ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Vol 11, Issue 7, July 2024

Scientific Literature Recommendation System Design for Optimizing On-time Students Graduation Using a Combination of Latent Semantic Analysis and Latent Dirichlet Allocation Text Mining Methods

[1] Pasha Dwi Mahendra*, [2] Aghus Sofwan, [3] Wahyul Amien Syafei

[1] [2] [3] Department of Electrical Engineering, Diponegoro University, Semarang, Indonesia
Corresponding Author Email: [1] dwimahendrapashax@gmail.com, [2] asofwan@elektro.undip.ac.id, [3] wasyafei@gmail.com

Abstract— ***This research aims to evaluate the effectiveness of the Latent Dirichlet Allocation (LDA) method in extracting keywords from abstracts of scholarly articles in Scopus, Elsevier, and Science Direct, highlighting the challenges in obtaining accurate and relevant information. To enhance text extraction performance, this study combines LDA with Latent Semantic Analysis (LSA), methods typically used separately. The method is tested through a comparison of Cosine Similarity and keyword overlap between the extraction results and manual extraction, GPT-4, Scopus, and author keywords. The results indicate that the calculation of Cosine Similarity and Keyword Overlap shows an increased accuracy and relevance score by 10% compared to the standard LDA method, with a more diverse distribution of keyword insights. The improvement in keyword accuracy and relevance scores is achieved with an insignificant increase in computational load time, which is less than 2%.***

Keywords: Cosine Similarity, Keyword Extraction, Latent Dirichlet Allocation, Latent Semantic Analysis.

I. INTRODUCTION

Higher education plays a strategic role in enlightening the nation and advancing science and technology while considering and applying humanistic values, as well as fostering and empowering the sustainable culture of the Indonesian nation [1].

Article 1, paragraph (1) of Law Number 20 of 2003 concerning the National Education System states that education is a conscious and planned effort to create a learning atmosphere and learning process so that learners actively develop their potential to have spiritual strength, self-control, personality, intelligence, noble character, and skills needed for themselves, society, the nation, and the state [1].

Education is the most important aspect for every individual, encompassing both formal and non-formal education, as humans fundamentally require education from childhood to old age. Thus, education is closely related to the development of human resources (HR) potential through teaching activities. Teaching activities are conducted at all levels of education, including compulsory basic education for 9 years, secondary education, and higher education [2].

Higher education in Indonesia still faces challenges in completing education on time. According to data from the Ministry of Research, Technology, and Higher Education in 2021, the graduation rate of undergraduate students within 4 years in Indonesia is only about 41.8%, while the rest still experience delayed graduation [3].

Factors causing student graduation delays include students' lack of insight into understanding lecture material, internal problems, and also due to the lack of monitoring and evaluation of student progress. Many students struggle to understand scientific literature related to the courses they take. This results in students spending more time studying lecture material and ultimately experiencing delayed graduation [2].

Numerous studies have been conducted to find solutions to the problem of student graduation delays. Some studies propose the use of technology to help students access the required scientific literature. Similar research has also been implemented by Dwi Astutik, who has implemented Simdalima and increased the percentage of on-time graduates from 49.55% to 56.65% in 2020 [4].

The problem of graduation delays can be reduced through the implementation of the fourth industrial revolution in education, involving Artificial Intelligence (AI), big data, automation, and the Internet of Everything (people, things, services, content, and education) [5]. This is supported by findings, which state that the implementation of technology assistance in the field of education is still very low, even though Jamaludin's research shows that the readiness index of Southeast Asian higher education institutions to implement the fourth industrial revolution is very high [6].

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Vol 11, Issue 7, July 2024

This research aims to create a journal search system utilizing the LDA, LSA, and LDA LSA combination methods to extract important information and keywords from abstracts of obtained articles. This system combines the LDA and LSA methods, which are generally used separately, as a new method novelty that yields improved accuracy and relevance of keywords through Cosine Similarity and Keyword Overlap calculations without causing a significant increase in computation time.

II. LITERATURE REVIEW

The literature review in this research includes a brief exposition on the methods used, namely Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), text mining, and natural language processing.

