Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 7, No.1, January 2025, pp. 0250117-01 ~ 0250117-08
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v7i1.1250

Hybrid Filtering for Student Major Recommendation: A Comparative Study

Nurtriana Hidayati[1*], Titin Winarti[2], Alauddin Maulana Hirzan[3]

[1]Faculty of Information Technology and Communication, Universitas Semarang, Jl. Soekarno-Hatta Tlogosari Semarang 50196, Central Java, Indonesia

[2]Faculty of Information Technology and Communication, Universitas Semarang, Jl. Soekarno-Hatta Tlogosari Semarang 50196, Central Java, Indonesia

[3]Faculty of Information Technology and Communication, Universitas Semarang, Jl. Soekarno-Hatta Tlogosari Semarang 50196, Central Java, Indonesia

*anna@usm.ac.id

Abstract. Choosing the right university major is an important decision for students, as delays or incorrect choices can harm their future careers and cause problems for academic departments. High dropout rates, which are frequently the result of poorly informed decisions, can be a considerable burden on faculty. This project aims to address these challenges by creating a recommendation system that provides individualized counsel to students based on their psychological profiles. A quantitative method was used, with questionnaires distributed to a large number of students. To verify the data's authenticity, replies were sought from students who were pleased with their selected majors rather than those who regretted their choices. The collected data formed the basis for a hybrid recommendation system that integrated Content-based Filtering and Collaborative Filtering methods. The system was then compared against standalone implementations of each filtering method to determine its usefulness in increasing suggestion accuracy. The results showed that the Hybrid Filtering strategy obtained a recommendation accuracy of 84.29%, outperforming Content-based Filtering at 81.43% and Collaborative Filtering at 78.57%. The proposed model is easy to implement in a school or a university, as long as the required data is available. Thus, the model can help a school or university to reduce dropout rates and boost academic outcomes.

Keywords: Hybrid Filtering, Recommendation Systems in Education, Student Major Selection

(Received 2024-11-13, Accepted 2024-12-31, Available Online by 2025-01-23)

1. Introduction

Selecting a university major is a critical decision that significantly impacts students' academic and professional futures[1], [2], [3]. For example in this matter, if someone wants to be a data analyst then he must graduate from an information system major. Choosing the incorrect university major will hinder his goals. Hence, he must consider this carefully. Traditional guidance methods often rely on subjective advice or generalized aptitude tests, which may not fully align with the individual's unique psychological profile and vocational interests[4]. Because of that, some university students face problems during the learning process[5]. This problem sometimes leads to increasing drop-out numbers[6], [7]. There are studies about student's satisfaction and regrets about their major choices[8], [9], [10]. Many students regret their major choice after entering university. Because of that, they suffer low learning performance and bad grades[11], [12]. These also affect their professional aim in the future. Thus, increasing the number of jobless unskilled graduates[13], [14].

To solve the problem in the traditional approach, many studies proposed recommendation models to recommend university majors based on individual qualities, like education or psychology. Based on the study in 2020, the article designed a recommendation system using a Collaborative Filtering algorithm which is implemented in a high-speed web framework for easier access[15]. However, this article only focuses on courses that indirectly affect the major course. In the same year, another researcher published a different model with a different algorithm. This model utilized an Expert System to determine university majors based on personal data[16]. The next year 2021, the researcher improved the model with different algorithms to recommend university majors. The model utilizes k-Nearest Neighbor to determine a suitable major for the students[17]. In 2023, the research still improves the existing model with different algorithms. The model uses Hybrid Feedback to determine student's major choice for Informatics and Non-Informatics majors[18], [19]. The latest recommendation in 2024 is to use the Fuzzy Inference System Mamdani. The model uses many different data like school exam grades, talents, and computer-based exams[20].

