Research article

Using Education Data Mining (EDM) and Tracer Study (TS) Data as Materials for Evaluating Higher Education Curriculum and Policies

Muhammad Arifin[1,2]*, Widowati[1], and Farikhin[1]

[1]Doctoral of Information System Department, Universitas Diponegoro Semarang, Indonesia
[2]Universitas Muria Kudus, Indonesia

Abstract.
The policy-making process is influenced by the results of a good evaluation process. However, to get valid results, the evaluation process requires supporting data and technology. Higher education produces huge amounts of data but is still lacking in turning it into useful information. With the rapid development of information technology, higher education data must be processed into information that can be used as an evaluation and basis for making quick and accurate policies. Educational Data Mining (EDM) is a technique used to measure students' academic achievement to improve their results, assess the learning process, evaluate the overall quality of education, etc. Besides, the relevance of the curriculum followed by a university and the competencies possessed can be analyzed using a tracer study (TS). The current study uses EDM to extract information from higher education data and tracer study results. The relationship between higher education data and TS data was analyzed. The results indicated that not all students with higher GPA scores get jobs per the competencies they had in college, however, those with higher GPAs and who were actively organized during college are easier to find work. The study results must be considered as material for curriculum evaluation and for making higher education policies.

Keywords: higher education, Educational Data mining development, Tracer Study, curriculum evaluation, policy making

Corresponding Author:
Muhammad Arifin; email:
arifin.m@umk.ac.id

1. Introduction

The development of technology, information and communication has made management perspectives and practices in all fields, including education. Information technology is revolutionizing from traditional education to modern education in the learning process and management. In terms of education management, the use of information technology-based management is very important so that the policies taken can be in accordance with the needs based on precise and accurate data. Conformity between needs and policies issued by management will bring education to the achievement

of the goals set. Technology and information as a data collection base have a very important role in supporting the achievement of educational goals.

Data in Education is very abundant, this is because every day there are activities that always produce data. But there are still few institutions that turn data into information that is useful for educational purposes [1]. If an educational institution does not get quality data and information support, it will certainly experience obstacles and difficulties, especially in the strategic decision-making process. This situation ultimately results in the failure of educational goals. Therefore, educational institutions must have a good, precise and accurate data collection system in order to provide good quality services and as a basis for decision making[2]. Ideally, policy making requires a valid data source and is easy to read. Valid and easy-to-read data is required as a consideration used in any time-limited policy making or decision-making. Therefore, the role of data in policy making is very important. If a policy is formulated without data, it can be ascertained that the policy will not be able to become a problem solver, but instead become a new problem.

Educational Data Mining (EDM), it is one of the methods used to evaluate education and learning processes[3]. In addition, EDM can also be used to improve the quality of education (1,2,4). To get good quality learning and institutional achievement required feedback from students. This feedback is one of the most powerful influences on learning and academic performance[4]. One aspect of higher education success is the relevance aspect of the curriculum. In this aspect of relevance, universities are required to be able to produce graduates who have competitiveness and are ready to take part in the business world and the industrial world and take a role in development.

The competitiveness of graduates is demonstrated through the waiting period for getting the first job, the success of graduates competing in selection, and the salaries earned. The competitiveness of graduates is also demonstrated through success in building a business or hiring others. The relevance (suitability) of this graduate's education is demonstrated through the job profile (sort and place of employment), the relevance of the job to the educational background, the benefits of programmed courses in employment, as well as graduate advice for the improvement of graduate competence[5].

In addition, the relevance of education is also shown through the opinions of graduate users about graduate user satisfaction, graduate competence and graduate user advice for the improvement of graduate competence. How much college graduates are able to take part in the business world and the industrial world and participation in development

that is in accordance with the relevance of education can be done search efforts on graduates (tracer study)[6].

Tracer Study is an approach that allows colleges to obtain information and feedback about deficiencies that may occur in the education and learning process. Tracer study (TS) data collected and processed can be used to measure the academic performance of an institution[7]. Based on the above, the research question in this study is how to process EDM and TS data into material (information) that can be used for curriculum evaluation and higher education policy making.

2. Methodology

2.1. Data

Data is a collection of facts to give a broad picture of a situation. A person who will take a policy or decision will generally use the data as a consideration. Through data one can analyze, describe, or explain a situation. In addition, data is collected through certain ways and then processed into information that is clear and easy for everyone to understand. Data is also needed in various purposes, such as sales, research to population. Data has its own type and function; this is in accordance with the needs of the data manager. In general, data is widely used for a particular study. However, over time, data is needed to meet various needs in various fields. Of course, this aims to provide clear and correct information after the data is processed further.