A. Latent Semantic Analysis Method

Latent Semantic Analysis (LSA) is a Natural Language Processing (NLP) method used to summarize a text by considering the frequency of occurrence or repetition. LSA takes text document strings as input and produces scores for each sentence as output. The process involves the removal of stopwords, text tokenization and cleaning, text vectorization, term-document matrix creation, Singular Value Decomposition (SVD), dimensionality reduction, and interpretation of results. The methodology flow of LSA consists of 7 processes:

1) **Stopwords removal**, which is the process of eliminating meaningless words, such as the, is, am, and are.
2) **Text tokenization and cleaning**, which is the process of breaking down text data into arrays of sentences and arrays of words. This process also cleans the text from numbers, punctuation marks, and normalizes the text to lowercase.
3) **Text vectorization** (bag of words, dictionary, TF IDF), which is the process of aggregating input words into a bag of words or a dictionary containing the frequency of word occurrences. TF-IDF is the vectorization method used by the author in this research to reduce the weight of frequently occurring words.
4) **Creating term-document matrix**, which is the process of arranging a matrix based on sentences and words generated from tokenization. This process produces a matrix with dimensions MxN, where M represents the number of sentences in the text and N represents the number of words in each sentence.
5) **SVD** (U S Vtranspose), which is the process of decomposing a matrix obtained using eigenvalues and singular values. Mathematically, SVD is guided by equations 3.1 to 3.9.

$$A = U \cdot S \cdot V^T \tag{3.1}$$

$$U = column\ matrix - left\ orthogonal\ (m \times m)$$

$$S = matrix\ diagonal\ singular\ value\ (m \times n)$$

$$V = row\ matrix\ - left\ ortogonal\ (n \times n)$$

$$U = [u_1\ u_2 \dots u_m] \tag{3.2}$$

$$(A - \lambda_x \cdot I) \cdot u_x = 0 \tag{3.3}$$

$$\det(A - \lambda \cdot I) = 0 \tag{3.4}$$

$$S = diag(\sigma_1\ \sigma_2 \dots \sigma_m) \tag{3.5}$$

$$\sigma_x = \sqrt{\lambda_x} \tag{3.6}$$

$$V = [v_1\ v_2 \dots v_n] \tag{3.7}$$

$$(A \cdot A^T - \lambda_x \cdot I) \cdot v_x = 0 \tag{3.8}$$

$$\det(A - \lambda \cdot I) = 0 \tag{3.9}$$

6) **Dimensionality reduction**, which is the process of reducing dimensions in the matrix accompanied by reducing features to simplify the information and insights generated by the matrix. Dimension reduction in the LSA summarization method aims to obtain the 5 sentences with the highest weight.
7) **Results Interpretation**, which is the process of obtaining insights into hidden patterns from the output of dimension reduction. Insights generated by the LSA method can include topics, top sentences, semantics, and sentiment. In this research, the LSA method is focused on obtaining top sentences.

B. Latent Dirichlet Allocation Method

Latent Dirichlet Allocation (LDA) is a probabilistic model used in topic analysis within a collection of documents. This model enables the grouping of documents into hidden topics based on the distribution of words appearing in them. LDA is based on the assumption that each document in the collection is a combination of several topics. Each topic is a distribution of words appearing in those documents. In the LDA process, the main goal is to identify the optimal topic distribution within the document collection, as well as the distribution of words in each topic.

Keyword extraction is performed using the LDA method, which takes text document strings as input and produces a string array consisting of a list of keywords. The process involves removing stopwords, text tokenization and cleaning, text vectorization, applying LDA, distribution of results, and interpreting the results. The LDA method begins with data cleaning, so the first 3 steps of LDA are similar to the first 3 steps of LSA. The methodology flow of LDA consists of 6 key processes:

1) **Stopwords removal**
2) **Text tokenization and cleaning**
3) **Text vectorization**
4) **Applying LDA** (number of topics, model training, iterations), which is the process of training the model through clustering using the number of topics as a reference for the number of clusters generated. Model training is done using all tokens used in all documents to show the association of each word in each cluster

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Vol 11, Issue 7, July 2024

generated.

