Research Topic Modeling in Informatics Engineering Study Program at Nusa Putra University using LDA method

Kamdan [a,1*], Ivana Lucia Kharisma [a,2], Gina Purnama Insany [a,3], Paikun [b,4]

[a] Department of Informatic Engineering , Nusa Putra University, Jl. Raya Cibolang Kaler No.21, Kab. Sukabumi 43152, Indonesia
[b]Department of Civil Engineering, Nusa Putra University, Jl. Raya Cibolang kaler No.21 Kab. Sukabumi, Indonesia
[1] kamdan@nusaputra.ac.id*; [2] ivana.lucia@nusaputra.ac.id; [3] gina.purnama@nusaputra.ac.id; [4] paikun@nusaputra.ac.id
* Corresponding Author email: kamdan@nusaputra.ac.id

Received 28 October 2022; revised 02 November 2022; accepted 05 November 2022

ABSTRACT

Writing research reports at the undergraduate level is one of the obligations that must be fulfilled by students as a fulfillment of graduation requirements at a university. One of the independent learning programs implemented at Nusa Putra University is through the research method, where students are required to conduct research as a graduation requirement in the Study Completion Program course. The growing development of information and communication technology provides opportunities for students to determine research themes. However, sometimes students take research themes that are not in accordance with the concentration in the study program. This research was conducted with the aim of identifying how the LDA topic modeling method can analyze research topic trends by modeling topics on research titles that have been taken by students at the Informatics Engineering Study Program, University of Nusa Putra. Latent Dirichlet Allocation (LDA) is one of the most popular topic modeling methods today. This research uses a dataset in the form of 159 titles of study completion program research reports and tittles of final assignment reports for students of the Informatics Engineering study program, University of Nusa Putra. This research is expected to be a reference in conducting research by students based on the topics that have been modeled

KEYWORDS
LDA
Topic Modelling
Research
Clustering
Research title data collection

1. Introduction

In the world of education, research is one of the important points that becomes a medium for students to develop attitudes, be critical of surrounding problems and find solutions to existing problems by applying the insights and knowledge gained [1]. The progress of research in an educational institution can be seen from the quantity and quality as well as the direction of the themes and research topics discussed [2]. The suitability of the research theme with the concentration of the department in a study program can be a benchmark for achieving the performance of the study program [3]. At Nusa Putra University, the Informatics Engineering Study Program has included the vision, mission and profile of the graduates produced in the study program. This graduate profile is formulated by taking into account the program curriculum, alumni tracking and analysis of future industry needs [4]. There are 5 graduate profiles in the Informatics Engineering study program at Nusa Putra University, namely software developers, computer network and computer security specialists, data scientists and data analysts, database specialists and IT academics. To see the themes and research topics that have been made by students, whether they are in accordance with the concentration of the study program or the achievement of the profile of graduates in a study program, it takes time and energy because you have to read the existing research one by one [5]. Therefore, an analysis is needed to find the concentration of research topics by modeling research topics by analyzing existing research titles so that the directions and trends of research carried out by students can be seen by applying the Latent Dirichlet Allocation (LDA) method [6].

The main topics are determined to be the distribution of word sets using statistical calculations [7]. The LDA method is quite popular as a method for determining themes or topics in documents. The LDA method is used in determining research topics in the Sinta indexed national journal repository about the field of nursing [8]. The Bayesian method and the LDA method are used in the analysis of news trends in the community related to the Surabaya bombing [9]. In this study the resulting topic probability value is 0.10057. Topic modeling on social media such as Twitter classifies existing opinions using Latent Dirichlet

Allocation (LDA) data obtained from Twitter [10]. The LDA method can also be applied to classify topics or themes semantically from consumer complaints submitted to the Consumer Financial Protection Bureau in America [11]. From the data contained on Twitter, it is possible to detect urban emergency data in the city of Virginia by applying the LDA method [12]. The research proposed by [13] uses the LDA method to analyze whether there is a match between the topic and the research domain at the University of Kentucky. Various studies have also been conducted on the development of existing topic modeling algorithms, which show the rapid impact of this LDA method in the field of Natural Language Processing (NLP) [14]. Modeling the topics contained in the Google Scholar link by JPTEI UNY lecturers was analyzed using the LDA method [15].

In this study, modeling the topic title of the research conducted by students of the Informatics Engineering study program at Nusa Putra University will use the Latent Dirichlet Allocation (LDA) method to see the suitability of the research with the concentration of the Informatics Engineering study program at Nusa Putra University. It is a state of the art of research [8], [13], [15]. The results of the research will show what research topics the students made.

The purpose of this study was to create a data collection system for themes and research titles that special students of the Informatics Engineering study program at Nusa Putra University had carried out. The results of this research will be helpful for study program management and students who will conduct research. In addition, this research can contribute to identifying the achievement of research themes based on the research roadmap of the Informatics Engineering study program at Nusa Putra University and provide references about this data collection to other universities internationally.

2. Method

In this study, the research stages used can be seen in Fig.1. The research stages were carried out by conducting a literature study, looking for references from various sources regarding the application of the LDA algorithm in topic modeling and the use of this method in solving problems related to natural language processing [16].

