Drop Out Students Identification Using Knowledge Base

Dara Kusumawati [1)*], Dini Faktasari[2)]

[1)2)]*STMIK AKAKOM*

Jln Raya Janti No 143, Yogyakarta, Indonesia

[1)]dara@akakom.ac.id

[2)]dini@akakom.ac.id

Abstract

The purpose of this study is modeling for the initial identification of college dropout students. Samples were taken from drop out student data for the past 4 years. Information from this sample will be acquired as a knowledge base in system modeling. The research aims to provide a knowledge-based system approach using Dempster Shafer for the management of student drop outs at universities especially in Yogyakarta. The symptoms of DO students are obtained from knowledge about DO that appears on campus in Yogyakarta. The system output is in the form of 3 groups of classification namely initial potential DO, enough potential and once potential. The results of the study produced a system that could help university managers deal with drop-out problems early.

I. INTRODUCTION

The main purpose of an education in general is to form good quality human beings. Human quality standards needed in the era of big data, especially Indonesia, are able to face increasingly fierce and faster competition with other countries. To produce the qualities of Indonesian people above, it is necessary to process management and organization of education that meets good standards and quality. Higher education which is one of the places for the development and formation of the nation's generation has many problems relating to students. The role of students in a tertiary institution is a basic substance that must be managed and considered properly. This is because students are closely related to their ability to translate dynamics and development of knowledge on an ongoing basis [1]. With the abilities possessed by students, students are potential subjects and are also objects in creativity improvement activities. This becomes an important point of students in developing self quality continuously. The quality parameters of students can be seen from the academic achievement every semester that shows the efforts of students [2].

Student academic achievement must be continuously evaluated to find out the learning outcomes that have been achieved each semester. Most academic problems that occur in tertiary institutions result in students resigning and dropping out. Determining the decision to drop out by a tertiary institution is not an easy thing to do because it must look at various academic variables or criteria clearly stipulated by the tertiary institution and be known to students from the beginning as a student. A decision to drop out from a tertiary institution can also cause problems at the tertiary institution because the level of the drop out phenomenon indicates the existence of a tertiary management process, especially in the academic field.

Dropout rates at tertiary institutions can be minimized through policies aimed at preventing student dropouts [3]. Form of prevention can be done by identifying or early detection of all students with symptoms that exist in a college. Detections can be made in the first year of all students who have the potential to drop out. The absence of an early detection system for students who have the potential to drop out can hamper tertiary institutions in implementing this policy. Yogyakarta, one of the education cities in Indonesia, has around 113 colleges, most of which are private universities. According to the Ministry of Education and Culture's higher education statistics in 2017, it showed that out of 372,480 students studying in Yogyakarta, there were as many as 12,425 students who had to drop out of college (DO).

Knowledge-based systems that are part of artificial intelligence can work like and as good as humans if the system is given good knowledge and reasoning skills. Knowledge-based systems have the ability to compute, store, think

processes, and store knowledge [4]. This system is more commonly referred to as an expert system. One method of uncertainty in an expert system is the dempster shafer method. The dempster shafer method can provide precise analysis and have stronger certainty without any changes or additions to its knowledge. The explanation above gives a very strong reason for the need to do a research on a knowledge-based system that can carry out early identification of students who have the potential to drop out at college, especially in private universities in the Kopertis V area using the Dempster Shafer method.

II. LITERATURE REVIEW

Topics and problems related to student academic failures, especially drop out, have been investigated by several researchers including research on the comparison between the Decision Tree algorithm, Bayesian classifiers, logistic models, rulebased learners and random forest to predict student dropouts with 648 student data sets with the result that the highest accuracy using a decision tree [5]. Another research conducted wanted to find out the factors that influence the dropout of students at tertiary institutions in East Java with the result that the factors of student intelligence and parents' income had a significant influence on the case of college drop-outs [6].

Other studies have compared four algorithms, namely logistic regression, decision tree, naive bayes, and neural network with the results showing that the decision tree is the most accurate algorithm, however the decision tree is not dominant to other algorithms. The logistic regression algorithm, decision tree, naïve bayes and neural network fall into the category of excellent classification [7]. Another study identified factors related to student dropouts using survival analysis with results that showed that the process of resigning students was influenced by the student's GPA. Faculty variables have different survival and hazard values but in general the values are almost the same so that the differences are not too large for each faculty [8].