5) **LDA Results** (words per topic and topics per document), which is the process of aggregating the output from the LDA application process. In this process, clusters are obtained that show the members of words along with their frequency of occurrence in each topic. This process can also be used to identify the composition of clusters or topics along with their weights in each document.
6) **Interpretation of results + topic interpretation + model evaluation** (e.g., adding stopwords), which is the process of obtaining insights from the output of the LDA processing. At this stage, the author can observe the composition of the top words in each topic to adjust stopwords. This is aimed at reducing noise and producing optimal keyword outputs.

C. Text Mining

Text mining is the process of extracting useful and meaningful information from unstructured or semi-structured text. Its goal is to convert text data into a form that can be understood by machines, so it can be processed and analyzed to gain valuable insights. By using techniques such as tokenization, text cleaning, vectorization, and statistical analysis, text mining can help identify patterns, trends, and relationships in large and complex text data. This enables better decision-making, sentiment analysis, document classification, and various other tasks involving text processing.

D. Natural Language Processing

Natural Language Processing (NLP) is a branch of computer science that focuses on the interaction between humans and human language in the form of text or speech. NLP encompasses a range of techniques and methods used to understand, analyze, and generate text used in human communication.

III. RELATED WORK

The issue of student graduation delays has been researched by [4] and [5]. Reference [4] proposes the use of technology to assist students in accessing the necessary academic literature. In their research, they implemented Simdalima and increased the percentage of on-time graduates from 49.55% to 56.65% in 2020 [4]. Meanwhile, [5] suggests that this issue can be reduced through the implementation of the 4.0 industrial revolution in the field of education, involving Artificial Intelligence (AI), big data, automation, and the Internet of Everything (people, thing, service, content, and education) [5]. Reference [5] research is supported by findings from Jamaludin, who states that the application of technology in the field of education is still very low compared to the readiness index of higher education institutions in Southeast Asia to implement the 4.0 industrial revolution [6].

The design of a recommendation system application with machine learning is conducted by [7]-[12]. Reference [7] investigates Adaptive Machine Learning, which has a forward feedback system for real-time system updates. Meanwhile, [8] studies spontaneous recommender with short processing time. References [10] and [11] create recommendation systems that only change the method of presenting data, thus narrowing down business choices that are often biased.

Text extraction research is conducted by [13]-[16]. Reference [13] utilizes the Cross Latent Semantic Analysis (CLSA) method to automatically summarize news documents filtered through RSS Feeds. References [14] and [16] develop a document similarity detection system where [14] combines LSA and Bayesian methods to prevent detecting similar documents when their sentence structures change, while [16] compares the Jaro-Winkler Distance method and the LSA method in creating a Plagiarism detection system. Reference [17] uses the LSA method for automating the summarization of legal documents.

LSA method is widely used for document summarization. Through document summarization, the LSA method can further be used to develop document similarity assessment systems or plagiarism detection systems.

The LDA method is generally used to map associations between each word in documents with other words. This is subsequently used to extract keywords and topic content in documents. The extraction of keywords and topics can be used to develop document recommendation systems based on the similarity of keyword and topic distributions in each document.

The combined LDA and LSA method in this research consists of document summarization to obtain the top n sentences to be used as input for LDA keyword extraction. The output of this combined method will produce keywords with higher reliability compared to the LDA method alone.

IV. PROPOSED DESIGN

This research proposes a system with input in the form of a String query keyword from the user. This input will then be used to fetch abstract data from an API, which will subsequently be processed using the LSA and LDA methods. The output of this system will be a list of keywords. The discussion in this section will be divided into data preparation and data processing. The general process can be seen in the Fig 1.

