Smart Techno (Smart Technology, Informatic, and Technopreneurship)
Volume 08, Number 01, 2026 pp. 97-108
E-ISSN : 2541-0679
Open Access: https://lppm.primakara.ac.id/jurnal/index.php/smart-techno/

Implementation of Latent Dirichlet Allocation in a Cookie-Based Final Project Topic Recommendation System

Fiddar Tahwifa Putri[1*], Rosita Yanuarti[2], Taufiq Timur Warisaji[3]
[1,2,3] Informatics Engineering, Universitas Muhammadiyah Jember, Jember, Indonesia

ARTICLE INFO
Article history:
Initial submission
09-12-2025
Received in revised form
19-12-2025
Accepted 14-01-2026
Available online 27-02-2026

Keywords:
Latent Dirichlet Allocation (LDA), Perplexity, Cookies,Recommendation System, Final Project

DOI:
https://doi.org/10.59356/smart-techno.v7i1

ABSTRACT

The selection of a final project topic is a crucial stage in the academic journey of students, as it determines the direction of research while serving as a means to apply the knowledge acquired during their studies. However, in practice, many students experience difficulties in choosing a topic that aligns with their interests and areas of expertise. This challenge is largely attributed to the absence of systems capable of providing personalized recommendations. To address this issue, this study develops a final project topic recommendation system by integrating the Latent Dirichlet Allocation (LDA) algorithm with a cookie-based approach to accommodate user preferences. The dataset used consists of 200 final project documents from the Informatics Engineering program, with titles and abstracts serving as the primary features for topic modeling during model training and perplexity evaluation. In addition, users' search histories and relevance feedback stored in cookie sessions are utilized as personalization features to generate more tailored recommendations. FastText is employed to produce more contextual word vector representations, while cosine similarity is applied to measure the closeness between search keywords and final project topic documents. Evaluation results based on perplexity indicate that the model with 22 topics yields the most statistically optimal performance. Furthermore, testing using Click-Through Rate (CTR) demonstrates that the combination of topic modeling and user preference personalization produces the highest relevance, achieving a CTR of 15.6%, which is significantly higher than the baseline CTR of 2.2%. These findings confirm that the proposed system is capable of delivering more relevant, adaptive, and user-oriented final project topic recommendations.

1. INTRODUCTION

The selection of a final project topic in the academic domain represents a crucial stage for students in pursuing higher education. This importance arises from the requirement to complete a final project as a graduation prerequisite to fulfill undergraduate learning outcomes, enabling students to systematically and logically apply their knowledge, skills, and ideas to specific problems within a particular field (Purwitasari et al., 2021). Therefore, every student is required to complete a final project as a prerequisite to maintain and ensure the quality of graduates who are expected to contribute meaningfully to society (Bangkalang, 2023). However, in practice, many students experience difficulties in determining a topic that aligns with their interests and areas of expertise (Kocimaheni et al., 2020; Magai & Putra, 2025). These difficulties may stem from limited access to references, insufficient understanding of research trends, and challenges in identifying relationships across different fields of study. To address these issues, a recommendation system is needed as a solution to assist students in selecting topics that match their interests. Accordingly, this study aims to develop a final project topic recommendation system capable of recommending relevant topics to students.

The advancement of machine learning and text mining technologies has progressed rapidly and can be utilized to support information retrieval processes, including the recommendation of research topics (Aleqabie et al., 2024; Magai & Putra, 2025). One applicable approach is Latent Dirichlet Allocation (LDA), a topic modeling technique that groups documents based on patterns of word distribution. The LDA algorithm is capable of identifying latent topics along with their prevalence within a given document collection (Anisatuzzumara, 2024). With this capability, LDA primarily aims to uncover hidden thematic structures embedded in textual data (Chauhan & Shah, 2022).

