

Scholarship Recommendation System Using Content-Based Filtering Approach

Ryan Ferdiansyah¹, Dwi Fatrianto Suyatno²

^{1,2} Universitas Negeri Surabaya, Surabaya, Indonesia

ryan.21073@mhs.unesa.ac.id, dwifatrianto@unesa.ac.id

ABSTRACT

The numerous prospective scholarship recipients who face difficulties in finding information that matches their educational profiles form the background of this research. The objective of this study is to design and develop a web-based scholarship recommendation system using the Content-Based Filtering (CBF) method and to determine the accuracy level of the Euclidean Distance algorithm through Precision testing. This research is a descriptive qualitative study utilizing the Rapid Application Development (RAD) approach, which includes Requirement Planning, User Design, Construction, and Cutover phases. Data sources consist of scholarship information from official institutions and user profile inputs. Data collection was conducted through literature studies, observations, and questionnaires. Functionality testing was performed using Black Box Testing, while recommendation accuracy analysis utilized the Precision method based on Euclidean Distance. The results showed that a web-based system using MERN Stack technology was successfully developed and capable of presenting a list of scholarships relevant to user profile metadata. Based on the testing, this algorithm achieved an average accuracy of 92%, proving that the system works optimally in prioritizing relevant scholarships and effectively overcomes the cold-start problem. In conclusion, this system is mathematically validated as a decision support tool for scholarship seekers. Suggestions for future development include integrating the system with external APIs such as PDDikti for automatic data validation and improving the text-matching mechanism to avoid rigid exact matches.

Keyword: Recommendation System, Scholarship, Content-Based Filtering, Euclidean Distance, RAD.

Article Info:

Article history:

Received July 09, 2026

Revised August 10, 2026

Accepted September 15, 2026

Corresponding Author

Ryan Ferdiansyah

Universitas Negeri Surabaya, Surabaya, Indonesia

ryan.21073@mhs.unesa.ac.id

1. INTRODUCTION

Education constitutes the primary foundation for building quality human resources and plays a strategic role in national development [1]. As a fundamental right guaranteed under Article 31 of the 1945 Constitution of the Republic of Indonesia, the provision of education is a state responsibility, a mandate reinforced by Law No. 20 of 2003 on the National Education System and Government Regulation No. 48 of 2008 on Education Funding, both of which require the state to provide scholarships for students from low-income families. This policy has proven impactful, benefiting more than 160,000 students every year [2]. Various scholarship programs have since been introduced to widen access to education across all levels of society,

ranging from the Smart Indonesia Card (*Kartu Indonesia Pintar/KIP*) for basic and secondary education to the Indonesia Endowment Fund for Education (*Lembaga Pengelola Dana Pendidikan/LPDP*) for higher education. Between 2012 and 2021 alone, LPDP distributed aid to 29,872 recipients [3], while by mid-2023 the combined number of beneficiaries from the Ministry of Education, Culture, Research, and Technology, the Ministry of Religious Affairs, and LPDP had reached 38,913, 26,575, and 7,605 respectively [4].

Despite this steady growth in scholarship distribution, many prospective recipients still struggle to locate information that matches their educational profile. The absence of a platform capable of aggregating scholarship information in a structured manner and generating recommendations based on individual preferences remains a central obstacle. Orgianus et al. argue that the lack of an integrated system forces students to search for information manually across multiple sources, a process that is both time-consuming and prone to inaccuracy [5]. This is compounded by findings that more than 65% of prospective scholarship applicants rely on social media and unofficial channels [6], which frequently yield outdated or unverified information rather than data from credible sources.

Addressing this gap requires a web-based scholarship recommendation system capable of guiding users toward the scholarships most relevant to their individual circumstances. Beyond merely consolidating scholarship listings on a single platform, such a system should apply a *Content-Based Filtering* (CBF) approach to generate personalized recommendations. CBF operates by matching user profile attributes — education level, field of study, grade point average (GPA), domicile, and economic condition — against the metadata of each available scholarship, so that the resulting suggestions are grounded in the characteristics of the user rather than the behavior of other users.