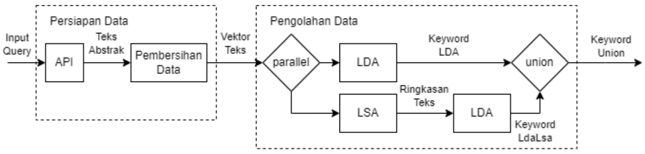


Fig 1. The result of merging search and abstract data

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Vol 11, Issue 7, July 2024

A. Data Preparation

This system accepts input in the form of a String query keyword. This query serves as a filter in fetching data from the API. The API then responds by providing between 0 to 25 article entries that meet the previously input query. Each entry is then cleaned using the first 3 steps of LDA and LSA to make it processable by the system. This process generates text vectors that are easier for machines to process and iterate as input for LSA and LDA processing.

B. Data Processing

This study modifies the function of the LSA method slightly in terms of weighting, which typically involves summing the information in sentences to averaging the information in sentences. This is done by changing the source code that originally used the sum method to the mean method in the dimension reduction process, or step 6 of the LSA method. The change from the sum method to the mean method aims to eliminate bias towards longer sentences as the sentences with the highest weight, thus the mean method compares the average weights within a sentence.

Furthermore, the output of LSA, consisting of a list of sentences with the highest weights, will be combined into a new text string as input for the LDA method. Referring to Figure 1, it can be seen that the parallel process involves the LDA method in each branch. The LDA method used in each branch does not have significant differences because it uses the same function. Therefore, the differences in output from each branch are only caused by differences in input entering the LDA method.

This system has 2 types of keyword outputs: LDA extraction keywords and LdaLsa extraction keywords. LDA extraction keywords are tokens or words that have high weights in a reading as a whole. Meanwhile, LdaLsa Extraction keywords are keywords generated from the text resulting from LSA summarization. The system does not recognize a phrase or several words that refer to one meaning, thus causing keyword extraction to focus on each word. Examples of keywords generated by the system can be observed in Fig 2.

```
lda_only

['capacity',
 'care',
 'customized',
 'groups',
 'health',
 'lies',
 'population',
 'precision',
 'prevention',
 'robust',
 'targeted',
 'utilize',
 'patient',
 'potential',
 'specific',
 'data']
```

(a)

```
lda_lsa

['automatically',
 'clinically',
 'considering',
 'identify',
 'learning',
 'machine',
 'relevant',
 'sources',
 'heterogeneous',
 'individuals',
 'potential',
 'subgroups',
 'data']
```

(b)

Fig 2. Keywords extracted using (a) LDA method and (b) LdaLsa method

V. RESULTS AND ANALYSIS

In this section, we discuss the analysis of keyword extraction from the LDA method and the LdaLsa method. The analysis conducted is a performance check that calculates the accuracy and relevance of keyword extraction, and a computation load that calculates the additional computational load that occurs when implementing text extraction.

A. Performance Check

The performance check is conducted by comparing the values of Cosine Similarity and Keyword Overlap between the extracted keywords and the reference keywords. The reference keywords used consist of manual keywords, GPT keywords, Scopus keywords, and Author keywords. The reference keywords used can be observed in Fig 3.

In addition to comparing the performance of the keywords extracted by the LDA and LdaLsa methods, the author also considers involving the union of system output keywords to observe potential improvements. This is intended to observe the level of difference in the keywords generated. The differences in the extracted keywords of each method can be combined either by intersect or union to determine whether the extraction results of each method can complement each other.

```
key_gpt.append([
  'Precision Population Health',
  'Patient Data',
  'Customized Prevention',
  'Care Targeting',
  'Specific Groups',
  'Machine Learning',
  'Clinically Relevant Subgroups',
  'Heterogeneous Data Sources',
  'Unsupervised Machine Learning (UML)',
  'Clinical Data Interpretation',
  'Coronary Artery Disease',
  'Aorta Dimensions',
  'Random Forest-Based Cluster Analysis',
  'Variables',
  'Participants (including gender breakdown)',
  'Unsupervised Clustering Approach',
  'Clinical Characteristics',
  'Flexible UML Algorithms',
  'Patient Data Processing',
  'Clinical Interpretation',
  'Risk Assessment',
])
```

