

Data Mining Power Determination Nurses Sultan Sulaiman Hospital With C4.5 Algorithm

Wenika Hidayati¹, Paska Marto Hasugian²

^{1,2}Informatics Engineering Study Program, STMIK Pelita Nusantara, Jl. Iskandar Muda 1 Medan, North Sumatra, Indonesia 20154

E-mail:wenikahidayati49@gmail.com¹, paskamarto89@gmail.com²

Abstract-*The hospital is an agency engaged in health services in the which there are a number of special professions that can provide health services to the community items, namely doctors, Midwives and nurses and other professes. In this discussion, Arise and problems that can be raised into case studies to find out the results and information of each process in data mining Carried out with the C4.5 algorithm items, namely nurses. However, there are Several obstacles to Determine the nurses who will be declared passed or failed and accepted to work and can provide health services to the community, especially Patients who come for treatment. Therefore we need a method to identify nurses in a hospital. Data Mining with c4. 5 Algorithm can be used to the make predictions or classifications of nurses who are eligible to perform health services in hospitals by making decision trees based on existing data. This study aims to apply the data mining algorithm C4.5 in Determining nurses based on four attributes of used items, namely Accreditation, GPA, Age, and the value of each criterion has been determined in advance. The results of the study in the form of a decision tree Obtained from the data mining process with the C4.5 algorithm will provide information on the determination of nurses in the Sultan Sulaiman Regional Hospital.*

Keywords: *Determination, Data Mining Algorithm C4.5.*

1. Introduction

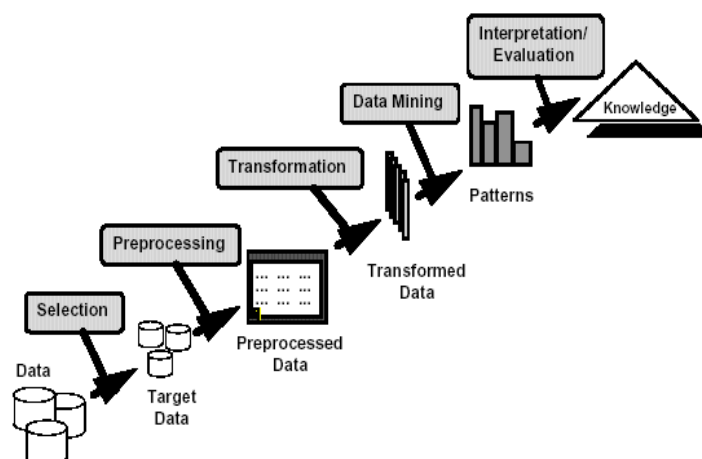
Data mining is the process to obtain a series of computer-based information that is useful from a large warehouse database.[1] Data mining can also be interpreted as extracting new information retrieved from large data chunks that help in decision making.[2]In data mining teknin there are many in the process, among which the algorithm C4.5, Naive Bayes and others. Data mining can be used as a determination of nurses based on the criteria that have been determined. The workings of this data mining cover all phases of making a problem, choose the relevant data and determine the modeling used in the process of resolving the problem, solving problems and solutions. Sultan Sulaiman Hospital is a hospital that is strategically located in Jl. SeiRampah Km.58 country. This hospital is a hospital grade C who develop health care professionals to the community. Sultan Sulaiman in hospitals, the number and types of nursing personnel by level of education, namely, D3 Nursing, Nursing S1, and the nursing profession nurses spread across work units outpatient and inpatient care, while the requirement for work and health services in hospitals Sultan Sulaiman is not only based on education, but it can also be determined from other aspects such as age, Accreditation, GPA and test scores in accordance with the criteria and hospital administrators. According to the data section and sexy nursing personnel department, is currently in Hospital Sultan Sulaiman using manual data to determine whether or not graduated nurses who work and health services in hospitals Sultan Sulaiman. GPA and the test results are in accordance with the criteria and hospital administrators. According to the data section and sexy nursing personnel department, is currently in Hospital Sultan Sulaiman using manual data to determine whether or not graduated nurses who work and health services in hospitals Sultan Sulaiman. GPA and the test results are in accordance with the criteria and hospital administrators. According to the data section and sexy nursing personnel department, is currently in Hospital Sultan Sulaiman using manual data to determine whether or not graduated nurses who work and health services in hospitals Sultan Sulaiman.