In addition to the LDA algorithm, this study also employs a cookie-based approach to be integrated into the existing system. Cookies are data stored in a web browser and transmitted with each request to the server, allowing the server to recognize user sessions and maintain state information (Bortz et al., 2011). Cookie sessions are used to store user activities, including search history and relevance feedback. These preference data serve as additional references for generating more personalized recommendations. The use of cookies is preferred because the system does not require user management through login or registration processes, which would demand more complex resource allocation and security management (Jovic & Krstovic, 2022). Through this cookie-based approach, the system can deliver personalized research topic recommendations, thereby enhancing the overall user experience.

Therefore, this study focuses on the development of a final project research topic recommendation system that utilizes Latent Dirichlet Allocation (LDA) as the topic classification model to provide recommendations that are relevant to users' needs. Cookies are employed to store session data containing search histories and user feedback, which also serve as references for topic matching based on frequently searched keywords, enabling more personalized recommendations. By integrating these two approaches, the developed recommendation system is expected to assist users particularly students and researchers in identifying research topics that align with their interests and fields of expertise.

2. LITERATURE REVIEW

2.1. Recommendation System

A recommendation system is a set of techniques used to predict user preferences or item ratings based on personal information and item characteristics (Huda et al., 2022). This system filters information to assist users in discovering relevant content, products, or services by leveraging preference histories and other related data (Anisatuzzumara, 2024). Unlike Decision Support Systems (DSS), which emphasize weighted analysis to generate optimal alternatives, recommendation systems are more flexible and personalized, providing suggestions without requiring complex evaluation processes (Katili et al., 2021). In the context of this study, the recommendation system plays a crucial role in delivering a personalized user experience, particularly in assisting users in identifying research topics that align with their preferences.

2.2. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a probabilistic topic modeling method in Natural Language Processing (NLP) that assumes each document is composed of a mixture of topics, and each topic is represented by a specific distribution of words (Erna et al., 2024). The LDA process is fundamentally based on three core formulations, namely topic sampling, word distribution per topic, and topic distribution per document (Griffiths & Steyvers, 2004).

$$P(z_i = j \mid z_{-i}, w) \propto \frac{n_{-i,j}^{(w_i)} + \beta}{n_{-i,j}^{(\cdot)} + W\beta} \times \frac{n_{-i,j}^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + T\alpha} \quad (1)$$

$$\phi_j^{(w)} = \frac{n_j^{(w)} + \beta}{n_j^{(\cdot)} + W\beta} \quad (2)$$

$$\theta_j^{(d)} = \frac{n_j^{(d)} + \alpha}{n_{\cdot}^{(d)} + T\alpha} \quad (3)$$

2.3 FastText

FastText is a word embedding model that represents each word as a single vector by decomposing words into subword units or character n-grams, where each subword is mapped to its own vector representation (Fajriah & Kurniawan, 2025).

2.4 Cosine Similarity

Cosine similarity is a similarity measurement used to compare two vectors in a multidimensional space by calculating the cosine of the angle between them, where a value of 1 indicates identical orientation, 0 indicates no similarity, and -1 indicates completely opposite orientation (Amalia, 2022).

2.5 Cookie

Cookies are data stored in a web browser and sent back to the server with each request. They function to store information that allows the server to recognize user sessions and maintain interaction states, even when users navigate between different pages within a website (Bortz et al., 2011).

2.6 Perplexity

Perplexity is a statistical measure used to evaluate the quality of a language model in predicting textual data. It essentially reflects the level of uncertainty of the model when predicting the next word in a sequence of text (Meister & Cotterell, 2021). The calculation is performed by computing the average log probability of words across the entire test document set and then converting it to an exponential scale (Santika, 2019). In the context of topic modeling, a lower perplexity value indicates better model performance (Khadijah & Cahyono, 2024). Therefore, perplexity serves as an important indicator for assessing how optimal a topic model is, particularly in determining the appropriate number of topics to achieve accurate modeling and good generalization.

$$Perplexity(D_{test}) = exp\left\{\frac{-\sum_d log\ p(w_d)}{\sum_d n_d}\right\} \qquad (4)$$

2.7 Click-Through Rate (CTR)

Click-Through Rate (CTR) is a metric used to evaluate the effectiveness of digital advertisements or recommendation systems by calculating the ratio between the number of clicks received and the total number of impressions or views (Liu et al., 2015). A higher CTR value indicates a stronger user interest and reflects the relevance of the recommendations provided to users (Murdiyanto & Himawan, 2022).

