

Implementation of Content-Based Filtering Method for Data Recommendation Of Computational Thinking Students-Based Informatics Subject

Zahratul Fitri¹, Safriana², Nurdin³


^{1,3}Department of Informatics, University of Malikussaleh, Indonesia

²Department of Education, University of Malikussaleh, Indonesia

ABSTRACT

Data recommendations are built by displaying the results of student subject recommendations based on students' computational thinking value. The process carried out is tokenization, stopword removal, stemming, and weighting. The extraction results were then compared using the cosine similarity approach. The greater the value of cosine similarity produced, the more similar the two data are, so that the material recommendations will be based on the smallest cosine similarity value between the extraction of student recommendation data. From the 535 data, several student data are included in 3 levels of material, namely recommendation 0 (low), recommendation 1 (medium), and recommendation 2 (high). Recommendation data was obtained from the results of students' computational thinking calculations by looking at decomposition value, pattern value, abstraction value, and algorithm value.

Keyword: Data Recommendation; Content-Based Filtering; Computational Thinking; Informatics Subject.

 This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Corresponding Author:

Zahratul Fitri,

Department of Informatics

Universitas Malikussaleh

Jl. Batam, Blang Pulo, Muara Satu – Lhokseumawe, Aceh, 24352, Indonesia.

Email : zahratulfitri@unimal.ac.id

Article history:

Received Dec 23, 2024

Revised Jan 15, 2025

Accepted Mar 10, 2025

1. INTRODUCTION

The development of the Industrial Revolution 4.0 currently has an impact on the rapid development of science and technology. All areas of life are directed towards the use of digital technology. One of them is in the field of education, entering the industrial revolution in the world of education is required to construct learning that involves technology. The education revolution 4.0 is a term used by education experts to describe how to implement technology into learning.

One of the improvement thinking skills that students must have in facing the 4.0 education era is Computational Thinking. According to (Tang et al., 2020). Computational thinking is defined as the thought process involved in formulating a problem, and its solution is represented in a form that can be done effectively in processing information. Thus, computational thinking must be embedded in the education system as a substantial learning goal to prepare students with future competencies. According to (Wiryasaputra et al., 2022), computational thinking is very important because, with the application of computational thinking, students are trained to think gradually, systematically, and creatively using computers and the internet. Computational thinking is a method to train how to think and solve problems using several techniques in the fields of computer science, science, and informatics. According to (Kuo & Hsu, 2020), the use of computational thinking in learning can increase students' desire to learn thereby increasing students' learning achievement. The use of computational thinking in learning can increase students' desire to learn thereby increasing students' learning achievement. This is shown from the results of research on the application of computational thinking in structural programming teaching materials that divide the learning group into two, namely, the study group that applies computational thinking in learning, and the group that does not use computational thinking. The result is that the group that uses computational thinking in learning can solve programming problems with significant goals compared to the group that does not.

The implementation of the Independent Curriculum is a curriculum that optimizes students to explore concepts and strengthen competencies and aims to restore education in Indonesia after the COVID-19 pandemic. Vocational High Schools (SMK) in Indonesia have implemented an independent curriculum since 2021, and this has caused several impacts that have become problems in educational

activities at vocational schools (Hidayat et al., 2022). Differentiated learning is a learning strategy that emphasizes the accommodation of individual differences of students in the learning process. In the context of the independent curriculum, differentiated learning is one of the effective strategies to accommodate individual student differences (Halimah Nurul et al., 2023). So far, in vocational schools, especially differential learning based on computational thinking, there has never been a system design that applies the concept of algorithms as a benchmark for providing material to students based on web applications so that students can be classified based on the ability to absorb material at each level of computational thinking.

Faisal, the Principal of SMK Negeri 3 Lhokseumawe City as well as the Chairman of MGMP ICT stated that the current student competency development activities have not focused on the development of Computational Thinking which focuses on student analysis. The activity only focuses on learning activities. This is evidenced by the fact in the field that the process of learning activities has not been maximized in the use of media and technology in learning. In addition, from the results of the study (Fitri & Utaminingsih, 2021) stated that especially in informatics materials, a number of weaknesses were still found, such as the importance of improving students' cognition to solve programming algorithm problems with a background in information technology. Therefore, it is necessary to make efforts to find out students' Computational Thinking abilities so that learning will be better in the future (Cacho, 2014). Furthermore, according to the chairman of MGMP ICT SMK, the assessment of student evaluation in the learning process is usually carried out with student learning evaluation activities which are routinely carried out at the end of each teaching semester with the aim of obtaining information related to the good and bad learning activities that have been implemented in the classroom (Safriana et al., 2023). Student learning evaluation should be used as a reflection material for teachers in developing and implementing the Computational Thinking component (Pulungtana & Dwikurnaningsih, 2020).