Other research that also discusses drop out is identifying the potential drop out of students that is based on academic grades, student performance, and socio-demographic students, all of these factors come from students' internals. Whereas other factors also have a significant influence on the prediction of potential drop outs for students, one of which is social behavior [9]. Therefore in this study a new model is proposed to identify student knowledge-based drop outs using the Damspter Shaper method. Because the variables used are knowledge-based, it needs the role of the academic department to define the variables that cause students to drop out in college.

A. Expert System

Expert systems (expert systems) are systems that seek to adopt human knowledge into computers, so that computers can solve problems as is usually done by experts, and good expert systems are designed to be able to solve a particular problem by imitating the work of experts. The expert system was first developed by the AI community in the mid 1960s. The expert system that first appeared was the General Purpose Problem Solver (GPS) developed by Newel & Simon.
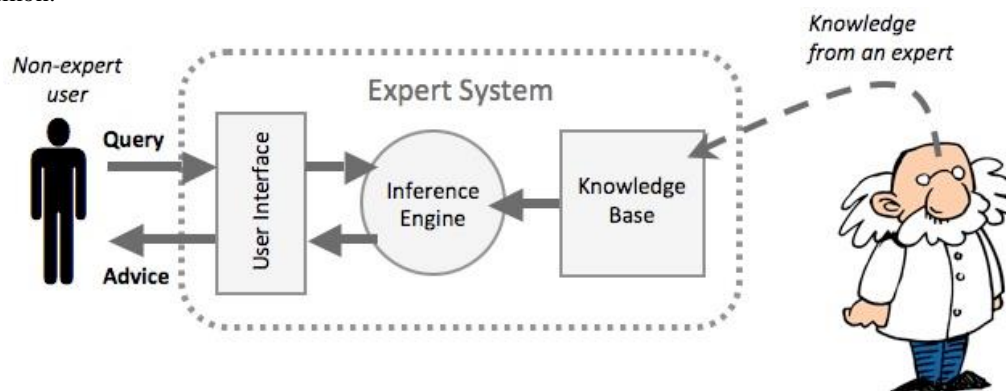

Figure 1. Expert System Component

Expert system itself is a software package or computer program package intended as a provider of advice and aids in solving problems in certain areas of specialization such as science, engineering, mathematics, medicine, education and so on. Expert systems are a subset of Artificial Intelligence. Basically an expert system is applied to support problem solving activities. Some of the problem solving activities in question:

1. Interpretation. Make conclusions or descriptions of raw data sets. Decision making from observations, including speech recognition, image analysis, signal interpretation, etc.
2. Predictions. Project the possible consequences of certain situations. Example: demographic predictions, economic predictions, etc.
3. Diagnosis. Determine the cause of malfunction in complex situations based on symptoms observed in medical, electronic, mechanical diagnoses, etc.
4. Design (design). Determine the configuration of system components that match specific performance goals that meet certain constraints. Example: circuit layout design, building.
5. Planning. Plan a series of actions that will be able to achieve a number of objectives with certain initial conditions. Example: financial planning, military, etc.
6. Monitoring. Comparing observations with expected conditions. Example: computer aided monitoring system.
7. Debugging. Determine and interpret ways to deal with malfunctions. Example: prescribe medication for failure.
8. Instructions. Detect and correct deficiencies in understanding the subject domain. Example: perform instructions for diagnosis and debugging.
9. Control. Manage the behavior of a complex environment. Example: control the interpretation, prediction, improvement and monitoring of system behavior.

B. Dempster Shafer

There are 2 important things in the Dempster-Shafer method, namely the belief function (trust function) and plausible reasoning (reasonable thinking). These two things are used to combine separate pieces of information (evidence) to calculate the probability of an event. The Dempster-Shafer theory is based on two ideas, namely the idea of obtaining degrees of trust from various subjective possibilities and the Dempster-Shafer rules themselves to combine degrees of trust based on the evidence obtained.

In general the Dempster-Shafer theory is written at intervals:

[*Belief*,*Plausibility*]

Belief (Bel) is a measure of the strength of evidence in supporting a set of propositions. If it has a value of 0 (zero) then it indicates that there is no evidence, and if it has a value of 1 indicates there is certainty.

