

PAPER • OPEN ACCESS

ID3 algorithm approach for giving scholarships

To cite this article: Ali Rohman *et al* 2019 *J. Phys.: Conf. Ser.* **1175** 012116

View the [article online](#) for updates and enhancements.

You may also like

- [Analysis of data mining classification by comparison of C4.5 and ID algorithms](#)
R. Sudrajat, I. Irianingsih and D. Krisnawan
- [Comprehensive Gas Characterization of a \$z = 2.5\$ Protocluster: A Cluster Core Caught in the Beginning of Virialization?](#)
Jaclyn B. Champagne, Caitlin M. Casey, Jorge A. Zavala et al.
- [Establishment and Improvement of Financial Decision Support System Using Artificial Intelligence and Big Data](#)
Cong Si and Wenquan Shi

BREATH
BIOPSY

Breath Biopsy® OMNI®

The most advanced, complete solution for
global breath biomarker analysis

TRANSFORM YOUR
RESEARCH WORKFLOW



Expert Study Design
& Management



Robust Breath
Collection



Reliable Sample
Processing & Analysis



In-depth Data
Analysis



Specialist Data
Interpretation

ID3 algorithm approach for giving scholarships

Ali Rohman¹, Isturom Arif¹, Isnin Harianti¹, Anang Surohman Hidayat¹, Wellem Anselmus Teniwut², Nuraini³, Inderawati⁴, Hasan Baharun⁵, Nurhidayah⁶, Asrianti Dja'wa⁷, Upik Rahmi⁸, Jef Rudiantho Saragih⁹, Jajang Burhanudin¹⁰, Gumgum Gumelar¹¹, Tri Listyorini¹², Tuti Agustin¹³, Lala Nailah Zamnah¹⁴, Moh Yamin¹⁵, M. Taufik Qurohman¹⁶, Aji Raditya¹⁷, Reza Rachmadtullah¹⁸, Agustinus Suradi¹⁹

¹Department of Informatics Engineering, Universitas Maarif Hasyim Latif, Indonesia

²Tual State Fisheries Polytechnic, Indonesia

³Universitas Hamzanwadi, Indonesia

⁴Universitas Halu Oleo, Indonesia

⁵Universitas Nurul Jadid Paiton Probolinggo, Indonesia

⁶Universitas Al Asyariah Mandar, Indonesia

⁷Universitas Halu Oleo, Indonesia

⁸Universitas Pendidikan Indonesia

⁹Simalungun University, Indonesia

¹⁰Politeknik LP3I Bandung, Indonesia

¹¹Universitas Negeri Jakarta, Indonesia

¹²Universitas Muria, Indonesia

¹³Universitas Sebelas Maret, Indonesia

¹⁴Universitas Galuh Ciamis, Indonesia

¹⁵Universitas Lambung Mangkurat, Indonesia

¹⁶Politeknik Harapan Bersama, Indonesia

¹⁷Universitas Muhammadiyah Tangerang Indonesia

¹⁸Universitas Negeri Jakarta Indonesia

¹⁹Universitas Widya Dharma, Indonesia

ali.edutask@gmail.com

Abstract. Universitas Maarif Hasyim Latif (UMAHA) is a College that organizes scholarship programs for its students. However, in the implementation of the program, some problems occur, namely the University has difficulty in making decisions, and a large number of student data applying for scholarships causes a long time to select the process. So, we need a tool/system to help solve the problem. The purpose of this study is to apply the ID3 decision tree algorithm in a decision support system that can help facilitate the decision making of scholarships by the UMAHA. Referring to Pressman, this study uses the Waterfall Method. Waterfall method is a systematic and sequential software approach starting from analysis, design, coding, and testing. The test results by applying the ID3 decision tree algorithm indicate that the decision support system built can help the effectiveness and efficiency of the scholarship selection decision making the process by UMAHA, and the decision support system built has a high level of decision-making accuracy. For the development of further research, decision making can be based on more criteria such as student activity. Then in addition to academic achievement in this



case using the GPA of the previous semester using non-academic achievements such as achievements in the fields of sports, arts, music.