In the study [10], the results of the study showed that P5 activities based on several indicators as the application of differentiated learning in the independent curriculum starting from the planning, implementation, management, assessment, evaluation and follow-up plans for the next activities ran very well and structured. Differentiated learning has been well implemented through the implementation of P5 activities and the use of learning methods that adjust to students' interests. P5 activities carried out as a form of differentiated learning in the independent curriculum have a positive impact on the development of students.

In the research (Nafaridah et al., 2023), a research study was carried out on the selection of courses in accordance with the interests and understanding of courses that have certain subjectivity and needs. In many cases, students have little knowledge about the purpose of a course so they tend to guess or follow the flow of lectures. In the study, an analysis was carried out using the Content Based Filtering algorithm to recommend courses that were in accordance with the basic knowledge of students and the results of the experiment proved that the proposed algorithm was accurate and effective. However, the study did not use the analysis of the level of computational thinking where students' abilities could be measured effectively.

In addition, research conducted by (Huang et al., 2022) found an increase in the value of using English teaching resources by using a collaborative filtering algorithm. The intelligent system model built in the study provides personalized learning content recommendations according to the student's learning background so that the intelligent system built with the algorithm has high-quality recommendations to determine students' abilities in learning needs. However, in this study, there is no recommendation for students to specialize in science and mathematics with analytical computational thinking designed in a system that will be built in this study so that students, especially in vocational schools specializing in informatics, can master aspects of students' fields of interest well.

In this study, data analysis was carried out with a content-based filtering algorithm to analyze the material according to the level of students' computational thinking skills through learning evaluation. This algorithm has also proven that the data from the recommendations can improve the quality of resources and provide more appropriate recommendations to improve the quality of student learning (Han, 2023).

2. RESEARCH STAGES

The method used in this study is using two research methods, namely the data collection method and the system development method. In the data collection method, several stages of data collection are carried out such as observation, interviews, questionnaires, document studies, case studies, and

secondary data collection. Meanwhile, in the system development method, the researcher conducts the stage of system needs analysis (Software and Hardware) and system testing analysis using the waterfall method. The following explains the system design flowchart used in this study has been illustrated in Figure 1:

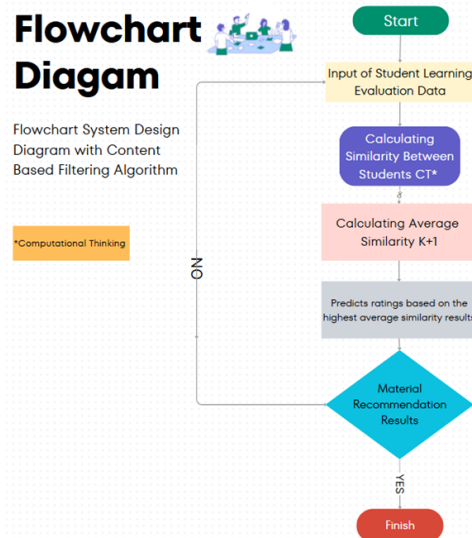


Figure 2. Flowchart System with Content-Based Filtering Algorithm

Based on the system design flowchart above, in the initial stage, data input is based on student learning evaluations and then processed with a content-based filtering algorithm where the input data will be calculated similarly based on the level of students' computational thinking. Then, the algorithm calculation will calculate the average similarity $K+1$, which is the similarity between students, and then predict the rating based on the highest average similarity results to get the results of recommendations for informatics materials that are by the level of computational thinking that students have.