(a)

```
key_manual.append([
  "precision population health",
  "customized prevention",
  "machine learning",
  "automatically identify",
  "clinically relevant subgroups",
  "unsupervised machine learning",
  "coronary artery disease",
  "range",
  "aorta dimensions",
  "Random Forest-Based Cluster Analysis",
  "clinical interpretation",
  "risk assessment",
])
```

(b)

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Vol 11, Issue 7, July 2024

```
key_author.append([
    "Aortic dimensions",
    "Clusterization",
    "Computed tomography coronary angiography",
    "Coronary artery disease",
    "Unsupervised learning",
])
```

(c)

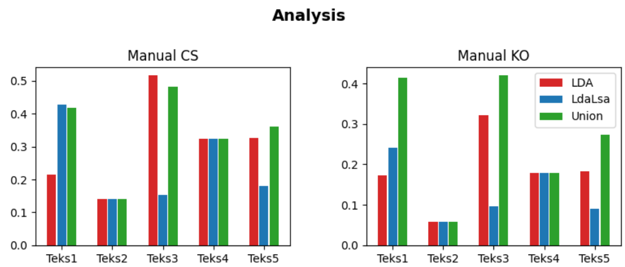
```
key_scopus.append([
    "Aortic dimension",
    "Clinical data",
    "Clusterization",
    "Computed tomography coronary angiography",
    "Coronary angiography",
    "Coronary artery disease",
    "Patient data",
    "Risk Identification",
    "Risk stratification",
    "Unsupervised machine learning",
])
```

(d)

Fig 3. Keyword comparison (a) GPT-4 (b) Manual (c) Author (d) Scopus

Based on the data displayed in Fig 4, the author found fluctuations in the accuracy of the LdaLsa combination method compared to the LDA method. This is indicated by both decreases and increases in accuracy, both below 10% and exceeding 10%. Through the figure, it can be seen that out of 40 comparisons, the LdaLsa method had superior performance 11 times, the LDA method was superior 10 times, and there were 19 instances where both methods had equal performance. Meanwhile, the union of the two methods resulted in superior performance 22 times, the LDA method 2 times, and there were 16 instances where both methods had equal performance. This indicates that the keyword outputs differ, allowing the generated keywords to be enhanced through the union method to achieve higher Cosine Similarity and Keyword Overlap scores.

Although sometimes the keyword extraction performance of the combination method experiences a decrease in accuracy, the variation in keywords generated can provide additional insights into the keywords obtained through the LDA method. This is demonstrated by the data in Fig 4, which shows that the accuracy level of the union tends to exceed the performance of both the regular LDA and the LdaLsa combination. Therefore, it can be concluded that the keywords generated by each method can complement each other to improve overall performance.



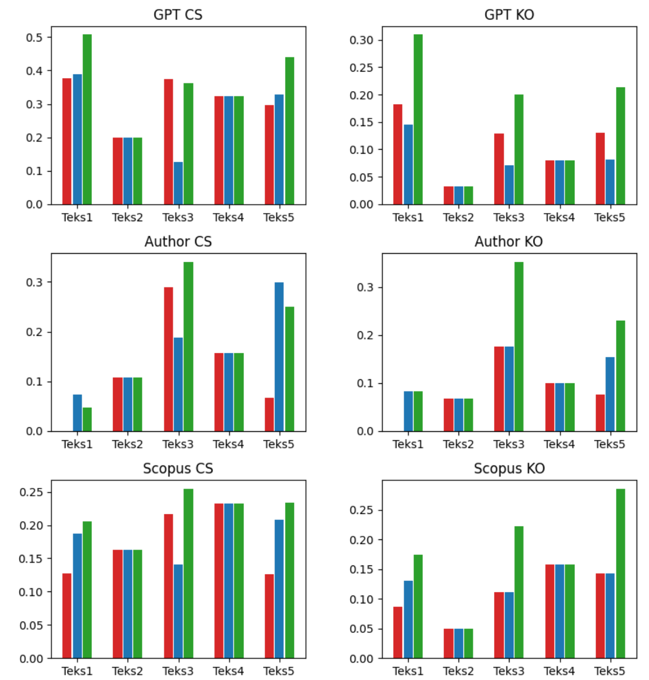


Fig 4. System performance mapping visualization

B. Computation Load

The computational load experienced by the server when processing the input obtained until it becomes the final output can be observed in Fig 5. At this stage, the author deployed the backend on the http://pythonanywhere.com server using the Python programming language with the Flask framework.