3. METHOD

This study adopts a software engineering approach, as it focuses on the development and implementation of a final project topic recommendation system that integrates the Latent Dirichlet Allocation (LDA) algorithm for topic modeling, FastText for text vector representation, cosine similarity for measuring vector similarity, and a cookie-based approach for user personalization. The system evaluation is conducted using two main metrics: perplexity to assess the performance of the topic model and Click-Through Rate (CTR) to measure the relevance of the recommended results based on user interactions. The research stages illustrated in Figure 1 include data collection, preprocessing, topic modeling, system implementation, and evaluation.

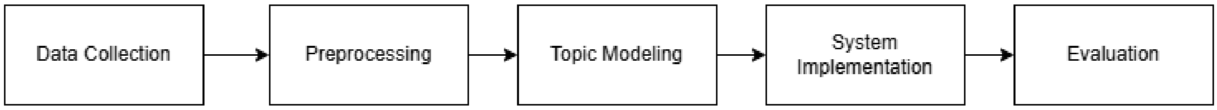

Figure 1. Research Methodology

3.1. Data Collection

The data collection stage began with obtaining and preparing data in the form of undergraduate thesis documents retrieved from the official repository of Universitas Muhammadiyah Jember via https://repository.unmuhjember.ac.id/. The dataset consists of 200 undergraduate theses from the Informatics Engineering Study Program, with attributes including author name, title, year, abstract, link, and keywords. Among these attributes, the title and abstract were used as the basis for training the topic extraction model.

3.2. Preprocessing

After all data had been collected, data preprocessing or cleaning was conducted using the following five stages:

3.2.1. Case Folding

This process converts all text characters into lowercase letters, for example, transforming the word “Implementasi” into “implementasi” (Bangkalang, 2023).

3.2.2. Cleansing

This process aims to remove text characters that are not required, such as numbers, symbols, and punctuation marks (Rosadi & Solichin, 2024).

3.2.3. Tokenizing

This process aims to split text sentences into smaller units called tokens, which are typically in the form of individual words (Bangkalang, 2023; Pamput et al., 2025).

3.2.4. Stopword Removal

This process removes commonly occurring words in the text that do not carry significant meaning, typically including prepositions, conjunctions, and pronouns (Bangkalang, 2023).

3.2.5. Stemming

This process is performed to convert each token into its root word by removing affixes, for example, transforming the word “pengembangan” into “kembang” (Pamput et al., 2025).

3.3. Topic Modeling

Topic modeling was performed on the previously cleaned dataset using the Latent Dirichlet Allocation (LDA) algorithm, which is a probabilistic approach in Natural Language Processing (NLP) that assumes each document consists of a mixture of topics represented by specific word distributions (Erna et al., 2024). At this stage, the process produces extracted topics from undergraduate theses, which are subsequently integrated into the system.

3.4. System Implementation

The system implementation stage was carried out by integrating all components, particularly the topic extraction results generated by LDA, into the recommendation system. Several additional components were incorporated to enhance system development, including the use of FastText, cosine similarity, and cookies. FastText was utilized to convert words or sentences into vector representations so that they could be processed by the algorithm, where word vectorization was applied to user search keywords and undergraduate thesis documents. After vectorization using FastText, the similarity between the two vectors was measured using cosine similarity. In addition, in this study, cookies were specifically used to store user IDs along with their activities, including user search history and relevance feedback.