The previous paragraph discusses the current state-of-the-art in university major recommendation systems. Typically, these systems rely on a single algorithm to suggest suitable majors for students, based on various types of data. However, a common limitation in existing models is their reliance on only one algorithm, which may yield reasonably good results but might not achieve optimal accuracy. Another limitation concerns the data used in these studies; most rely on exam scores or academic performance in specific subjects, which may not fully capture a student's personality or interests. Thus, this problem potentially leads students to pursue majors that are not the best fit for their abilities, interests, or long-term goals. Additionally, the narrow scope of data, often limited to exam scores or academic performance, overlooks critical factors like personality traits and personal interests, essential for aligning students with majors that resonate with their aspirations and motivations. This can result in decreased satisfaction, higher dropout rates, and less effective career alignment, undermining the purpose of these systems.

To address these problems, this study proposes a Student Major Recommendation System that integrates content-based and collaborative filtering approaches. The content-based filtering is used as a recommendation based on existing data. This algorithm is implemented for music or movie recommendation systems[21], [22]. Meanwhile, collaborative filtering uses participants' favourites to recommend the choices. This algorithm is often used as a recommendation system in learning content or common products[23], [24]. While content-based filtering offers personalized recommendations, it may struggle with diversity, whereas collaborative filtering can introduce more variety but may face challenges with new users or items[25]. The next consideration is about the data. The system leverages data from psychological assessments, specifically the OCEAN model of personality traits and Holland's vocational theory, to provide personalized recommendations[26], [27].

Existing research on university major recommendation systems highlights significant gaps. Current models rely on single algorithms, limiting accuracy and personalization, and use narrow datasets focused on academic performance, neglecting factors like personality and interests. While content-based and collaborative filtering algorithms show promise in other domains, their combined potential for major recommendations is underexplored. This study addresses these gaps by proposing a hybrid approach that integrates both techniques, balancing personalization and diversity to improve recommendation quality. By combining these theoretical frameworks with advanced filtering techniques, the system aims to enhance the accuracy and relevance of major recommendations, thereby supporting students in making more informed decisions.

2. **Methods**

In this section, this study explains the research process. The first step is to find the problem with the university major selection. This is a prevalent situation where a student has to choose his university major without an outsider telling him what to do. In some cases, there is a student who follows his friend around and does not consider his professional future. These problems cause most students to regret about their choices.

The impact of this problem is the increase in dropouts. This event will be a problem for the faculty and university. The next step is to gather the data for the proposed system. The system will not work without data to feed hybrid algorithms. In this case, this study gathers the required data by giving questionnaires to Information System (IS) and Information Technology students that satisfied with their choices equally. Thus, this technique will ensure that the data from both side is balanced. With this method, the students will give their consent about their psychological information. However, this study only takes initials as the identity to comply with data privacy concerns[28].

The questions consist of the OCEAN (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism) personality test[29], [30] and Holland's Theory of Vocational Choice[31]. For the OCEAN category, this study creates three questions with the Likert scale as the answer for each element. The total number of questions for this category is 15 questions. Meanwhile, the other category has six questions where three questions are intended for Information Systems and the rest for Information Technology. These questions will help to determine whether the participant is an Information System or Information Technology. This study has gathered more than 50 data rows where half of them are Information System students and the rest are Information Technology.

After obtaining the data from the questionnaire, the next process is to pre-process the data. Since the form requires the student to fill all the required columns, data cleaning process was not required. However, data pre-processing is still required after obtaining the data. All data in OCEAN columns are formatted on a Likert scale from 1 to 5. Thus, data pre-processing is required to simplify the Likert scale into a simple form that is easy to use with the proposed model. This study utilized MinMax preprocessing to simplify the result from OCEAN columns. After data preprocessing is done, then the data is ready to be used inside the hybrid model. The proposed hybrid model is configured with 50:50 weight for both Content-based and Collaborative Filterings. Assigning a 50:50 weight to content-based and collaborative filtering ensures a balanced approach, combining the personalization of content-based methods with the diversity of collaborative filtering. This balance mitigates the limitations of each method while providing accurate, robust, and holistic recommendations for university majors. This study used a web-based platform as the recommendation system. This technology utilizes the Python Web Framework, HTML and CSS. Then, the system will be uploaded to a cloud server for easier access.

