CASE BASED REASONING: RECOMMENDATIONS OF HIGH SCHOOL STUDENT PROGRAM

Anisa Lutfiyani and Fatchul Ariffin
Electronics and Informatics Engineering Education Postgraduate,
Yogyakarta State University, Yogyakarta, Indonesia

Email: an.lutfiyani@gmail.com, fatchul@uny.ac.id

Abstract. Study Program Recommendation System is a computer-based information system that aims to provide an overview or recommendation of study program based on students' interests or character from testing through Case Based Reasoning method. The main purpose of this article is to develop an expert system as study program recommendation for High School student through Case Based Reasoning. This system is made because many parents tend to force their health when choosing a college for their children. It can be impact on the children's developing. Determining the recommendations of a student's study program with a certain character, often not enough for the expert's knowledge, but need to pay attention to cases or events similar to those of the student. A concept held in the field of artificial intelligence that uses case-based or CBR reasons for analysis and decision making. Decisions are determined by calculating the similarities between the new cases and the expert case that has occurred. This study is observation and literature study. The results of this study showed that an engineering inference expertise with the aim of obtaining a decision of study program recommendations can be done using the CBR method based on the characteristics of students.

1. Introduction

Article 1 of the National Education System Law of 2003 explains that Senior High School is a formal education that prioritizes the expansion of knowledge by students to continue to a higher level.

High school / vocational high school students in class XII who plan to continue their education to higher education are sometimes confused to choose the appropriate study program. Often the choice of students does not match their actual interests, so it is unfortunate that a very large desire to enter college, is not mature enough in choosing study programs. Finally it is often heard that quite a lot of new students fail midway or drop out of school when they are accepted into college.

Data from Menristekdikti, there are several higher education institutions whose dropout rates are very large, namely: DKI Jakarta contributed 40,584 students dropping out of school in 2017 or 3.7% students in this province. West Java Province came in second place, with 21,146 dropouts or 2.5% students. Furthermore, North Sumatra, with a total of 12,418 students dropping out of school, 2.9% of the total number of students in the province. Meanwhile, DI Yogyakarta is in tenth position. Of 372.48 students studying in this city in 2017, there were 12,425 or 3.3% students who had to drop out of school.

Meanwhile, according to research conducted by Educational Psychologist from Integrity Development Flexibility (IDF) Irene Guntur, M. Psi., CGA was quoted from okezone.com (Harahap, 2014), as many as 87% of students in Indonesia experienced the phenomenon of taking the wrong study program, its extreme impact was the forerunner of unemployment. Thus, high school / vocational high school students grade XII need to have educational goals in the future and strive to realize real goals in
the future. Therefore there is a need for actions that change the way of thinking and encourage prospective students to be more careful in choosing the right study program with their desires and interests and talents.

Based on the above problems, an expert system is needed to provide solutions to high school / vocational high school students who want to go to college without being confused about what study program they will choose based on their talent or characteristics through tests. This system will be built web-based with the method used is Case Base Reasoning.

Case Base Reasoning is a method of reasoning in the field of artificial intelligence (artificial intelligence) which is used to find the best solution of problems related to choosing one solution from many solutions (Daqing & Burrell, 2001). CBR can overcome all the problems mentioned earlier. Case-based CBR is easier to obtain than rules (Salem & HodHod, 2005). CBR can reuse previous results, from the beginning does not require a step by step reason, but their effectiveness will continue to develop by solving new problems (Aamodt & Plaza, 2009). The reasoning used in this study is the character of students while the conclusion is the results of the study program recommendations.

With this research, it is expected to develop an expert system to provide study program recommendation to high school students using case-based reasoning method with similarity measurement method nearest neighbor.

2. Research Method

2.1 Data Collection

This research was conducted in several stages, namely:

2.1.1 Interviews, conducted with experts namely psychology, this is to get basic knowledge about the characteristics of students needed in the process of making and developing systems.

2.1.2 Literature, carried out by collecting data or written texts in the form of scientific journals, research reports, e-books, and other relevant sources and collecting and studying reference books or literatures that discuss student characteristics and Case Based Reasoning methods.

2.2 Testing Method

2.2.1 Case-based Reasoning (CBR) Process

To determine the results of the study program recommendations, there are several stages of the process carried out, this stage is carried out to find the similarity of new cases with standard cases that are already stored in the database. The stages of the process that occurred at CBR can be seen in Figure 1, as follows:
In general, the case based reasoning method consists of 4 steps, namely:

1. **Retrieve**
   In this process is a problem recognition process, namely by identifying new problems.