2.1. Content Based Filtering Algorithm

Content-based filtering is one of the popular methods in the data recommendation and analysis system, this algorithm focuses on the characteristics or content of the item to be recommended or (Javed et al., 2021). To provide relevant recommendations, content-based filtering considers the interests and preferences of each user. This method uses the features and attributes of items to meet the user's desires so that the recommendations given will be better tailored to the user's needs and interests (Pratama et al., 2019). The following explanations have been illustrated in Figure 2.

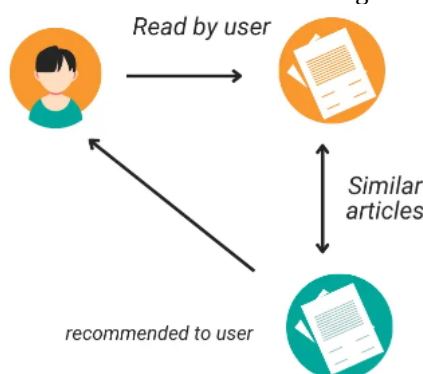


Figure 2. Content-Based Filtering

(Source: <https://dqqlab.id/content-based-filtering-dalam-algoritma-data-science>)

Based on the figure above, three types of content-based filtering are used in various contexts and applications, namely text-based content-based filtering, image-based content-based filtering, and video-based content-based filtering. It can be concluded that content-based filtering is used in the

recommendation system to recommend content to users based on their characteristics or content similarities.

In building a recommendation system using the content-based filtering algorithm, it can be done through the following stages:

1. Profile Item Formation
Profile Items are formed from the data that has been collected, namely based on the level of Computational Thinking (CT) of students and are hereinafter referred to as student profiles that are the content of the research (Anis Budiono & Sri Eniyati, 2023).
2. Preprocessing
Preprocessing is a stage of the process of changing the form of unstructured data into structured data so that it is suitable for processing to the next stage. Several stages of preprocessing are carried out, namely case folding, tokenization, elimination, filtering, and stemming (Mondi et al., 2019).
3. Cosine Similarity
Cosine similarity is a calculation that is often used to calculate similarities between items. In general, the similarity function accepts two object profiles in the form of real numbers (0 and 1) and returns the similarity value between the two objects in the form of real numbers. The cosine similarity function between item A and item B is shown as follows (Amelia et al., 2023).

$$sim(A, B) = \frac{n(A \cap B)}{\sqrt{n(A)n(B)}} \tag{1}$$

Information:

sim(A, B): similarity values of items A and B

n(A): many features of item A content

n(B): many features of item B content

n(A∩B): the number of content features contained in item A and also found in item B

If the result of the calculation obtains that the two objects have a similarity value of 1, then the two objects are said to be identical and vice versa. The greater the result of the similarity function, the more similar the two objects are considered, and vice versa.

4. Top-N Recommendation

The user selects from the items generated from the similarity calculation. Only N items with the highest similarity value are selected, and it is considered that the user is likely to choose similar items with a higher similarity value.

5. Confusion Matrix

Confusion matrix is one of the methods that can be used to measure the performance of a system. The confusion matrix contains information that compares the classification results carried out by the system with the classification results that should be, precise and accurately formulated with equations (Adawiyah & Nuddin, 2021).

Table 1. The performance of prediction value

		True Value	
		TRUE	FALSE
Prediction Value	TRUE	True Positive (TP)	False Positive (FP)
	FALSE	False Negative (FN)	True Negative (TN)

In this study, a Content-Based Filtering algorithm is used which focuses on student objects based on the ability of students who have a level of computational thinking to then be used as a rating matrix in the calculation of methods in the system.

2.2. Computational Thinking in Informatics Materials

Computational Thinking learning carried out with informatics learning can be done by preparing the necessary steps according to the instructions on each algorithm. There are many variations in patterns in case resolution. One example is learning to make a simple mathematics-based algebra program by arranging steps according to instructions (Adawiyah & Nuddin, 2021). The stages of completing the instruction in the application of computational thinking are as follows:

1. Model functions in algebra through programs (compare them with functions in programs).
2. Write down an algorithm (or the exact sequence of steps) about how to do matrix multiplication or how to solve a quadratic equation.
3. Use decomposition to solve word problems.
4. Generalization (as an algebraic representation) by identifying patterns.