Among the recommendation approaches commonly discussed alongside CBF are *Collaborative Filtering* and *Hybrid Filtering*. Collaborative Filtering, however, depends heavily on historical user-interaction data, which makes it less effective for new users who have not yet generated any interaction history. CBF was therefore chosen for its ability to produce recommendations independently from profile characteristics alone, effectively resolving this *cold-start* problem. Its effectiveness is well documented in recent quantitative studies: a web-based makeup recommendation system achieved a Precision score of 80.8% [7], an e-book recommendation system reported an accuracy of 78.4% [8], and a coffee-selection application recorded a user-satisfaction rate of 91.1% [9]. These figures indicate that CBF is not only conceptually suitable but also empirically reliable for problems characterized by limited initial interaction data.

Even so, most existing applications of CBF remain concentrated in consumer-oriented domains such as cosmetics, e-books, and coffee preference, leaving its use in scholarship recommendation largely unexplored [7]–[9]. Decision support systems built specifically for scholarship allocation, meanwhile, have generally relied on additive scoring methods such as Simple Additive Weighting, which rank alternatives against fixed criteria rather than tailoring results to the interaction between an individual applicant's profile and each scholarship's specific requirements [10]. The present study addresses this gap by combining a CBF approach with the *Euclidean Distance* algorithm to measure profile-to-scholarship similarity, an application that, to the authors' knowledge, has not previously been examined in the scholarship domain.

Building on this gap, the present study develops a web-based Decision Support System (DSS) for scholarship recommendation using the CBF method, targeting three intended outcomes. The first is relevance, ensuring the system presents scholarship lists aligned with each user's profile metadata. The second is accuracy, measured quantitatively through a

Confusion-Matrix-based Precision test to validate the performance of the applied Euclidean Distance algorithm. The third is efficiency, reflected in the reduction of manual search time through automated recommendation generation. Accordingly, the system is designed not merely as a repository of scholarship information but as a mathematically validated decision-support tool that helps prospective applicants identify the educational funding opportunities best suited to them.

2. METHODS

This study adopts a descriptive qualitative research design aimed at designing and building a web-based scholarship recommendation application through a *Content-Based Filtering* (CBF) approach. Rather than testing a hypothesis in the conventional sense, the research traces a design-and-build process: from problem formulation, through literature review and data collection, to system development, testing, and evaluation of the resulting prototype.

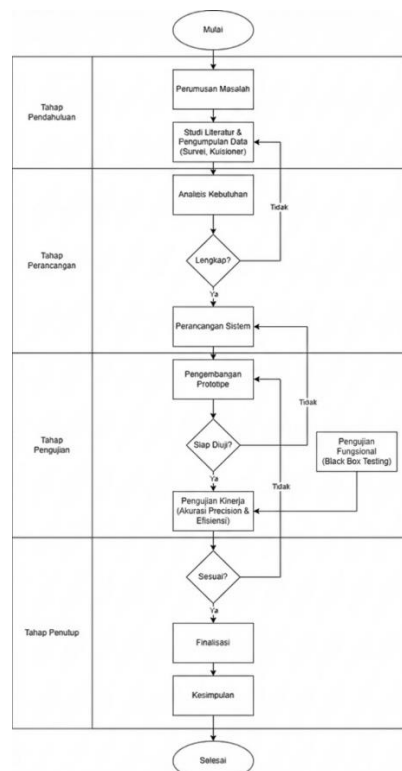


Figure 1. Research Flow Diagram

As illustrated in Figure 1, the process begins by identifying the difficulty prospective scholarship applicants face in locating information relevant to their educational background, followed by a literature study on recommendation-system theory and the *Rapid Application Development* (RAD) approach, before proceeding to requirement analysis, system design, development, and functional testing.

Preliminary data were gathered through three complementary techniques: a literature study of academic references and technical documentation related to recommendation systems and CBF; direct observation of existing scholarship-information platforms to understand how scholarship data is structured and presented; and an online questionnaire distributed to 70 respondents. Respondents were predominantly students at the senior-secondary and undergraduate level, and 82.8% reported having searched for a scholarship at some point, with

37.1% doing so frequently and 45.7% occasionally. The same proportion, 82.8%, also reported difficulty in locating suitable scholarship information — 35.7% frequently and 47.1% occasionally — indicating that although scholarship-seeking behavior is common, the search process itself remains inefficient.

