step is to convert the tweet data above into a dictionary and bag of words. Map unique words to numeric IDs by utilizing the corpora gesim module by calling the Dictionary module. Next, the bag of word indexes each word and calculates occurrences or probabilities based on unique words specified in the dictionary process. After that, it continued with Latent Direchlet Allocation (LDA) modeling.

Topic Modeling Results With LDA

The parameters used as a reference to produce the best topic model are number of topics and words in topic. Number of topic is the number of topics obtained in one document, while words in topic is the number of words that make up the topic. In this study, Number of topic = 5 and word of topic = 10 were used. The output of topic modeling produces probability values from several words. Probability is the number of times a word appears in a document. Word selection is taken as many as the number of words (word of topic) that have the highest probability. The following Table 2 is the topic output generated from the topic modeling.

DOI: 10.5281/zenodo.11257494

Table 2: Topic Modeling Output

Topik	Probability * Word
0	0.034*"gizi" + 0.014*"makan" + 0.014*"sehat" + 0.011*"nutrisi" + 0.009*"baik" + 0.008*"at" + 0.006*"kurang" + 0.006*"jadi" + 0.005*"rumah" + 0.005*"tidak"
1	0.032*"sehat" + 0.019*"nutrisi" + 0.019*"tidak" + 0.014*"gizi" + 0.010*"yang" + 0.008*"jaga" + 0.007*"makan" + 0.006*"asupan" + 0.006*"buat" + 0.005*"butuh"
2	0.030*"gizi" + 0.026*"nutrisi" + 0.013*"sehat" + 0.013*"baik" + 0.012*"kalau" + 0.010*"tidak" + 0.008*"aku" + 0.006*"makan" + 0.006*"sama" + 0.005*"jadi"
3	0.027*"sehat" + 0.022*"yang" + 0.017*"nutrisi" + 0.011*"gizi" + 0.008*"buat" + 0.007*"baik" + 0.007*"jadi" + 0.006*"protokol" + 0.006*"tidak" + 0.005*"alhamdulillah"
4	0.045*"sehat" + 0.018*"gizi" + 0.010*"moga" + 0.010*"anda" + 0.010*"hari" + 0.010*"beri" + 0.009*"nutrisi" + 0.009*"jaga" + 0.009*"baik" + 0.007*"anak"

From the results of the topic modeling above, researchers conduct an analysis to compile topics based on the probability value of words that arise from the model that has been made. The following in table 3 is the result of the researcher's analysis of the topic modeling results table above.

Table 3: Analysis Topic Modeling Results

Topik	Analysis Results
0	The Importance Of Eating Healthy Food, Nutrition And Good Nutrition At Home
1	Keep The Intake Of Nutrients And Nutrients In Food So That The Body Remains Healthy
2	Do Not Eat The Same Food, So That The Body Gets Good Nutrition And Nutrition
3	Making Good Protocols Helps Maintain Nutrition And Nutrition So Healthy.
4	May Children Always Be Healthy And Maintain Their Nutrition By Providing Good Nutrition Every Day