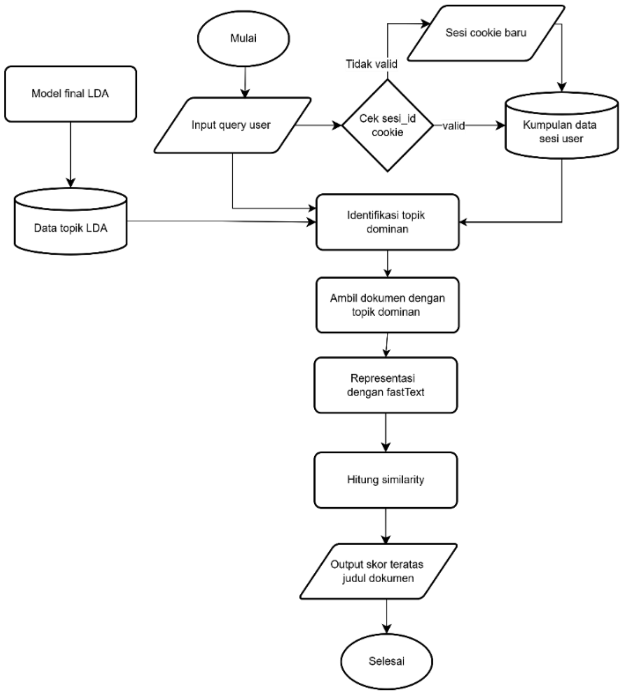

Figure 2. Workflow of the Undergraduate Thesis Topic Recommendation System

The system workflow illustrated in Figure 2 shows that when a user submits a search query to the system, the processing is divided into two paths, whose results are then combined with an equal weighting of 50:50. The linear path processes the query directly: the query topic is identified using LDA, relevant documents are retrieved, and both the query and documents are represented as FastText vectors to subsequently calculate similarity scores with documents of dominant topics. Meanwhile, the personalization path utilizes cookies to recognize user sessions. Cookies are implemented after the query is submitted to store search activities without authentication and to facilitate personalized recommendation results. The system considers managed user preferences, including keywords and relevance feedback, which are also stored in the database. This set of preferences is identified for its dominant topics using LDA, transformed into vector representations, and its similarity is calculated against related documents. Finally, the results from both the linear and personalization processing paths are merged and presented to the user as personalized undergraduate thesis topic recommendations.

3.5. Evaluation

The evaluation was conducted to measure the model performance and the effectiveness of the generated recommendations using perplexity and Click-Through Rate (CTR) techniques.

3.5.1. Perplexity

Perplexity serves as an important indicator for evaluating how optimal the topic modeling is, such as in determining the appropriate number of topics to produce accurate modeling and good generalization. Therefore, the model identified as optimal based on perplexity testing is integrated into the system and applied to each recommendation generation process

3.5.2. Click-Through Rate (CTR)

However, CTR has limitations because a click does not always indicate that a topic is truly relevant to the user. To address this issue, an additional approach in the form of explicit feedback is required, namely feedback consciously provided by users to express their preferences toward an item. The advantage of explicit feedback lies in its ability to provide more definitive guidance and help the system understand item relevance more accurately. In this study, CTR is applied to evaluate the recommendation system by monitoring the frequency with which users click on research topics recommended by the system. Therefore, the general CTR equation presented in Equation (5) is modified into Equation (6) by incorporating this additional factor.

$$CTR = \frac{(Total\ Number\ of\ Clicks)}{(Number\ of\ Impressions)} \tag{5}$$

$$CTR = \frac{(Number\ of\ Relevant\ Clicks)}{(Number\ of\ Impressions)} \tag{6}$$

The attribute “number of relevant clicks” in Equation (6) is obtained from recorded user activities when providing relevance feedback, while the “number of impressions” attribute is derived from the system-generated recommendation results.

4. RESULT AND DISCUSSION

4.1. Data Processing

This study utilizes 200 undergraduate thesis documents from the Informatics Engineering Study Program, covering the period from 2021 to 2024, which were manually collected from the university repository. The dataset was not explicitly divided into training and testing sets, as this approach is commonly applied in unsupervised topic modeling. All data first underwent preprocessing stages before being used in training the LDA model to generate topic extraction. This text cleaning process aims to improve data quality so that the model can perform optimally. The preprocessing stages were applied to two types of data: a combination of thesis title and abstract attributes, and user-entered search query data. The results of preprocessing the undergraduate thesis data are presented in Figure 3.

```
                                            Datasets  \
0  klasifikasi penyakit stroke menggunakan algori...
1  penerapan support vektor machine (svm) untuk k...
2  arsitektur microservice untuk resiliensi siste...
3  implementasi nondeterministic finite-state mac...
4  penerapan metode single exponential smothing d...

                                     Processed Texts
0  klasifikasi sakit stroke algoritma knn optimas...
1  terap support vektor machine svm klasifikasi s...
2  arsitektur microservice resiliensi sistem web ...
3  implementasi nondeterministic finitestate mach...
4  terap single exponential smothing sistem dukun...
```