1. Introduction

In the current era the costs used for education are increasingly unreachable or more expensive, especially for those who have a middle to lower economy. One of the highest levels of education is higher education, with the highest level of education causing higher education costs. As a result for some underprivileged students this can be an obstacle in continuing to study in college so that students can often apply for academic or dropout leave. This is what makes Universitas Maarif Hasyim Latif (UMAHA) provide scholarship programs for outstanding students and students who are classified as having underprivileged economies. But to get a scholarship, the students must fulfill the rules set by the university with various criteria.

The criteria that have been set for this case study are the academic achievement index, semester, activeness in the campus organization, student transfer status, employment, total income which includes parent income and personal income when working, etc. Therefore, not all students who enroll in the scholarship program will be accepted by the university, only for students who meet certain criteria who will get the scholarship [1]–[3]. With the many existing criteria and data processing procedures students applying for scholarships conducted by the University include data collection, grouping, sorting, manual calculations or own estimates, and in the end the data is arranged in a number of report forms. So that the process of determining the scholarship award was made quite a long time.

Similar problems were also discussed in research in J.K's research. Using Data Mining Technique for Scholarship Disbursement which concludes that applications that are built based on the decision tree with ID3 algorithm can be applied to decision support systems in scholarships [4], [5]. Application testing that is built with the decision tree classification technique shows the results that the application is effective, efficient and able to overcome the problems of the existing system. Based on the considerations described above, the ID3 algorithm will be applied to the recommendation for scholarships at the UMAHA. Because the ID3 algorithm has the advantage of being able to extract hidden information in a set of data, dividing a set of data into smaller sets and the results of an analysis in the form of tree diagrams that are easy to understand [6], [7].

2. Research Method

The determination of scholarships at UMAHA is still done through manual calculation by the Dean of the student affairs department itself so that the service of the University to the students who apply for the scholarship requires a long time. This research was conducted at the UMAHA which is located at Jl. Ngelom Megare No 30 Sepanjang Taman Sidoarjo from June to July 2018. and the object is data from students of UMAHA by conducting interviews with the Dean of UMAHA to find out the criteria limitations of obtaining scholarships later used as a reference in collecting student data by giving questionnaires to students. then the use of data mining techniques to process student data that was previously obtained by questionnaire.

3. Result and Discussion

3.1. Data Description

In distributing questionnaires to each questionnaire that has been given to the respondents and it is expected that respondents can fill in the questions submitted on the questionnaire in accordance with the actual situation. Of the 150 questionnaires distributed to students, only 115 questionnaires can be used and can be processed into useful data for the continuation of this research and from the data 100 data will be used as testing data.

3.2. Data Mining

The steps to be taken in the data mining process are as follows:

- Data Cleaning. Data cleaning is done by selecting relevant student data for scholarship applicants in accordance with the criteria for determining scholarships.
- Data Integration. Integration or merging of data is done by combining data received from scholarship questionnaires at UMAHA, in the form of *NIM* data, name, address, gender, study program, semester, student transfer status, previous semester GPA, number of campus organizations followed, total income of parents, employment status. For some of the initial data, the questionnaire support system for scholarship decision support participants can be seen in figure 1.