Visualization of Topic Modeling Results

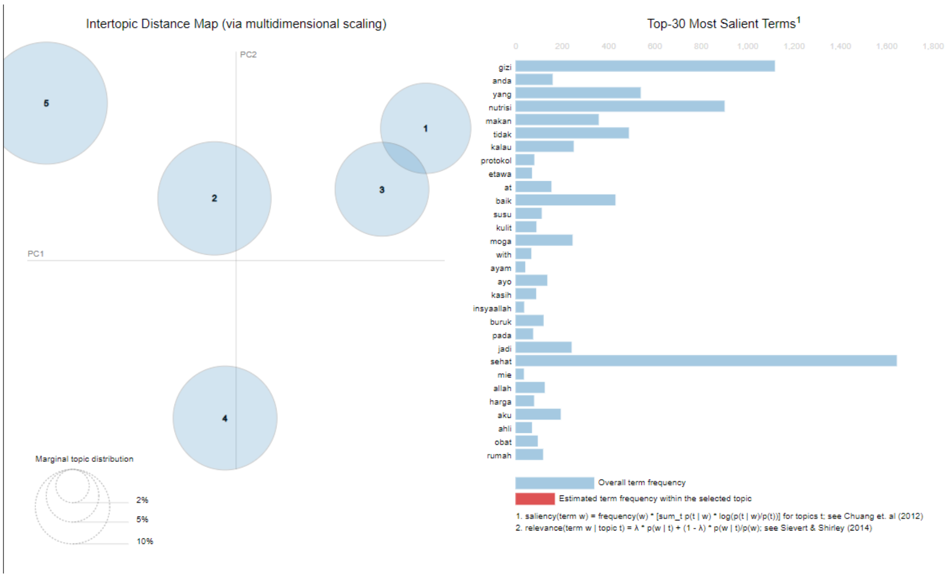


Figure 6: Topic Model Visualisation

The Seybold REPORT
ISSN 1533-9211
DOI: 10.5281/zenodo.11257494

Figure 6 is a result of the topic modeling visualized with the pyLDAvis library. There is a circle depicting the topic. The larger the circle shape indicates that the topic is very influential or often appears in the document. While the words in the right panel are the dominant words discussed in the topic. It consists of 30 terms or words and displays the percentage of word tokens for each topic. The words consist of nutrition, you, which, nutrition, eat, no, kalua, protocol, etawa, at, good, milk, skin, moga, with, chicken, come, love, inshaallah, bad, on, so healthy, noodles, allah, price, me, expert, medicine and home. If the words in the side panel are selected, the circle shape multiplies the shape change. The larger the shape of the circle, indicating that the word appears dominant on the topic.

Grouping Tweets

Here is a grouping of several documents or tweets from preprocessed data, using the LDA model formed above. Documents or tweets used as many as 5 documents or tweets. And the results can be seen in the following table 4:

Table 4: Document or Tweet classification

Tweets	Topik
Masuk angin..perut kembung...mata berkunang2..nutrisi otak tekkor *efficiency effect http://myloc.me/iO1lf	1
Km dulu berapa? RT @mutiarafarah: @Nutrisi_Bangsa: Usia 2 th tinggi bdn anak mnggmbrkan (cont) http://t.co/yui5Jay8	3
Keinget kata katanya kevin dimas pas selesai Sooca pulang kerumah itu memperbaiki nutrisi : D	2
Menguap menunjukkan bahwa otak dan tubuh kita membutuhkan oksigen dan nutrisi	0
hanya dengan juz nutrisi ini berat badan ideal dan kesehatan bisa kalian peroleh http://t.co/UGmm9KmtMO	4

Spread of Topics

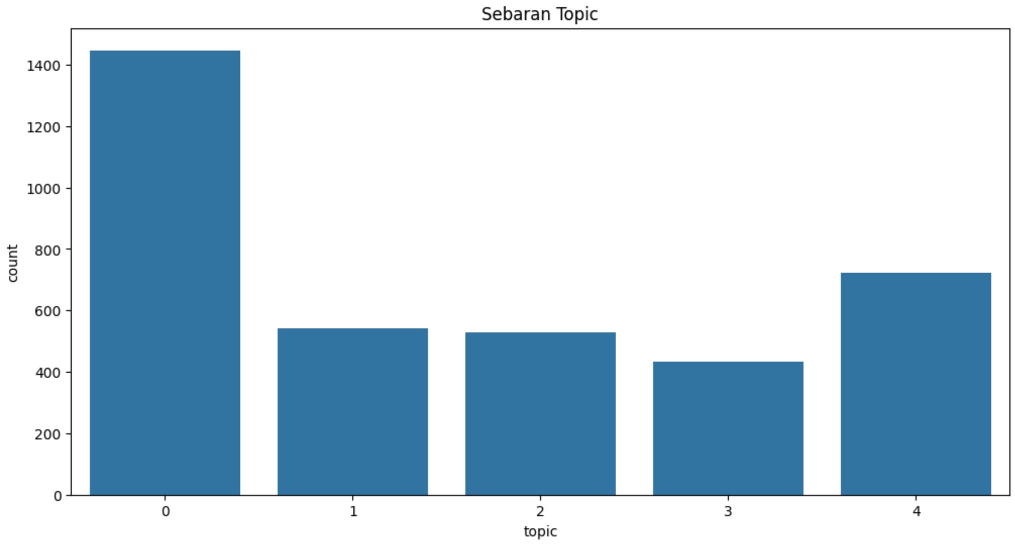


Figure 7: Visualize the Spread of Each Topic on Tweets Data

ISSN 1533-9211

DOI: 10.5281/zenodo.11257494

In figure 7 the following is a visualization or overview of the distribution of data from each topic. The y-axis or ordinate is the number of topic spreads, and the x-axis or abscissa is the number of topics corresponding to various topics. Topic 0 "The importance of eating healthy food, nutrition and good nutrition at home" has a very high data distribution of 1447 Tweets. Topic 1 "Keep your nutritional intake and nutrients in food to keep your body healthy" has a data distribution of 532 tweets. Topic 2 "Not eating the same foods, so that the body gets good nutrition and nutrition" has a data distribution of 528 tweets. Topic 4 "Creating good protocols helps maintain nutrition and nutrition so healthy" had the lowest data distribution of any other topic at 434 tweets. Meanwhile, topic 5 "May children always be healthy and maintain nutrition by providing good nutrition every day" has a data distribution of 723 tweets.

