Selection of Scholarship Recipient by Implementing Genetic Algorithm and Fuzzy Logic

Yohanssen Pratama*, Monalisa Pasaribu, Joni Nababan, Dayani Sihombing, Dicky Gultom
Faculty of Informatics and Electrical Engineering, Institut Teknologi Del, Laguboti 22381, Indonesia
*yohanssen.pratama@del.ac.id

Abstract. PPA (Academic Achievement Improvement) scholarship acceptance Institut Teknologi Del is open annually for students who have improved academic performance but are economically disadvantaged. Every year there is significant increasing of the number of students to apply for the scholarship. As a result, it becomes more difficult to determine students who are entitled to receive the PPA scholarship since the procedure was done manually. As a result, the selection time needed to produce a decision is longer. In response to this, the authors propose the implementation of Genetic Algorithms and Fuzzy Logic in determining scholarship recipients as well as using data from the Institut Teknologi Del Information System (CIS) using student data of 2016, 2017 and 2018. The data obtained is done by data pre-processing, then being input into the system. In the system the state has been set with CR = 0.9, MR = 0.1 and fuzzy parameters will be used. Following the genetic algorithm stages of population initialization, fitness evaluation, selection, crossover, mutation, and elitism, the fuzzy logic process is carried out in the form of fuzzyfication, inference and defuzzyfication. After the fuzzy process is completed, 10 scholarship recipient student data are obtained. Since the results of the membership function provided by the genetic algorithm always changed, it caused different result in each experiment. Four data similar to the original data of the scholarship recipient was finally shown in the 37th experiment from the total of 50 experiments carried out.

1. Introduction

Scholarships are usually awarded by governments, companies and foundations [1]. In Institut Teknologi Del, a higher-education institute located in North Sumatra, various kinds of scholarships are offered to students. However, in the selection of scholarship recipients, the process currently is still done manually. The initial process is student fills out the application form. After that, the student affairs will process the scholarship applicants' data and then select the scholarship recipients. In determining the scholarship recipients, candidate must meet the criteria for scholarship recipients, namely parents salary, number of dependents, maximum behavior score of 9, and a minimum GPA of 3.00 semester 1 and 2. Due to the large number of student data that will be processed to determine scholarship recipients, it obviously will take a lot of time to do this whole process. Therefore, solutions are needed for this problem.

Alfiani Fitri and Wayan Firdaus Mahmudy have previously conducted research on determining scholarship recipients. In this study, there is a method used to determine scholarship recipients, namely fuzzy logic where the problem of optimizing the Tsukamoto fuzzy membership function was genetic algorithm and chromosome repair using the hill climbing algorithm. The calculation result of system accuracy used the Spearman correlation while in the optimization of Tsukamoto fuzzy membership, the experiment used a genetic algorithm with a chromosome repair mechanism was
0.986 [2]. Research on the determination of scholarship recipients was also conducted by Emirza Wira Saputra by comparing the results of determining the recipients of fuzzy logic scholarships without using genetic algorithms with the results of determining scholarships using fuzzy logic with the help of genetic algorithms. The results obtained in this study are done by comparing the Fuzzy Mamdani method without optimization of 0.8801 and the Fuzzy Mamdani method with optimization using the Genetic Algorithm of 0.9172. The measurement was done through Spearman correlation. This indicated that the use of the genetic algorithm can optimize the correlation level of 0.0371 [3].

Based on several studies that have been carried out, the solution to the problem of determining scholarship acceptance at the Institut Teknologi DeI is by building a system that will make it easier to determine scholarship recipients using genetic algorithms and fuzzy logic. The fuzzy logic is chosen because this algorithm is a form of decision support model where in determining the membership function, it is still based on human perception. It is relevant since human will be the experts in the problem of choosing scholarships which reflects how fuzzy membership works. It provides probability with gap value of [0,1] yet it has significant difference in data interpretation. While the genetic algorithm is a form of the type of evolution algorithm (EA) which is most popular because of its ability to solve various complex problems and the genetic algorithm is a random search but leads to a better solution. Therefore, it is expected that by implementing a genetic algorithm, it will easier to determine the fuzzy membership function since it is done automatically [3].