2. Theory

2.1 Data Mining

Data Mining is the set of processes

Data Mining is one of a series Knoeledge Discovery in Databases (KDD). KDD associated with integrase and penemuau scientific technique, interpretation and visualization of the number of data patterns. Data Mining can be divided into several stages as shown in the figure below:[3]



Picture 1. KDD process

2.2 C4.5 algorithms

General cara C4.5 algorithm to construct a decision tree is as follows:[4]

- a. Select an attribute as root.
- b. Create a branch for each value.
- c. For the case of the branches.
- d. Repeat the process for each branch until all cases the branches have the same class.

C4.5 algorithm has major advantages is that it can produce a model in the form of a tree or a rule that is easily interpreted, have an acceptable level of accuracy, can handle the type of discrete and numerical attributes.[5]C4.5 algorithm is an algorithm used to construct a decision tree based on criteria decision shaper. Formula C4.5 algorithm is divided into two formulas, the first formula to find Gain[6]

3. Research methods

The research method is a series of structured or systematic way used by researchers with the aim of getting the right answer on what is the question on the object of research. This method describes the method used in this research, the methodology used. The methodology is the science of the framework to carry out research that applying on a way or method, the branch of logic that deals with the general principles of the formation of knowledge (Knowledge).

To select an attribute as roots, is based on the highest gain value from the existing attributes. To calculate the gain, should we have to find the value of entropy beforehand.

The formula is as the following equation.[7]

$$Entropy(S) = - \sum_{j=1}^k P_j \log_2 P_j$$

Information :

S : The set of cases

k : The number of partitions S

p_j : The probability derived from the number (Yes / No) divided by the total cases

Meanwhile, calculating the value of the gain can be seen in equation 2 below.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} Entropy(S_i)$$

Information:

S : The set of cases

A : attribute

n : The number of partitions attribute A

$|S_i|$: Number of cases in the partition to-i

$|S|$: The number of cases in S

3.1 Framework Research

To assist in the preparation of this study, the need for the composition of the framework (Frame Work) is clear in its phases. This framework is the steps that will be done to solve the problem to be discussed. The following steps are carried out in accordance with the study design workflow author as follows