The data used in this study is data from one of the universities in Central Java, Indonesia. The data used is 3,231 data, and the data is graduate data in the period 2018-2019 which is then called EDM data. While the TS data of the 2018-2019 graduate period amounted to 1,896. EDM data has variables; year of entry, student parent number, name, study program, GPA, Organizational status. While the TS data has variables; study program code, year of graduation, student parent number, name, mobile phone number, email, job status, waiting period to get a job, workplace, type of workplace, workplace name, competency suitability, required level of education, competence, etc.

2.2. Design

Figure 1 shows the methods of retrieving and processing EDM and TS data to obtain information that can be used for curriculum evaluation and higher management policies.
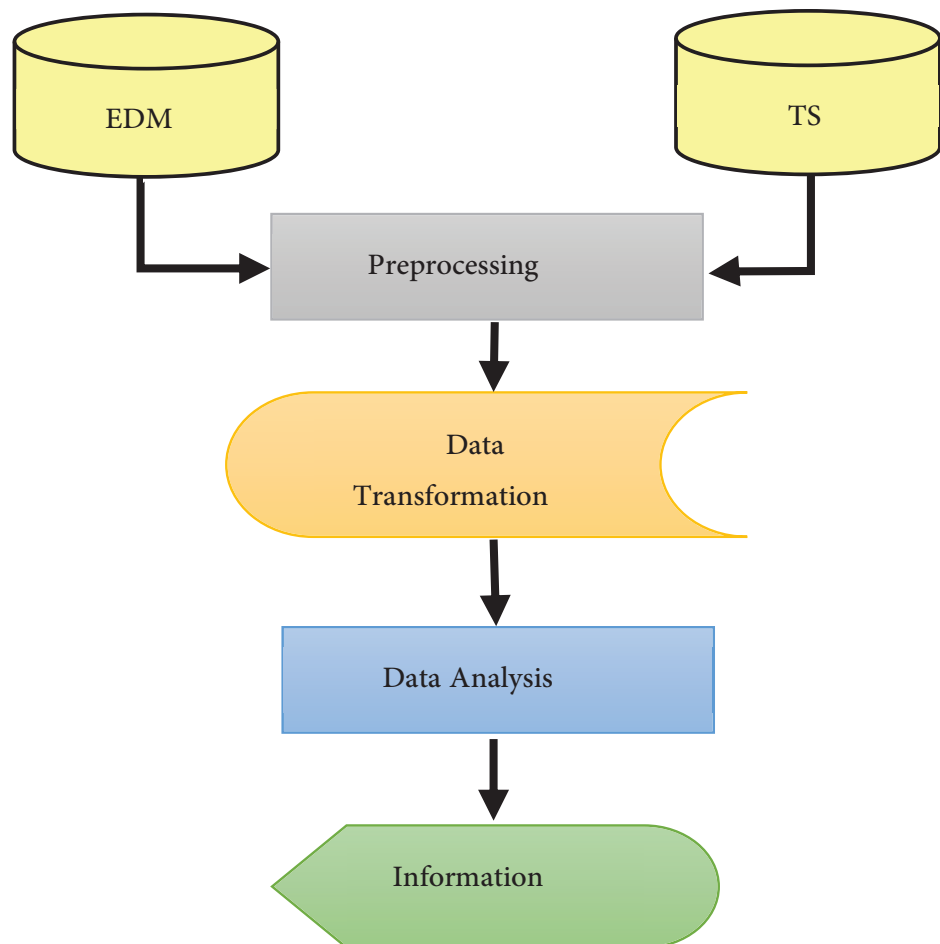
Figure 1: Design Data extract process to obtain higher education information.

2.3. Preprocessing

At this stage consists of several processes, namely Data Cleansing, the process where incomplete, error-containing and inconsistent data is discarded from the data collection. Also know data lifecycle management to find out about data processing. Data Integration is the process of data integration where repeating will be combined. Selection process or selection of data relevant to the analysis to be received from the existing data collection. Data Transformation, the process of transforming data that has been selected into the form of mining procedures through means and aggression of data.

2.4. Data Transformation

Data transformation is the stage at which data is transformed and consolidated into a form suitable for mining[8]. According to [9] in data transformation, there are several operations / techniques to perform data transformation, namely smoothing, attribute

construction, normalization, aggregation, and discretization. Smoothing operation is used to overcome data is noise / values that are not valid for the mining process with regard to the values of the company [9]. Attribute Construction is the process of constructing or adding new attributes to improve the accuracy of the mining process. Normalization is the process of grouping attributes into relationships that are well structured and free from anomalies [10]. Normalization is used to transform a scaled numeric attribute in a smaller range such as -1.0 to 1.0. The technique used for this operation is Z-Score Normalization. Aggregation is an operation for summary that is used for numerical data by using roll up operations. Discretization is used to reduce a set of values contained in a continuous attribute, by dividing the range of attributes into intervals. The operation used in discretization is Binning.