No	NIM	Nama	Jenis Kelamin	Alamat	Jurusan	Semester	IPK	Organisasi	Penghasilan Orangtua	Pekerjaan	Penghasilan pribadi	Pindahan PT lain	Beasiswa
1	143213124	Rahmad Hidayat Indra	Laki-laki	Tanjung Trosobo Rt. 04 Rv. 05	Teknik Informatika	Semester 8	3.00	0	Rp3.000.000,00	Belum Bekerja	Rp0,00	Tidak	Belum memenuhi syarat
2	241216135	Angga Bagus S	Laki-laki	Jalan Raya Katagan	Manajemen	Semester 4	3.04	0	Rp4.000.000,00	Sudah Bekerja	Rp1.500.000,00	Tidak	Belum memenuhi syarat
3	143214519	Adiyono	Laki-laki	Dh. Kebon Anom Rt. 01 Rv. 07	Teknik Informatika	Semester 8	2.8	0	Rp2.700.000,00	Sudah Bekerja	Rp2.900.000,00	Tidak	Belum memenuhi syarat
4	242216016	Dewi Intan Sari	Perempuan	Dus. Panggang Rt. 10 Rv. 03	Akuntansi	Semester 4	3.63	0	Rp3.000.000,00	Belum Bekerja	Rp0,00	Tidak	Belum memenuhi syarat
5	242216039	Euphrasia Ferry D	Perempuan	Rusunawa Nyalom	Akuntansi	Semester 4	2.84	0	Rp0,00	Belum Bekerja	Rp0,00	Tidak	Belum memenuhi syarat
6	242216017	Linda Purwati N	Perempuan	Tanjungsari Doyorogo	Akuntansi	Semester 4	3.19	0	Rp0,00	Belum Bekerja	Rp0,00	Tidak	Belum memenuhi syarat
7	241216050	Ya Sinta Moi	Perempuan	Il. Maatip No.189 Karangpilang	Manajemen	Semester 4	3.18	0	Rp7.000.000,00	Belum Bekerja	Rp0,00	Tidak	Belum memenuhi syarat
8	143216901	Raiza Esa A.N.S	Perempuan	Il. Brijen Katamaso 3 No. 75	Teknik Informatika	Semester 8	3.24	0	Rp2.000.000,00	Belum Bekerja	Rp0,00	Tidak	Belum memenuhi syarat

Figure 1. Part of the initial table of participants in the SPK scholarship questionnaire

- Data Selection (Data Selection). Data in the database is often not all data used, therefore only the data that is suitable for analysis will be taken from the database. The data used are *NIM*, name, department, semester, GPA, organization, parental income, employment, personal income, transfer status from other universities, then on income attributes because there are two, it will be combined into one attribute that is the total income attribute. Whereas for sex data and addresses of students will be removed from the table because the two data from these attributes will not be used in the data mining process. So that the list of attributes used in the data mining process can be seen in table 1.

Table 1. Attribute selection.

Attribute	Using
<i>NIM</i>	✓
<i>Nama</i>	✓
<i>Jenis Kelamin</i>	X
<i>Alamat</i>	X
<i>Jurusan</i>	✓
<i>Semester</i>	✓
<i>IPK</i>	✓
<i>Organisasi</i>	✓
<i>Penghasilan Orangtua</i>	✓
<i>Pekerjaan</i>	✓
<i>Penghasilan Pribadi</i>	✓
<i>Pindahan dari PT lain</i>	✓

No	NIM	Nama	Jurusan	Semester	IPK	Organisasi	Total Penghasilan	Pekerjaan	Pindahan PT lain	Beasiswa
1	143213124	Rahmad Hidayat Indra	Teknik Informatika	Semester 8	3.00	0	Rp3.000.000,00	Belum Bekerja	Tidak	Belum memenuhi syarat
2	241216135	Angga Bagus S	Manajemen	Semester 4	3.04	0	Rp5.500.000,00	Sudah Bekerja	Tidak	Belum memenuhi syarat
3	143214519	Adiyono	Teknik Informatika	Semester 8	2.8	0	Rp5.600.000,00	Sudah Bekerja	Tidak	Belum memenuhi syarat
4	242216016	Dewi Intan Sari	Akuntansi	Semester 4	3.63	0	Rp3.000.000,00	Belum Bekerja	Tidak	Belum memenuhi syarat
5	242216039	Euphrasia Ferry D	Akuntansi	Semester 4	2.84	0	Rp0,00	Belum Bekerja	Tidak	Belum memenuhi syarat
6	242216017	Linda Purwati N	Akuntansi	Semester 4	3.19	0	Rp0,00	Belum Bekerja	Tidak	Belum memenuhi syarat
7	241216050	Ya Sinta Moi	Manajemen	Semester 4	3.18	0	Rp7.000.000,00	Belum Bekerja	Tidak	Belum memenuhi syarat
8	143216901	Raiza Esa A.N.S	Teknik Informatika	Semester 8	3.24	0	Rp2.000.000,00	Belum Bekerja	Tidak	Belum memenuhi syarat
9	143214507	Aditya Darmawan	Teknik Informatika	Semester 8	3.6	2	Rp6.500.000,00	Sudah Bekerja	Tidak	Belum memenuhi syarat
10	143214508	Azizah	Teknik Informatika	Semester 8	3.0	2	Rp6.800.000,00	Sudah Bekerja	Tidak	Belum memenuhi syarat
11	143214505	Rico Agung Prabowo	Teknik Informatika	Semester 8	2.98	0	Rp2.200.000,00	Sudah Bekerja	Tidak	Belum memenuhi syarat
12	143214520	M. Rifai	Teknik Informatika	Semester 8	3.10	0	Rp2.500.000,00	Sudah Bekerja	Tidak	Belum memenuhi syarat

