Increasing Prediction the Original Final Year Project of Student Using Genetic Algorithm

Rijois Iboy Erwin Saragih ¹, Mardi Turnip ², Delima Sitanggang ², Mendarissan Aritonang ¹, Eva Harianja ¹

¹Universitas Methodist Indonesia, Jl. Hang Tuah No. 8 Medan, 20152, Indonesia
²Faculty of Technology and Computer Science, Universitas Prima Indonesia, Indonesia

E-mail: 0105128001@methodist.ac.id

Abstract. Final year project is very important for graduation study of a student. Unfortunately, many students are not seriously doing their final projects. Many of students ask for someone to do it for them. In this paper, an application of genetic algorithms to predict the original final year project of a student is proposed. In the simulation, the data of the final project for the last 5 years is collected. The genetic algorithm has several operators namely population, selection, crossover, and mutation. The result suggest that genetic algorithm can do better prediction than other comparable model. Experimental results of predicting showed that 70% was more accurate than the previous researched.

1. Introduction
Higher Education has the goal of producing qualified graduates [1]. The best graduates are not created in a night. It required a long process, year by year, course by course to write final Year Project. Student’s final year project is one of the most important aspects of his/her engineering degree. Final year project is an important element as a final step of the education journey of a student [2]. A long process has to be passed by a student starting from the first semester until the last semester to get a proposal seminar and the defence time. The real problem which is faced by some universities in Indonesia is plagiarism of final year project of a student. Plagiarism, collusions and other forms of academic misconducts have always been regrettable but unavoidable aspect of academic existence [3]. This situation has become a habit among the students; even more it has already become a bad culture. Students are willing to pay someone to make their final year project. This practice has put a bad image on education. Not only student involves in this situation but also some lecturers that offer a service for final year project of a student. This situation has become a big problem for universities.

In refer to previous worked [4] used different method for the prediction original final year project of student. This system used the apriori algorithm that found the correlation between various factors. The output of this system was the accuracy prediction of the original final year. In this study, GPA, presence, practical work, and mentoring were used as input variables. Prediction accuracy using this system revealed about 50%. Therefore, in this research, approaches are done to predict the original final year of student with data collecting techniques. Data collecting is the solution to this serious problem. Data collecting is an essential step in the process of knowledge discovery in databases. Thus, it refers to extracting or finding knowledge from large amounts of data [5]. GPA, presence,
practical work, and mentoring variables are used to predict the original final year project of student. These variables can be obtained from historical data of the students in the last few years. From the explanation above, this research obtained the historical data of the students in the last few years by implementing data collecting techniques [6]. The purpose is to get new knowledge or information from those data regarding to the research problems. The same data will be used from previous researched on this paper with different algorithm which is genetic algorithm.

2. Methodology
There are several steps of process to solve optimization problems in genetic algorithm. The steps are: population, selection, crossover and mutation. Those steps are continually repeated until finding the expected result [7]. Prediction approaches are to obtaining the available data and then using genetic algorithm to get new information from the process [8-15]. The population consists of ten chromosomes or populations, the chromosome are ordered by gens which contain scores. In the Genetic Algorithm, the problem to be solved is represented by a list of parameters which drives an evaluation procedure, called chromosomes or genomes. Chromosomes can be defined as simple strings of data and instructions. In the first step of the algorithm, such chromosomes are generated randomly or heuristically to form an initial pool of possible solutions called first generation pool. The proposed algorithm can be described as [16]:

1. Firstly suppose we have a population of N size, with chromosomes generated randomly;
2. Apply fitness to each chromosome or genomes of population;
3. Make new chromosomes or genomes through crossings of selected chromosomes of this population. Apply recombination and mutation in these chromosomes;
4. Eliminate old population members, so that there is enough space to insert new chromosomes, keeping the population with the same N chromosomes;

Further explanation, a population of genetic algorithm also called chromosomes that are randomly distributed in the solution space is selected as the starting point of the search. The goal of fitness function is to numerically encode the performance of the chromosomes. Crossover causes to form a new offspring between two randomly selected good parents. In this researched is used one point crossover. The crossover occurs only with some probability. The last step is mutation, this operator play important role to ensures that the probability of reaching any point in the search is never zero [17].