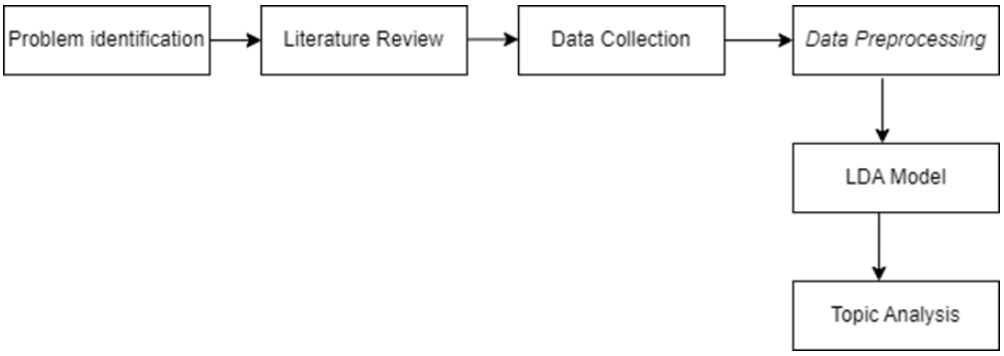


Fig. 1. Research Stages

2.1. Identification of Problems

The importance of a research activity in the academic world, which is carried out by students either as a learning activity in writing a research report or as a condition for completing lectures, makes the importance of this research being carried out [17]. By analyzing the title of research conducted in a study program, trends and research topics will be obtained [18]. By knowing the research topic, the study program can evaluate and assess whether there is compatibility between the research conducted and the concentration of the study program [19]. The results of this evaluation can be used as a reference for improving learning activities in the study program [20]. Improvement activities can be in the form of changes or revisions of existing courses to be adapted to existing technological developments, especially in the field of information technology which is developing very fast [21].

2.2. Literature Review

In this study, data and library sources were collected from various sources in the form of journals, documentation books, the internet, and libraries. References related to the definition of the topic modelling method with Latent Dirichlet Allocation, as well as the application of this method in various fields.

2.3. Data Collection

The data taken from this study were taken from research titles written by students of the Informatics Engineering study program at Nusa Putra University, both those which were the titles of research activity reports on the Study Completion Program chosen by students, as well as from the title of thesis writing. The data collected is taken from document sources stored in the study program data storage in the form of soft files and report results in the form of hard files. Data is stored in excel worksheet format with 159 research titles.

2.4. Pre-Processing Data

Data Pre-Processing is a step that must be carried out to process raw data that we have obtained from data sources into data that is of high quality or has significant value. The irregularity of the data format is also the reason for pre-processing the data [22], the process is needed to produce the right data clustering. The data pre-processing stages carried out in this study include case folding, tokenizing, and stop words [23].

2.4.1. Case Folding

This stage the process of changing capital letters to lowercase letters is carried out [24].

2.4.2. Tokenizing

This stage is the process of cutting a sentence or string into words which in the process of cutting are done on whitespace or spaces and removing punctuation marks [25].

2.4.3. Stop word

This stage the process of selecting words that are considered to represent documents from the token results by eliminating words that are considered meaningless such as question words, interjections, conjunctions. After the stop word process, an additional process is carried out, namely lemmatization, the process of turning a word into a basic word by knowing the context of the word, then changing the composition of a sentence into a trigram form, for example: "Design and Build E-Learning Applications Based on Progressive Web Apps", when changed to bigram: [design, build, app], [progressive, web, apps] [26].

2.5. LDA Modeling

Topic modeling is an unsupervised machine learning method that applies clustering to discover latent variables from large text data. The most popular method for topic modeling is Latent Dirichlet Allocation (LDA) which was introduced by Blei and Jordan, described as a generative probabilistic model to look for the semantic structure of a corpus set based on hierarchical Bayesian analysis [27]. An LDA is a collection of mixed-topic documents that contain words with a certain probability. The workflow of LDA can be seen in Fig. 2 below [28]:

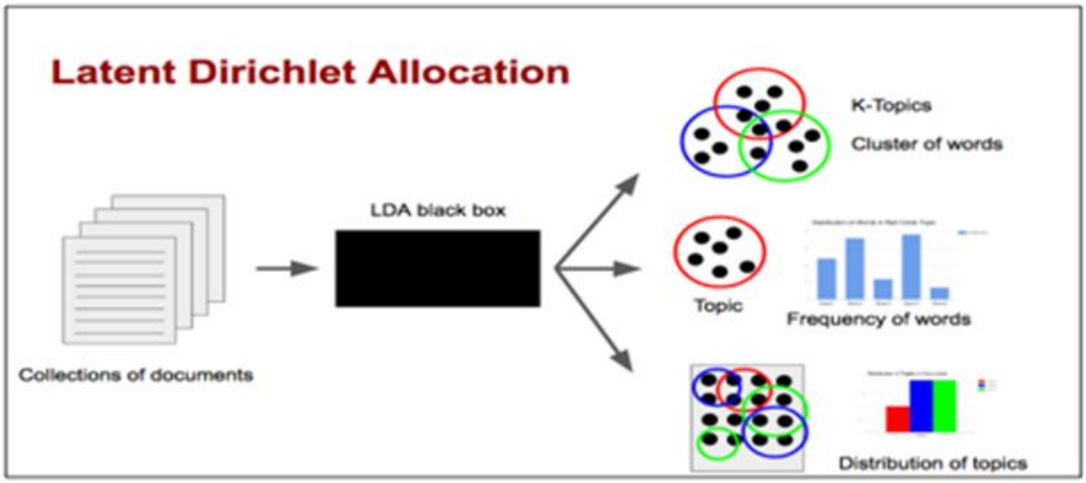


Fig. 2. LDA Modeling

The procedure for how LDA works is as follows: (1) Initialize several parameters, including the number of documents, topics, and iterations. In LDA, the most important parameter is the number of k topics. (2) Assign words to certain topics randomly according to the Dirichlet distribution. (3) Repeat each process flow for all words in the corpus. The parameters used when calculating the LDA are as follows:

(1) Random state: 100; (2) Update Every: 1; (3) Chunk Sizes: 10; (4) Passes: 10; (5) Alpha: Symmetric; (6) Iterations: 100; (7) Per Word Topics: True [29]. This study used the Latent Dirichlet Allocation (LDA) method to see the suitability of the research with the concentration of the Informatics Engineering study program at Nusa Putra University.

2.6. Topic Analysis

This research builds knowledge of student research networks in the Informatics Engineering study program at Nusa Putra University, which is then further analyzed to see trends in a research topic. The output data from the topic modeling process will be analyzed subjectively. The output data will be a reference for adjusting existing documents. This process produces informative topic descriptions about things that can represent the contents of each of these topics.