2. Methodology

Below is research design that we used in this experiment:

![System Design](image.png)

Figure 1. System Design

As indicated in Figure 1, the process of determining scholarship recipients using genetic algorithms and fuzzy logic is explained as follows:

The process of determining scholarship recipients uses data that has been previously processed according to the scholarship variables or criteria used in determining who is entitled to receive the scholarship, namely parents’ salary, number of dependents, GPA of semester 1 and 2, and behavior scores to make it easier to determine the scholarship recipient. PPA scholarships are only given to students who show progress in their academic achievement but are economically disadvantaged in continuing their studies. The quota of this scholarship is 10 students per year [4]. This scholarship has the following requirements:

a. Parents’ salary
b. Number of dependents of parents
c. Maximum behavior score of 9

d. Minimum GPA of 3.00

The duration of scholarships for active students is based on the current fiscal year period and will be awarded for the first time for at least 6 months. PPA Scholarships can be terminated if students [4]:
a. Already passed other scholarship
b. Resign or drop out
c. Receive academic sanctions from universities
d. Do not meet the predetermined requirements
e. Provide incorrect data
f. Pass away

2.1. Data Set

The data used in the study is taken from the Information Resources Unit of Institut Teknologi Del (SDI IT Del). This study uses student data from the 2016, 2017 and 2018 batches using only semester 1 and 2 data. The data used shows parents salary, number of dependents, GPA (semester 1 and 2) and behavior scores. The data used is divided into training data, namely data for batches 2016 and 2017 because batches of 2016 and 2017 have passed that semester and have participated in the PPA scholarship recipient selection. While for the testing process the data used is based on students in batch 2018 because they have just passed semester 1 and 2.

2.2. Data Preprocessing

Data pre-processing is a data processing stage so that it will produce good data and in accordance with the required data. Data mining is one of the KDD (Knowledge Discovery in Databases) processes. The preprocessing stages carried out on student data are data selection, data cleaning, and data transformation. The KDD (Knowledge Discovery in Databases) explanation is used, namely:

a) Data Selection

The data selection stage is the stage of selecting data or variables to be used and fulfilling the criteria for scholarship recipients. Student data that has been collected is student data from semester 1 to 5, so that the data will be selected and produce data according to the needs in the study, namely student data in semesters 1 and 2. Then an examination is carried out on student data in the form of unused variables so that the data will be deleted, such as variables of father's income range, mother's income range, father's job, mother's job, father's job description and mother's job description.

b) Data Cleaning

The data cleaning stage is the stage for handling values that are missing. At this stage data is filled in the missing value for data that is empty in a variable. The method used is to use the average or median of a variable to fill in any blank data on that variable.

c) Data Transformation

The data transformation stage is the process of changing the data format. At the data transformation stage, using the Attribute Construction method, it creates a new attribute using mathematical expressions on existing attributes. The variables processed at this stage are the GPA in semester 1 and 2, father's salary, mother's salary and the number of dependents as a reference for family expenses. Following is the chart of student data comparison for batch 2016, 2017 and 2018 after pre-processing.
Figure 1. Average Salary Histogram

Figure 2 depicts the average salary of each class, namely 2016, 2017 and 2018. The histogram shows the average salary in 2016 of 1,444,574, in 2017 it was 1,456,483 and in 2018 it was 1,755,056. The figure also shows that the highest paid salary is in 2018 and the lowest average salary is in 2016.

Figure 2. Average Behaviour Score Histogram

Figure 3 illustrates the average score of behavior from each class, namely 2016, 2017 and 2018. The histogram shows the average score of behavior in 2016 of 4 and in 2017 and 2018 of 3. In the picture it also looks average - The highest average value of student attitudes in 2016.

Figure 3. Average GPA Histogram

Figure 4 illustrates the average GPA of students from each class, namely 2016, 2017 and 2018. The histogram shows the average GPA in 2016 of 3.24 and in 2017 of 3.22 and 2018 of 3.19. The picture also shows the highest average GPA of students in 2016 and the lowest average GPA in 2018.

