International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I6P218
Volume 70 Issue 6, 144-154, June 2022

Original Article

Selection of Student Extracurricular using Hybrid Multi-Criteria Recommendation System and Particle Swarm Optimization

Arif Budiman Harahap[1], Antoni Wibowo[2]

[1,2] *Computer Science Department, BINUS Graduate Program, Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480.*

[1]arif.harahap@binus.ac.id

Received: 31 March 2022 Revised: 17 May 2022 Accepted: 05 June 2022 Published: 27 June 2022

Abstract - *Along with the times, the latest technology can help users find needs according to their conditions and interests by using a recommendation system. The development of user needs also makes the techniques used in the recommendation system more varied. One of the benefits can be used to determine extracurricular according to the talents and interests of students. Extracurricular is an additional activity at school that can be a means to hone students' talents and interests. Therefore, students need to be able to identify the appropriate talents and interests from an early age so that the talents possessed by students can develop properly. Several studies have been done previously, one of which is to combine the hybrid method and MCRS with the GA method. However, other studies have been carried out and found that the PSO method can produce better outputs than GA. Therefore, a hybrid method between the MCRS and PSO methods is combined with student extracurricular recommendations. As a result, the proposed method produces a longer execution time of 19.108 seconds but produces a better error percentage of 2.45% compared to the hybrid MCRS and GA methods.*

Keywords - *Genetic Algorithm, Hybrid, Multi-Criteria, Particle Swarm Optimization, Recommendation System.*

1. Introduction

Extra curriculars are additional activities in schools that can be a means to hone the talents and interests of students whose goals have been listed in Permendiknas No. 39 of 2008, namely (1) developing the potential of learners optimally and integrated, which includes talents, interests, and creativity; (2) solidify the personality of learners to realize school resilience as an educational environment to avoid negative efforts and influences and contrary to educational goals; (3) actualize the potential of learners in achieving superior achievements according to talents and interests; (4) prepare students to become citizens of a society that is noble, democratic, and respects human rights to realize civil society [1]. The school provides various extracurricular activities to support this goal, including science, art, language, organization, sports, and others. Therefore, students need to choose extracurriculars that suit their interests and talents so that students potential is growing and can help them in the future.

Along with the times, the latest technology can help students determine extracurriculars that suit their talents and interests by using the recommendation system, which has been widely used in everyday life because it can help users find needs according to their conditions and interests [2]. A recommendation system is an intelligent application that reduces the risk of information overload problems by filtering information according to the user's needs [3]. The development of user needs also makes the techniques used in the recommendation system also become more varied.

In general, there are 3 methods commonly used in recommendation systems, namely, Content-Based (CB), Collaborative Filtering (CF), and Hybrid [2]. The CB method provides recommendations based on data owned by the user, CF provides recommendations based on the user's relationship with other users, and the Hybrid method provides recommendations based on the user's profile and relationship with other users. The Hybrid method is the best recommendation system because it can be used for wider and more complex conditions [4].

Research on the recommendation system has previously been conducted by [5] to assist students in choosing extracurriculars; research with the Hybrid method uses Genetic Algorithm (GA) and Multi-Criteria Recommender System (MCRS) and uses datasets in the form of student data at the University of Cordoba. The study resulted in an improvement in Root Mean Square Error (RMSE) to 0.971. MCRS techniques are used to improve the accuracy and performance of the recommendation system by applying more criteria choices used as the basis for calculating recommendations and weighting with GA so that the recommendations given are more following the criteria owned by the user. GA is a computational search technique for finding approaches to optimization and search problems. Techniques used in GA are based on evolutionary biology techniques such as inheritance, mutation, selection, and crossover [6].

Other research on recommendation systems using Particle Swarm Optimization (PSO) has been conducted by [7]. The study used CF and PSO methods and a dataset in the form of data from MovieLens in the form of 100,000 ratings, 943 users, and 1,682 films. The study improved the Mean Absolute Error (MAE) to 0.7547 with a standard deviation of 5.067e-03. Fuzzy C-Means (FCM) algorithms are used to find interconnected users and improvised using PSO and K-Means to produce better precision and accuracy. It was mentioned that the study chose to use PSO compared to GA because GA has limitations on undirected variable mutations that lead to slower calculation results than PSO.

Other research on the comparison between GA and PSO has been conducted by [6]. The study results stated that the difference between GA and PSO lies in the purpose of data exploration and performance. GA is used to explore broader data but consequently results in slower performance than PSOs that produce faster performance but with less data exploration. It uses data mutations based on the previous best position (pbest) and global best position (gbest).

Based on previous research, it can be known that hybrid methods can fix weaknesses in CB and CF methods. MCRS can be used to make recommendations better by applying more recommendation calculation criteria to better match users' data. Optimization methods of both GA and PSO have their advantages and disadvantages. GA can be used to explore larger data but with slower performance, and PSOs have better performance but with less data exploration. Therefore, the research to be conducted is to improve the accuracy of the recommendation system by applying hybrid recommendation methods and using PSO as an optimization method, to see how well PSO performs and its accuracy if used on larger exploration data with MCRS, and to use student extracurricular data as a research dataset.