Figure 3. Text Preprocessing Results

4.2. Latent Dirichlet Allocation (LDA) Modeling

Topic modeling using Latent Dirichlet Allocation (LDA) was conducted in two main phases: initial exploration and fine-tuning. During the initial exploration phase, the number of topics was determined exploratively. Given the relatively small dataset size, this study evaluated topic numbers of 5, 10, 15, 20, and 25, each with 100 iterations. The results of this initial exploration were then evaluated using standard metrics for language model assessment, producing various topic groups. Based on this evaluation, the range of topic numbers that demonstrated the most optimal performance indicated by the lowest perplexity value was selected as a reference for the fine-tuning stage. This fine-tuning stage is crucial, as it aims to determine the optimal number of topics to be integrated and utilized in the recommendation system. The process is similar to the initial exploration phase but focuses on the topic range obtained from the evaluation of the initial exploration.

Figure 4. Word Cloud of Dominant Terms Across All Topics

Based on the dominant terms produced by LDA-based topic extraction shown in Figure 4, the following are the topic interpretations derived from the dominant terms associated with each topic, as presented in Table 1.

Table 1. Interpretation of Undergraduate Thesis Topics

Topic	Dominant Keywords	Interpretation
Topic 1	fruit, temperature, internet, c, sensor, of, learning, device, cm, campus	Internet Of Things (IoT)
Topic 2	algorithm, k-nearest, neighbor, accuracy, classification, disease, hepatitis, scenario, community service, recall	K-Nearest Neighbor (KNN) Algorithm for Disease Classification
...	...	...
Topic 12	feasibility, housing, liver, algorithm, basic, testing, disease, livable, double, learning	Housing Feasibility & Disease Analysis
...	...	...
Topic 21	system, disease, expert, plant, diagnosis, chaining, forward, ownership, error, factor	Expert System for Disease Diagnosis
Topic 22	accuracy, neighbor, disease, classification, k, testing, k-nearest, disorder, autism, algorithm	K-Nearest Neighbor (KNN) for Disease Classification

4.3. Cookie Implementation

In this recommendation system, cookies are used to facilitate personalization by storing user preferences, such as search history and relevance feedback. This preference information is crucial because it enables the system to learn user interest tendencies based on search intensity and interaction patterns, thereby allowing the presented undergraduate thesis topic recommendations to become more personalized and better aligned with individual user needs.

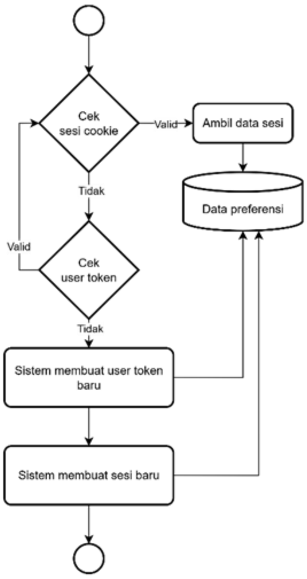

Figure 5. Cookie Management Workflow in the System

Technically, the cookie management mechanism follows the workflow illustrated in the flowchart shown in Figure 5. The process begins with checking the validity of the cookie session. If the session is valid, the system directly retrieves the stored preference data. However, if the cookie session is not valid, the system proceeds to check the user token, which serves as a unique identifier. If the token is also unavailable, the system automatically generates a new token and session to store subsequent user interaction data.

Figure 6. Recommendation Results Based on User Preferences

This recommendation system is designed by integrating topic extraction results from LDA and cookie-based personalization techniques. The system accepts user search keywords as input, which are then processed to generate undergraduate thesis title recommendations. The recommendation results are divided into two main categories: topic-based search recommendations shown in Figure 6 and personalized preference-based recommendations shown in Figure 7.