2. **Reuse**
   In the Reuse process, the system will search for old problems in the database through identifying new problems. Then reuse the old problem information that has similarities to complete the new problem.

3. **Revise**
   In this process, information will be re-evaluated to overcome problems that occur in new problems. Then the system will issue a new problem solution.

4. **Retrain**
   In this last process, the system will save new problems into knowledge-based to solve future problems.

Based on the description above, CBR works by comparing case standards with new cases, if case standards are related to new cases, CBR will provide standard case solutions for new cases. Each problem contains a problem and a solution. If there is no match, CBR will divert or revise by inserting new cases into the database of cases, so that knowledge will increase indirectly.

### 2.2.2 Inference Engine

Processes in Case Based Reasoning (CBR) can use various techniques, including the Nearest Neighbor Algorithm. The nearest neighbor algorithm is an approach to look for cases with the closeness between new cases and standard cases, which is based on the weight matching of a number of existing features (Kusrini, 2009). This method looks for distances to the destination of data that has been previously stored. After getting the distance, then the closest distance is searched. The shortest distance is used to find the destination solution.
The formula used in the calculation of closeness (similarity) is as follows (Lubis, 2016):

\[
\text{Similarity (problem, case)} = \frac{\sum_{i=1}^{N} w_i S_i + w_i S_i + \cdots + w_i S_i}{N_1 + N_2 + \cdots + N_N}
\]

where:
- \( S \) = Similarity (similarity value)
- \( W \) = Weight (weight given)
- \( N \) = Amount of Input New Entry

3. Results and Discussion

3.1 Flowchart Case Based Reasoning

![Flowchart Case Based Reasoning](image)

**Figure 2.** Flowchart Case-Based Reasoning (Yulianti, 2016).

3.2 Determination of Similarity Solutions

The basis of CBR is solving the information that has been stored in the previous case. Based on the stages in the CBR system, three main steps are needed in determining the similarity, (Lubis, 2016) namely:
3.2.1 Build a case base

Build a case base that will be used as storage. Each case stored in the case base is divided into 2 parts, namely:

| Table 1. Part Factors in Each Case |
|-----------------------------------|
| Part factor in each case          |
| Student character                 |
| Study Program Recommendations     |

This division is done to facilitate the retrieval of data in accordance with the new case, as well as to store new case data into the case base.

3.2.1.1 Student character, this section contains the character of students who determine student interest. The examples of some student characters are as follows:

| Table 2. Character of students |
|-------------------------------|
| Code | CHARACTER                   |
| K001 | Like to play computers and discover how computers work. |
| K002 | Likes to do systematic and thorough work. |
| K003 | Like playing games that use strategies or sharpen the brain, such as chess, mystery games, puzzles, logic, etc. |
| K004 | Easily remember numbers or statistics such as soccer scores, height of the tallest building etc. |
| K005 | Happy to describe visual information in the form of graphics, tables or images. |
| K006 | Like to use technology to solve mathematical problems. |
| K007 | Like to play and find ways to improve electronic equipment. |
| K008 | Like to think critically in analysis. |
| K009 | Like to study or work alone. |
| K010 | Like to do research. |
| K011 | Like to think about the future and goals to be achieved. |
| K012 | Like at home and spend time alone. |
| K013 | Like to contemplate or think about life or myself. |
| K014 | Knowing my strengths and weaknesses. |
| K015 | Like writing diaries to express ideas, memories, Feelings of events that occur. |
| K016 | Like to think or consider the positive or negative side when encountering problems or making decisions. |
| K017 | Like math or counting |
| K018 | Not always using fingers or tools (such as calculators etc.) in counting. |
| K019 | Like to find opportunities that may or can be found based on formulas, formulas that already exist. |
| K020 | Like to read books |
| K021 | ................................................................. |
3.2.1.2 *Recommended Study Program*, this section contains recommendations for study programs that determine student interest. This section has several study programs, namely:

| CODE | NAME PRODY                        |
|------|-----------------------------------|
| J001 | Informatics / Computer Engineering|
| J002 | Electrical Engineering            |
| J003 | Psychology                        |
| J004 | Mathematics                       |
| J005 | Language                          |
| J006 | Health                            |
| J007 | Physics                           |
| J008 | Economic/accounting               |
| J009 | Music Art                         |
| J010 | Sports                            |
| J011 | Geography                         |
| J012 | Law                               |
| J013 | History                           |
| J014 | Chemistry                         |
| J015 | Biology                           |

3.2.2 *Determine The Similarity Function*

The process of determining the similarity of new cases by adapting existing case standards, there are 4 processes in determining similarity, namely:

3.2.2.1 *Retrieve.* At the beginning of the process of determining the recommendation of a study program, the user fills in the student biodata, after that it is continued with the user inputting the answer from the questionnaire to be answered. The user fills out the questionnaire consisting of 65 questionnaires in the form of student characters, the user finishes answering the questionnaire, the user will see the results of the study program recommendations that have passed the case-based reasoning process, search for similarities with the previous case standards, in the case memory to find out the recommendations of the study program. Then the system will do the weighting process by matching one by one between the characters that have been entered with those in the knowledge base. After carrying out the weighting process, the highest similarity value is determined for the results in determining the study program. The weighting carried out by the system is displayed in the calculation of the following nearest similar similarity algorithm (Lubis, 2016):

**Parameter weight (w):**
- Selected features (equal) = 1
- Selected features (not the same) = 0
- Unselected features = 0

**Similarity (problem, case)** = \( \frac{S_1w_1+S_2w_2+\ldots+S_nw_n}{N_1+N_2+\ldots+N_n} \)  \( (2) \)

- **S** = Similarity (similarity value)
- **W** = Weight (weight given)
N = Amount of Input New Entry
N is filled with student characters For example:
K001, K004, K011, K014, K020, K028, K040 then n counts as many as 7 new case entries.

3.2.2.1.1 Old Case Tables
In order for the new case to be calculated as close to the old case standard, it is necessary to define the weight of the feature. Defining the weight of CBR features is shown as shown below:

Table 4. Defining the Informatics Engineering Study Program

| Code | Character                                                                 | Similarity | Similarity Value |
|------|---------------------------------------------------------------------------|------------|-----------------|
| K001 | Like to play computers and discover how computers work.                   | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K002 | Likes to do systematic and thorough work.                                 | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K003 | Like playing games that use strategies or sharpen the brain, such as chess, mystery games, puzzles, logic, etc. | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K004 | Easily remember numbers or statistics such as soccer scores, height of the tallest building etc. | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K005 | Happy to describe visual information in the form of graphics, tables or images. | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K006 | Like to use technology to solve mathematical problems.                     | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K009 | Like to study or work alone.                                              | Same       | 1               |
|      |                                                                          | Different  | 0               |

Study Programs
Informatics Engineering/Computer

Table 5. Defining of Electrical Engineering Study Programs

| Code | Character                                                                 | Similarity | Similarity Value |
|------|---------------------------------------------------------------------------|------------|-----------------|
| K007 | Like to play and find ways to improve electronic equipment.               | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K008 | Like to think critically in analysis.                                    | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K009 | Like to study or work alone.                                             | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K010 | Like to do research.                                                     | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K002 | Likes to do systematic and thorough work                                 | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K005 | Happy to describe visual information in the form of graphics, tables or images. | Same       | 1               |
|      |                                                                          | Different  | 0               |
| K006 | Like to use technology to solve mathematical problems.                    | Same       | 1               |
|      |                                                                          | Different  | 0               |

Study Programs
Electrical Engineering
Table 6. Defining the Psychology Study Program

| Code  | Character                                                                 | Similarity   | Similarity Value |
|-------|---------------------------------------------------------------------------|--------------|------------------|
| K011  | Like to think about the future and goals to be achieved.                  | Same         | 1                |
|       |                                                                           | Different    | 0                |
| K012  | Like at home and spend time alone.                                        | Same         | 1                |
|       |                                                                           | Different    | 0                |
| K013  | Like to contemplate or think about life or myself.                        | Same         | 1                |
|       |                                                                           | Different    | 0                |
| K014  | Knowing my strengths and weaknesses.                                      | Same         | 1                |
|       |                                                                           | Different    | 0                |
| K015  | Like writing diaries to express ideas, memories, Feelings of events that occur. Like to think or consider the positive or negative side when encountering problems or making decisions. | Same         | 1                |
|       |                                                                           | Different    | 0                |
| K016  | Likes to think of solutions to problems that are happening.               | Same         | 1                |
|       |                                                                           | Different    | 0                |
| K047  |                                                                           | Same         | 1                |
|       |                                                                           | Different    | 0                |

3.2.2.1.2 New Case Table

An example of a new case table, containing student characters that have not been stored in the case memory. Examples of new cases that will be matched to look for similarities with the standard old case in recommending the study program are:

Table 7. New Case Tables

| Code  | Character                                                                 |
|-------|---------------------------------------------------------------------------|
| K001  | Like to play computers and discover how computers work.                   |
|       | Easily remember numbers or statistics such as soccer scores, height of the tallest building etc. |
| K004  | Like to use technology to solve mathematical problems.                    |
| K014  | Knowing my strengths and weaknesses.                                      |
| K020  | Like to read books                                                        |
| K028  | Nice to do scientific experiments so that I can train my creativity.      |
| K040  | Prefer to move, to learn something that makes it easier for me to remember. |