2.1 Population
Genetic algorithm is generally started with an initial population that is generated randomly. Some researches have been conducted by using special techniques to produce a higher quality initial population. Thus, an approach is designed to give the GA a good start and speed up the evolutionary process [18].

| Table 1. Population. |
|----------------------|
| Population           |
| 10 6 5 4 1 7 3 8 2 9 |
| 10 9 7 1 3 8 5 2 4 6 |
| 5 8 2 10 3 9 6 1 7 4 |
| 4 7 1 9 3 2 6 5 8 10 |
| 9 2 10 8 7 3 5 4 6 1 |

Every score in the table is the chromosome gen which represents the students. For example number 10 is the student who is chosen before number 10. This method is based on data factor in all variables such as GPA, presence, practical work and mentoring.
2.2 Fitness
The fitness function makes certain that evolution forward to optimization with accumulating for each individual on population [19 - 21].

\[ F = V_1 + V_2 + V_3 + V_4 \]

Where,
- \( F \) = Fitness value
- \( V_1 \) = GPA
- \( V_2 \) = Presence
- \( V_3 \) = Practical Work
- \( V_4 \) = Mentoring

- **2.2.1 GPA**
  GPA is Grade Point Average in term in academic system. GPA plays an important role to evaluate the students.

| No | Scale       | Value |
|----|-------------|-------|
| 1  | 2.0 – 2.7   | 4     |
| 2  | 2.7 – 3.0   | 3     |
| 3  | 3.0 – 3.5   | 2     |
| 4  | 3.5 – 4.0   | 1     |

- **2.2.2 Presence (P)**
  A presence is an evidence which shows the students’ presence. This is also important because it shows the attitude of the students.

| No | Scale          | Value |
|----|----------------|-------|
| 1  | P < 40%        | 3     |
| 2  | 40% > P < 70%  | 2     |
| 3  | P > 70%        | 1     |

- **2.2.3 Practical Work (PW)**
  To know the students in computer study program have skills in computer by practical work. This practical work has to be done by students with good grade.

| No | Scale          | Value |
|----|----------------|-------|
| 1  | PW<60%         | 3     |
| 2  | 40% > PW<70%   | 2     |
| 3  | PW> 70%        | 1     |

- **2.2.4 Mentoring(M)**
Mentoring time must be followed by student so that lecturer or supervisor can lead them to the right way how to write a good paper, or how to do good research. The student must bring a card of mentoring that show their presence by lecturer signature.

| No | Scale                | Value |
|----|----------------------|-------|
| 1  | PW< 60%              | 3     |
| 2  | 40% > PW < 70%       | 2     |
| 3  | PW > 70%             | 1     |

3. Discussion and Result
For this step, it will be discussed about the process and the result of the increasing prediction of the original final year project using genetic algorithm. The parameters and values of genetic algorithm on this research are:
1. Population = 10
2. Generation = 100
3. Crossover probability = 0.7
4. Mutation probability = 0.1

| Experiment | Generation | Prediction | Reality |
|------------|------------|------------|---------|
| 1          | 99         | 7          | 5       |
| 2          | 92         | 9          | 8       |
| 3          | 87         | 10         | 7       |
| 4          | 70         | 11         | 7       |
| 5          | 95         | 9          | 9       |
| 6          | 90         | 7          | 9       |
| 7          | 85         | 12         | 5       |
| 8          | 90         | 7          | 5       |
| 9          | 89         | 10         | 7       |
| 10         | 95         | 7          | 5       |

On Table 6 is seen that prediction and the reality result over all the experiment are different. This experiment is done in ten times. The result shows that on generation 90 the prediction is 7 students but in reality is 9 students. It means that 2 score different compare to other generation which is average above 2 score. Therefore, percentage result between 7 and 9 is about 70% accurate predicted.

| Previous Researched | Recent Research |
|---------------------|-----------------|
| Implementation of Data Mining | Increasing Prediction the Original Final Year Project of Student |
| Association Rules to know the Original Final year Project of Student | Using Genetic Algorithm |
In the table 7 shows the comparison to previous work and recent work on the same context or data research. The different is seen in the prediction accuracy of 50% and 70%, it means that recent work shows a better result than the last research. This result is based on the prediction accuracy with the same variable and data. It took 10 times experiment with the same genetic algorithm parameter.