3. Results and Discussion

In this chapter, the results of the research based on the steps taken in the previous chapter will be explained. In carrying out the topic modelling analysis process, steps are needed in a coherent manner to produce good modelling results. Following are the results of the implementation of the stages of the research method carried out:

3.1. Data Collection

According to the explanation at the data collection stage, taken from the research title, both the results of thesis writing and the results of student research on Student Completion Program (SCP) activities. As many as 159 data were collected from research from 2020 – 2022. The data has 3 attributes, including the year which was the year the research was written, student name and title. The focus of the research is on the title attribute. An example dataset is shown in Table 1.

Table 1. Research Title Dataset

Year	Student	Title
2022	DINI AGNIA HARDIANTY	RANCANG BANGUN APLIKASI E-LEARNING BERBASIS PROGRESSIVE WEB APPS UNTUK MENUNJANG PEMBELAJARAN ONLINE MENGGUNAKAN METODE PROTOTYPING"
2022	Rizwan Gustama	SISTEM INFORMASI GEOGRAFIS (SIG) PEMETAAN KRIMINALITAS BERBASIS WEB
2022	Isma Mahdaniyah	RANCANG BANGUN APLIKASI UDADI DI PT.OMIND MUDA BERKARYA INDONESIA
2022	Andriana	PEMANFAATAN CUSTOMER RELATIONSHIP MANAGEMENT (CRM) PADA INDEKOS
		DALAM MENINGKATKAN PELAYANAN PELANGGAN DIMASA PANDEMI COVID-19 BERBASIS WEB

There is a non-uniformity of the writing on the author's name and research title as in the example in Table 1 this is the original data in the Informatics Engineering study program at Nusa Putra University, so it needs to be processed so that it becomes uniform.

3.2. Pre-Processing Data

- The first step is to change uppercase to lowercase using Case Folding. The result of the Case Folding process is that all uppercase letters in the title data are changed to lowercase. The results of data changes after the case folding process is shown in Table 2.

Table 2. Data after Case Folding Process

Year	Student	Title
2022	DINI AGNIA HARDIANTY	rancang bangun aplikasi e-learning berbasis progressive web apps untuk menunjang pembelajaran online menggunakan metode prototyping
2022	Rizwan Gustama	sistem informasi geografis (sig) pemetaan kriminalitas berbasis web di kota sukabumi
2022	Isma Mahdaniyah	rancang bangun aplikasi udadi di pt.omind muda berkarya indonesia

2022	Andriana	pemanfaatan customer relationship management (crm) pada indekos dalam meningkatkan pelayanan pelanggan dimasa pandemi covid-19 berbasis web

Example of data from research titles contained in Table 2, which consists of uppercase letters have been changed to lowercase letters in the case folding stage. Changes in uppercase letters to lowercase letters are only made on the data contained in the title column, because the data in column student name is not processed in the research of topic modelling.

- After the Case Folding stage, the title is then pre-processed with the tokenizing stage, where the titles in the form of sentences will be separated into entities called tokens, which can be words, numbers, symbols. The results of the Tokenizing process are shown in Table 3:

Table 3. Data after the Tokenizing Process

Year	Student	Title
2022	DINI AGNIA HARDIANTY	'rancang', 'bangun', 'aplikasi', 'learning', 'berbasis', 'progressive', 'web', 'apps', 'untuk', 'menunjang', 'pembelajaran', 'online', 'menggunakan', 'metode', 'prototyping'
2022	Rizwan Gustama	'sistem', 'informasi', 'geografis', 'sig', 'pemetaan', 'kriminalitas', 'berbasis', 'web', 'di', 'kota', 'sukabumi'
2022	Isma Mahdaniyah	'rancang', 'bangun', 'aplikasi', 'udadi', 'di', 'pt', 'omind', 'muda', 'berkarya', 'indonesia'
2022	Andriana	'pemanfaatan', 'customer', 'relationship', 'management', 'crm', 'pada', 'indekos', 'dalam', 'meningkatkan', 'pelayanan', 'pelanggan', 'dimasa', 'pandemi', 'covid', 'berbasis', 'web'

It can be seen in table 3, that the title data which was previously in the form of sentences has separated into words where each word has a single meaning.

- After going through the case folding and tokenizing stages, the title data will go through the preprocessing stage which consists of Stop Word Process, Lemmatization and Trigrams. Stop Words aims to remove words that are less meaningful for topic modeling such as question words and conjunctions. After going through the stop words stage, lemmatization which is the process of turning a word into a basic word by knowing the context of the word implemented in the title data, then the data is processed using the Trigram technique which will combine 3 items in sequence. The results of the stop words, lemmatization and trigram processes are shown in Table 4

Table 4. Data after Stop Words Process, Lemmatization, Trigrams

Data
['geografis', 'berbasis', 'web', 'sukabumi'],
['omind', 'muda', 'berkarya'],
['text', 'recogniter'],
['learn', 'berbasis', 'progressive', 'web', 'app'],

In table 4, the results of the data preprocessing which consists of the stop words, lemmatization and trigram stages show that there is a grouping of title data that has been selected for words that are less meaningful in the stop words process, changing words taking into account the context of meaning and grouping 3 words sequentially.

3.3. LDA Topic Modeling

Modeling uses Latent Dirichlet Allocation using a dictionary derived from words derived from existing document data, then using genism tools in order to provide good functionality in terms of API and computational efficiency [14], [30]. The number of topics that are dominant in the research titles of Informatics Engineering study program students, by determining the number of topics as much as 5 based on the achievements of graduates from the Nusa Putra University Engineering study program to look for correlations between research titles whether they have included the concentration of study programs to achieve these graduate achievements. Following are the graphical results of the coherence value of each word produced which can be seen in Table 5 below, where the coherence value of the word shows the value of the level of semantic similarity between words in one topic.