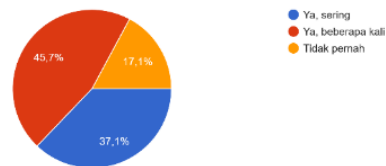


Figure 2. Questionnaire 1 Results Diagram

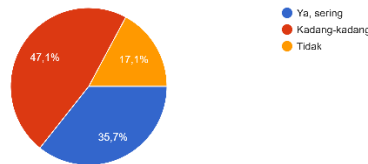


Figure 3. Questionnaire 2 Results Diagram

These findings confirmed the need for a structured, profile-based recommendation mechanism rather than a simple keyword-based search.

System development followed the RAD method, selected for its capacity to support fast, iterative software delivery through continuous collaboration between developers and users [11]. RAD comprises four phases — *Requirement Planning*, *User Design*, *Construction*, and *Cutover*.

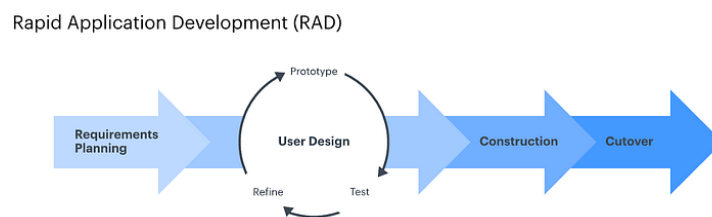


Figure 4. RAD method

During *Requirement Planning*, functional needs (personalized CBF-based recommendation, user-profile management, admin content management, JWT-based authentication, and an administrative monitoring dashboard) and non-functional needs (response speed, data security, scalability, availability, resource efficiency, and maintainability) were defined. The *User Design* phase translated these requirements into interface flows for three user roles — general user, contributor, and admin — covering registration, profile completion, scholarship browsing, and recommendation display. In the *Construction* phase, the system was implemented using the *MERN* stack: *MongoDB* for data storage, *Express.js* and *Node.js* for backend services and API handling, and *React.js* for a responsive, interactive frontend. Each component was tested iteratively as it was built. Finally, the *Cutover* phase covered deployment and real-time monitoring of server performance and user behavior to catch any remaining functional or security issues before full release.

During the system design stage, Unified Modeling Language (UML) was used to model system functionality and the interaction among the platform's three user roles: User, Contributor, and Admin.

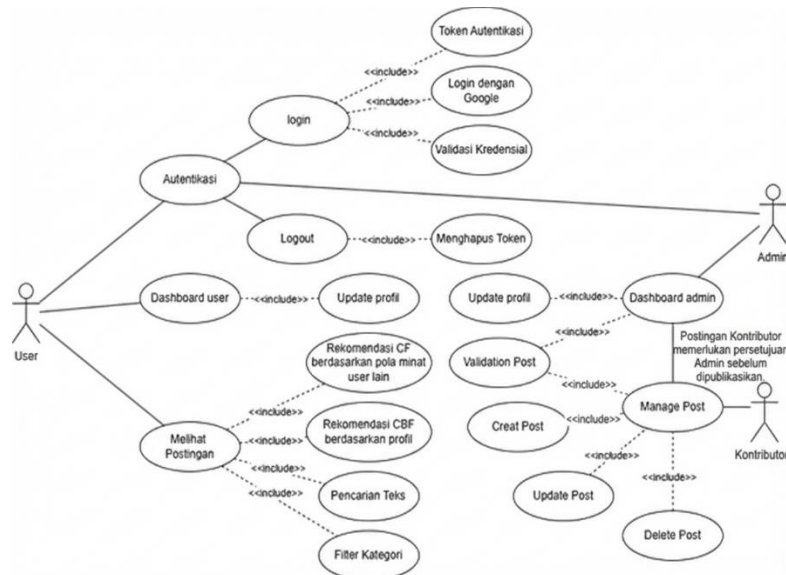


Figure 5. Use Case Diagram

As shown in Figure X, all three roles interact through a shared authentication gateway covering *Login* — supported by credential validation, Google sign-in, and authentication-token generation — and *Logout*, which revokes the active token. Users and Contributors share access to a personal dashboard for profile updates and to a scholarship-browsing feature equipped with text search, category filtering, and CBF-based recommendations. Admins, in turn, have exclusive access to an administrative dashboard for managing their own profile and validating incoming submissions. Content management itself is shared between the Admin and Contributor roles: both may create, update, and delete postings, but a submission created by a Contributor first enters a pending state and requires Admin approval before it is published to the platform.