2. Related Work

There are 3 types of methods are most commonly used in recommendation systems. One of them is Content-Based (CB). CB makes recommendations based on content data that exists in the user. Compared to other methods, this method is the best method to overcome user privacy problems because it does not use user history or habits. The Association Rule Mining algorithm is the most commonly used CB method because it is useful for connecting relationships between users and items to be recommended [8]. Such as research by [9] on book recommendation systems in digital libraries using CB methods with association rules algorithms and datasets in the form of 65,521 transactions from January 2012 - February 2014. The study compares the proposed model in the form of User, Category, Loan, and Title itemsets to the traditional itemset model of User, Loan, and Title. The results show that the model proposed in the study gets a precision value of 92% compared to the traditional model of 91.22%.

Other research on CB methods was conducted by [10] to help students choose extracurricular activities. The study used the Naive Bayes algorithm to find the relationship between student attributes and extracurricular activities. This study used 158 data, 78 training data, and 80 testing data. The results showed an improvement in recommendation results of 0.664. Other CB methods research has been conducted by [11] on movie recommendations using datasets from MovieLens using Neuro-Fuzzy and Deep Neural Network (DNN). The method produced effectiveness of 98.8% with more than 1000 users.

Another study using the CB method was conducted by [12] to compare Machine Learning methods using 2000 datasets on airline companies. The methods compared were Logistic Regression, SGD Classification, and Random Forest Classifier methods. The results showed that the SGD Classification method produced the best accuracy of 88%, then Logistic Regression with an accuracy of 86%, followed by the Random Forest Classifier method with an accuracy of 80.25%. Another study using the CB method was conducted by [13] using Feature Selection, Vector Generation, and Softmax Regression on Computer Science publication data and obtained from 28 journals and 38 conferences. The results obtained an accuracy of 61.37%.

The next method in the recommendation system is the Collaborative Filtering (CF) method. This method recommends items to users based on the relationship between users and other users. This method is usually used in e-commerce systems to show item recommendations to users with the same relationship or interest. The advantage of using this method is that it produces better recommendations than the CB method because it uses the relationship between users to produce recommendations. The disadvantage is that this method is very dependent on reviews and ratings, so the recommendation results are not good for users who have just registered (cold start problem).

Research on recommendation systems using the CF method has been carried out by [14] with Genetic Algorithm (GA) and Yahoo! Movie as a dataset. The proposed method is to apply the multi-criteria rating to find relationships between users and calculate weights with GA. The method resulted in a Mean Absolute Error (MAE) improvement of 33.3% and a Root Mean Square Error (RMSE) improvement of 31.3%.

Other research with the CF method has been conducted by [15] to recommend doctors to patients by using Matrix Factorization (MF) to find the interaction relationship between doctors and patients. The dataset used was 1 million consultation record data containing 382,817 patients and 314 family doctors in 16 hospitals between 2012 - 2017. The study added a trust attribute which indicates that the patient's trust in the doctor has increased and will subsequently choose the doctor again. The results showed that the accuracy of patient recommendations with suitable doctors increased to 3% compared to the CF method that did not use the trust attribute.

Other research using the CF method was conducted by [16]. Using the Singular Value Decomposition (SVD) algorithm. The research was conducted using book sales data from the e-commerce website https://book.douban.com. The results obtained an MAE improvement of 0.04 compared to the Property-based method and an improvement of 0.06 compared to the Item-based method. Other research using the CF method was conducted by [17] using the k-NN and Cosine Similarity algorithms. The research was conducted using a dataset from MovieLens. The result is an MAE improvement of 0.017 compared to the Model-based method and an improvement of 0.021 compared to the CB method.

Other research using the CF method has been conducted by [18] to calculate how much the CF method improves using the AHP algorithm with the addition of the time attribute. The research was conducted using a dataset from the Aliyun Tianchi e-commerce company with 6000 purchase data. The results obtained by adding the time attribute proved to produce better accuracy with an RMSE improvement result of 0.0178 compared to not using the time attribute. Another study using the CF method was conducted by [19] using 10,000 MovieLens datasets using the Ensamble k-NN algorithm and then compared with Item-based k-NN and User-based k-NN. The results show that using the Ensamble k-NN algorithm obtained an increase in accuracy of 0.01 compared to User-based k-NN and an increase in accuracy of 0.06 compared to Item-based k-NN.

The next method is the Hybrid method, a combination of content-based and collaborative filtering methods. Hybrid methods are methods that provide item recommendations to new users or recommend users a new item based on the history of decisions that have been made by users [14]. Hybrid methods take advantage of CB methods and combine them with CF methods to produce methods that are better than both [4].

Research on recommendation systems using the Hybrid method has been conducted by [25] using datasets from MovieLens. The Hybrid method proposed is by finding the number of similarities between the two films, which are then processed using the weight calculation method. The weight obtained is then combined with the movie rating matrix from the user that has been processed using Collaborative Filtering and the Pearson Correlation algorithm. The results are then evaluated by measuring the Mean Absolute Error (MAE) and compared with the content-based method using Singular Value Decomposition and the pure Collaborative Filtering method. The results showed an improvement in MAE, and sparsity problems can be improved by 1% - 2% depending on the data used.

Other research using hybrid methods has been conducted by [5] to help students in choosing extracurricular activities, research with the Hybrid method using Genetic Algorithm (GA) and Multi-Criteria Recommender System (MCRS) and using datasets in the form of student data at the University of Cordoba. This research resulted in an improvement in RMSE to 0.971. Other research using hybrid methods has been conducted by [21] on 250 patients data with a history of heart disease. The research was conducted by combining Deep Learning methods, namely Multiple Kernel Learning (MKL) and Adaptive Neuro-Fuzzy Inference System (ANFIS). The results produced an MSE of 0.01 on the KEGG Reaction dataset.