ISSN 2620-9632
International Journal of Engineering and Applied Technology (IJEAT)
Vol. 5., No. 2, November 2022, pp. 24-35

Table 5. . Results of Coherence Values on Each Topic

Topic	Word
Topic 0	Word : '0.226*"berbasis" + 0.140*"web" + 0.034*"waterfall" + 0.022*"app" + 0.021*"laporan" + 0.019*"tenaga" + 0.019*"pendampe" + 0.019*"medium" + 0.019*"multimedia" + 0.014*"siakad"
Topic 1	Word : '0.036*"cipher" + 0.027*"scale" + 0.027*"siakad" + 0.022*"bucket" + 0.022*"htb" + 0.022*"hierarchical" + 0.022*"token" + 0.022*"router" + 0.021*"analisis" + 0.019*"membangun"
Topic 2	0.169*"sistem" + 0.058*"website" + 0.030*"informasi" + 0.028*"smk" + 0.027*"kabupaten" + 0.024*"architecture" + 0.020*"management" + 0.018*"codeigniter" + 0.018*"ujian" + 0.018*"framework"
Topic 3	'0.107*"monitor" + 0.056*"untuk" + 0.033*"pemakaian" + 0.030*"server" + 0.028*"dengan" + 0.026*"fuzzy" + 0.024*"desa" + 0.023*"sukamanis" + 0.021*"bkynk" + 0.020*"aplikasi"
Topic 4	'0.125*"berbasis" + 0.115*"android" + 0.082*"internet" + 0.075*"thing" + 0.055*"pada" + 0.034*"sukabumi" + 0.030*"smart" + 0.023*"service" + 0.015*"simple" + 0.013*"recognition"

On topic 0, the word that has the highest level of coherence is based, which means that the document included in topic 0 has the highest semantic similarity with the word based. For topic 1, the highest coherence value is in the word cipher, for topics 2, 3 and 4, the highest coherence value is in the word system, monitor and based. Even though there are similarities in the coherence results in topic 0 and topic 4, we can see the coherence values in the following words which have different meanings

3.4. LDA Modeling Visualization

After the topic modeling stage with the LDA method, visualization is carried out using the pyLdavis library. With this library, 30 important words appear in the corpus [8]. The panel on the right shows these important words. We can see that the right panel shown in Figure 3 forms a visualization of each topic and the words that appear the most in research titles include based, system, web, android, monitor, internet, thing, website, on, for, cipher, sukabumi, waterfall, usage, information, server, smart, smk, siakad, district, with, scale, fuzzy, architecture, village, token, htb, router, hierarchical, bucket appear in corpus.

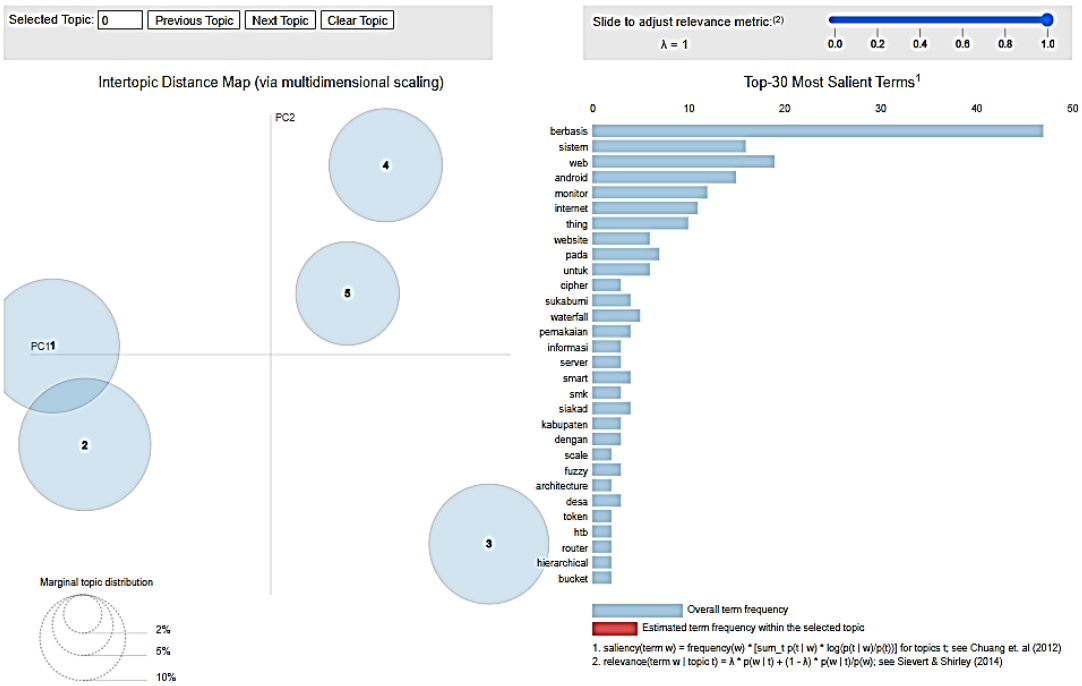


Fig. 3. . Topic Visualization with Pyldavis

In Figure 3 represents a visualization of the topic modeling results where it can be seen that for the selected topic 0, the right panel will contain the order of words with the highest to lowest level of coherence in the topic selected.