Figure 7. Recommendation Results Based on Search

User input is not only processed to generate recommendations based on LDA topics but is also utilized to produce more personalized recommendations through cookie sessions. By storing cookie sessions, the system is able to record user activities and generate individually relevant suggestions.

4.4. Perplexity Testing

Perplexity testing was conducted in two phases of the LDA topic extraction process, namely during the initial exploration stage and the fine-tuning stage. After the number of topics for model training was determined exploratorily within a specific range, the resulting topic groups were evaluated using the perplexity metric with the formula presented in Equation (4). The initial exploration stage was carried out to identify a range of models that demonstrated relative stability. The perplexity results from the first testing phase are presented in Figure 8.

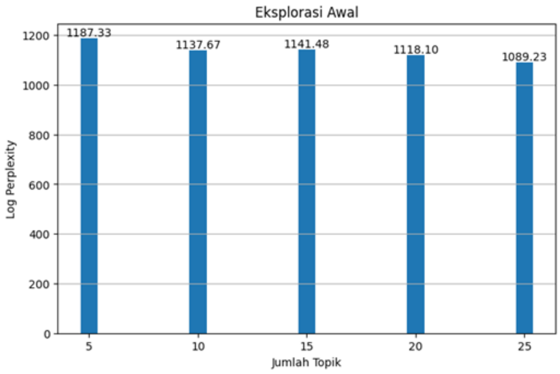

Figure 8. Visualization of Initial Exploration Perplexity Results

As indicated by the perplexity values shown in Figure 8, the lowest points were observed in the topic range of 20 and 25. Therefore, this range of topic numbers was selected for the fine-tuning stage. Although the testing procedure was similar to that of the initial exploration, fine-tuning was conducted with finer intervals between 20 and 25 topics to determine the most optimal number of topics.

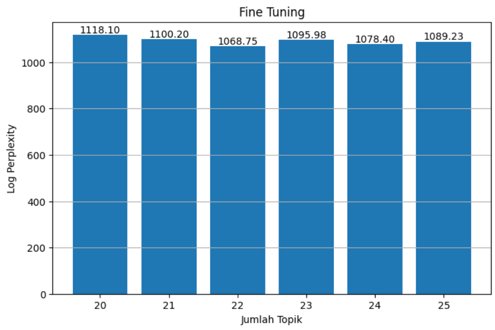

Figure 9. Visualization of Fine-Tuning Perplexity Results

As shown in Figure 9, the fine-tuning process produced the lowest perplexity value of 1068.75 at 22 topics. Therefore, 22 topics were determined as the optimal number of topics to be implemented in the recommendation system.

4.5. Click-Through Rate (CTR) Testing

This metric was selected to evaluate the level of user engagement with the presented topic results. CTR testing was conducted through a series of experiments aimed at comparing the performance of the recommendation system developed in this study. The comparison involved a baseline system that displayed recommendations based solely on topic modeling (LDA), as well as a system that generated recommendations based only on user preferences, which served as an additional benchmark.

Table 2. CTR Testing Scheme

Component	Description
Participants	5 students from the Informatics Engineering Study Program
Observation Duration	9 days, 3 search sessions per day
Recorded Data	recommendation_logs & relevance_feedback

Based on the system testing scheme shown in Table 2, observations were conducted on five students over a total period of nine days. This observation period included the use and evaluation of the developed system, the baseline system (topic-based), and the comparison system (preference-based). The following are the CTR results obtained from each of these evaluations:

Table 3. CTR Testing Results

Test Condition	Relevant Clicks	Total Recommendations	CTR (%)
Baseline	10	450	2.2%
Personalisasi	27	447	6.0%
Combined (Topik + Preferensi)	70	450	15.6%
Combined_Topic	28	450	6.2%
Combined_Pref	42	450	9.3%

Table 3 presents the Click-Through Rate (CTR) results for each test condition. The results indicate that the recommendation system developed in this study, which combines topic-based and preference-based recommendations, achieved the highest CTR of 15.6%. In contrast, the lowest CTR value was observed in the baseline test, at 2.2%.