The core recommendation logic applies CBF by matching user-profile attributes — education level, field of study, GPA, domicile, and parental income — against the metadata of each listed scholarship. Categorical attributes such as field of study and domicile were first converted into numerical values and normalized to a common scale prior to computation. Similarity between a user profile P and a scholarship B was calculated using the *Euclidean Distance* metric [12], [13]:

$$d(P, B) = \sqrt{\sum (P_i - B_i)^2}, \text{ for } i = 1 \text{ to } n \quad (1)$$

where n denotes the number of attributes used, and P_i and B_i represent the value of the i -th attribute for the user profile and the scholarship, respectively. A smaller distance value indicates a closer match between the user's profile and the scholarship's requirements, placing that scholarship higher in the recommendation list.

System performance was evaluated quantitatively using a *Confusion Matrix* [14], which classifies each recommendation outcome into four categories: True Positive (TP), a relevant scholarship correctly recommended; False Positive (FP), an irrelevant scholarship mistakenly recommended; True Negative (TN), an irrelevant scholarship correctly excluded; and False Negative (FN), a relevant scholarship mistakenly excluded. Based on this classification, recommendation accuracy was measured using *Precision* [15], defined as the ratio of relevant items retrieved to the total items retrieved:

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

A higher Precision value indicates fewer irrelevant recommendations reaching the user, and thus more reliable recommendation output.

Functional verification was carried out using *Black Box Testing*, which evaluates system behavior from the user's perspective by comparing actual output against expected output for a given input, without inspecting the underlying code. Testing scenarios covered user registration and authentication, profile completion, scholarship browsing, and the accuracy of the CBF-generated recommendation list, and were repeated at each development iteration so issues could be corrected early.

3. RESULTS AND DISCUSSION

3.1 Landing Page and Core Navigation



Figure 6. Home Page View

The landing page serves as the system's main entry point, built with *React.js* and styled using *Tailwind CSS*. It is arranged into four vertical segments: a hero section introducing the "Gerbang Beasiswa" platform and its recommendation capability, a value-proposition segment explaining the CBF-based mechanism, a short how-it-works guide, and a dynamically loaded list of the most recent scholarships retrieved from the database. Two primary calls to action are placed side by side in the hero section — one directing first-time users to complete their profile and receive personalized recommendations, and another allowing users to browse the full scholarship archive without registering.

3.2 Scholarship Search and Recommendation Interface

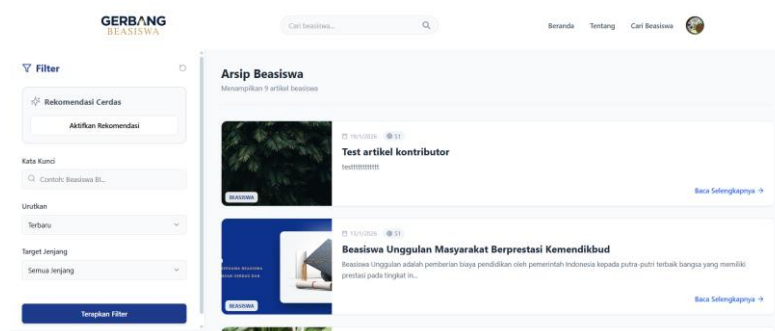


Figure 7. Scholarship Archive Page View with Filter and Search

This page functions as the system's main interaction hub, combining a sticky filter sidebar with a scrollable result list. Users can filter scholarships by keyword, education level, and posting recency, or toggle a "Smart Recommendation" switch to activate CBF-based ranking. When this mode is active, matching scholarship cards are visually distinguished with a highlighted border and a badge indicating "Very Suitable" or "Meets Criteria," allowing users to gauge relevance at a glance before opening the full detail page.