Other research using hybrid methods has been conducted by [22] using datasets from JD.com. The research was carried out by analyzing food comments in the comments column using the Semantic Orientation Pointwise Mutual Information (SO-PMI) algorithm. Then the results were compared with the CB and CF methods. The result is that the MAE on the hybrid is better by 0.000008 compared to the CB method and better by 0.023311 compared to the CF method. Other research using hybrid methods has been conducted by [23]. The research was conducted to compare the effect of sentiment analysis variables on 6 million review data on the Douban Movie website. Sentiment analysis calculation applies the Term Frequency - inverse Document Frequency (TF-iDF) method. The result obtained a better F1-score value when using the hybrid method (CB + CF) and sentiment analysis of 0.189 compared to not applying sentiment analysis.

Other research on hybrid methods has been conducted by [24], which combines data query methods based on keywords and N-grams methods on 300-course data at Crescent University using the Ontology model. The study obtained accuracy results using hybrid methods better by 5.4% compared to the N-grams method alone and better by 20.14% compared to the keyword query method alone. Based on the literature review, the results of the literature review summary can be seen in Table 1.

Table 1. Literature review summary

No	Publication	Method	Dataset	Result
1	[5]	Hybrid, with MCRS and GA	University of Cordoba student data	RMSE improvement to 0.971.
2	[9]	Content-based user profile and Association Rule	65.521 transaction	Improved precision to 92%.
3	[10]	Content-based with Naive Bayes	158 dataset	The recommendation result is 0.664.
4	[11]	Content-based with Neuro-Fuzzy and Deep Neural Network	MovieLens	Effectiveness result of 98.8% on more than 1000 users.
5	[12]	Content-Based with Logistic Regression, SGD Classification, and Random Forest Classifier	20.000 dataset review from airlinequality.com	Classification Accuracy of SGD is 88%, Logistic Regression is 86%, and Random Forest Classifier is 80.25%.
6	[13]	Content-Based with Feature Selection, Vector Generation, and Softmax Regression	Web Crawling 28 Jurnal publications and 38 conferences	Accuracy of 61.37%.
7	[14]	Collaborative Filtering with a multi-criteria rating and Genetic Algorithm	Yahoo! Movie	MAE accuracy improvement of 33.3% and RMSE of 31.3%.
8	[15]	Collaborative Filtering with trust attribute	1 million patient and doctor consultation records	Accuracy improvement of 3% compared to not using the trust attribute.
9	[16]	Collaborative Filtering and Singular Value Decomposition (SVD)	Book data sourced from https://book.douban.com	MAE improvement of 0.04 compared to Property-based and 0.06 compared to Item-based.
10	[17]	Collaborative Filtering using k-NN and Cosine Similarity	MovieLens	MAE improvement of 0.017 over Model-based and 0.021 over CB method.
11	[18]	Collaborative Filtering with a time attribute	6000 transaction data e-commerce Aliyun Tianchi	RMSE improvement of 0.0178 compared to not using the time attribute.
12	[19]	Collaborative Filtering and Ensamble k-NN	10.000 dataset MovieLens	Improvement of RMSE by 0.01 compared to User-based k-NN and by 0.06 compared to Item-based k-NN.
13	[25]	Hybrid, with Naive Bayes and Pearson Correlation	MovieLens	MAE improvement and 1-2% sparsity improvement.
14	[21]	Hybrid with Multiple Kernel Learning (MKL) and Adaptive Neuro-Fuzzy Inference System (ANFIS)	250 patient data with heart disease	MSE of 0.01 on the KEGG Reaction dataset.

15	[22]	Hybrid with algoritma Semantic Orientation Pointwise Mutual Information (SO-PMI)	food comments data on JD.com	MAE of 0.131704. It is 0.000008 better than the CB method and 0.023311 better than the CF method.
16	[23]	Hybrid and Sentiment Analysis with TF-iDF	6 Million review data on Douban Movie	F1-score is better by 0.189 compared to not applying sentiment analysis.
17	[24]	Hybrid combine query Keywords and N-grams on Ontology model	300 data matakuliah pada Crescent University	Accuracy of 95.25% using the hybrid method.

Based on the literature review, there are 3 methods commonly used in recommendation systems: CB, CF, and hybrid. Of the three methods, the Hybrid method is the best because the Hybrid method combines the advantages of the CB method and the CF method, or it can be said that the Hybrid method takes and applies both important aspects of the recommendation system, namely content description, relationships between users, and user assessment of items.

Based on research that has been done before, the methods and improvements produced in the recommendation system, especially the extracurricular recommendation system, are quite good using the Hybrid method, namely GA and MCRS. The research on student extracurricular recommendation systems that will be carried out aims to improve the weaknesses of the GA algorithm by applying PSO because GA has limitations on undirected variable mutations that cause slower calculation results compared to PSO. So that the research conducted can improve the accuracy and performance of the student extracurricular recommendation system.

3. Proposed Method

Based on the result of the study in the literature review, the Hybrid method combined with MCRS and PSO to improve the accuracy of the student extracurricular recommendation system. The initial stage is the collection of student data and extracurricular data. Student data has some information that can be used as a basis for assessment to predict the talents and interests of these students. These data include previous extracurricular data, student majors data, student gender data, and student subject value data. In extracurricular data, several data can be used to predict extracurriculars based on the ratings possessed by these extracurriculars, including extracurricular rating data and extracurricular category data.