The recommendation system in this study uses a form-shaped interface replicating the data-gathering process in the previous explanation. However, this system has additional functions to calculate Content-based Filtering and Collaborative Filtering. The results from both algorithms were then merged based on each weight. For the evaluation process, this study will invite some different students to try the system by entering their personalities. This system then calculates the recommended major based on their input. The result between the recommendation and the real result will be calculated with a confusion matrix to obtain accuracy[32], [33]. With the result, this study can compare between Content-based only, Collaborative only, and Hybrid Filterings' accuracies.

3. Results and Discussion

In this section, this study explains the recommendation system's result based on the testing process. After the testing process, this study has gathered data from as many as 70 rows that consist of 35 IS and 35 IT students. The result of the system is shown in Table 4:

Table 4. Recommendation System Result based on Student Input

Id	Content-based		Collaborative		Recommendation		Result	
	IS	IT	IS	IT	IS	IT	Recommend	Real
1	0.69	0.58	0.94	0.00	0.81	0.29	IS	IS
2	0.68	0.52	0.92	0.00	0.80	0.26	IS	IS
3	0.67	0.55	0.00	0.92	0.33	0.73	IT	IS
36	0.58	0.69	0.00	0.93	0.29	0.81	IT	IT
37	0.46	0.70	0.00	0.91	0.23	0.80	IT	IT
38	0.54	0.73	0.00	0.95	0.27	0.84	IT	IT

Table 4 presents the results of testing for Content-Based Filtering, Collaborative Filtering, and Hybrid Filtering. A higher value between IS and IT indicates a stronger recommendation for that major. After calculating individual results for Content-Based and Collaborative Filtering, Hybrid Filtering was computed with a balanced 50:50 weight ratio, yielding results similar to Content-Based Filtering. The highest percentage corresponds to the preferred major recommendation. The "Major" column lists the student's current major, while the "Recommendation" column shows the recommended major. Overall, the system successfully identifies the recommended major, though some recommendations were inaccurate.This study evaluated the recommendation system's accuracy by compiling the results into a confusion matrix. Inside the matrix, there are two predictions and two data classes. With this matrix, this study can calculate the accuracy with the following equation:

$$Acc(\%) = \frac{(S_c + T_i)}{(S_c + S_i + T_c + T_i)} x100\% \qquad (1)$$

Equation 1 calculates the accuracy of the confusion matrix, represented as Acc (%) in percentage form. In this equation, S and T represent Information Systems and Information Technology, respectively, with ii denoting incorrect predictions and cc correct predictions. To calculate accuracy, the correct predictions $(S_c + T_i)$ are summed and then divided by the total data count. This value is then multiplied by 100% to express accuracy as a percentage. Before calculating accuracy, confusion matrices were created for each filtering method—Content-Based, Collaborative, and Hybrid. Each confusion matrix includes two classes: Information Systems (IS) and Information Technology (IT). The table below presents the confusion matrix and accuracy for Content-Based Filtering only.

Table 5. Confusion Matrix Result

Filtering Type	True Positive	True Negative	False Positive	False Negative	Total
Content-based	33	24	11	2	70
Collaborative	35	20	15	0	70
Hybrid	33	26	2	9	70

According to Table 5, the recommendation system for content-based only correctly recommends information systems for as many as 33. Meanwhile, information technology has as many as 24. However, this algorithm suffered incorrect recommendations as many as 2 for Information Systems and 11 for Information Technology. The recommendation accuracy for this system was 81.43%. Different from the previous result, Collaborative Filtering method correctly recommended 35 information systems and 20 information technology. However, this algorithm also suffered incorrect recommendations as many as 15 recommendations. The accuracy of this algorithm was 78.57%.

Meanwhile, the result of Hybrid Filtering correctly recommended 33 students for Information Systems and 26 for Information Technology. Although some recommendations were incorrect, the combined system had fewer errors than using a single algorithm, achieving an accuracy of 84.29%.