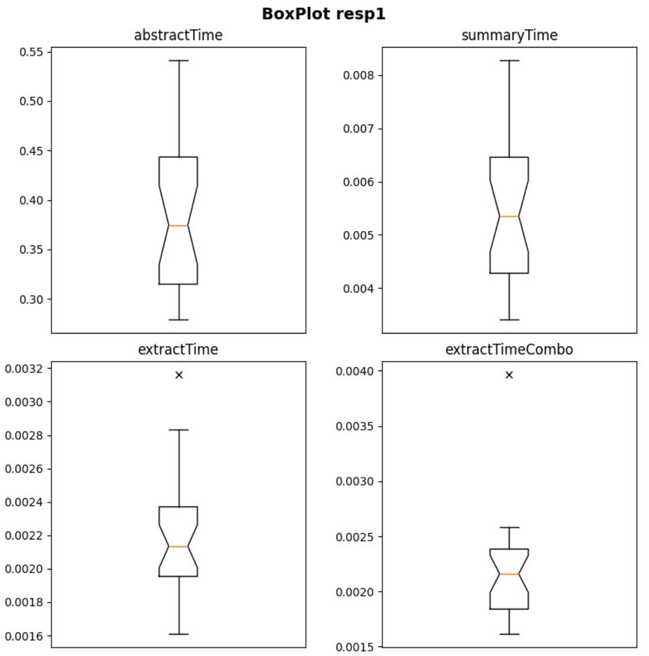


Fig 5. The computation time experienced by the server

In Fig 5, the average processing time for an article ranges from 300ms to 450ms. The computational load recorded by the server dashboard indicates 32% CPU usage and 10% file

IFERP Explore Your Research Journey...

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

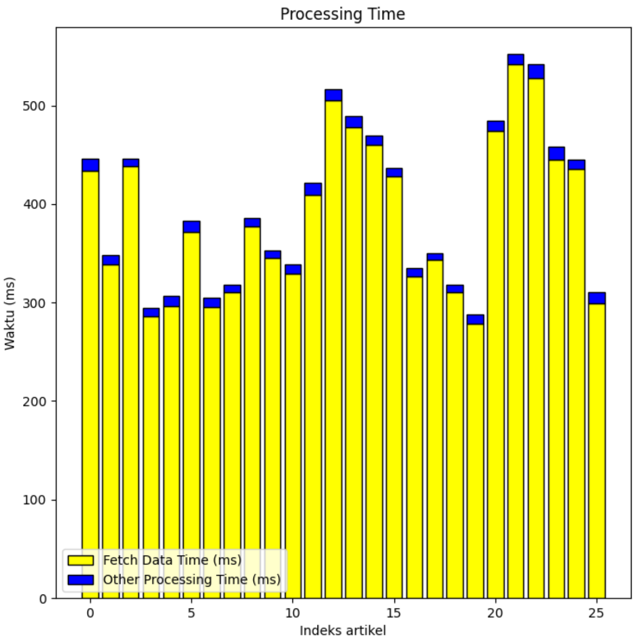
Vol 11, Issue 7, July 2024

storage. This suggests that the use of free resources still allows for small-scale deployment.

Fig 6 shows the overall distribution of load time experienced by the system. It can be observed that the average processing time for an article is around 400 ms. However, in reality, there is a gap between the initial processing and data presentation of 50.98 seconds. This is because only some entries obtained from the API have abstracts. Therefore, even though the system displays 25 articles, it actually processes 125 articles. The "pii" field indicates that these articles are owned by Scopus, Science Direct, or Elsevier, and the API cannot display detailed article abstracts without pii.