3.3 Match Analysis Page

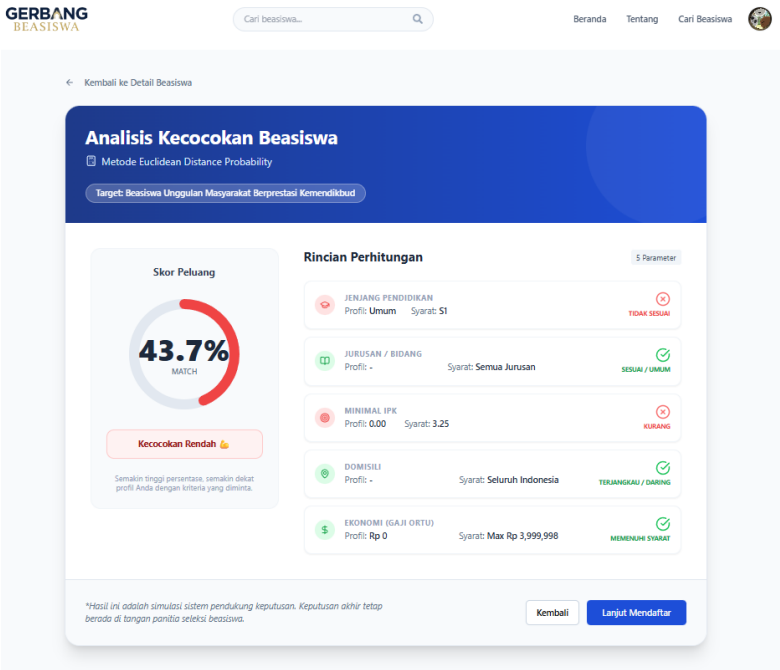


Figure 8. Scholarship Recommendation Match Analysis Page View

To make the recommendation process transparent, each scholarship detail includes a match-analysis view presenting a real-time compatibility score as a gauge chart, color-coded green, yellow, or red depending on the match percentage. Below the chart, a side-by-side comparison lists the user's profile value against each scholarship's required value for all five matching criteria, with a check or cross icon indicating whether that specific criterion is satisfied. This design allows users to see exactly which attributes support or limit their eligibility, rather than receiving a single opaque score.

3.4 Role-Based Content Management

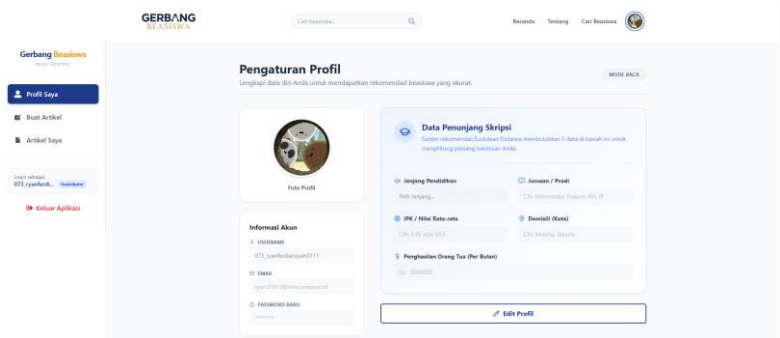


Figure 9. Profile Dashboard Page View

The system distinguishes three access levels: general users, who may only browse scholarships and receive recommendations; contributors, who may submit new scholarship listings that enter a "Pending" state until reviewed; and administrators, who manage the full

scholarship database and approve or reject contributor submissions through a dedicated moderation queue. This tiered structure keeps scholarship data centrally curated while still allowing community contribution.

3.5 Algorithm Verification

To confirm that the deployed algorithm matches its mathematical definition, a manual verification was performed using a sample user profile (S1, Informatics Engineering, GPA 3.40, domiciled in Surabaya, parental income equivalent to a normalized value of 5.0) against three scholarship records.

Table 1. Sample Scholarship Data Used for Verification

Criteria	Scholarship 1	Scholarship 2	Scholarship 3
Education level	S1	S1	S1
Field of study	All majors	Informatics	All majors
Minimum GPA	2.50	3.25	2.80
Domicile	Nationwide	Nationwide	Surabaya
Parental income cap	4.0 (normalized)	15.0 (normalized)	10.0 (normalized)

Categorical attributes were converted using an exact-match rule (0 for a match, 1 for a mismatch, and an automatic 0 for "all majors"/"nationwide" entries), while numeric attributes used the direct difference between the user's value and the scholarship's threshold. Applying Equation (1) manually to each pair produced distances of $d = 1.345$ for Scholarship 1, $d = 5.035$ for Scholarship 3, and $d = 10.001$ for Scholarship 2 — placing Scholarship 1 first due to its narrow economic gap, Scholarship 3 second, and Scholarship 2 last despite its close academic match, owing to its far larger income gap. This ranking was identical to the system's actual output, confirming that the CBF–Euclidean Distance logic was implemented correctly in the application code.