The following subsection discusses the results in detail. The first point addresses the significance and implications of this study. This study proposed a hybrid model that utilized content-based and collaborative filtering to recommend students' majors based on their psychology. Thus, this model will help the students to determine their future major. The second discussion is about the comparison with the previous models. This model utilized two filtering methods to create a recommendation for the students. This approach is different from previous models that only utilize a single algorithm. Thus, the proposed model has better recommendations compared to previous models. Based on the evaluation, the proposed model reached an accuracy of up to 84.29%, whereas the other models only reached 78.57% (for Collaborative Filtering only) and 81.43% (for Content-based Filtering only). The third discussion is about the impact of the proposed model. This proposed model is easy to implement in other schools or universities with minor adjustments. As long as the used data is credible and valid, the proposed model will automatically adapt to the situation. Thus, the proposed model can help a school or university to decrease dropout rates. The fourth discussion is about the strengths and weaknesses of the proposed model. This model has been evaluated with high accuracy, but it does not mean that the model is powerful. There is a limitation where this model is data-dependent. This limitation means that incorrect or unreliable data, such as feedback from students with regrets, could reduce accuracy and increase errors. Therefore, ensuring data accuracy is essential. Less data also reduces the accuracy of the model. Besides that, there is another limitation of this model where students' academic achievements were not considered and only focused on psychological aspects. The last discussion is about the future direction of this study. Several aspects can be improved like adding academic achievements as consideration, using a different algorithm to improve accuracy, and implementing a more user-friendly interface.

4. Conclusion

University majors are crucial things for the students to decide earlier. The late decision will affect not only their professional future but also the faculty, especially the departments. The increase in drop-outs will become a burden for the faculty. The solution to this matter is to create a recommendation system. Many have tried to design a recommendation system with many different algorithms. However, the results were not satisfying. This study designed a recommendation by combining Content-based and Collaborative Filtering to help recommend students based on their psychological conditions. After testing and evaluation processes, this study found that Hybrid Filtering has the highest recommendation accuracy of 84.29% followed by Content-based Filtering at 81.43% and Collaborative Filtering at 78.57%. This result proved that combining Content-based Filtering and Collaborative Filtering increased the accuracy compared to a single algorithm system. Although the proposed model successfully creates a recommendation for the students, it does not mean that the model is perfect. Several aspects of the model can be improved. For example, adding academic achievements to the model since this model does not include that kind of data. Either way is to use different algorithms like Random Forest to improve the recommendation's accuracy.

Acknowledgements

This work was supported by *Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Semarang* under *Penelitian Hibah Kompetisi Internal USM* Grand number 016/USM.H7.LPPM/L/2024.

References

[1] N. Tastanbekova, B. Abenova, M. Yessekeshova, Z. Sagalieva, and G. Abildina, "Development of Professional Skills in the Context of Higher School Dual Education," *Int. J. Emerg. Technol. Learn. IJET*, vol. 16, no. 10, pp. 179–193, May 2021.