2.3. Fuzzy Logic

Figure 4. Fuzzy Logic Flow [16]
Fuzzy logic (cryptic logic) is considered as an approach in mapping an input (input) space into an output (output) space. This logic operates a fuzzy set to obtain the expected output [16]. Groups that represent certain conditions in fuzzy variables can be called fuzzy sets. Meanwhile, fuzzy variables are variables that will be processed in a fuzzy system, for example, salary load, behavior score, performance index and others [17]. A curve that maps input points to membership values at intervals of 0 to 1 can be called a membership function [18]. Fuzzy membership provides a measure of opinion or decision, while probability indicates the proportion of the frequency with which an outcome is true in the long run [7]. The fuzzy logic design process can be seen in Figure 2.

The system has set cr = 0.9, mr = 0.1 and the parameters to be used in fuzzy logic. The initial population is generated using a genetic algorithm. The population initials consist of a collection of individuals using a chromosome representation that is a direct representation. One chromosome represents the fuzzy curve parameters used in the fuzzy variables (load salary, GPA, and behavior score) and fuzzy set. Where the problem will be solved using fuzzy logic. The length of the chromosomes to be made is 27 and the population to be raised is 10.

2.4. Genetic Algorithm

Generating the initial population is a stage to generate a number of random individuals (figure 6). The determination of the population to be raised depends on the type of genetic operator and the problem to be handled. In raising the population must meet the requirements well [7]. Then followed with determining the fitness value. Fitness value is a form of representation of a value that is used as a benchmark to see whether or not there is a result or solution (individual). The genetic algorithm aims to find the maximum fitness value or maximize the fitness value. The fitness value used is adding up each gene on a chromosome in a population [7]. The formula for the fitness value used refers to "(2)".

\[
\text{Fitness Value} = \sum_{i=1}^{n} V_i
\]  

Information:
- \( V_i \) : the ith gene on a chromosome
- \( n \) : the number of deep in a chromosome
- \( i \) : the ith index on a chromosome
We calculated the fitness value with formula 1 by counting all chromosomes in one population. The selection of two parent chromosomes to be crossed is carried out by the roulette wheel method. After selecting the best individual as a parent, the crossover stage is carried out with the uniform crossover type. Then the random mutation stage is carried out. This method is used so that the mutated gene will be replaced with another gene, where the replacement gene is randomly selected. Elitism is a method in which one or more chromosomes with the highest fitness value are copied directly to the next generation without being manipulated.

Then at this stage the system stops from the maximum iteration, which is 100 iterations. After getting the output of the GA process in the form of parameters for each variable that will be used in the fuzzy process, the next step is the fuzzification process. The fuzzification process is a process to get the degree of membership from an input numeric value (crisp). Inference is the process of making rules for fuzzy input obtained in the fuzzification process so that this process produces fuzzy output. Defuzzification is the last step in a fuzzy logic system which converts the results from inference to real numbers, where the result is an eligibility value. Then the feasible values obtained from the defuzzification are sorted based on the highest to low feasibility values. Then the 10 highest eligibility scores become scholarship recipient candidates.

3. Result and Discussion

This section describes the results of the determination of the scholarship recipients using genetic algorithms and fuzzy logic with 50 trials. Figure 7 is the GUI of our application which is a display of the system after carrying out the process of determining and issuing student scholarship recipient data:

![Figure 6. Graphical User Interface](image)

After implementing fuzzy logic and genetic algorithms and carrying out 50 experiments, the 37th experiment obtained the results of 4 recipient data which were the same as the original data of the scholarship recipients. using the membership function which can be seen in Figure 8 so that the data appears. Because the data obtained is raw data where some student data is incomplete, it affects the process carried out for implementation using fuzzy logic and genetic algorithms. Moreover, the membership function given by genetic algorithms is different for each process and this impacts the fuzzy output that is always changing in every result.
4. Conclusion

After the implementation of fuzzy logic and genetic algorithms in this final project, it can be concluded that the genetic algorithm and fuzzy logic can produce scholarship recipient data, within 50 experiments to reach the expected result. In the 37th experiment, it showed that the similarity of the scholarship recipient data has a different set of 4 recipient data compared to the previous experiments. This happens because the student data obtained from the student affairs section is incomplete so that during the preprocessing stage each incomplete or redundant student data will be deleted even though some of the student data that is deleted is a PPA scholarship recipient candidate, so this has an impact on making membership functions and rules on fuzzy logic. Then the membership function given by the genetic algorithm is different for each process and this also has an impact on the fuzzy output which is always changing in each result.