Formula Giarratano dan Riley function *belief* can define is :

$$Bel(X) = \sum_{Y \subseteq X} m(Y) \qquad \text{...(1)}$$

whereas Plausibility (Pls) is denoted as :

$$Pls(X) = 1 - Bel(X') = 1 - \sum_{Y \subseteq X'} m(X') \qquad \text{..(2)}$$

where:

Bel(X) = *Belief* (X)
Pls(X) = *Plausibility* (X)
m(X) = *mass function* dari (X)
m(Y) = *mass function* dari (Y)

Some possible ranges between Belief and Plausibility are in Table 1.

Tabel 1. *Range Belief* dan *Plausibility*

Kemungkinan	Keterangan
[1,1]	Semua Benar
[0,0]	Semua Salah
[0,1]	Ketidakpastian
[Bel,1] where 0 < Bel < 1	Cenderung Mendukung
[0,Pls] where 0 < Pls < 1	Cenderung Menolak
[Bel,Pls] where 0 < Bel ≤ Pls < 1	Cenderung Mendukung dan Menolak

In the Dempster-Shafer theory also known as the existence of a frame of discernment denoted by. This FOD is the universe of discussion from a set of hypotheses so that it is often referred to as the environment,

With function $\Theta = \{\theta 1, \theta 2, ... \theta n\}$

where:

$\Theta =$ FOD atau *environment*

$\theta 1....\theta n =$ elemen/unsur bagian dalam *environment*

The environment contains elements that describe possibilities as answers and there is only one that will match the answers needed. This possibility in Dempster-Shafer theory is called the power set and is notated with P (). each element in this power set has an interval value between 0 to 1.

$$m = P(\Theta) \longrightarrow [0,1]$$

III. METHODS

This research makes a framework model for early identification of dropout students. The system was developed using a knowledge base approach from data and dropout student problems in Indonesia, especially in Yogyakarta. The system was built using a web programming language so that it will be easier for users to identify knowledge-based student drop outs [10]. The system model that will be built can be seen in Figure 2. The processed knowledge comes from 3 external factors, namely tertiary institutions, students and the environment that is around students and the location of higher education. All data from external factors are collected into complete information. From this complete information, knowledge acquisition is carried out to get basic knowledge about the variables that cause drop out in tertiary institutions.

The knowledge that is processed comes from 3 external factors, namely Universities, students and the environment that is around students and location of higher education. All data from external factors are collected into complete information. From this complete information, knowledge acquisition is carried out to get basic knowledge about the variables that cause drop out in tertiary institutions. The next stage is to make a rule model that consists of variables that greatly influence the drop out and drop out classification. In this model, the drop out class is defined as high potential drop out, medium drop out potential and low drop out potential. Each class has different rules.
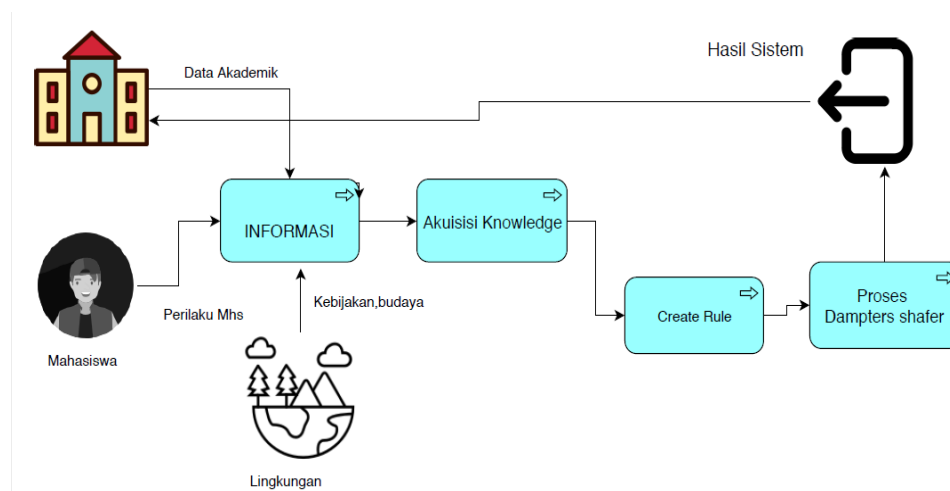