3.6 Scholarship Search and Recommendation Interface

Recommendation accuracy was then tested across five distinct user-profile scenarios, each varying a different attribute, to evaluate ranking consistency.

Table 2. Precision Testing Results

No	Scenario	Items Displayed	Relevant (Top 5)	Precision
1	Education level	9	5	100%
2	Field of study	9	4	80%
3	GPA	9	5	100%
4	Domicile	9	4	80%
5	Parental income	9	5	100%
Average accuracy				92%

The two scenarios scoring 80% involved scholarships that technically met the stated criteria but carried a slightly wider distance score due to minor economic-parameter differences, which pushed them just outside the top-5 relevant set. Overall, the 92% average Precision indicates that the CBF-based ranking mechanism performs reliably across varying profile conditions.

3.7 System Testing

Black Box Testing was carried out across the system's core workflows, covering authentication, content management, search and filtering, profile validation, recommendation execution, and session handling.

Table 3. Black Box Testing Summary

Module	Expected Result	Result
Login authentication	Valid credentials grant access via JWT session	Successful
Login validation	Invalid credentials are rejected with an error message	Successful
Scholarship create/update	Admin-submitted data is saved and reflected immediately	Successful
Search and filter	Results match the selected category and keyword in real time	Successful
Profile validation	Recommendation is blocked until required profile fields are complete	Successful
Recommendation execution	Scholarships are ranked correctly by Euclidean Distance	Successful
Logout	Session ends and access reverts to the login state	Successful

All tested scenarios produced the expected outcome, indicating that the system's core functionalities operate reliably under normal usage conditions.

3.8 Discussion

The 92% average Precision achieved by "Gerbang Beasiswa" is consistent with, and slightly exceeds, the accuracy levels reported by prior CBF-based recommendation systems in other domains — 80.8% for makeup recommendation [7] and 78.4% for e-book recommendation [8] — and approaches the 91.1% satisfaction rate reported for a coffee-selection application [9]. This difference is plausibly explained by the nature of the matching problem itself: unlike product-description domains that rely on text-similarity techniques

such as *TF-IDF* or *Cosine Similarity* to compare unstructured content, scholarship eligibility is defined by a small set of well-structured categorical and numeric attributes, which the *Euclidean Distance* metric is naturally suited to compare with low ambiguity.

The findings also address the gap identified earlier: prior scholarship-specific decision support systems have generally applied additive scoring methods such as Simple Additive Weighting [10], which rank alternatives against fixed criteria weights but do not compute a personalized distance between an individual applicant's own profile and each scholarship's requirements. By combining CBF with Euclidean Distance, this study demonstrates that a *cold-start-resistant*, profile-driven matching approach — previously validated mainly in consumer-product domains — is equally applicable to, and effective within, the scholarship-allocation context.

At the same time, the system's accuracy remains bounded by three practical limitations observed during testing and implementation. First, recommendation quality depends entirely on the honesty and accuracy of user-submitted profile data, since no external verification (e.g., integration with the national higher-education database, PDDikti) is currently in place. Second, the exact-match rule used for categorical attributes such as field of study can undercount genuinely relevant scholarships when program names are phrased differently but refer to the same field (e.g., "Informatics Engineering" versus "Informatics"). Third, the system remains rule-based and does not yet incorporate machine-learning techniques capable of learning from user browsing or click behavior over time. These limitations point directly to concrete directions for future development, discussed further in the Conclusion.

CONCLUSION

This study successfully designed and developed "Gerbang Beasiswa," a web-based scholarship recommendation system built using the *Rapid Application Development* (RAD) method and the *MERN Stack*. Black Box Testing confirmed that all core functionalities — from scholarship data management by administrators to search, filtering, and recommendation execution by end users — operated correctly and as specified, addressing the first research question regarding the system's functional feasibility.