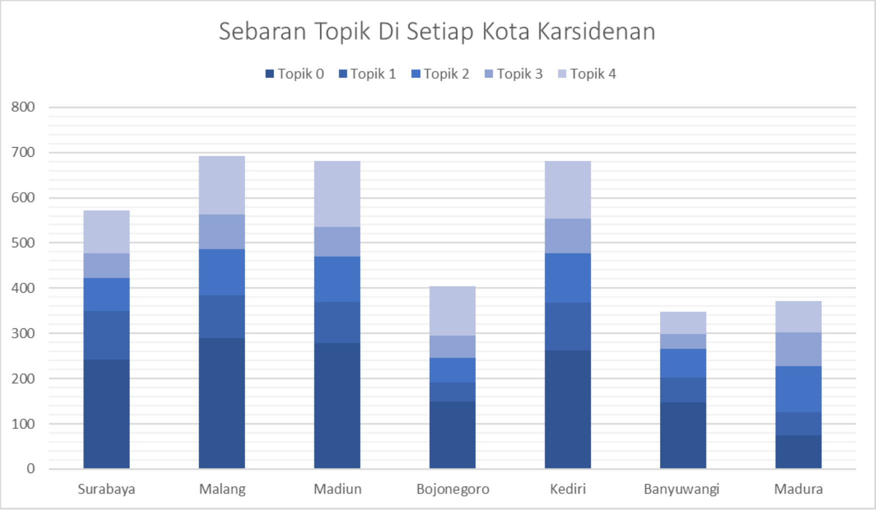


Figure 8: Visualization of the Distribution of Each Topic in Each Carsidenan City

In figure 8, the following is a visualization or overview of the distribution of data from each topic in each carsidenan city in East Java province. The y-axis or ordinate is the sum of the topic spreads, and the x-axis or abscissa is the name of the carsidenan city. Almost all carsidenan cities are Surabaya, Malang, Madiun, Bojonegoro, Kediri and Banyuwangi. Topic 0 ranked highest because it had the highest data distribution and Topic 4 had the lowest data distribution. While in Madura, topic 2 has the highest distribution data and topic 1 has the lowest distribution data of other topics.

CONCLUSIONS

Based on the results of testing and analysis that has been done, it can be concluded that the LDA model can represent topics in tweets documents. So that the LDA method can be used to search or find topics that are being discussed by the community through tweets on twitter (X). With the number of topics found as many as 9 in this study. However, the highest coherence score is in the number of topics as many as 5, with the highest topic coherence score of 0.410638. This shows that the selection of the number of topics is a parameter in finding the best coherence value. Determining the appropriate number of topics will produce a valid and

ISSN 1533-9211

DOI: 10.5281/zenodo.11257494

optimal coherence scor value so that it can be interpreted. This study aims to analyze the opinions of East Java people regarding health, nutrition, and nutrition based on Twitter data. The results showed that there are five main topics related to health, nutrition, and nutrition in East Java. The study's findings have several important implications for the government to assist in helping decide health policies and programs in East Java.

References

1) V. Berliana dan T. A. Arsanti, “Analisis Pengaruh Self-efficacy, Kapabilitas, dan Perilaku Kerja Inovatif terhadap Kinerja,” *J. Maksipreneur Manajemen, Koperasi, dan Entrep.*, vol. 7, no. 2, hal. 149, 2018, doi: 10.30588/jmp.v7i2.364.

2) S. Malian, *Kebijakan Publik Dalam Negara Hukum*, 1 ed. Yogyakarta: Kreasi Total Media, 2021.

3) D. M. Blei, “Probabilistic Topic Models,” vol. 55, hal. 77–84, 2012, doi: 10.1145/2133806.2133826.

4) J. Fernandez, Y. Song, M. Padua, dan P. Liu, “A Framework for Urban Parks: Using Social Media Data to Assess Bryant Park, New York,” *Landsc. J.*, vol. 41, no. 1, hal. 15–29, 2022, doi: 10.3368/lj.41.1.15.

5) W. L. Lim, C. C. Ho, dan C. Y. Ting, “Sentiment analysis by fusing text and location features of geo-tagged tweets,” *IEEE Access*, vol. 8, no. September, hal. 181014–181027, 2020, doi: 10.1109/ACCESS.2020.3027845.

6) M. Alsaqer, S. Alelyani, M. Mohana, K. Alreemy, dan A. Alqahtani, “applied sciences Predicting Location of Tweets Using Machine Learning Approaches,” 2023.

7) H. Jelodar *dkk.*, “Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey,” 2018.