2.5. Data Analysis

Data analysis is the process of processing data with the aim of finding useful information that can be used as a basis in decision making for the solution of a problem. This analysis process includes grouping data based on its characteristics and creating data models to find important information from that data.

2.6. Information

Based on the understanding of information according to experts can be concluded that information is a set of facts that have been processed into a form of data, so it can be more useful and can be used by anyone who needs these data as knowledge or can be used in decision making. Information can be said to be knowledge gained from learning, experience or instruction. In computer science, information is data that is stored, processed or transmitted. Experts examine the concept of information as knowledge gained from learning, experience and instruction. From other notions of information is data that has been given meaning.

3. Result and Discussion

Preprocessing stage starts from the merger of EDM data and TS data. EDM data that is not in the TS data is eliminated, this is because the data needed is a combination of these data. The process of merging and trimming data generated 1,896 data. Furthermore, the process of cleaning data from missing value, remove duplicate or irrelevant

observations, fix structural errors, Filter unwanted outliers, handle missing data, Validate and QA. After cleaning the data in addition to the process of retrieving variables in accordance with the needs of the research. In this research the required variables are GPA, Active organization, job status, waiting period to get a job, and conformity of competence.

The next step is data transformation. At this stage starts from the data classification stage, where GPA data is divided into 3 classes, first class 2.00-2.75, second class 2.76-3.50 and final class 3.51-4 classification GPA[5]. The variable waiting period for getting a job is divided into 2 classes, namely before 6 months and after 6 months. Likewise, the active variables of the organization, job status and conformity of competence are each made into 2 classes, and are reformed by the rules if the data contains "No" then it is reformed to 0 and if the data is worth "Yes" then it becomes 1. The next preprocessing data is stored in memory which is then ready for analysis.

Data analysis process, at this stage the process of data analysis using SPSS software, analysis is used to look for the influence of each variable on the affected variables. In addition, the results of the analysis also show the number of data distributions in accordance with the categories on the affected variables. Data analysis also shows some correlations between variables. The results of data analysis are presented in the tables below.

GPA * Working Status Crosstabulation

			Working Status		Total
			No	Yes	
GPA	2,00-2,75	Count	105	285	390
		% within Working Status	26.4%	19.0%	20.6%
	2,76-3,50	Count	198	767	965
		% within Working Status	49.9%	51.2%	50.9%
	3,51-4,00	Count	94	447	541
		% within Working Status	23.7%	29.8%	28.5%
Total		Count	397	1499	1896
		% within Working Status	100.0%	100.0%	100.0%

Figure 2: Tabulation GPA with working status.

Table 1. showing that GPA 2.76-3.5 most get a job compared to the GPA below it or GPA above it. But there is a possibility that alumni who have GPA above 2.75-3.5 do not work because they continue study.

Table 2. It illustrates that not all students who actively organize during college after graduation quickly get a job, but the data above shows that students who are actively organizing during college they get a job twice when compared to students who are

KnE Social Sciences

Active Organization * Working Status Crosstabulation

				Working Status		Total
				No	Yes	
Active Organization	No	Count		137	512	649
		% within Active Organization		21.1%	78.9%	100.0%
	Yes	Count		260	987	1247
		% within Active Organization		20.9%	79.1%	100.0%
Total		Count		397	1499	1896
		% within Active Organization		20.9%	79.1%	100.0%

Figure 3: Tabulation Active organization with Working status.

not actively organizing. Thus, it can be concluded that students who actively organize during college have a greater 2 times the chance compared to students who are not actively organizing.

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Active Organization, GPA[b]	.	Enter

a. Dependent Variable: Working Status
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.079[a]	.006	.005	.406	.006	5.999	2	1893	.003

a. Predictors: (Constant), Active Organization, GPA

Figure 4: Correlation between variable.

Table 3.

Active Organization * Getting a job before 6 months Crosstabulation

				Getting a job before 6 months		Total
				No	Yes	
Active Organization	No	Count		209	440	649
		% within Active Organization		32.2%	67.8%	100.0%
	Yes	Count		436	811	1247
		% within Active Organization		35.0%	65.0%	100.0%
Total		Count		645	1251	1896
		% within Active Organization		34.0%	66.0%	100.0%

Figure 5: Tabulation Active organization with getting a job before 6 months.