Regarding recommendation accuracy, the implementation of *Content-Based Filtering* combined with the *Euclidean Distance* algorithm proved effective, achieving an average Precision of 92% across five distinct user-profile testing scenarios. Manual verification further confirmed that the system's ranking output matched its mathematical formulation exactly, demonstrating that the recommendation logic was implemented correctly rather than merely appearing accurate by chance. Taken together, these results answer the second research question and indicate that the system is reliable in prioritizing scholarships most relevant to a user's academic and economic profile, substantially reducing the manual effort previously required to search for suitable opportunities.

Beyond its practical outcome, this study contributes to the growing body of literature on *Content-Based Filtering* by demonstrating its applicability to a domain — scholarship allocation — that has so far been dominated by additive multi-criteria methods such as Simple Additive Weighting. Whereas such methods rank alternatives against fixed criteria weights, the profile-to-item distance approach used here personalizes recommendations to each applicant's specific combination of academic and economic attributes, while remaining unaffected by the *cold-start* problem that limits collaborative approaches for first-time users.

Nevertheless, the system's current design leaves room for further refinement. Future work may integrate *Natural Language Processing* or string-similarity techniques such as *Levenshtein Distance* to reduce the rigidity of the exact-match rule currently applied to categorical attributes like field of study. Connecting the system to an external academic database such as PDDikti would also allow automatic verification of user-submitted GPA and enrollment data, reducing reliance on self-reported information. Finally, incorporating machine-learning techniques capable of learning from user interaction patterns — such as click history and time spent viewing specific listings — could complement the current rule-based profile matching with a more adaptive, behavior-aware layer of personalization.

REFERENCES

- [1] M. Sa'adah, "Studi komparatif reformasi pendidikan di Singapura dan Indonesia," *Jurnal Pembangunan Pendidikan: Fondasi dan Aplikasi*, vol. 7, no. 1, pp. 70–79, 2020, doi: 10.21831/jppfa.v7i1.26499.
- [2] Suntono, "Penawaran Program Beasiswa Pusbindiklatren Bappenas Tahun 2024 Tahap 2," Kementerian PPN/Bappenas, 2023.
- [3] N. Adisty, "Berapa Jumlah Penerima Beasiswa LPDP Tiap Tahun?" Goodstats, May 25, 2022.
- [4] S. Erlina, "Ini Jumlah Penerima Beasiswa Pemerintah Indonesia hingga Pertengahan 2023," Databoks, Jun. 27, 2023.
- [5] Y. Orgianus, F. R. I. Lapalanti, H. Tarmizi, and H. Oemar, "Optimizing scholarship distribution: a management information system approach," *Acta Logistica*, vol. 11, no. 2, pp. 197–209, 2024, doi: 10.22306/AL.V11I2.497.
- [6] S. Gunani, "Pengumuman Bantuan Pemerintah Program Pembelajaran Internasional 2023," Kementerian Pendidikan Tinggi, Sains, dan Teknologi, Apr. 6, 2023.
- [7] P. M. S. Utami, I. N. P. Trisna, and W. O. Vihikan, "Web-based makeup recommendation system using hybrid filtering," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 3, 2025.
- [8] X. Fabiano et al., "Implementasi sistem e-book rekomendasi berbasis web sederhana menggunakan Content-Based Filtering," *AICOMS 2025*, vol. 4, no. 1, pp. 15–24, 2025, doi: 10.58466/aicoms.v3i2.1789.
- [9] N. A. Naziron, N. B. A. Shah, and M. H. Raji, "A constructive Content-Based Filtering recommendation application: Optimizing coffee selection based on user preferences," *Information Management and Business Review*, vol. 17, no. 1, 2025.
- [10] I. Maryani, V. Ma'arif, and N. S. Kristiana, "Sistem pendukung keputusan penentuan beasiswa berbasis web menggunakan metode Simple Additive Weighting (SAW)," *Jurnal Sains dan Manajemen*, vol. 8, no. 2, 2020.
- [11] M. A. Fauzi, H. Tribiako, A. Moniva, I. Khalid Ilyas, and E. Utami, "Systematic literature reviews on rapid application development information system," *Bulletin of Computer Science and Electrical Engineering*, vol. 4, no. 1, pp. 57–64, 2023.
- [12] E. Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta, Indonesia: Andi, 2014.
- [13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.
- [14] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*. Berlin, Germany: Springer, 2011.
- [15] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2008.