Some things will be solved faster by humans and some will be solved faster with computers, we as humans are the ones who decide. In practice, there is a set of abilities that are part of computational thinking. Google for Education discusses how a computational thinking approach should be. Google for Education has four basic steps consisting of decomposition, pattern recognition, abstraction, and algorithm design (Kidd, Lonnie R, & Morris, Jr., 2017). Here is the definition of the four basic steps (BBC Bitesize, n.d.)

3. RESULTS AND DISCUSSION

The analysis of recommendation data was carried out using the results of the evaluation of the learning of Lhokseumawe City Vocational School students who had been grouped based on the level of Computational Thinking, students and calculated the rule base using the Tsukamoto Fuzzy Inference method. In data grouping, data is generated in the form of weighting. The results of data processing are carried out data processing to produce similarity values between informatics materials. Figure 5. Content-based filtering algorithm scheme.

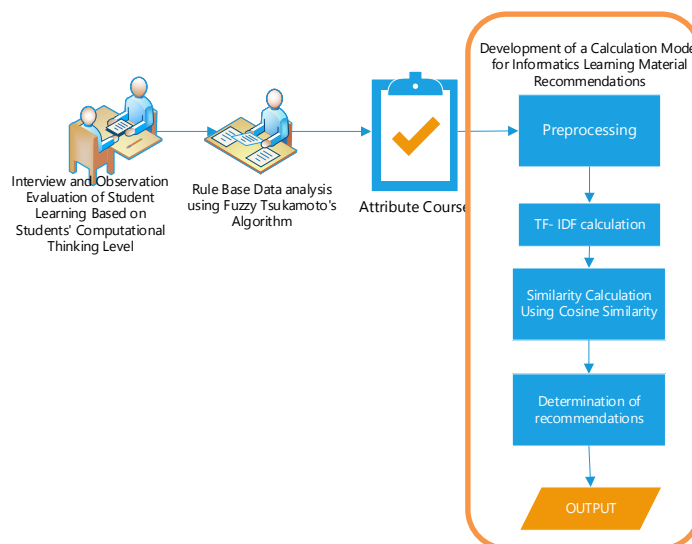


Figure 5. Content-Base Filtering Algorithm Scheme.

Based on the image above, it is explained as follows:

1. Interview and Observation Evaluation of Student Learning Based on Students' Computational Thinking Level: In this process, data observations are carried out in the form of students' computational thinking scores taken from teachers who have been appointed as research assistants. The teacher gave an instrument test in 5 classes, the results of which were measured by 4 categories of students' computational thinking.
2. Rule Base Data analysis using Fuzzy Tsukamoto's Algorithm: From the results of students' computational thinking data, it was analyzed again with rule base data using Fuzzy Tsukamoto's algorithm. The data rule base is grouped into 4 categories.
3. Attribute Course: Once the attributes are obtained, the attributes will be processed by the content-based filtering algorithm.
4. Development of a Calculation Model for Informatics Learning Material Recommendations: The model calculation process is carried out to provide material recommendations that are in accordance with the students' computational thinking criteria.
5. Output: recommendation results and recommendation accuracy test.

The following are 535 data on the computational thinking of students with each of them having a decomposition value, pattern value, abstraction value, and algorithm value. This value is the range result generated from the calculation of the Fuzzy Tsukamoto method.

Table 2. The result of computational thinking student

NO.	Name	Student ID	decomposition value	pattern value	abstraction value	algorithm value
1.	Aan syukralillah	424111	High	Medium	High	High
2.	Abdil mubaraq	424112	Medium	Medium	Medium	Medium
3.	Abdul hasyib	424113	Low	Low	Low	Low
4.	Abidal raihannur iman	424114	High	Medium	Medium	Medium
5.	Abrar azizi	424115	High	High	High	Medium
6.	Adinda lathifa	424116	High	Medium	High	High
7.	Afdhalul zikri	424117	Medium	Medium	Medium	Medium
8.	Agus ramadhan	424118	Low	Low	Low	Low
9.	Ahmad nazuar	424119	High	Medium	Medium	Medium
10.	Ahmad zaki	424120	High	High	High	Medium
11.	Ahya nazira	424121	High	Medium	High	High
...
535	Zulfandi	424645	High	Medium	High	High

In the results of the calculation of Informatics material recommendations using the Content-Base Filtering method, there are 3 (three) levels of material recommendations, namely low, which is 0, medium, which is 1, and high, which is 2. and there were 11 total materials that were tested for similarity. based on the following table.