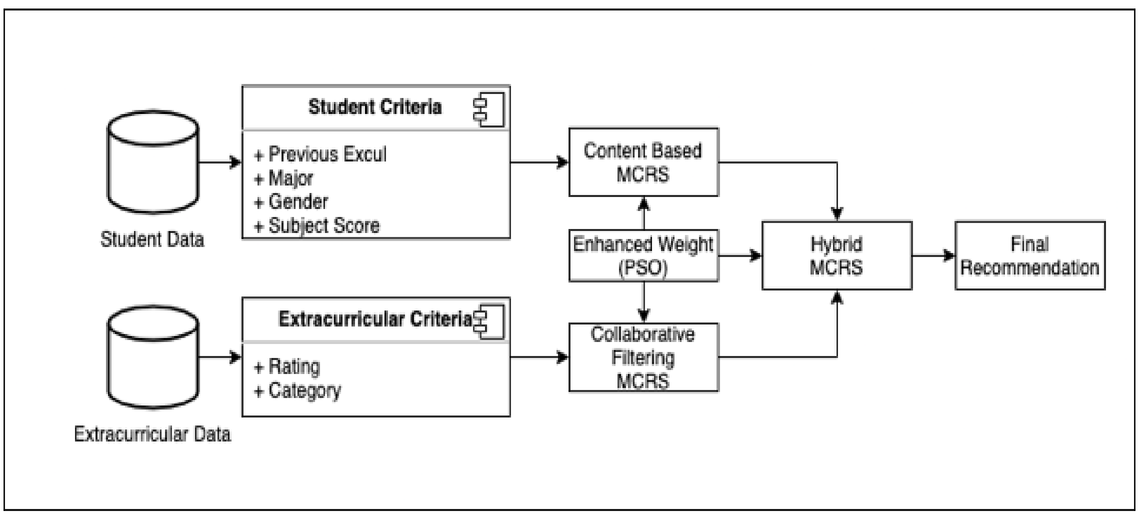

Fig. 1 Proposed method

After collecting data, as shown in Figure 1, the prediction calculations are carried out for each data. CB recommendation system method is used to predict extracurricular based on student data, while the CF method is used to predict extracurricular based on extracurricular data. The two recommendation systems are then combined into a Hybrid recommendation method. The three methods, CB, CF, and hybrid, are then optimized using the PSO algorithm to increase the final value of the recommendations. The final result of calculating the hybrid method with the PSO algorithm is then used as the basis for extracurricular recommendations to students.

3.1. Assessment Criteria Methods

Based on the predetermined criteria, a different calculation method is used for each criterion because the data type for each criterion is different. The assessment criteria used are then categorized into 2 types of recommendations, namely CB and CF, as shown in Table 2.

Table 2. Method of criteria calculation

Type	Criteria	Method
CB	Previous Excul	Dot Product
	Major	Dot Product
	Gender	Dot Product
	Score	Jaccard Similarity
CF	Rating	KNN & Cosine Similarity
	Category	KNN & Cosine Similarity

The data included in the CB recommendation type are derived from student profile data. Meanwhile, rating and category data is a type of CF recommendation because it comes from extracurricular data used by students.

3.2. PSO Configuration

Determination of the PSO weight is carried out with the initial initialization values of the PSO parameters w = 0.5, C1 = 1, and C2 = 1.5. The results of the calculation of the recommendations are then multiplied by the weights that have been obtained for each criterion to obtain the final recommendations for each criterion. The search for fitness values can be done based on Equation (1).

$$f_{(w,x,y,z)} = (k_w \times w) + (k_x \times x) + (k_y \times y) + (k_z \times z) \quad (1)$$

where,
f = fitness value,
v = velocity value of each dimension,
(w,x,y,z) = position value of each dimension

The dimensions determined as in Equation (1) use 4 dimensions for CB and 2 for CF according to the criteria described in Table 2. For more details regarding the steps for calculating the PSO algorithm as in Algorithm 1:

Algorithm 1: PSO Algorithm

Result: best position of weight for each criterion

a. Determine the initial position of the particle with a random position;
   Adjust the position so that it always equals the total position;
   Determine the initial velocity of the particle with a random value;
   Calculate the fitness value based on a predetermined random position as in (1);
   Check the fitness value. If it is better than the general fitness value, then update the general fitness value;

b. Determine the initial value of the weights, w = 0.5, C1 = 1, and C2 = 1.5;

c. **while** iteration approximately equal to 70 **do**
   Repeat for each particle to find a new velocity;
   Determine below limit = 0 and upper limit = 1;
   if the new position is outside the upper and below limit, **then**
      update position to old position;
   end
   Adjust the position so that it always equals the total position;
   Determine the fitness value for the new position;
   Update the fitness value if the latest fitness value is better than the previous fitness value;
   Update the general fitness value if the latest fitness value is better than the general fitness value;
   end

3.3. Data Collection

The data used in this study is data sourced from schools managed by the Sokrates System, which currently consists of 11 school foundations. Some of them are Binus School Bekasi, Sekolah Harapan Bangsa (SHB), Puhua, Strada, YPII, SMN, Santa Angela, Kanisius, Pangudi Luhur, Regina Pacis, and Santa Maria.

The data collected is 50 data, with 30 data designated as training data and 20 data intended as testing data. The data consisted of 20 students in grade 7, 15 in grade 8, and 15 in grade 9. It was done to provide a variety of data for the research based on the profiles of grade 7, grade 8, and grade 9 students.