4. Conclusions

Based on the discussion and implementation which have been done on this research, therefore the conclusion shows that the implementation of genetic algorithm to predict the original final year project of student shows better result than previous research in term of time and accuracy. The result describes that generation = 90, prediction = 7, and reality = 9. Based on that result it means 7 to 9 differentiate 2 numbers which is about 70% accuracy predicted. The contribution from this research is that genetic algorithm with the proper parameter setting and formulated the problem can give good result in prediction system.

References

[1] De Corte, E., (2003), Excellence in Higher Education, Portland Press, London.
[2] Wanli X, Petakovic G R and Sean G 2015 Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. Computer in Human Behavior. Vol. 47 Pg. 168-181
[3] Carroll, and Appleton, J (2001) “Plagiarism: A Good Practice Guide”. JISC publication.
[4] Saragih R I E and Sembiring H 2015. Penerapan Data Mining Association Rules. MajalahIlmiahMethoda. Vol. 5 No. 2
[5] Sivanandam S.M, and Deepa, S.N. 2008. Introduction to Genetic Algorithms, Springer, Berlin:Germany.
[6] Guo X, Shao andHua N. 2009. A hierarchical Text clustering algorithm with cognitive situation Dimensions. Second International Workshop on Knowledge discovery and Data Mining
[7] David de la Fuente, Alejandro Garrido, Jaime Laviada, Alberto Gomez,(2006) “Genetic Algorithms to Optimize the Time to Make Stock Market Investment”. GECCO 2006, Vol.2, pp.1857-1858.
[8] Naik L R, Ramesh, D, Manjula B and Govardhan, A2012. Prediction of Stock Market Index Using Genetic Algorithm. Computer Engineering and Intelligent System. Vol. 3 No. 7
[9] Ratnakar S, Rajeswari K and Jacob R 2013Prediction Of Heart Disease Using Genetic Algorithm For Selection of Optimal Reduced Set Of Attribute. International Journal of Advance Computational Engineering and Networking. Vol. 1 Issue 2
[10] Ahmadi F, Amjadian S andPardegi, K. 2012. New Approach to Bankruptcy Prediction Using Genetic Algorithm. International Journal of Computer Application. Vol. 44 No. 4
[11] Guo X, Shao Z and Hua N. 2009. A hierarchical Text clustering algorithm with cognitive situation Dimensions. Second International Workshop on Knowledge discovery and Data Mining
[12] Hu J Xiong and Shu J Zhou X., Zhu J. 2009. A novel Text clustering method based on TGSOM and Fuzzy K-means “First International Workshop on Educationa Technology and Computer Science
[13] Zang Z and Cheng H. 2008. Clustering Aggregation based on genetic algorithm for Document clustering. IEEE Congress on Evolutionary Computation.
[14] Cyril Schoreels, Brian Logan, Jonathan M.Garibaldi. 2004“Agent based Genetic Algorithm Employing Financial Technical Analysis for Making Trading Decisions Using Historical Equity Market Data”, IAT04.
[15] L.S. Duke and Long, J., (1993) “Neural network futures trading - A feasibility study,” Adaptive Intelligent Systems, ElsevierScience Publishers.

[16] Reeves C, dan Rowe, J. (2002). Genetic Algorithms Principles and Perspectives. Kluwer, Norwell, MA.

[17] Naik R. L, Ramesh D, Manjula B, and Govardhan A. 2012 “Prediction of Stock Market Index Using Genetic Algorithm” Computer Engineering and Intelligent System. Vol. 3 No. 7

[18] K F Man, K S and Tang, S. Kwong, “Genetic Algorithms: Concept and Designs”, Springer, Chapter 1-10, Pg 1-348.

[19] Coley D A (1999). An introduction to genetic algorithms for scientists and engineers. World scientific.

[20] Smith J E 2002. Handbook of Global Optimization. Volume 2, 275-362

[21] Reeves C, and Rowe, J. (2002). Genetic Algorithms Principles and Perspectives. Kluwer, Norwell, MA. Coley, D. A. (1999). An introduction to genetic algorithms for scientists and engineers. World scientific.

Acknowledgments

Thanks to all the authors and the supervisors for all the times and supports.