Table 3. Shows the influence of active variables and GPA on work status. The significant value of F Change of 0.003. This shows that alumni in obtaining jobs are influenced by the activeness of organizing during college and GPA owned when graduating.

On Table 4. It is conveyed that actively organizing has a great opportunity in getting a job after graduation when compared to students who are not actively organizing, while in table 5. It was shown that actively organized students have a greater chance of getting a job in less than 6 months compared to students who are not actively organizing.

GPA * Getting a job before 6 months Crosstabulation

			Getting a job before 6 months		Total
			No	Yes	
GPA	2,00-2,75	Count	138	252	390
		% within GPA	35.4%	64.6%	100.0%
	2,76-3,50	Count	315	650	965
		% within GPA	32.6%	67.4%	100.0%
	3,51-4,00	Count	192	349	541
		% within GPA	35.5%	64.5%	100.0%
Total		Count	645	1251	1896
		% within GPA	34.0%	66.0%	100.0%

Figure 6: Tabulation GPA with Getting a job before 6 months.

GPA * Conformity with competence Crosstabulation

			Conformity with competence		Total
			No	Yes	
GPA	2,00-2,75	Count	84	201	285
		% within GPA	29.5%	70.5%	100.0%
	2,76-3,50	Count	220	547	767
		% within GPA	28.7%	71.3%	100.0%
	3,51-4,00	Count	100	347	447
		% within GPA	22.4%	77.6%	100.0%
Total		Count	404	1095	1499
		% within GPA	27.0%	73.0%	100.0%

Figure 7: Tabulation GPA with Competence.

In the calculation of GPA with competence, the data used is less than the previous data, this is because the data of alumni who do not work is not used in measuring the suitability of competence. Table 6. Showing that the amount of GPA does not greatly affect the suitability of competencies needed in their work. But based on the numbers, it is shown that GPA 2.76-3.5 works the most in accordance with the competencies possessed.

4. Conclusion

By extracting information from EDM and TS it can be found that students with high GPA scores do not always get jobs in accordance with the competencies they have in college, students with high GPAs and who are actively organized during college are easier to find work. Furthermore, the results of the research can be used as material for curriculum evaluation and in making higher education policies.

Acknowledgments

References

[1] Morris LV. Mining data for student success. Innovative Higher Education. 2016;41[3]:183–185. https://doi.org/10.1007/s10755-016-9367-6

[2] Abu SA, Al-Emran M, Shaalan K. Factors affecting students' performance in higher education: A systematic review of predictive data mining techniques. Technology, Knowledge and Learning. 2019;24(4); 567–598. https://doi.org/10.1007/s10758-019-09408-7

[3] Bakhshinategh B, Zaiane OR, ElAtia S, Ipperciel D. Educational data mining applications and tasks: A survey of the last 10 years. Education and Information Technologies. 2018;23[1]:537-53. https://doi.org/10.1007/s10639-017-9616-z

[4] Hattie J, Timperley H. The power of feedback. Review of Educational Research. 2007;77[1]:81–112. https://doi.org/10.3102/003465430298487

[5] Rahmawati E, Saputra O, Saftarina F et al. Hubungan gaya belajar terhadap indeks prestasi kumulatif (IPK) mahasiswa fakultas kedokteran universitas lampung association of learning styles with grade point average (GPA) medical students of Lampung University. Medula. 2018;8[1]:7–11.

[6] Albina AC, Sumagaysay LP. Employability tracer study of information technology education graduates from a state university in the Philippines. Social Sciences & Humanities Open. 2020;2[1]:1-6 https://doi.org/10.1016/j.ssaho.2020.100055

[7] Sira F, Bakar NAA. Identifying patterns of students academic performance from tracer evaluation using descriptive data mining. International Journal of Recent Technology

and Engineering. 2019;8[2]:187–91. https://doi.org/10.35940/ijrte.B1034.0782S219

[8] Han J, Kamber M, Pei J. Data mining. 3rd ed. Han J, Kamber M, Pei J, editors. Morgan Kaufmann; 2012. Data preprocessing;83-124 Netherland. https://doi.org/10.1016/B978-0-12-381479-1.00003-4

[9] Junaedi H, Budianto H, Maryati I, Melani Y. Data transformation pada data mining. Prosiding Konferensi Nasional Inovasi Dalam Desain Dan Teknologi-IDeaTech. 2011;7:93–9.

[10] Lee H. Justifying database normalization: A cost/benefit model. Information Processing & Management. 1995;31[1]:59–67. **https://doi.org/10.1016/0306-4573[95]80006-F**