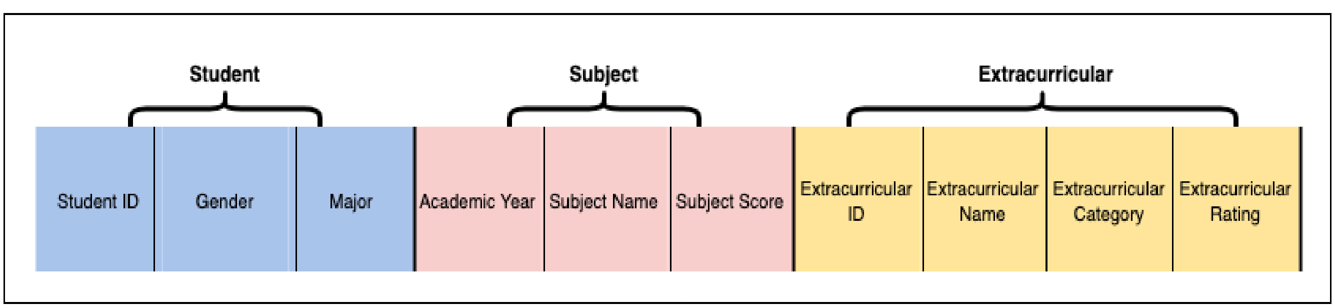

Fig. 2 Data collection

This study combines student, subject, and extracurricular data to form a data matrix. In student data, the data used are student ID, gender, and major. Data on gender and majors were used to assess students' extracurricular recommendations based on these criteria. Furthermore, the subject data, academic year data, subject names, and subject scores are used to assess extracurricular recommendations based on the value of subjects owned by students. In extracurricular data, extracurricular ID, extracurricular name, extracurricular category, and extracurricular rating are used to assess extracurricular recommendations based on rating criteria and extracurricular categories.

3.4. Data Processing

The data processing used in this research is the data cleansing method. The cleansing method is used to validate the data to be used.

Table 3. Data processing

Data Column	Before	After
Gender	Male	M
	Female	F
Score	88,625	89
	92,33	92
Grade	Natural Science	IPA
	Social Science	IPS

The validation is intended to clean up empty data, data with the duplicate values, and data to make uniform data types used. The string, boolean, char, and text data types are standardized into strings, and the integer, float, and double data types are standardized into integer data types.

4. Results and Discussion

4.1. Calculation of Multi-Criteria Recommendation System

At this stage, a recommendation system calculation is carried out on 6 predetermined data criteria. The initial step of the calculation is calculated by calculating each extracurricular's probability value and how much the extracurricular data appears. Based on the results of calculations on each criterion in Table 2, the final results of recommendations with the highest value for all criteria can be seen in Table 4.

Table 4. Best prediction value from each criterion

Hasil Kriteria Penilaian	Student ID: 110027	
Previous Extracurricular	Prediksi	Rohani
	Nilai	2,5
Major	Prediksi	Paskibra
	Nilai	5
Gender	Prediksi	Tari
	Nilai	5
Score	Prediksi	Rohani
	Nilai	5
Extracurricular Rating	Prediksi	English Club
	Nilai	3
Extracurricular Category	Prediksi	English Club
	Nilai	2,5

4.2. PSO Model

This stage is carried out to determine the weight of each predetermined criterion. Weight calculations are carried out based on preferences and existing data on student data and extracurricular data. The sum of the weights based on the type of recommendation must add up to 1 to obtain which criteria determine more student recommendations. PSO weight calculations are carried out by following the flow described in Table 4. One of the results of fitness and position evolution values can be seen in Figure 3.

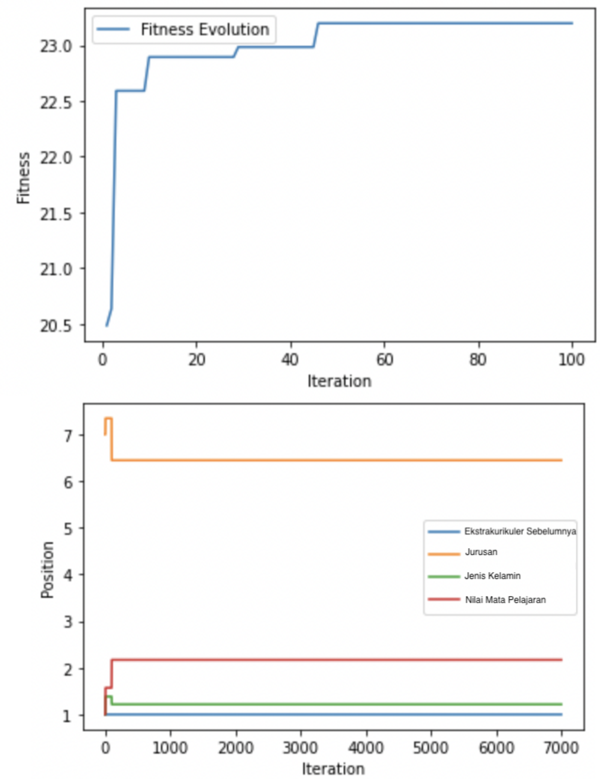

Fig. 3 Fitness and evolution value

Based on the fitness and position values obtained from Table 4, the result is a weight value that can be seen in Table 5.