Figure 2. Identification Model System Drop Out

The relationship between external factors represented by actors with all system functionalities is shown in Figure 2. There are 2 main actors of the system namely the admin who manages the existing system and data management and the second actor is the college part represented by the ADAK section. The system is described in the form of a number of system functionalities consisting of 6 usecases namely login, input symptoms / DO variables, input categories, input data or load student data, process a dampster shafer and display the results.
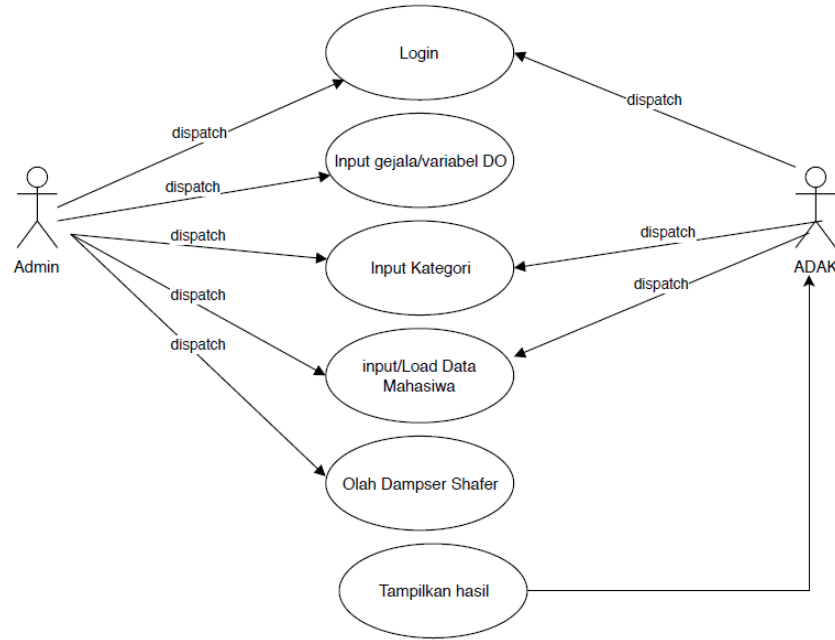
Figure 3. Usecase Diagram System

IV. RESULTS

The implementation of the usecase diagram of Figure 3 in the study contained several functionalities that are presented in Figure 4 through Figure 6. All the functionality that is on the system is displayed in the student dashboard shown in Figure 4.
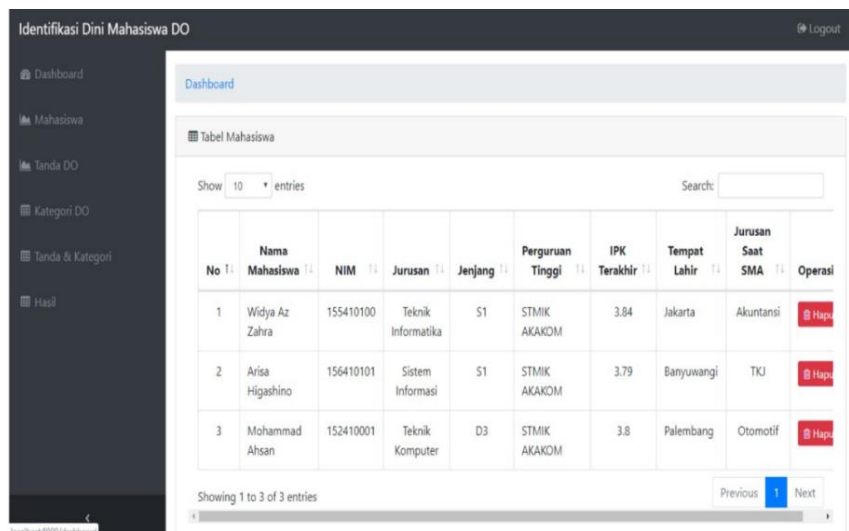