Table 3. The Result of Recommendations Material Student Grouping

NO.	Student Group	ID	Informatics Material	Recommendations
1.	ABDUL HASYIB	111	Breaks down big problems into smaller, manageable parts.	0
2.	AAN SYUKRALILLAH ABDIL MUBARAQ ABIDAL RAIHANNUR IMAN ABRAR AZIZI ADINDA LATHIFA AFDHALUL ZIKRI AHMAD NAZUAR AHMAD ZAKI DARA SINTA	211	Create a clear and structured set of steps to solve the problem.	1
3.	AAN SYUKRALILLAH ABDIL MUBARAQ ABIDAL RAIHANNUR IMAN ABRAR AZIZI ADINDA LATHIFA AFDHALUL ZIKRI AHMAD NAZUAR AHMAD ZAKI DARA SINTA	212	Translating algorithms into programming languages.	1
4.	AAN SYUKRALILLAH ABDIL MUBARAQ ABIDAL RAIHANNUR IMAN ABRAR AZIZI ADINDA LATHIFA AFDHALUL ZIKRI AHMAD NAZUAR	213	Create a simple model of a system to predict behavior or outcomes.	1

	AHMAD ZAKI DARA SINTA			
5.	AAN SYUKRALILLAH ABIDAL RAIHANNUR IMAN ABRAR AZIZI ADINDA LATHIFA AHMAD NAZUAR AHMAD ZAKI DARA SINTA	311	Analyze data to find relationships, patterns, or trends.	2
6.	AAN SYUKRALILLAH ABIDAL RAIHANNUR IMAN ABRAR AZIZI ADINDA LATHIFA AHMAD NAZUAR AHMAD ZAKI DARA SINTA	312	Creating algorithms that are not only correct, but also efficient in the use of time and memory.	2
7.	AAN SYUKRALILLAH ABIDAL RAIHANNUR IMAN ABRAR AZIZI ADINDA LATHIFA AHMAD NAZUAR AHMAD ZAKI DARA SINTA	313	Use object and class concepts to build more complex and structured programs.	2
8.	AAN SYUKRALILLAH ABIDAL RAIHANNUR IMAN ABRAR AZIZI ADINDA LATHIFA AHMAD NAZUAR AHMAD ZAKI DARA SINTA	314	Create computer models that can learn from data without being explicitly programmed.	2

4. CONCLUSION

Data analysis is built by displaying the results of student material recommendations based on students' computational thinking. The process carried out is tokenization, stopword removal, stemming, and weighting. The extraction results were then compared using the cosine similarity approach. The greater the value of cosine similarity produced, the more similar the two data are, so that the material recommendations will be based on the smallest cosine similarity value between the extraction of student recommendation data. From the 535 data, there are several student data that are included in 3 levels of material, namely recommendation 0 (low), recommendation 1 (medium) and recommendation 2 (high). Recommendation data was obtained from the results of students' computational thinking calculations by looking at decomposition value, pattern value, abstraction value and algorithm value

REFERENCES

- Adawiyah, A., & Nuddin, S. R. (2021). Penerapan Algoritma Content Based Filtering dan Frequent Pattern Growth pada Sistem Rekomendasi Program Mahasiswa Wirausaha di Universitas Negeri Surabaya. *Journal of Informatics and Computer Science (JINACS)*, 3(02), 123–130. <https://doi.org/10.26740/jinacs.v3n02.p123-130>
- Amelia, T., . P., & Pambudi, A. (2023). Rekomendasi Jurusan Kuliah Berdasarkan Minat dan Kemampuan Menggunakan Metode Content Based Filtering. *Technologia: Jurnal Ilmiah*, 14(3), 245. <https://doi.org/10.31602/tji.v14i3.11336>
- Anis Budiono, & Sri Eniyati. (2023). Sistem Rekomendasi Dosen Pembimbing Tugas Akhir Menggunakan Content Based Filtering. *Jurnal Elektronika Dan Komputer*, 16(1), 64–71.
- Cacho, R. M. (2014). TPCK assessment of pre-service teachers toward enhancing teacher educators' modeling. *Asian Journal of Education and E-Learning*, 2(5), 349–356.
- Fitri, Z., & Utaminingsih, E. (2021). Penerapan Metode Computational Thinking Pada Kurikulum Aceh Untuk Mencapai Kognitif. *Jurnal MathEducation Nusantara*, 4(1), 60–73.