Table 5. Weight value from each criterion

Type	Criteria	Student ID			
		1100 26	1100 27	1100 28	1100 29
CB	Previous Extracurricular	0,3	0,1	0,4	0,5
CB	Major	0,2	0,7	0,2	0,1
CB	Gender	0,2	0,1	0,2	0,1
CB	Score	0,3	0,1	0,2	0,3
CF	Rating	0,7	0,1	0,6	0,7
CF	Category	0,3	0,9	0,4	0,3

Based on the weight calculation results in Table 5, it can be seen that the weights for each criterion owned by students have different results, depending on the preferences and data owned by each student. For example, the highest weight owned by students with ID 110027 is the Extracurricular Category criteria in the CF method because these students have given good ratings to several extracurricular activities compared to other students in the calculation results in Table 5. The calculation is carried out based on each criterion's type of recommendation system so that if all weights are summed up based on the recommendation system, it will result in a value of 1.

For the calculation of weights on the hybrid recommendation system, further calculations are carried out to determine each weight on CB and CF using PSO. It is done to produce a more optimal weight between CB and CF based on preferences and data contained in students and extracurricular activities. The summation result between CB and CF weights is 1. Some of the calculation results can be seen in Table 6.

Table 6. Weight value based on recommendation type

Type	**Student ID**			
	110026	**110027**	**110028**	**110029**
CB	0,6	0,6	0,2	0,1
CF	0,4	0,4	0,8	0,9

The weight calculation results based on the recommendation system type in Table 6 show the preferences of each student and extracurricular data. For example, students with ID 110027 weight 0.6 on Content-Based because the data contained in CB is better and more complete than in CF. Meanwhile, students with ID 110029 give good ratings to several extracurricular activities, so the CF's weight is greater than that of CB.

4.3. Implementation of Hybrid MCRS and PSO

The recommendation system calculation is based on 6 predetermined data criteria using PSO weights at this stage. The calculation is done by multiplying the value of the recommendation results that have been obtained in Table 4 and then multiplying by the value of the weight results that have been obtained in Table 5. Some of the results of the recommendation system calculation based on the assessment criteria can be seen in Table 7.

Table 7. Recommendation value based on PSO weight

Criteria	**ID Siswa = 110027**	
Previous Extracurricular	Prediksi	Rohani
	Nilai	0,25
Major	Prediksi	Paskibra
	Nilai	3,5
Gender	Prediksi	Tari
	Nilai	0,5
Score	Prediksi	Rohani
	Nilai	0,5
Rating	Prediksi	English Club
	Nilai	0,3
Category	Prediksi	English Club
	Nilai	2,25

The results in Table 7 are the results of each assessment criteria multiplied by each weight calculated previously. These results obtain the best recommendation value for each criterion from the matrix of student and extracurricular data results. Some student recommendation data produce recommendations with a value of 0. It is because the data on these criteria is empty.

The final hybrid recommendation results are obtained by summing up each result based on the criteria according to the type of recommendation system and then multiplying by the PSO weights previously obtained in Table 5. Some of the results of hybrid extracurricular student recommendations can be seen in Table 8.

Table 8. The final recommendation by recommendation type

Criteria	**ID Siswa = 110027**	
CB	Prediksi	Paskibra
	Nilai	2,85
CF	Prediksi	English Club
	Nilai	1,02
Hybrid (Final Recommendation)	Prediksi	Paskibra
	Nilai	3,87

To get results based on the type of recommendation, the results of each criterion are summed up based on the type of recommendation system. The summation results are then multiplied by the weight of the recommendation system for each student to obtain prediction results based on the recommendation system, both CB and CF. The CB and CF recommendation results are then summed up to obtain the final result of the hybrid recommendation. The hybrid recommendation value is then used as a reference for extracurricular recommendations for students. The higher the final result obtained, the more extracurricular activities are recommended.

4.4. Evaluation

4.4.1. MAPE Evaluation

In the next stage, calculating the recommendation system is evaluated using the MAPE method. MAPE method can show how many percent errors are generated by the prediction results with actual data. The data in this study only ranges from dozens of student and extracurricular data, so it is more appropriate to use the MAPE method than the MAE or MSE methods.

The calculation is done by sorting the recommendation values according to the recommendation method performed. The sorting value of the recommended value is then used as the predicted value, and the actual value is obtained from the extracurricular value obtained by the actual student. The results of the error value obtained are then aggregated according to each recommendation method carried out. Then the aggregate value for each recommendation is then averaged, which then results in an error value for each recommendation method performed. The results of the final calculation of the error value for each recommendation method can be seen in Table 9.

Table 9. Calculation time and error

Method	Execution Time (s) (smaller better)	Error (%) (smaller better)
Content Based	1,188	13,63
Content Based dan GA	5,016	12,84
Content Based dan PSO	12,370	13,98
Collaborative Filtering	11,231	23,25
Collaborative Filtering dan GA	13,451	23,25
Collaborative Filtering dan PSO	20,355	23,25
Hybrid (CB + CF)	12,419	8,74
Hybrid (CB + CF) dan GA	23,151	8,65
Hybrid (CB + CF) dan PSO	42,259	6,20

The calculation results in Table 9 are the calculation results of three recommendation system methods, namely Content-Based, Collaborative Filtering, and Hybrid, and then compared the results of execution time and percentage error between normal methods, weight improvisation with GA, and weight improvisation with PSO.