4.6. Discussion

Based on the evaluation results using perplexity, the optimal number of topics was found to be 22, which yielded the lowest perplexity value compared to other configurations. This finding is particularly important considering the relatively small dataset used, consisting of 200 undergraduate thesis documents from a single study program. In such conditions, using an excessively large number of topics could increase the risk of overfitting, where the model tends to adjust too closely to the training data. Therefore, selecting the number of topics based on perplexity serves as a control mechanism to ensure that the LDA model remains representative and maintains good generalization capability despite the small dataset. Click-Through Rate (CTR) testing was conducted on the model optimized with 22 topics, the number determined through the previous perplexity evaluation. The results presented in Table 3 indicate significant performance differences across the tested schemes. The baseline scheme, which relied solely on general topic-based recommendations, achieved a CTR of only 2.2%, highlighting the low relevance when personalization is omitted. The application of personalization alone increased the CTR to 6.0%. However, the highest performance was observed in the system that combined topic-based and preference-based recommendations, achieving a CTR of 15.6%. In detail, personalization contributed more dominantly (9.3%) compared to topic-only recommendations (6.2%), demonstrating that integrating both approaches is key to the system’s success.

Relevance in this system is defined on a personal basis, determined by user search history and feedback. In other words, a recommendation is considered relevant if the suggested topic aligns with the recorded interest patterns and receives positive user feedback. This implies that, although the research dataset is static, the system is capable of generating dynamic recommendations. The variation in recommendations is not driven by document updates, but rather by the system’s automatic adaptation to the latest user preferences, such as search history and incoming feedback over time.

5. CONCLUSION

The undergraduate thesis topic recommendation system was successfully developed through the integration of LDA-based topic extraction and cookie-based personalization, supported by FastText for vector representation and cosine similarity for text matching. Perplexity evaluation confirmed that 22 topics represent the most optimal configuration for generating recommendations. Click-Through Rate (CTR) testing practically demonstrated the system's effectiveness; the combination of topic-based and user preference-based recommendations achieved significantly higher relevance, with a CTR of 15.6%, compared to the baseline CTR of only 2.2%. Therefore, the developed model is shown to be not only statistically optimal but also effective in enhancing user engagement.

For future research, it is recommended to integrate dynamic datasets so that the recommendation system can adapt to evolving topics in real-time. Additionally, subsequent development could explore combining LDA with other algorithms, such as K-Means or K-Nearest Neighbor (K-NN), to further strengthen the clustering and quality of recommendation results.

6. REFERENCES

Aleqabie, H. J., Aleqabie, H. J., Sfoq, M. S., Albeer, R. A., & Abd, H. (2024). A Review Of TextMining Techniques : Trends , and Applications In Various Domains A Review Of Text Mining Techniques : Trends , and Applications In Various Domains. 5(1).

Amalia, J. (2022). Membangun Slang Dictionary Untuk Normalisasi Teks Menggunakan Pre-Trained Fasttext Model. JSR : Jaringan Sistem Informasi Robotik, 6(2), 250–256. https://doi.org/10.58486/jsr.v6i2.184

Anisatuzzumara. (2024). Implementasi Latent Dirichlet Allocation (LDA) dan K-Nearest Neighbors (KNN) pada Sistem Rekomendasi Jurnal Terindeks Garuda. Universitas Sultan Agung Semarang (2024).

Bangkalang, D. H. (2023). Analisis dan penerapan topic modeling pada judul tugas akhir mahasiswa menggunakan metode latent dirichlet allocation (lda). 8(4), 1275–1287.

Bortz, A., Barth, A., & Czeskis, A. (2011). Origin Cookies : Session Integrity for Web Applications. Web 2.0 Security & Privacy 2011, 1(7), 8. http://www.w2spconf.com/2011/papers/session-integrity.pdf

Chauhan, U., & Shah, A. (2022). Topic Modeling Using Latent Dirichlet allocation: A Survey. ACM Computing Surveys, 54(7), 1–7. https://doi.org/10.1145/3462478

Erna, A., Jailani, A. K., Informasi, S., & Dipa, U. (2024). Identifikasi Potensi Topik Penelitian Baru : Analisis Judul Skripsi Teknik Informatika (2022-2024) menggunakan Latent Dirichlet Allocation (LDA). XIII(2), 159–167.