Word cloud, which is also called tag cloud, is also used in the visualization of LDA modeling in this study. With the word cloud, you can see the metadata description of a collection of keywords or keywords or tags in a document to provide visualization of topics in text form. The word cloud visualization in this study is based on the results of the LDA modeling in Table 5 so that it is illustrated that there are 5-word cloud visualizations shown in Fig. 4 below:

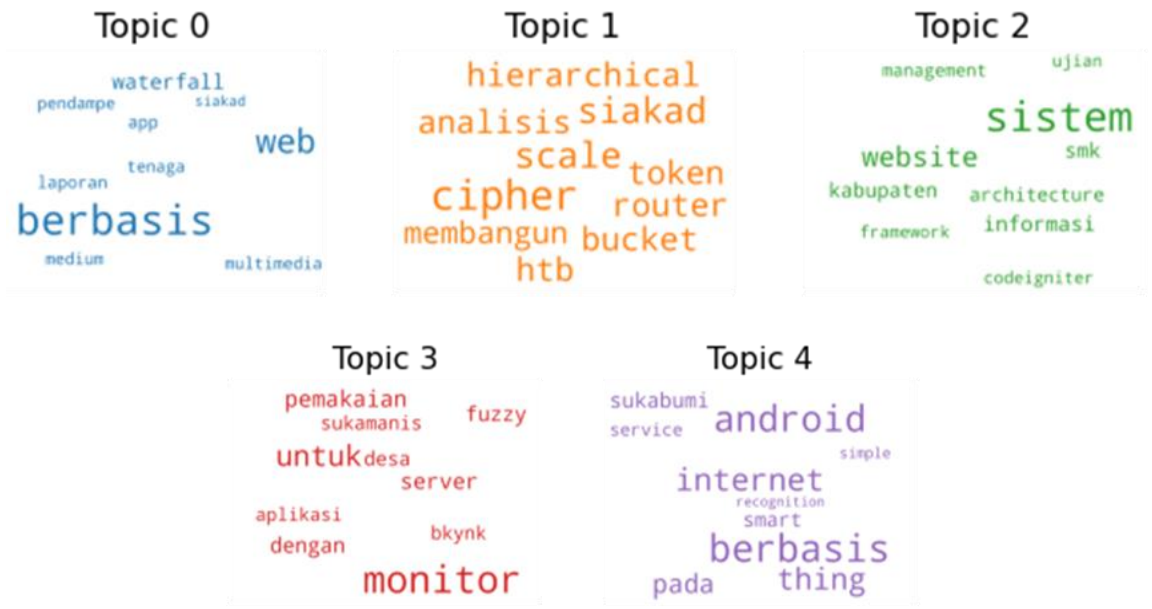
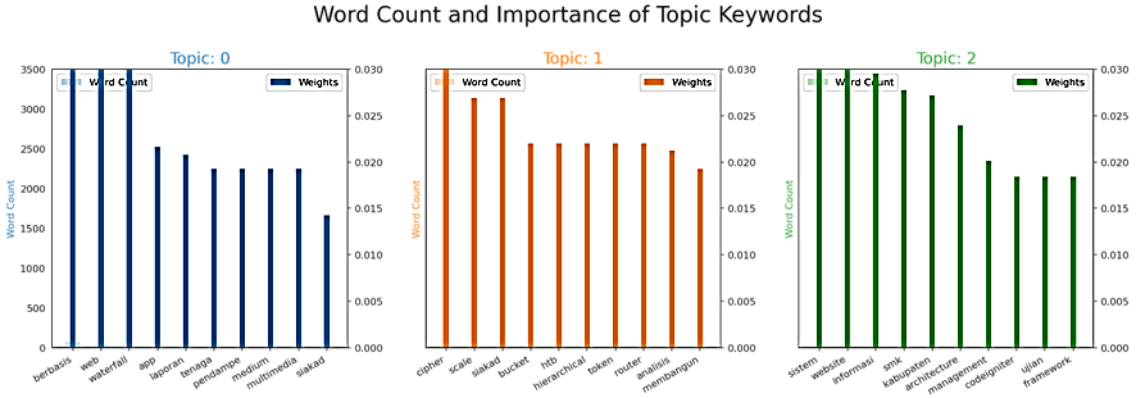


Fig. 4. Visualization with Wordcloud

From the visualization shown in Fig.4, it can be seen that topic 0 contains words about based, waterfall, siakad, web, reports, personnel, medium, multimedia. Important words related to the Informatics Engineering study program are based, waterfall and web. While topic 1 contains the words hierarchical, siakad, analysis, scale, cipher, token, router, build. Some important words that can be associated with the study of Informatics Engineering are hierarchical, analysis, construct. Topic 2 contains words management, exam, system, website, architecture, framework, information. Some important words that can be associated with the study of informatics engineering are system, information, website. In topic 3 we can see words like use, fuzzy, for, like sweet, server, application, monitor. Some important words that we can relate to the study of informatics engineering are fuzzy, application and server. Whereas in topic 4 there are the words Sukabumi, service, android, internet, based, things. Some important words that can be associated with the study of informatics engineering are android and Internet of Things (IoT).

The above interpretation is strengthened by visualizing the number of each word that appears on each topic shown in Fig. 5.



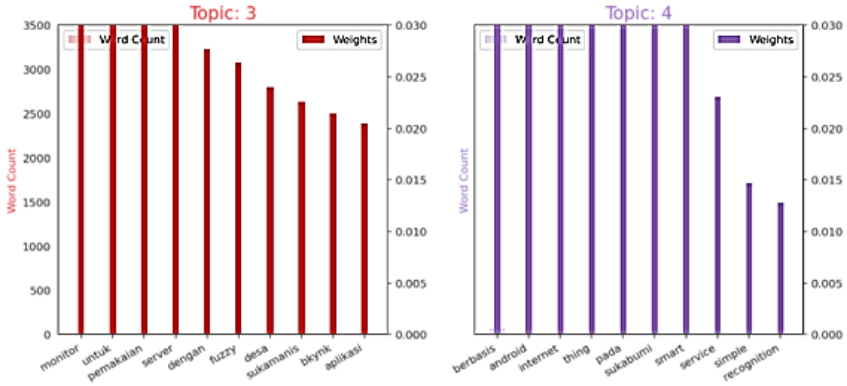


Fig. 5. Word Count Visualization

From Fig.5 we can see a visualization of the number of words that appear most frequently on each topic. The word that has the greatest number of occurrences in each cluster includes Cluster 1 (Topic 0) is Based, Web, Waterfall. The 2nd cluster (Topic 1) is Cipher, Scale, Siakad. The 3rd cluster (Topic 2) is System, Website, Information. The 4th cluster (Topic 3) is monitor, for, usage, server. The 5th cluster (Topic 4) is Based, Android, Internet, Thing, Pada, Sukabumi, Smart. Of all the clusters as a whole we can conclude that the words that are most often in research titles and become topics in research at the Informatics Engineering study program at Nusa Putra University are Based, Waterfall, Web, Siakad, System, Information, monitor, server, android and internet of thing.