3.2.2.1.3 Calculations using the Nearest Neighbor Retrieval Algorithm

At this stage a case calculation will be carried out in which the calculation of the new case with this old case aims to find similarities, namely using the weighting process using the Nearest Neighbor Retrieval formula.
Calculation of Case I

Similarity \( (x, K_1) \)

\[
\frac{(1+1)+(0+0)+(0+0)+(1+1)+(0+0)+(1+1)+(0+0)}{1+1+1+1+1+1+1} = \frac{6+6+6+6+6}{7} = \frac{3}{7} = 0.43
\]

Calculation of Case II

Similarity \( (x, K_2) \)

\[
\frac{(0+0)+(0+0)+(0+0)+(0+0)+(0+0)+(1+1)}{1+1+1+1+1+1+1} = \frac{0+0+0+0+0+0+1}{7} = \frac{1}{7} = 0.14
\]
Calculation of Case III

\[
\text{Similarity}(x, K3) = \frac{(0+0)+(0+0)+(0+0)+(1+1)+(0+0)+(0+0)+(0+0)}{1+1+1+1+1+1+1} = \frac{10}{7} = 0.14
\]

After entering the value, the new case will be compared with each standard case as in table 4. The results of the calculation for the similarity of each case stored in the knowledge base with the new case are as follows:

| Case Base | Total |
|-----------|-------|
| K1        | 3/7 = 0.43 |
| K2        | 1/7 = 0.14 |
| K3        | 1/7 = 0.14 |

3.2.2.1.4 Data Retrieval

Criteria for selecting cases are cases that have the highest similarity value with new cases submitted as a solution. The following are the similarity criteria for the division, namely:

| Similar Decimal Value | Similarity Value |
|-----------------------|------------------|
| 0.8 – 1               | High             |
| 0.4 – 0.79            | Medium           |
| 0 – 0.39              | Low              |
4. Conclusion

Based on the table of similarity criteria, each case on the basis of the case has similarity criteria with the new case as follows:

1. K1 has a similarity of 0.43, the value of similarity of Medium criteria
2. K2 has a similarity of 0.14, the value of similarity Low criteria
3. K3 has a similarity of 0.14, the value of the similarity of Low criteria.

Therefore, the results of the criteria for the similarity of new cases will be the suggested solution for new cases, if the level of similarity reaches a minimum of 80%. Because the criteria for the similarity of new cases with the old case K1 are Medium, while K2 and K3 are Low so the solution of each old case is not recommended for the solution of existing new cases, in other words, the new case is probably not the same as the case on the knowledge base. But in this case the value of the K1 Informatics Engineering / computer is the highest value compared to the other cases.

5. References

[1] Aamodt, A., dan Plaza E., (1994), “Case Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches”. Journal of Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. Al Communications. IOS Press, Vol. 7: 1, pp. 39-59.

[2] Daqing. C. and Burrerel, P., (2001), Case Based Reasoning System and Artificilan Neural Network: A review, Springer-Verlag.

[3] Harahap, R. F., “okezone.com,” MNC, 25 February 2014, [Online], Available: http://news.okezone.com/read/2014/02/24/373/945961/duh-87-mahasiswa-indonesia-salah-jurusan. [Diakses 2019].

[4] Kemenristekdikti, (2017), Statistik Perguruan Tinggi (Higher Education Statistical Year Book), Jakarta: Pusdatin Iptek Dikti.

[5] Kusrini dan Emha, (2009), Algoritma Data Mining. Yogyakarta : C.V Andi Offset.

[6] Law of the Republic of Indonesia Number 20, (2003), Concerning the National Education System.

[7] Lubis,A. (2016), Case-Based Pregnancy Consultation Expert System Using Case Based Reasoning (CBR) Method, Universitas Pasir Pengairan, Riau Journal Of Computer Science Vol. 2/No.1/2016 : 66 – 75.

[8] Salem . M., Roushdy. M and HodHod. R.A., (2005), A Case Based Expert System for Supporting Diagnosis of Heart Diseases, AIML, Journal, Volume (5), Issue 1, March.

[9] Yulianti, W., (2016), Case-Based Reasoning Aptitude Testing in Expert Systems to Determine Interest and Talent of Elementary School Students, Jurnal Teknologi dan Sistem Informasi UNIVRAB Vol. 1 No. 2, Juli 2016.