Figure 4. Dashboard Sistem

On this dashboard there are several function menus that can be easily accessed by the user. The first function is managing student master data. This data becomes a very important part because it will be managed to identify drop outs that occur at tertiary institutions. Student master data is used as a basic database for the DO identification system which must be complete, this data can be taken from the academic data of students in the ADAK section of tertiary institutions. Other master data besides students are the DO category shown in Figure 5. The campus can update the data related to DO symptoms. There are 3 categories: High Potential, Medium Potential and Low Potential.
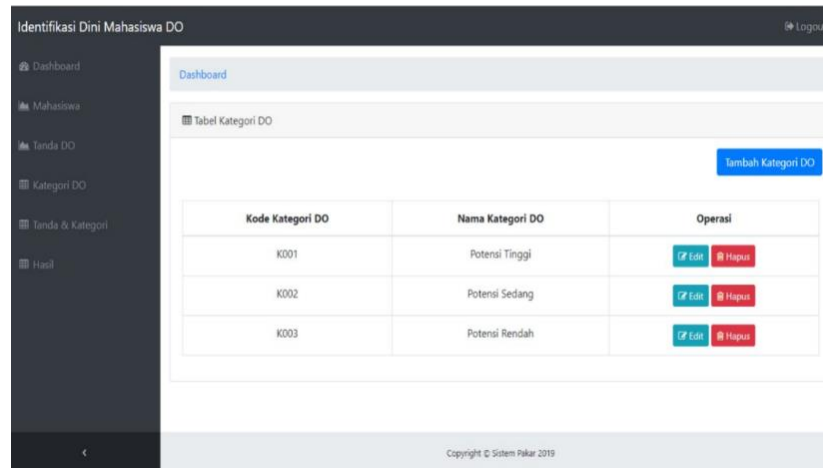
Figure 5. Category Drop Out Management

After the sign data and DO categories are completed, the next step is to give weight to each predetermined rule. Giving weights will determine the results of identification carried out by the system. The weights entered for each symptom will give the final result of the calculation of the dempster shafer value. The weight data entered by the university becomes the basis for identification of student drop outs at a college. The weight value for each university is different because it depends on the case in that college. The weight value is obtained from the academic section based on data available at each college.
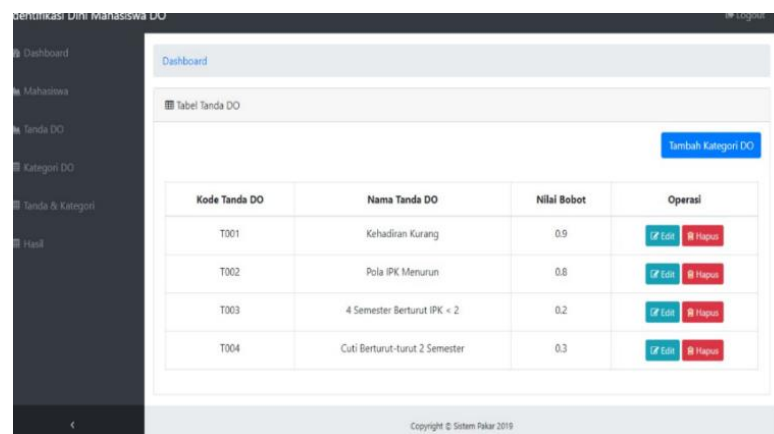

Figure 6. Manajemen Bobot Rule

The data in table 2 will have different values for each tertiary institution because the problems that arise are different. This situation shows that rule modeling will determine the results of early identification of students who have the potential to drop out.

Table 2. Weights Symptoms and rules

TANDA	KATEGORI			BOBOT
	POTENSI TINGGI	POTENSI SEDANG	POTENSI RENDAH	
Kehadiran Kurang		X	X	0,7
IPK dengan Tren Turun (3 semester berturut-turut)	X			0,8
IPK dibawah standar 2 semester berturut-turut	X	X		8,8
Cuti Ijin 2 semester		X	X	0,6
Pelanggaran Kode Etik	X			0,5
Masa Studi Maksimal tinggal 1 semester	X	X		0,8
Nilai D dan E masih 30% setiap semester	X			0,9
Cuti tanpa ijin 2 semester	X	X	X	0,7
Kegiatan Organisasi dominan (4 semester)		X	X	0,4
Kuliah sambil Kerja (lbh dari 3 semester)	X	X		0,6
Salah pergaulan	X	X	X	0,7
Tidak bisa menyesuaikan kegiatan kampus		X	X	0,5