The following are some examples of research titles that can be seen in each cluster or topic. Research titles on the five topics can be seen in Tables 6, 7, 8, 9 and 10. Research titles included in Cluster 1 (Topic 0) can be seen in Table 6 below:

Table 6. Research Title in Cluster 1

Cluster 1 (Topic 0)
RANCANG BANGUN APLIKASI E-LEARNING BERBASIS PROGRESSIVE WEB APPS UNTUK MENUNJANG PEMBELAJARAN ONLINE MENGGUNAKAN METODE PROTOTYPING
SISTEM INFORMASI GEOGRAFIS (SIG) PEMETAAN KRIMINALITAS BERBASIS WEB DI KOTA SUKABUMI
Perancangan sistem informasi geografis pemetaan pertahanan desa berbasis web dengan metode waterfall
Sistem Informasi Manajemen Rekrutmen Karyawn Di PT Pratama Abadi Industri (JX)
Sistem informasi layanan mahasiswa untuk surat menyurat berbasis web menggunakan extreme programing di program studi teknik informatika universitas nusa putra

The research titles included in the 2nd Cluster (Topic 1) can be seen in Table 7 below:

Table 7. Research Title in Cluster 2

Cluster 2 (Topic 1)
RANCANG BANGUN APLIKASI UDADI DI PT.OMIND MUDA BERKARYA INDONESIA
"PENENTUAN PENERIMAAN BEASISWA DI UNIVERSITAS NUSA PUTRA MENGGUNAKAN METODE PROFILE MATCHING"
Analisis sentimen untuk opini olah raga bulu tangkis dan sepak bola pada microblogging menggunakan k-nearest neighbour
Perancangan Single Sign On (Studi Kasus SMK AR Rahmah)
Analisis sentimen terhadap review bank digital pada googel play store menggunakan metode support vector machine (SVM)

Table 8. Research Title in Cluster 3

Cluster 3 (Topic 2)
Sistem Informasi Geografis Pemetaan Bencana Pergerakan Tanah Kabupaten Sukabumi Menggunakan Metode Prototype
sistem pendistribusian semen di CV. Indosatu dengan pendekatan supply Chain management
Pengembangan Aplikasi Speech berbasis android menggunakan metode Learning Vector Quantization Dalam Optimalisasi Komunikasi Tunarungu
Implementasi Gamifaction pada Sistem Informasi Perpustakaan Berbasis Web Guna meningkatkan Minat Baca Siswa (Studi Kasus SMAN 1 Parungkuda)
Rancang bangun sistem informasi E-Voting pemilihan ketua osis SMAN 1 Warungkiara menggunakan metode rapid Application Development (RAD) Berbasis Website

The research titles included in the 4th Cluster (Topic 3) can be seen in Table 9 below:

Table 9. Research Title in Cluster 4

Cluster 4 (Topic 3)
Sistem Monitoring dan Pengatur Suhu Otomatis Menggunakan Metode Fuzzy Logic untuk kandang Ayam di desa Sukamanis Berbasis Internet of things
Klasifikasi Citra untuk pengelompokan Sampah dengan menggunakan covolution Neural Network
Aplikasi Pemilihan Jalur evakuasi di tempat Patiwisata dengan algoritma Floyd Warshall berbasis web studi kasus Bukit belendung Desa Cisarua Kecamatan Sukaraja
Rancang Bangun Sistem monitoring Pemakaian daya listrik pada alat ruma tangga berbasis IOT dan Android
Implementasi network monitoring system dengan menggunakan raspbery PI sebagai server monitoring

The research titles included in the 5th Cluster (Topic 4) can be seen in Table 10 below:

Table 10. Research Title in Cluster 5

Cluster 5 (Topic 4)
Rancang bangun automatic Liquid Filling Machine berbasis Internet Of Things Menggunakan Nodemcu dan telegram
Rancang Bangun Smart plant Pada Tanaman Blacksapote Berbasis Intenet Of Things
PERANCANGAN SISTEM MONITORING PENANAMAN HIDROPONIK MENGGUNAKAN SISTEM NUTRIENTI FILM TECHNIQUE BERBASIS INTERNET OF THINGS
PERANCANGAN ALAT KONTROL PADA SUTTER KAMERA DSLR BERBASIS INTERNET OF THINGS
SISTEM KONTROL KETINGGIAN AIR TANDON BERBASIS INTERNET OF THINGS (IOT)

The grouping of 5 clusters shows a visualization of the relationship between clusters and topics. Visualization of data grouping is shown by a scatter plot as shown in Fig. 6 below:

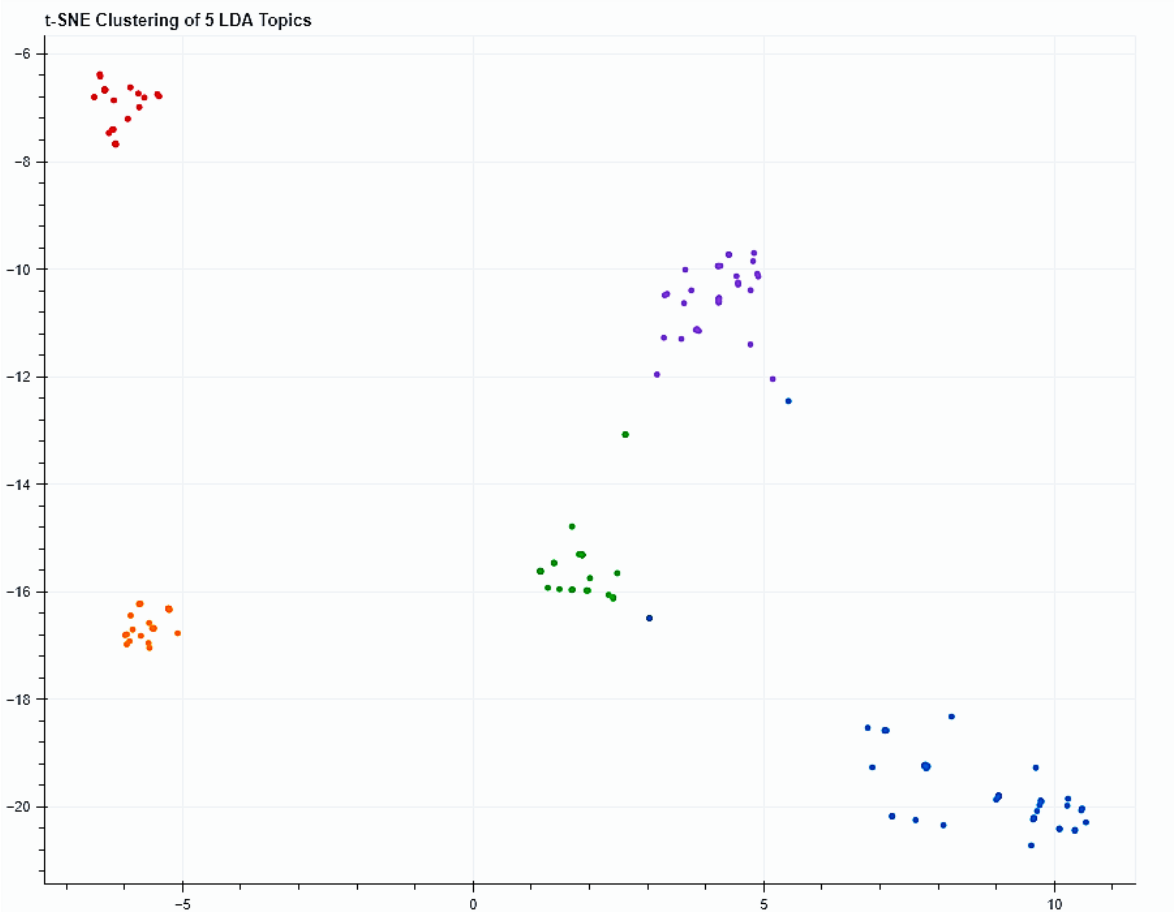


Fig. 6. Visualization of Clusters with Scatter Plots

4. Conclusion

Modeling research topics from research title data either in the form of final assignments or theses as well as student research activities in the Informatics Engineering study program can be well modeled using the Latent Dirichlet Allocation method. Modeling can also be visualized using word clouds, scatter plot and pyLDAvis. The results of the topic of word distribution on research titles in the informatics engineering study program are about web-based, waterfall, hierarchical, system, information, monitor, internet of things, applications, and android. Based on the spread of these words, it can be concluded that the titles and research topics taken by students are in accordance with the concentration of the Informatics Engineering study program. The weakness of this research so that it can be improved for further research is the need for a special approach for non-Indonesian texts in research titles in the pre-processing process.

References

[1] P. F. Lazarsfeld, "Remarks on administrative and critical communications research," *Zeitschrift für Sozialforsch.*, vol. 9, no. 1, pp. 2–16, 1941. Doi. 10.5840/zfs1941912

[2] N. A. Ghani *et al.*, "Bibliometric analysis of global research trends on higher education internationalization using Scopus database: Towards sustainability of higher education institutions," *Sustainability*, vol. 14, no. 14, p. 8810, 2022. Doi. 10.3390/su14148810

[3] D. Sharma, R. Taggar, S. Bindra, and S. Dhir, "A systematic review of responsiveness to develop future research agenda: a TCCM and bibliometric analysis," *Benchmarking An Int. J.*, 2020. Doi. 10.1108/BIJ-12-2019-0539

[4] A. C. Albina and L. P. Sumagaysay, "Employability tracer study of Information Technology Education graduates from a state university in the Philippines," *Soc. Sci. Humanit. Open*, vol. 2, no. 1, p. 100055, 2020. Doi. 10.1016/j.ssaho.2020.100055

[5] M. DeJonckheere and L. M. Vaughn, "Semistructured interviewing in primary care research: a balance of relationship and rigour," *Fam. Med. community Heal.*, vol. 7, no. 2, 2019. Doi. 10.1136/fmch-2018-000057

[6] R. Huang, H. Taubenböck, L. Mou, and X. X. Zhu, "Classification of settlement types from Tweets using LDA and LSTM," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 6408–6411. Doi. 10.1109/IGARSS.2018.8519240

[7] N. N. Hidayati, P. Damayanti, and A. Z. Arifin, "Identification of Traffic Information on Twitter Data using Topic Modeling and Entity Recognition," *J. Linguist. Komputasional*, vol. 4, no. 1, pp. 1–7, 2021. Avaliable at Google Scholar

[8] Y. Sahria, N. I. Febriarini, and P. D. Oktavianti, "PEMODELEN TOPIK PENELITIAN BIDANG KEPERAWATAN INDONESIA PADA REPOSITORY JURNAL SINTA MENGGUNAKAN METODE TOPIC MODELLING LDA (LATENT DIRICHLET ALLOCATION)," in *SEMASTER: Seminar Nasional Teknologi Informasi & Ilmu Komputer*, 2020, vol. 1, no. 1, pp. 90–102. Avaliable at Google Scholar