Figure 2. Part table student data after data selection process.

- Data Transformation. Data that has previously been discarded or merged into the appropriate format hereinafter to be processed in the ID3 data mining decision tree. Because the data from

several attributes obtained from the questionnaire in the form of numerical numbers, it is necessary to change the attributes into several intervals. In this case the organization needs to be changed into two attributes, namely between those who are active in organizations if students follow at least 1 or more organizations that are at UMAHA or who are not active in the organization if the students do not follow the organization at all, while for semester attributes there are many attributes will be divided into 3 categories, namely the initial semester which represents 1st semester, mid semester representing 2nd semester, 3rd semester, 4th semester, 5th semester, and also 6th semester, as well as the final semester representing 7th semester and 8th semester, for categories the GPA attribute will be divided into two categories, namely GPA 3.00 and above, or GPA below 3.00, and the attributes of total income are also divided into 2 categories, namely the total income below 3 million and also the total income of 3 million and above. So after the data transformation process is carried out to produce categories that can be useful to facilitate the data mining process later because the above attributes have been categorized according to their respective values. For the results of the student questionnaire data table, the data transformation can be seen in figure 3.

No	NIM	Nama	Jurusan	Semester	IPK	Organisasi	Total Penghasilan	Pekerjaan	Pindahan PT lain	Beasiswa
1	143213124	Rahmad Hidayat Ind	Teknik Informatika	Semester akhir	3.00 Keatas	Tidak Aktif	3 Juta Keatas	Belum Bekerja	Tidak	Belum memenuhi syarat
2	241216135	Angga Bagus S	Manajemen	Semester pertengahan	3.00 Keatas	Tidak Aktif	3 Juta Keatas	Sudah Bekerja	Tidak	Belum memenuhi syarat
3	143214519	Adyoso	Teknik Informatika	Semester akhir	Dibawah 3.00	Tidak Aktif	3 Juta Keatas	Sudah Bekerja	Tidak	Belum memenuhi syarat
4	242216016	Dewi Intan Sari	Akuntansi	Semester pertengahan	3.00 Keatas	Tidak Aktif	3 Juta Keatas	Belum Bekerja	Tidak	Belum memenuhi syarat
5	242216039	Euphrasia Fenny D	Akuntansi	Semester pertengahan	Dibawah 3.00	Tidak Aktif	Di bawah 3 Juta	Belum Bekerja	Tidak	Belum memenuhi syarat
6	242216017	Linda Purwati N	Akuntansi	Semester pertengahan	3.00 Keatas	Tidak Aktif	Di bawah 3 Juta	Belum Bekerja	Tidak	Belum memenuhi syarat
7	241216050	Ya Sinta Moi	Manajemen	Semester pertengahan	3.00 Keatas	Tidak Aktif	3 Juta Keatas	Belum Bekerja	Tidak	Belum memenuhi syarat
8	143216901	Raiza Esa A.N.S	Teknik Informatika	Semester akhir	3.00 Keatas	Tidak Aktif	Di bawah 3 Juta	Belum Bekerja	Tidak	Belum memenuhi syarat
9	143214507	Aditya Darmawan	Teknik Informatika	Semester akhir	3.00 Keatas	Aktif	3 Juta Keatas	Sudah Bekerja	Tidak	Belum memenuhi syarat
10	143214508	Azzizah	Teknik Informatika	Semester akhir	3.00 Keatas	Aktif	3 Juta Keatas	Sudah Bekerja	Tidak	Belum memenuhi syarat
11	143214505	Rico Agung Prabowo	Teknik Informatika	Semester akhir	Dibawah 3.00	Tidak Aktif	Di bawah 3 Juta	Sudah Bekerja	Tidak	Belum memenuhi syarat

Figure 3. Part table student data after data transformation process.