It is known that the Scopus website takes 8.71 seconds to display 25 articles, so the average time for each article is around 348 ms. The LdaLsa Search application takes 50.98 seconds to display 25 articles out of 125 processed articles, resulting in an average time of around 408 ms per article. Based on the summarized data in Fig 5, the average processing time for abstract extraction is 398 ms. Therefore, the author assumes a loss of 10 ms as the sum of processing losses for both backend and frontend of each server used.

Before implementing the LDA and LdaLsa methods, the average processing time was 348ms. The implementation of the LDA method increased the computation time by 2ms, while the combined LdaLsa method increased the computation time by 2.3ms. Thus, the implementation of this system increased the total computation by less than 2% compared to the original computation time.



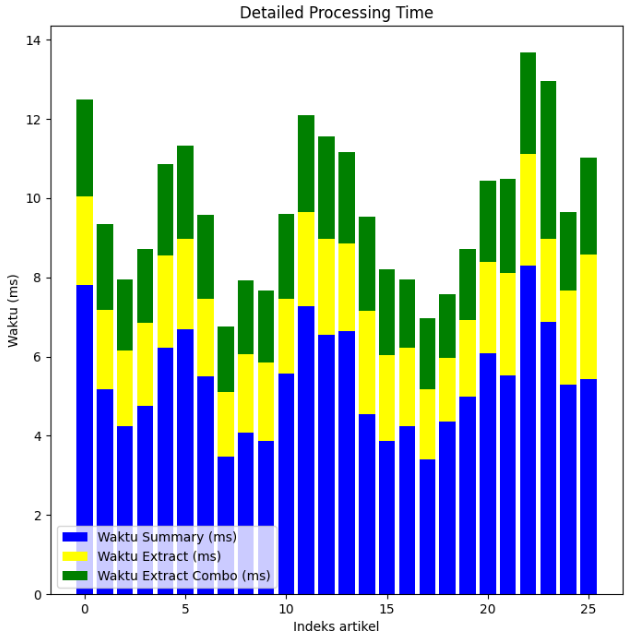


Fig 6. Processing time of each article

VI. CONCLUSION

This study demonstrates the performance of the LDA method and the combination of LdaLsa. The study's reliability was tested using 5 articles, 4 reference keywords, and 2 accuracy calculation methods. The reference keywords used consist of manual, GPT, Scopus, and Author keywords. Meanwhile, the accuracy calculation methods include Cosine Similarity and Keyword Overlap. Through 40 calculations, it was found that the LdaLsa method outperformed the LDA method 11 times, the LDA method outperformed the LdaLsa method 10 times, and there were 19 instances where both methods performed equally. It was also found that the output keywords had differences, allowing them to be enhanced through the union method to achieve higher Cosine Similarity and Keyword Overlap values. The improvement of the union method resulted in 22 instances of superior performance, 2 instances of superior performance for the LDA method, and 16 instances of equal performance for both methods.

Additionally, this study observed changes in computation. Before implementing the LDA and LdaLsa methods, the average processing time was 348ms. The implementation of the LDA method increased the computation time by 2ms, while the combination method of LdaLsa increased the computation time by 2.3ms. Thus, the implementation of this system increased computation by less than 2% compared to the original computation time.

REFERENCE

[1] Undang-undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi. Indonesia, 2012.

[2] O. S. Sari, “Faktor Keterlambatan Penyelesaian Studi

Mahasiswa Prodi Bimbingan Konseling Islam Institut Agama Islam Negeri Bengkulu," (Skripsi Inst. Agama Islam Negeri Bengkulu), p. 50, 2018.

[3] BPS-Statistics Indonesia, "Labour Force Situation in Indonesia February 2022," no. February 2022, p. 229, 2022.

[4] D. Astutik, S. Untari, and Susetyaningsih, "Laporan Peningkatan Persentase Lulusan Tepat Waktu Mahasiswa Fakultas Ilmu Sosial Universitas Negeri Malang Melalui Pembuatan Aplikasi Sistem Monitoring Kendali Mahasiswa 'Simdalima'" Malang, 2020.