[9] M. Fajriyanto, "Penerapan Metode Bayesian dalam Model Latent Dirichlet Allocation Di Media Sosial," *J. Kaji. dan Terap. Mat.*, vol. 7, no. 4, pp. 74–78, 2018. Avaliable at Google Scholar

[10] M. Fitriasih and R. Kusumaningrum, "Analisis Klasifikasi Opini Tweet Pada Media Sosial Twitter Menggunakan Latent Dirichlet Allocation (LDA)." Universitas Diponegoro, 2017. Avaliable at Google Scholar

[11] K. Bastani, H. Namavari, and J. Shaffer, "Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints," *Expert Syst. Appl.*, vol. 127, pp. 256–271, 2019. Doi. 10.1016/j.eswa.2019.03.001

[12] Y. Wang and J. E. Taylor, "DUET: Data-driven approach based on latent Dirichlet allocation topic modeling," *J. Comput. Civ. Eng.*, vol. 33, no. 3, 2019. Doi. 10.1061/(ASCE)CP.1943-5487.0000819

[13] S. Joo, I. Choi, and N. Choi, "Topic analysis of the research domain in knowledge organization: A latent Dirichlet allocation approach," *KO Knowl. Organ.*, vol. 45, no. 2, pp. 170–183, 2018. Doi. 10.5771/0943-7444-2018-2-170

[14] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, 2019. Doi. 10.1007/s11042-018-6894-4

[15] A. Nurlayli and M. A. Nasichuddin, "Topik Modeling Penelitian Dosen JPTEI UNY pada Google Scholar Menggunakan Latent Dirichlet Allocation," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 4, no. 2, pp. 154–161, 2019. Doi. 10.21831/elinvo.v4i2.28254

[16] M. Hajjem and C. Latiri, "Combining IR and LDA topic modeling for filtering microblogs," *Procedia Comput. Sci.*, vol. 112, pp. 761–770, 2017. Doi. 10.1016/j.procs.2017.08.166

[17] M. Blake and V. Gallimore, "Understanding academics: a UX ethnographic research project at the University of York," *New Rev. Acad. Librariansh.*, vol. 24, no. 3–4, pp. 363–375, 2018. Doi. 10.1080/13614533.2018.1466716

[18] B. J. Kim, S. Jeong, and J.-B. Chung, "Research trends in vulnerability studies from 2000 to 2019: Findings from a bibliometric analysis," *Int. J. Disaster Risk Reduct.*, vol. 56, p. 102141, 2021. Doi. 10.1016/j.ijdrr.2021.102141

[19] M. Schneider and F. Preckel, "Variables associated with achievement in higher education: A systematic review of meta-analyses.," *Psychol. Bull.*, vol. 143, no. 6, p. 565, 2017. Doi. 10.1037/bul0000098

[20] A. Irons and S. Elkington, *Enhancing learning through formative assessment and feedback*. Routledge, 2021. Doi. 10.4324/9781138610514

[21] R. M. Simamora, "The Challenges of Online Learning during the COVID-19 Pandemic: An Essay Analysis of Performing Arts Education Students," *Stud. Learn. Teach.*, vol. 1, no. 2, pp. 86–103, 2020. Doi. 10.46627/silet.v1i2.38

[22] T. Z. Dessiaming, S. Anraeni, and S. Pomalingo, "ANALISIS DATA AKADEMIK PERGURUAN TINGGI MENGGUNAKAN TEKNIK," vol. 3, no. 5, pp. 1203–1212, 2022. Avaliable at Google Scholar

[23] W. Wahyudin, "APLIKASI TOPIC MODELING PADA PEMBERITAAN PORTAL BERITA ONLINE SELAMA MASA PSBB PERTAMA," in *Seminar Nasional Official Statistics*, 2020, vol. 2020, no. 1, pp. 309–318. Doi. 10.34123/semnasoffstat.v2020i1.579

[24] R. A. Pane, M. S. Mubarok, and N. S. Huda, "A multi-lable classification on topics of quranic verses in english translation using multinomial naive bayes," in *2018 6th International Conference on Information and Communication Technology (ICoICT)*, 2018, pp. 481–484. Doi. 10.1109/ICoICT.2018.8528777

[25] R. Narayan and A. Tidström, "Tokenizing coopetition in a blockchain for a transition to circular economy," *J. Clean. Prod.*, vol. 263, p. 121437, 2020. Doi. 10.1016/j.jclepro.2020.121437

[26] A. Schofield, M. Magnusson, and D. Mimno, "Pulling out the stops: Rethinking stopword removal for topic models," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, short papers*, 2017, pp. 432–436. Doi. 10.18653/v1/E17-2069

[27] C. Jiang, W. Li, S. Wu, and Q. Bai, "OMT: An Operate-Based Approach for Modelling Multi-topic Influence Diffusion in Online Social Networks," in *International Conference on Web Information Systems Engineering*, 2021, pp. 542–556. Doi. 10.1007/978-3-030-90888-1_41

[28] A. F. Hidayatullah, S. K. Aditya, and S. T. Gardini, "Topic modeling of weather and climate condition on twitter using latent dirichlet allocation (LDA)," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 482, no. 1, p. 12033. Doi. 10.1088/1757-899X/482/1/012033

[29] Y. Li, B. Rapkin, T. M. Atkinson, E. Schofield, and B. H. Bochner, "Leveraging Latent Dirichlet Allocation in processing free-text personal goals among patients undergoing bladder cancer surgery," *Qual. Life Res.*, vol. 28, no. 6, pp. 1441–1455, 2019. Doi. 10.1007/s11136-019-02132-w

[30] C. Sharma and S. Sharma, "Latent DIRICHLET allocation (LDA) based information modelling on BLOCKCHAIN technology: a review of trends and research patterns used in integration," *Multimed. Tools Appl.*, pp. 1–27, 2022. Doi. 10.1007/s11042-022-13500-z