- Mining Process. In the mining process, what will be done is to form a decision tree rule with ID3 algorithm, then the Entropy and Information Gainnya are calculated. The following is a tree that is formed after the calculation process of Entropy values and Information Gain of 100 student data with 13 positive resolution data and 87 negative resolution data.

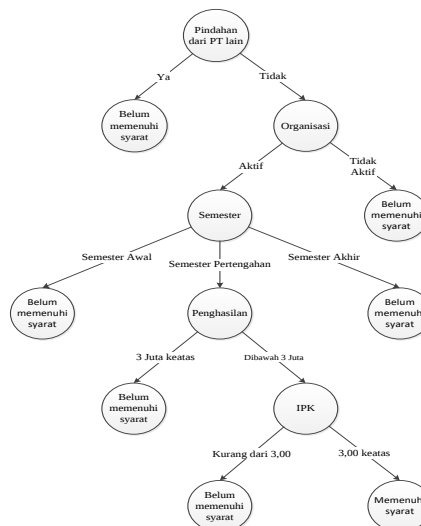


Figure 4. Rule Decision Tree ID3 For SPK Scholarships

3.3. Data Mining

Rule accuracy testing is done to determine the accuracy of the decision results by the rule in the decision support system. In this study the authors tested 100 data on student training for scholarship applicants using the RapidMiner application. Following are the results of testing decision support systems.

☒ Table View
 ☐ Plot View

accuracy: 99.00% +/- 3.00% (micro average: 99.00%)

	true Belum memenuhi syarat	true Memenuhi syarat	class precision
pred. Belum memenuhi syarat	86	0	100.00%
pred. Memenuhi syarat	1	13	92.86%
class recall	98.85%	100.00%	

Figure 5. Rule ID3 accuracy calculation results.

4. Conclusions

Based on the results and discussion of the Decision Support System for Giving Scholarships at Maarif Hasyim Latif University, it can be concluded that the Decision Support System built has a high level of decision-making accuracy, and the decision tree with ID3 algorithm can be used to gain knowledge in the field of education especially in giving scholarship.

References

- [1] G. N. Rayasad, "Association Rule Mining In Educational Recommender Systems."
- [2] J. Mesarić and D. Šebalj, "Decision trees for predicting the academic success of students," *Croat. Oper. Res. Rev.*, vol. 7, no. 2, pp. 367–388, 2016.
- [3] I. A. Khan and J. T. Choi, "An application of educational data mining (EDM) technique for scholarship prediction," *Int. J. Softw. Eng. Its Appl.*, vol. 8, no. 12, pp. 31–42, 2014.
- [4] J. K. Alhassan and S. A. Lawal, "Using Data Mining Technique for Scholarship Disbursement," *World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng.*, vol. 9, no. 7, pp. 1748–1751, 2015.
- [5] R. Rahim *et al.*, "Prototype File Transfer Protocol Application for LAN and Wi-Fi Communication," *Int. J. Eng. Technol.*, vol. 7, no. 2.13, pp. 345–347, 2018.
- [6] M. Balamurugan and K. Viji, "Predicting Eligible Educator Category for Disability Student Welfare using Decision Tree Method," *Int. J. Comput. Appl.*, vol. 119, no. 12, 2015.
- [7] Q. A. Al-Radaideh, A. Al Ananbeh, and E. Al-Shawakfa, "A classification model for predicting the suitable study track for school students," *Int. J. Res. Rev. Appl. Sci*, vol. 8, no. 2, pp. 247–252, 2011.