[2] E. Lehtinen, “Can simulations help higher education in training professional skills?,” *Learn. Instr.*, vol. 86, p. 101772, Aug. 2023, doi: 10.1016/j.learninstruc.2023.101772.
[3] N. A. Khumaira and M. T. Safirin, “Impact of Employee Placement, Motivation, and Career Development on Performance and Productivity at Bank XYZ Using PLS-SEM,” *Adv. Sustain. Sci. Eng. Technol.*, vol. 7, no. 1, pp. 0250104–0250104, 2025.
[4] D. A. Leontiev, E. N. Osin, A. K. Fam, and E. Y. Ovchinnikova, “How you choose is as important as what you choose: Subjective quality of choice predicts well-being and academic performance,” *Curr. Psychol.*, vol. 41, no. 9, pp. 6439–6451, Sep. 2022, doi: 10.1007/s12144-020-01124-1.
[5] N. Zhang, Q. Li, S. X. Wu, J. Zhu, and J. Han, “A Novel Influence Analysis-Based University Major Similarity Study,” *Educ. Sci.*, vol. 14, no. 3, 2024, doi: 10.3390/educsci14030337.
[6] K. Koo, I. Baker, and J. Yoon, “The First Year of Acculturation: A Longitudinal Study on Acculturative Stress and Adjustment Among First-Year International College Students,” *J. Int. Stud.*, vol. 11, no. 2, pp. 278–298, Apr. 2021, doi: 10.32674/jis.v11i2.1726.
[7] S. A. Raza, W. Qazi, and S. Q. Yousufi, “The influence of psychological, motivational, and behavioral factors on university students’ achievements: the mediating effect of academic adjustment,” *J. Appl. Res. High. Educ.*, vol. 13, no. 3, pp. 849–870, Jan. 2021, doi: 10.1108/JARHE-03-2020-0065.
[8] F. Martínez-Roget, P. Freire Esparís, and E. Vázquez-Rozas, “University Student Satisfaction and Skill Acquisition: Evidence from the Undergraduate Dissertation,” *Educ. Sci.*, vol. 10, no. 2, 2020, doi: 10.3390/educsci10020029.
[9] A. Sofroniou, B. Premnath, and K. Poutos, “Capturing Student Satisfaction: A Case Study on the National Student Survey Results to Identify the Needs of Students in STEM Related Courses for a Better Learning Experience,” *Educ. Sci.*, vol. 10, no. 12, 2020, doi: 10.3390/educsci10120378.
[10] Y. Hsu and Y. Chi, “Academic major satisfaction and regret of students in different majors: Perspectives from Self-Determination Theory,” *Psychol. Sch.*, vol. 59, no. 11, pp. 2287–2299, Nov. 2022, doi: 10.1002/pits.22563.
[11] L. Arthur, “Evaluating student satisfaction - restricting lecturer professionalism: outcomes of using the UK national student survey questionnaire for internal student evaluation of teaching,” *Assess. Eval. High. Educ.*, vol. 45, no. 3, pp. 331–344, Apr. 2020, doi: 10.1080/02602938.2019.1640863.
[12] A. Kanwar and M. Sanjeeva, “Student satisfaction survey: a key for quality improvement in the higher education institution,” *J. Innov. Entrep.*, vol. 11, no. 1, p. 27, Mar. 2022, doi: 10.1186/s13731-022-00196-6.
[13] F. Rahmita, S. Purwaningsih, A. Andriawan, R. F. Febriani, Winda, and I. Izmuddin, “The Effect Of Education Level And Labor Absorption On Unemployment In Indonesia,” *Adpebi Sci. Ser.*, Jan. 2023, [Online]. Available: http://adpebipublishing.com/index.php/AICMEST/article/view/199
[14] S. Chaudhary and A. K. Dey, “Influence of student-perceived service quality on sustainability practices of university and student satisfaction,” *Qual. Assur. Educ.*, vol. 29, no. 1, pp. 29–40, Jan. 2021, doi: 10.1108/QAE-10-2019-0107.
[15] J. Li and Z. Ye, “Course Recommendations in Online Education Based on Collaborative Filtering Recommendation Algorithm,” *Complexity*, vol. 2020, no. 1, p. 6619249, Jan. 2020, doi: 10.1155/2020/6619249.
[16] C. P. Lee, Z. B. Ng, Y. E. Low, and K. M. Lim, “Expert System for University Program Recommendation,” in *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Sep. 2020, pp. 1–6. doi: 10.1109/IICAIET49801.2020.9257822.
[17] N. Rachburee, P. Sunantapot, D. Ounjit, P. Panklom, P. Porking, and W. Punlumjeak, “A Major Recommendation System in Educational Mining,” in *2021 1st International Conference On*