- <https://jurnal.pascaumnaw.ac.id/index.php/JMN>
- Halimah Nurul, Hadiyanto, & Rusdinal. (2023). Analisis Pembelajaran Berdiferensiasi Sebagai Bentuk Implementasi Kebijakan Kurikulum Merdeka. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, 8(1), 5019–5033.
- Han, J. (2023). A Collaborative Filtering Recommendation Algorithm Model Based on User Feature Transfer. *More than One Article*, 18(November 2023), 164–187. <https://doi.org/https://doi.org/10.3991/ijet.v18i22.43987>
- Hidayat, T., Purnomo, S., Hadi, S., & Setuju. (2022). ANALISIS MODEL PEMBELAJARAN UNTUK MENINGKATKAN HASIL BELAJAR MERDEKA BELAJAR SISWA SMK. 003(xx), 1–16.
- Huang, C., Chen, R., & Lin, L. (2022). English Teaching Resource Recommendation Algorithm Based on Collaborative Filtering Technology. 4, 34–40. <https://doi.org/10.23977/aduhe.2022.040606>
- Javed, U., Shaukat, K., Hameed, I. A., Iqbal, F., Alam, T. M., & Luo, S. (2021). A Review of Content-Based and Context-Based Recommendation Systems. *International Journal of Emerging Technologies in Learning*, 16(3), 274–306. <https://doi.org/10.3991/ijet.v16i03.18851>
- Kuo, W. C., & Hsu, T. C. (2020). Learning Computational Thinking Without a Computer: How Computational Participation Happens in a Computational Thinking Board Game. *Asia-Pacific Education Researcher*, 29(1), 67–83. <https://doi.org/10.1007/S40299-019-00479-9/METRICS>
- Mondi, R. H., Wijayanto, A., & Winarno. (2019). Recommendation System with Content-based Filtering Method for Culinary Tourism in Mangan Application. *ITSMART: Jurnal Ilmiah Teknologi Dan Informasi*, 8(2), 65–72.
- Nafaridah, T., Ahmad, A., Maulidia, L., Ratumbusang, M., & Kesumasari, E. M. (2023). Analisis Kegiatan P5 sebagai Penerapan Pembelajaran Berdiferensiasi pada Kurikulum Merdeka Era Digital di SMA Negeri 2 Banjarmasin. *Seminar Nasional PROSPEK II, Prospek II*, 84–97.
- Pratama, D. R., Nugraha, G. S., & Dwiyanaputra, R. (2019). SISTEM REKOMENDASI TOPIK PENELITIAN MENGGUNAKAN METODE CONTENT-BASED FILTERING (STUDI KASUS PROGRAM STUDI TEKNIK INFORMATIKA) RESEARCH TOPIC RECOMMENDATION SYSTEM USING CONTENT-BASED FILTERING METHOD (CASE STUDY OF INFORMATICS ENGINEERING STUDY PROGRAM. x(36), 1–9.
- Pulungtana, J. R., & Dwikurnaningsih, Y. (2020). Evaluasi Kinerja Mengajar Guru IPS Dalam Mengimplementasikan TPACK. *Jurnal Ilmu Sosial Dan Humaniora*, 9(1), 146–155. <https://doi.org/10.23887/IISH-UNDIKSHA.V9I1.24672>
- Safriana, Fitri, Z., & Ginting, F. W. (2023). Implementation of Fuzzy Tsukamoto Method in Analyze Science Teacher ' s Technological Pedagogical Content Knowledge. 11(2), 276–287.
- Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers and Education*, 148, 103798. <https://doi.org/10.1016/j.compedu.2019.103798>
- Wiriasaputra, R., Salomo, A., Sevani, N., & Seruni. (2022). Peningkatan Pola Berpikir Komputasi pada Siswa/i SMAK MATER DEI Melalui Bahasa Pemrograman Java dan Python. *Servirisma*, 2(2), 127–145. <https://doi.org/10.21460/servirisma.2022.22.28>