The results of the dempster shafer method calculation are shown in Figure 7. These results will show how much the value of a student has the potential dropout value at a tertiary institution because of some of the signs or symptoms that exist.
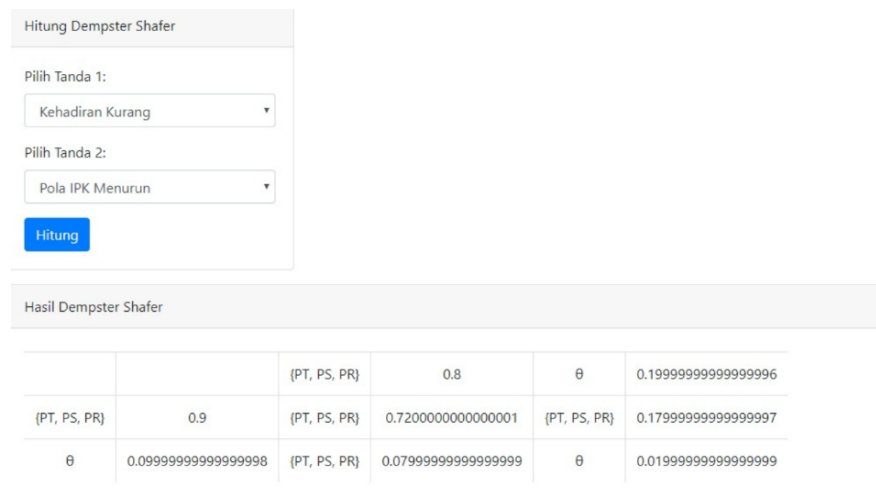

Figure 7. Dempster Shafer Menu

V. DISCUSSION

The results of the identification of students dropping out at a knowledge-based tertiary institution using the Dempster Shafer method show that the variables that influence the identification results are the symptoms that appear, the weight of each symptom and the rules that contain the symptoms. The identification results will produce choices consisting of high potential, medium potential and low potential. The weight inputted by the academic department is based on the results of data processing of a sample of students dropping out of 100 data from one of the universities

in Yogyakarta. In this study accuracy is not counted because the dempster shafer method is not part of machine learning so there is no prior learning process.

VI. CONCLUSIONS

Some things that are made conclusions from the results of research, including:
1. Dempster Shafer can be used as an alternative method in the knowledge-based system to identify student drop outs.
2. Variables that can affect the results of identification include symptoms, the weight value of each symptom and the rules that appear for the target class.
3. Identification is divided into 3 classes, namely high potential, medium potential and low potential.

VII. REFERENCES

[1] S. Harahap, Harahap, Penegakan moral akademik di dalam dan di luar kampus, RajaGrafindo Persada, 2006.

[2] A. Sobur, Psikologi Umum, Bandung: Pustaka Setia, 20016.

[3] G. W. Dekker, M. Pechenizkiy and J. M. Vleeshouwers, ""Predicting students drop out: A case study," EDM'09 - Educ. Data Min," in *2nd Int. Conf. Educ. Data Min*, 2009.

[4] J. Bayer, H. Bydzovská, J. Géryk, T. Obsivac and L. Popelínský, "Predicting drop-out from social behaviour of students," in *Proc. 5th Int. Conf. Educ. Data Min*, 2012.

[5] D. Goldblatt, "Foreword," *Semin. Neurol,* vol. 14, no. 1, pp. 241-249, 1994.

[6] M. M. Hidayat, D. Purwitasari and H. Ginardi, "Analisis Prediksi Drop Out Berdasarkan Perilaku Sosial Mahasiswa Dalam Educational Data Mining," *J. IPTEK,* vol. 17, no. 2, pp. 109-119, 2013.

[7] F. Imran, B. Susetyo and H. A. Wigena, "Identifikasi Faktor-Faktor Yang Berhubungan Dengan Mahasiswa Putus Kuliah Di Ipb Angkatan 2008 Menggunakan Analisis Survival," *Xplore: Journal of Statistics,* vol. 1, no. 2, 2013.

[8] A. Andriani, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," in *Semin. Nas. Mat*, 2012.

[9] T. F. Prasetyo, KNOWLEDGE BASED SYSTEM, 2016.