Cyber Management And Engineering (CyMaEn), May 2021, pp. 1–5. doi: 10.1109/CyMaEn50288.2021.9497279.
[18] T. Kim and J. Lim, "Developing an Intelligent Recommendation System for Non-Information and Communications Technology Major University Students," *Appl. Sci.*, vol. 13, no. 23, 2023, doi: 10.3390/app132312774.
[19] S. Patil, M. Bhosale, and R. Kamble, "Program Recommendation System for Students or Coder through View Histories and Feedback Systems," in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, Oct. 2020, pp. 185–187. doi: 10.1109/ICSIDEMPC49020.2020.9299652.
[20] A. M. N. Azhar, D. Pradeka, and D. A. R. Agustini, "Study Program Selection Recommendation System Using the Fuzzy Inference System Mamdani," *J. Sist. Cerdas*, vol. 7, no. 1, pp. 13–25, Apr. 2024, doi: 10.37396/jsc.v7i1.384.
[21] S. M. Sakti, A. Laksito, B. W. Sari, and D. Prabowo, "Music Recommendation System Using Content-based Filtering Method with Euclidean Distance Algorithm," *2022 6th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE*, vol. null, pp. 385–390, 2022, doi: 10.1109/ICITISEE57756.2022.10057753.
[22] D. H. Kusuma and Moh. N. Shodiq, "Sistem Rekomendasi Destinasi Pariwisata Menggunakan Metode Hibrid Case Based Reasoning dan Location Based Service Sebagai Pemandu Wisatawan di Banyuwangi," *INTENSIF J. Ilm. Penelit. Dan Penerapan Teknol. Sist. Inf.*, vol. 1, no. 1, pp. 28–34, Feb. 2017, doi: 10.29407/intensif.v1i1.540.
[23] A. F. Hidayat, D. D. J. Suwawi, and K. A. Laksitowening, "Learning Content Recommendations on Personalized Learning Environment Using Collaborative Filtering Method," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*, Jun. 2020, pp. 1–6. doi: 10.1109/ICoICT49345.2020.9166371.
[24] C. Bharathipriya, D. Aswini, X. F. Jency, R. Kirubakaran, and B. Swathi, "Product Recommender System Using Collaborative Filtering Technique," *2021 2nd Int. Conf. Emerg. Technol. INCET*, vol. null, pp. 1–7, 2021, doi: 10.1109/INCET51464.2021.9456160.
[25] G. Parthasarathy and S. Sathiya Devi, "Hybrid Recommendation System Based on Collaborative and Content-Based Filtering," *Cybern. Syst.*, vol. 54, no. 4, pp. 432–453, May 2023, doi: 10.1080/01969722.2022.2062544.
[26] B. W. Roberts and H. J. Yoon, "Personality Psychology," *Annual Review of Psychology*, vol. 73, no. Volume 73, 2022. Annual Reviews, pp. 489–516, 2022. doi: https://doi.org/10.1146/annurev-psych-020821-114927.
[27] D. Maestripieri and B. B. Boutwell, "Human nature and personality variation: Reconnecting evolutionary psychology with the science of individual differences," *Neurosci. Biobehav. Rev.*, vol. 143, p. 104946, Dec. 2022, doi: 10.1016/j.neubiorev.2022.104946.
[28] E. Durnell, K. Okabe-Miyamoto, R. T. Howell, and M. Zizi, "Online Privacy Breaches, Offline Consequences: Construction and Validation of the Concerns with the Protection of Informational Privacy Scale," *Int. J. Human–Computer Interact.*, vol. 36, no. 19, pp. 1834–1848, Nov. 2020, doi: 10.1080/10447318.2020.1794626.
[29] M. A. Khan *et al.*, "Medical Student Personality Traits and Clinical Grades in the Internal Medicine Clerkship," *Med. Sci. Educ.*, vol. 31, no. 2, pp. 637–645, Apr. 2021, doi: 10.1007/s40670-021-01239-5.
[30] N. Salankar, D. Koundal, and Y.-C. Hu, "Impact on the personality of engineering students based on project-based learning," *Comput. Appl. Eng. Educ.*, vol. 29, no. 6, pp. 1602–1616, Nov. 2021, doi: 10.1002/cae.22412.
[31] P. van Huizen, R. Mason, and B. Williams, "Exploring paramedicine student preferences using Holland's vocational theory: A cross-sectional study," *Nurs. Health Sci.*, vol. 23, no. 4, pp. 818–824, Dec. 2021, doi: 10.1111/nhs.12870.
[32] M. Makhtar, D. C. Neagu, and M. J. Ridley, "Binary Classification Models Comparison: On the Similarity of Datasets and Confusion Matrix for Predictive Toxicology Applications," in

Information Technology in Bio- and Medical Informatics, C. Böhm, S. Khuri, L. Lhotská, and N. Pisanti, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 108–122.

[33] G. Canbek, T. Taskaya Temizel, and S. Sagiroglu, "BenchMetrics: a systematic benchmarking method for binary classification performance metrics," *Neural Comput. Appl.*, vol. 33, no. 21, pp. 14623–14650, Nov. 2021, doi: 10.1007/s00521-021-06103-6.