In the CB method, the smallest percentage error result is obtained when added with the GA method, but the best time result is obtained when the normal method is used without adding other methods. It is reasonable because additional PSO or GA methods require additional calculation time. In the Collaborative Filtering method, the results of the percentage error obtained are the same. It is because there is some data on the empty rating criteria, causing the percentage error results to be large. For execution time, Collaborative Filtering without additional methods produces faster time, but when compared between GA and PSO, the results of PSO execution time are better than those of GA. The smallest percentage error result is obtained using the PSO method in the hybrid method. Still, the acquisition of execution time with PSO is the largest execution time when compared to no additional method, even with the GA method. However, the difference in the percentage error results of PSO is better by 2.45 compared to GA is a good result that can be done to minimize the percentage error value.

4.4.2. T-TEST Evaluation

In addition to evaluating with the MAPE method, an evaluation with the T-Test method is also carried out to determine how much influence the optimization has on the recommendation results. Before the t-test is carried out, a homogeneity test or equality of data variants is first carried out to test the equality of data variants used in the study using the F Test or Levene Test. The formula used in the F Test can be seen in Equation (2).

$$W = \frac{(N - k)}{(k - 1)} \frac{\sum_{i=1}^{k} N_i (\bar{Z}_i - \bar{Z}_{..})^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_{i.})^2} \tag{2}$$

The hypothesis used in the F Test is as follows:
- H_0 = Both data variants are the same
- H_1 = Both data variants are different

After testing the data using the F test, the result obtained is 0.4286 and compared with the predetermined significance value of 0.05. Because the value obtained is greater than the predetermined significance value (0.4286 > 0.05), the H_0 hypothesis is accepted with the conclusion that the two variants used are the same.

The test used in this study is the Paired T-Test with a significance level of 5% or 0.05. Before testing the T-Test, the hypothesis used is first determined, namely:

- H_0 = There is a significant increase in the results of recommendations using the PSO / GA optimization algorithm.

- H_1 = There is no significant increase in the results of recommendations using the PSO / GA optimization algorithm.

After the hypothesis is carried out, the Paired T-Test is tested on the data, with the results of statistical calculations can be seen in Table 10.

Table 10. Result of t test value

Variable	Mean	Std. Dev	Std. Err Mean	95% Conf	Interval
PSO	0,0638	0,0876	0,0085	0,0467	0,0808
GA	0,0935	0,1361	0,0133	0,0670	0,1200

Based on the values obtained in table 10, it is used to find the t value with the following formula:

$$t = \frac{\sum d}{\sqrt{\frac{n(\sum d^2) - (\sum d)^2}{n - 1}}} \tag{3}$$

where:
d = the difference of each paired value,
n = number of data,

So based on Equation (3), the obtained t value is 0.0626. With a significance value of 0.05, the t value is greater (0.0626> 0.05), so the H_0 hypothesis is accepted. So the conclusion is that there are significant results in the results of the recommendations with the addition of the PSO / GA optimization algorithm.

4.5. Discussion

In the results of the research that has been done, the recommendations for each student with the best results are using the hybrid MCRS method with PSO. It is obtained because of the weight optimization that has been done on each criterion. The result of a longer execution time of 19.108 seconds in PSO compared to GA is more because, in PSO, there is a process of finding fitness values and position values for each criterion, so the more criteria used, the longer the PSO process results. In GA, the best solution value is obtained by mutating and sorting based on the highest fitness value, so the number of criteria does not affect the overall processing time.

However, the error evaluation results show that PSO recommendations get a better value of 2.45% compared to GA. It is because the weights obtained using PSO are calculated based on the pbest and gbest values of each criterion and particle, compared to GA, which calculates weights based on information owned by individuals or chromosomes and then spread to other individuals with crossover and mutation techniques.

The hybrid method (CB + CF) and PSO also produces a better error evaluation value compared to the CB and PSO method alone or CF and PSO alone. The hybrid (CB + CF) and PSO methods are better by 7.78% against the CB and PSO method and better by 17.05% against the CF and PSO method. However, both the CB and PSO method and the CF and PSO method are better in execution time because the CB method calculation is carried out on 4 criteria, and the CF method is carried out on 2 criteria only, as explained in Table 2.

5. Conclusion

Based on the research that has been conducted, the following conclusions are obtained:

1. The MCRS hybrid recommendation system produces better accuracy by 4.89% compared to the CB MCRS recommendation and better by 14.51% compared to the CF MCRS recommendation. However, from the performance results, the MCRS hybrid recommendation produces a longer time than the CB MCRS recommendation and the CF MCRS recommendation.
2. The PSO method produces better accuracy by 2.45% on hybrid MCRS than the GA method. However, the performance obtained by the PSO method is longer than the GA method on hybrid MCRS.
3. Gender assessment criteria are the criteria that have the most influence on the results of extracurricular recommendations in general. Then followed by subject value criteria and major criteria.

Acknowledgments

Author 1 and Author 2 contributed equally to this work.