References

[1] Pusat Pembinaan dan Pengembangan Bahasa, *Kamus Besar Bahasa Indonesia*. Jakarta: Depdikbud, 1990.
[2] A. Fitri and W. F. Mahmudy, “Optimasi Keanggotaan Fuzzy Tsukamoto Menggunakan Algoritma Genetika pada Penentuan Prioritas Penerima Zakat,” *Optimasi Keanggotaan Fuzzy Tsukamoto*, vol. 1, no. 2, pp. 125–138, 2017.
[3] E. W. Saputra, “Optimasi Fungsi Keanggotaan Fuzzy Mamdani Menggunakan Algoritma Genetika Untuk Penentuan Penerima Beasiswa,” *J. Simada*, vol. 02, no. 02, 2019.
[4] Kemahasiswaan Institut Teknologi Del, *Pedoman Pemberian Beasiswa*. Sitoluama, 2018.
[5] D. E. Myori, “Kombinasi Logika Fuzzy dan Algoritma Genetika untuk Masalah Penjadwalan Perkuliahan,” pp. 284–292, 2018.
[6] D. L. Robie, “Correction to Genetic Algorithms and Applications to Data Compression,” pp. 1–30, 1996.
[7] W. Lestari, L. Benar, and L. Boolean, “Untuk Menentukan Mahasiswa Yang Berhak Menerima Beasiswa Di Stt Telkom Analysis and Implementation of Genetic Algorithm and Fuzzy Logic To Select Students Who Deserve Scholarships in,” 2006.
[8] L. Budin, D. Jakobović, and M. Golub, “Genetic algorithms in real-time imprecise computing,” *J. Comput. Inf. Technol.*, vol. 8, no. 3, pp. 249–257, 2000.
[9] R. Sivaraj, “a Review of Selection Methods in Genetic Algorithm,” \textit{Int. J. Eng. Sci. Technol.}, vol. 3, no. 5, pp. 3792–3797, 2011.

[10] A. Shukla, H. M. Pandey, and D. Mehrotra, “Comparative review of selection techniques in genetic algorithm,” \textit{2015 1st Int. Conf. Futur. Trends Comput. Anal. Knowl. Manag. ABLAZE 2015}, no. February, pp. 515–519, 2015.

[11] “Fuzzy Logic Laboratorium Linz-Hagenberg,” 2004.

[12] “Genetic algorithms in search, optimization, and machine learning,” \textit{Choice Rev. Online}, vol. 27, no. 02, pp. 27-0936-27–0936, 1989.

[13] U. A.J. and S. P.D., “Crossover Operators in Genetic Algorithms: a Review,” \textit{ICTACT J. Soft Comput.}, vol. 06, no. 01, pp. 1083–1092, 2015.

[14] D. Sastry, K., & Goldberg, “Genetic algorithms,” pp. 97–105, 2005.

[15] I. G. Agung, M. Wibhu, and C. Asrawan, “Analisis Optimasi Dengan Algoritma Genetika Pada Audio Watermarking Berbasis Discrete Wavelet Transform,” vol. 4, no. 1, pp. 457–471, 2016.

[16] Setiono and S. Marwoto, “Pemodelan Logika Fuzzy Terhadap Kerusakan Jembatan Beton,” \textit{Media Tek. Sipil}, vol. X, no. 1, pp. 28–35, 2010.

[17] A. Saelan, \textit{Logika Fuzzy}, vol. 1. 2009.

[18] Prihatini, “Metode Ketidakpastian dan Kesamaran dalam Sistem Pakar,” 2011.

[19] L. W. Trimartanti, “Penerapan Sistem Fuzzy Untuk Diagnosis Campuran Bahan Bakar Dan Udara Pada Mobil F15 Gurt,” \textit{Fak. Mat. dan Ilmu Pengetah. AlamUniversitas Negeri Yogyakarta}, pp. 7–37, 2011.

[20] Suyanto, \textit{Data Mining}. Bandung: Informatika Bandung, 2017.