Fajriah, R., & Kurniawan, D. (2025). Optimalisasi Model Klasifikasi Naive Bayes dan Support Vector Machine Dengan Fast Text dan Chi Square. 17(4), 334–345. https://doi.org/10.30998/faktorexacta.v17i4.24751

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America, 101(SUPPL. 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Huda, A. A., Fajarudin, R., & Hadinegoro, A. (2022). Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita. Building of Informatics, Technology and Science (BITS), 4(3), 1679–1686. https://doi.org/10.47065/bits.v4i3.2511

Jovic, K., & Krstovic, S. (2022). Cookies Implementation Analysis and the Impact on User Privacy Regarding GDPR and CCPA Regulations.

Katili, M. Z., Amali, L. N., & Tuloli, M. S. (2021). Implementasi Metode AHP-TOPSIS dalam Sistem Pendukung Rekomendasi Mahasiswa Berprestasi. Jambura Journal of Informatics, 3(1), 01–10. https://doi.org/10.37905/jji.v3i1.10246

Khadijah, U., & Cahyono, N. (2024). Analisis Topic Modelling Pariwisata Yogyakarta Menggunakan Latent Dirichlet Allocation (LDA). The Indonesian Journal of Computer Science, 13(4), 6075–6086.

Kocimaheni, A. A., Aminin, Z., & Kartika, A. D. (2020). IDENTIFIKASI KESULITAN MAHASISWA FAKULTAS BAHASA DAN SENI UNIVERSITAS NEGERI SURABAYA DALAM PENYELESAIAN STUDINYA. 7(2), 139–146.

Liu, Y., Yuan, P., Liu, W., & Li, X. (2015). What drives click-through rates of tourism product advertisements on group buying websites? Procedia Computer Science, 55(Itqm), 221–230. https://doi.org/10.1016/j.procs.2015.07.037

Magai, E., & Putra, R. P. (2025). KLASIFIKASI TOPIK SKRIPSI TEKNIK INFORMATIKA BERDASARKAN PEMINATAN MENGGUNAKAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR (K-NN). 4(2), 149–157.

Meister, C., & Cotterell, R. (2021). Language Model Evaluation Beyond Perplexity. ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference, 5328–5339. https://doi.org/10.18653/v1/2021.acl-long.414

Murdiyanto, A. W., & Himawan, A. (2022). Identification Of Keywords That Impact Of Increasing The Click Through Rate Of Online Advertising On Search Engines. Telematika, 19(1), 77. https://doi.org/10.31315/telematika.v19i1.6450

Pamput, J. P., Muthmainnah, A. R., Surianto, D. F., & Fadilah, N. (2025). Perbandingan Cosine Similarity dan Weighted Jaccard Similarity dalam Pengembangan Mesin Pencari Perpustakaan Digital. 10(4), 907–919. https://doi.org/10.30591/jpit.v10i4.8773

Purwitasari, D., Aida Muflichah, Novrindah Alvi Hasanah, & Agus Zainal Arifin. (2021). Pemodelan Topik dengan LDA untuk Temu Kembali Informasi dalam Rekomendasi Tugas Akhir. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 5(3), 421–428. https://doi.org/10.29207/resti.v5i3.3049

Rosadi, Z., & Solichin, A. (2024). Topic Modeling Tugas Akhir Mahasiswa Menggunakan Metode Latent Dirichlet Allocation Dengan Gibbs Sampling. 13(September), 38–44.

Santika, T. (2019). Evaluasi Perplexity Untuk Pemodelan Topik Diskusi Agama Islam Di Media Sosial Twitter Indonesia Tahun 2006-2018 Menggunakan Latent Dirichlet Allocation [UIN Syarif Hidayatullah Jakarta]. https://repository.uinjkt.ac.id/dspace/handle/123456789/52331