References

[1] Menteri Pendidikan, Student Development. Regulation of the Minister of National Education of the Republic of Indonesia Number 39 (2008).
[2] M. H. Mohamed, M. H. Khafagy, and M. H. Ibrahim, Recommender Systems Challenges and Solutions Survey, in 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), (2019) 149–155. doi: 10.1109/ITCE.2019.8646645.
[3] F. Ricci, L. Rokach, and B. Shapira, Introduction to Recommender Systems Handbook BT - Recommender Systems Handbook, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, (2011) 1–35. doi: 10.1007/978-0-387-85820-3_1.
[4] G. Geetha, M. Safa, C. Fancy, and D. Saranya, A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System, Journal of Physics: Conference Series, 1000 (2018) 12101. doi: 10.1088/1742-6596/1000/1/012101.
[5] A. Esteban, A. Zafra, and C. Romero, Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization, Knowledge-Based Systems, 194 (2020) 105385. doi: https://doi.org/10.1016/j.knosys.2019.105385.
[6] Y. Gong et al., Genetic Learning Particle Swarm Optimization, IEEE Transactions on Cybernetics, 46(10) (2016) 2277–2290.
[7] R. Katarya and O. P. Verma, A Collaborative Recommender System Enhanced with Particle Swarm Optimization Technique, Multimedia Tools and Applications, 75(15) (2016) 9225–9239. doi: 10.1007/s11042-016-3481-4.
[8] X. Zhao, A Study on E-commerce Recommender System Based on Big Data, in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), (2019) 222–226. doi: 10.1109/ICCCBDA.2019.8725694.
[9] P. Jomsri, Book Recommendation System For Digital Library Based On User Profiles by using Association Rule, In Fourth Edition Of The International Conference on the Innovative Computing Technology (INTECH 2014), (2014) 130–134. doi: 10.1109/INTECH.2014.6927766.
[10] F. Firmahsyah and T. Gantini, Penerapan Metode Content-Based Filtering Pada Sistem Rekomendasi Kegiatan Ekstrakulikuler (Studi Kasus Di Sekolah ABC), Jurnal Teknik Informatika dan Sistem Informasi, 2(3) (2016). doi: 10.28932/jutisi.v2i3.548.
[11] T. Rutkowski, J. Romanowski, P. Woldan, P. Staszewski, R. Nielek, and L. Rutkowski, A Content-Based Recommendation System Using Neuro-Fuzzy Approach, in 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), (2018) 1–8. doi: 10.1109/FUZZ-IEEE.2018.8491543.
[12] P. K. Jain and R. Pamula, Content-Based Airline Recommendation Prediction Using Machine Learning Techniques, in Machine Learning Algorithms for Industrial Applications, S. K. Das, S. P. Das, N. Dey, and A.-E. Hassanien, Eds. Cham: Springer International Publishing, (2021) 185–194. doi: 10.1007/978-3-030-50641-4_11.

[13] D. Wang, Y. Liang, D. Xu, X. Feng, and R. Guan, A content-based recommender system for computer science publications, Knowledge-Based Systems, (157) 1–9. doi: https://doi.org/10.1016/j.knosys.2018.05.001.
[14] H. Mohamed, L. Abdulsalam, and H. Mohammed, Adaptive Genetic Algorithm for Improving Prediction Accuracy of a Multi-Criteria Recommender System, in 2018 IEEE 12th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC), (2018) 79–86. doi: 10.1109/MCSoC2018.2018.00025.
[15] Q. Han, Í. M. de R. de Troya, M. Ji, M. Gaur, and L. Zejnilovic, A Collaborative Filtering Recommender System in Primary Care: Towards a Trusting Patient-Doctor Relationship, in 2018 IEEE International Conference on Healthcare Informatics (ICHI), (2018) 377–379. doi: 10.1109/ICHI.2018.00062.
[16] W. Hong-xia, An Improved Collaborative Filtering Recommendation Algorithm, in 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), (2019) 431–435. doi: 10.1109/ICBDA.2019.8713205.
[17] M. Gupta, A. Thakkar, Aashish, V. Gupta, and D. P. S. Rathore, Movie Recommender System Using Collaborative Filtering, in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), (2020) 415–420. doi: 10.1109/ICESC48915.2020.9155879.
[18] Z. Su and C. Chen, Implementation of collaborative filtering algorithm with time attribute, in 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 9 (2020) 1554–1558. doi: 10.1109/ITAIC49862.2020.9338992.
[19] P. Venil, G. Vinodhini, and R. Suban, Performance Evaluation of Ensemble based Collaborative Filtering Recommender System, in 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), (2019) 1–5. doi: 10.1109/ICSCAN.2019.8878777.
[20] Adel H. Al-Mter, Songfeng Lu, Yahya E. A. Al-Salhi, Arkan A. G. Al-Hamodi, A Particle Swarm Optimization Based on Multi Objective Functions with Uniform Design SSRG International Journal of Computer Science and Engineering, 3(10) (2016) 1-7. Crossref, https://doi.org/10.14445/23488387/IJCSE-V3I10P104
[21] Y. Afoudi, M. Lazaar, and M. al Achhab, Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network, Simulation Modelling Practice and Theory, 113 (2021) 102375. doi: https://doi.org/10.1016/j.simpat.2021.102375.
[22] C. Song, Q. Yu, E. Jose, J. Zhuang, and H. Geng, A Hybrid Recommendation Approach for Viral Food Based on Online Reviews, Foods, 10(8) (2021) doi: 10.3390/foods10081801.
[23] Y. Wang, M. Wang, and W. Xu, A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework, Wireless Communications and Mobile Computing, (2018) 8263704. doi: 10.1155/2018/8263704.
[24] Z. Gulzar, A. A. Leema, and G. Deepak, PCRS: Personalized Course Recommender System Based on Hybrid Approach, Procedia Computer Science, 125 (2018) 518–524. doi: https://doi.org/10.1016/j.procs.2017.12.067.
[25] A. Pal, P. Parhi, and M. Aggarwal, An improved content based collaborative filtering algorithm for movie recommendations, in 2017 Tenth International Conference on Contemporary Computing (IC3), (2017) 1–3. doi: 10.1109/IC3.2017.8284357.