[5] N. Ma, X. Yao, and K. Wang, "Current status and prospect of future internet-oriented wisdom manufacturing," Zhongguo Kexue Jishu Kexue/Scientia Sin. Technol., vol. 52, no. 1, pp. 55–75, 2022, doi: 10.1360/SST-2021-0232.

[6] R. Jamaludin, E. McKAY, and S. Ledger, "Are we ready for Education 4.0 within ASEAN higher education institutions? Thriving for knowledge, industry and humanity in a dynamic higher education ecosystem?," J. Appl. Res. High. Educ., vol. 12, no. 5, pp. 1161–1173, Jan. 2020, doi: 10.1108/JARHE-06-2019-0144.

[8] E. D. R. S. Gonzalez et al., "Making real progress toward more sustainable societies using decision support models and tools: Introduction to the special volume," J. Clean. Prod., vol. 105, pp. 1–13, 2015, doi: 10.1016/j.jclepro.2015.05.047.

[9] R. N., S. P., T. G., and A. G., "Selecting a suitable Cloud Computing technology deployment model for an academic institute," Campus-Wide Inf. Syst., vol. 31, no. 5, pp. 319–345, Oct. 2014, doi: 10.1108/CWIS-09-2014-0018.

[10] M. T. Islam and N. Huda, "Material flow analysis (MFA) as a strategic tool in E-waste management: Applications, trends and future directions," J. Environ. Manage., vol. 244, no. February, pp. 344–361, 2019, doi: 10.1016/j.jenvman.2019.05.062.

[11] M. ZAMANI, M. MAEEN, and M. HAGHPARAST, "Implementation of Business Intelligence to Increase the Effectiveness of Decision Making Process of Managers in Companies Providing Payment Services," vol. 11, no. 2, pp. 1–24, 2017, [Online]. Available: https://www.icommercecentral.com/open-access/implementation-of-business-intelligence-to-increase-the-effectiveness-of-decision-making-process-of-managers-in-companies-providing-payment-services.php?aid=85788&view=mobile.

[12] A. A. Tole and N. C. Matei, "Executive information system (EIS) structure and their importance in decision-making: A comparison between decision support computer systems," J. Inf. Syst. Oper. Manag., vol. 10, no. 1, pp. 194–207, 2015, [Online]. Available: http://www.rebe.rau.ro/RePEc/rau/jisomg/SU16/JISOM-SU16-A19.pdf.

[13] R. B. Velasco, I. Carpanese, R. Interian, O. C. G. Paulo Neto, and C. C. Ribeiro, "A decision support system for fraud detection in public procurement," Int. Trans. Oper. Res., vol. 28, no. 1, pp. 27–47, 2021, doi: 10.1111/itor.12811.

[14] G. Mandar and G. Gunawan, "Peringkasan dokumen berita Bahasa Indonesia menggunakan metode Cross Latent Semantic Analysis," vol. 3, no. 2, pp. 94–104, 2017, [Online]. Available: https://doi.org/10,26594/register.v3i2.1161.

[15] D. W. Wicaksono, M. I. Irawan, and A. M. Rukmi, "Sistem Deteksi Kemiripan Antar Dokumen Teks Menggunakan Model Bayesian Pada Term Latent Semantic Analysis (LSA)," vol. 3, no. 2, 2014, doi: 10.12962/j23373520.v3i2.7998.

[16] T. Tinaliah and T. Elizabeth, "Perbandingan Hasil Deteksi Plagiarisme Dokumen dengan Metode Jaro-Winkler Distance dan Metode Latent Semantic Analysis," J. Teknol. dan Sist. Komput., vol. 6, no. 1, pp. 7–12, 2018, doi: 10.14710/jtsiskom.6.1.2018.7-12.

[17] M. Rusbandi, I. F. Rozi, and K. S. Batubulan, "Otomatisasi Peringkasan Teks Pada Dokumen Hukum Menggunakan Metode Latent Semantic Analysis," J. Inform. Polinema, vol. 7, no. 3, pp. 9–16, 2021, doi: 10.33795/jip.v7i3.515.