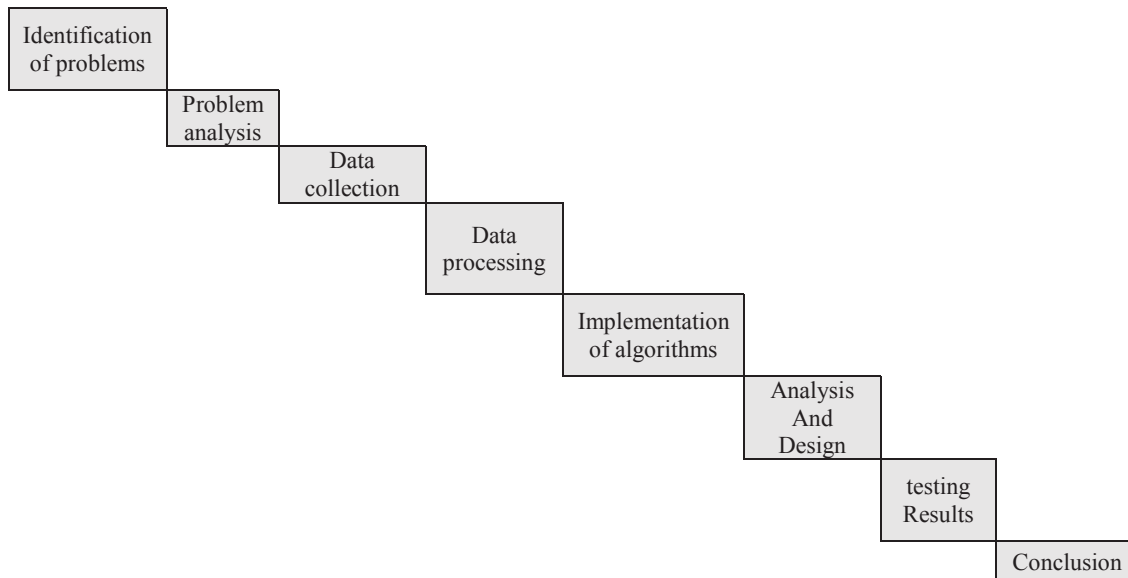


Figure 2. Framework Research

4. Ana#lysis

4.1 ana#lysis

Analysis of the system is an early stage in the design and development of a system to be designed, for this phase will be measured and evacuated on the performance of the system designed. Identification of existing problems and measures for the design needs to be expected.[8]

table 1
 The results of the Data Cleaning Power Nurses

No.	name nurse	Accreditation	Age	GPA	Score
1	AfrianiYulma	26	3.19	B	83
2	Mariska Nurhabibi	22	3.63	B	92
3	Nurul BARIAH	25	3:01	B	79
4	Halim Fadila	23	3.69	B	85
5	AprilinSyahputraLubis	26	3.42	B	84
6	Mariska Nurhabibi	22	3.63	B	92
7	Dedek Kurniawan	21	3.02	A	78
8	Asiya	25	3.46	A	90
9	AyuAnnisa	24	3:41	B	80

From the above process produced a table that is ready for the process of classification, attributes and attribute values listed in Table 2 below were obtained attribute table.

Table 2.
 attributes nurse

No.	Attribute	Information
1	Accreditation	An attribute of campus accreditation is "A" and "B".
2	Age	Is written evidence that can be seen from the KTP
3	GPA	It is a written value given by a faculty / university that can be seen from a diploma or transcript..
4	Score	Is the value obtained from the test results of each participant

From the list of items that are above, then use that data in accreditation A and B, age 21 s / d 28 years, GPA 2.50 s / d 4:00, value > 55 s / d 100 for it's taken some of the data to be sampled the determination of nurses in hospitals Sultan Sulaiman District. Serdang Bedagai. Here's a list of names of nurses who used sampeldi Hospital Sultan Sulaiman.Counting the number of cases for the decision to pass or fail, the number of cases to pass or fail decision, Entropy of all cases and cases were divided based on attributes of the number of nurses,

nurses eligible to be accepted, and a nurse who does not deserve to be received. After that, do the calculation of the gain for each attribute. The calculation result is shown by Table 3.

Table 3.

C4.5 Calculation (1)

variable name	Information	Sum (Score)	Sum (Yes)	Sum (No)	entropy	gain
Accreditation	A	2	2	0	0	
	B	7	3	4	0.985	0,225
Age	<= 25	6	5	1	0.65	
	> = 25	3	0	3	0	.558
GPA	<= 3.50	8	4	4	1	
	> = 3:51	1	1	0	0	0.102
Score	<75	1	0	1	0	
	> 75	8	5	3	.954	0.143

entropy:

- a. Accreditation A = $(- (2/2) \times \log_2 (2/2) + (- (0/2) \times \log_2 (0/2)) = 0$
 Accreditation B = $(- (3/7) \times \log_2 (3/7) + (- (4/7) \times \log_2 (4/7)) = 0.985$
 Gain for accreditation:
 $= 0.99107 - ((2/9) * 0) + ((7/9) * 0.985)$
 $= 0,225$

entropy:

- b. Age <= 25 = $(- (5/6) \times \log_2 (5/6) + (- (1/6) \times \log_2 (1/6)) = 0.65$
 Age > = 25 = $(- (0/3) \times \log_2 (0/3) + (- (3/3) \times \log_2 (3/3)) = 0$
 Gain for accreditation:
 $= 0.99107 - (6/9) * 0.65 + ((3/9) * 0)$
 $= 0.558$

entropy:

- c. GPA <= 3.50 = $(- (4/8) \times \log_2 (4/8) + (- (4/8) \times \log_2 (4/8)) = 1$
 GPA > = 3.51 = $(- (1/1) \times \log_2 (1/1) + (- (0/1) \times \log_2 (0/1)) = 0$
 Gain for accreditation:
 $= 0.99107 - (4/9) * 1 + ((1/9) * 0)$
 $= 0.102$

entropy:

- d. Value of <= 75 = $(- (0/1) \times \log_2 (0/1) + (- (1/1) \times \log_2 (1/1)) = 0$
 Value > = 75 = $(- (5/8) \times \log_2 (5/8) + (- (3/8) \times \log_2 (3/8)) = 0.954$
 Gain for accreditation:
 $= 0.99107 - (4/10) * 0.81127812445 + ((6/10) * 0)$
 $= 0.99107$ to 0.32451124978
 $= 0.143$

Based on the results seen that the attributes with the highest Gain Age, which amounted to 0.558 Thus Age can be the root node, the attribute value is already classified the cases into one, namely the decision PASS, so no need for further calculations.

Counting the number of cases, the number of cases for decision PASS, FAIL number of cases for decision, and the entropy of all cases and cases divided by the Accreditation attribute, value, GPA can be the root node of the attribute value. After that do the calculations Gain for each attribute. The results are shown in the following table:

Table 4.

C4.5 calculation (2)

variable name	Information	Sum (Score)	Sum (Yes)	Sum (No)	entropy	gain
Accreditation	A	2	2	0	0	
	B	4	3	1	.811	.109
Score	<= 75	0	0	0	0	
	> = 75	6	5	1	0.65	0
GPA	<= 3.50	5	4	1	0.722	
	> = 3:51	1	1	0	0	0,048

- a. Accreditation A = $(- (2/2) \times \log_2 (2/2) + (- (0/2) \times \log_2 (0/2)) = 0$
 Accreditation B = $(- (3/4) \times \log_2 (3/4) + (- (1/4) \times \log_2 (1/4)) = 0.811$
 Gain for accreditation:



- $= 0.99107 - ((2/4) * 0) + ((2/4) * 0.811)$
 $= 0.109$
- b. Value of $\leq 75 = (- (0/0) \times \log_2 (0/0) + (- (0/0) \times \log_2 (0/0) = 0$
 Value $> 75 = (- (5/6) \times \log_2 (5/6) + (- (1/6) \times \log_2 (1/6) = 0.65$
 Gain for accreditation:
 $= 0.99107 - (0/6) * 0) + ((0/6) * 0)$
 $= 0$
- c. $GPA \leq 3.50 = (- (4/5) \times \log_2 (4/5) + (- (1/5) \times \log_2 (1/5) = 0.722$
 $GPA > 3.51 = (- (1/1) \times \log_2 (1/1) + (- (0/1) \times \log_2 (0/1) = 0$
 Gain for accreditation:
 $= 0.99107 - (5/1) * 0.722) + ((5/1) * 0)$
 $= 0,048$

Based on the results seen that the attributes with the highest Gain Accreditation, amounting to 0.109. Thus Accreditation can be a branch node of the attribute value. There are two attributes of Accreditation grades A and B. From the second to the attribute values, the value of A is already classified the cases into PASS and attribute values B already classified the cases into FAIL, then these things do not need to do further calculations, but for the value attribute B still needs to be calculated again. From these results can be described as a decision tree as shown below:

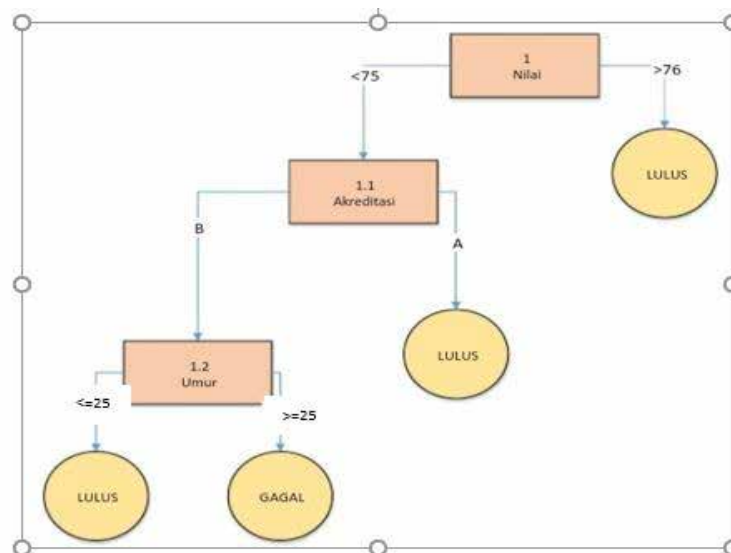


Figure 3. Decision Trees

- If Age <25 Accreditation B, GPA > 3.51 = Passed
- If Age <25 Accreditation B, GPA > 3.50, Value <85 = Pass
- If Age <25 Accreditation B, GPA > 3.50, Value > 85 = Pass
- If Age <25 Accreditation = Passed
- If Age > 25 = Fail

Counting the number of cases, the number of cases for decision PASS, FAIL number of cases for decision, and the entropy of all cases and cases were divided based on attributes, and status that can be the root node of the attribute value. By paying attention to the decision tree in the previous figure, note that the case has been entered in the class. Thus the final decision tree is formed.

5. Conclusion

From the analysis, design, and implementation can be obtained the following conclusions:

- The system is built with C4.5 algorithms can form a decision tree from the root to the last branch is formed.
- A system that can be used to help mengklasifikasikanperawat yangmemenuhi requirements based on criteria that have been determined.
- After calculating the entropy and gain use of the existing data, AlgoritmaC4.5 telahdapatmengimplementasikanklasifikasi nurses based on the assessment criteria

- d. The system can calculate the number of cases with PASS and FAIL decisions based on the value of the entropy of all cases based on attributes.

6. Reference

- [1] P. M. Hasugian, "Penguujian Algoritma Apriori Dengan Aplikasi Weka Dalam," *J. Mantik Penusa*, 2017.
- [2] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor -Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017.
- [3] A. S. Sunge, "Prediksi Kompetensi Karyawan Menggunakan Algoritma C4 . 5 (Studi Kasus : PT Hankook Tire Indonesia)," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, pp. 23–24, 2018.
- [4] G. F. Mandias, "Pemanfaatan Data Mining Untuk Prediksi Mahasiswa Peserta ' Capping Day ' di Universitas Klabat Data Mining for Student Prediction Participant of " Capping Day " at Universitas Klabat Using the C4 . 5 Algorithm," pp. 86–93.
- [5] K. Zaman, " Penerapan Data Mining Menggunakan Algoritma C4 . 5 Untuk Menentukan Kelayakan PENERIMA Bantuan Rehabilitas Sosial Rumah Tidak Layak Huni (Studi Kasus Di Pemerintahan Kabupaten Solok Selatan)," vol. 3, no. 2, pp. 12–24, 2016.
- [6] G. Syahputra, "Penerapan Algoritma C4 . 5 Dalam Analisa Kelayakan Penerima Bonus Tahunan Pegawai (Studi Kasus : PT . Multi Pratama Nauli Medan)," *J. Mantik Penusa*, vol. 16, no. 2, 2014.
- [7] L. N. Rani, "Kl asifikasi Nasabah Menggunakan Algoritma C4 . 5 Sebagai Dasar Pemberian Kredit," 2016.
- [8] R. V Palit, Y. D. Y. Rindengan, and A. S. M. Lumenta, "Rancangan Sistem Informasi Keuangan Gereja Berbasis Web Di Jemaat GMIM Bukit Moria Malalayang," vol. 4, no. 7, pp. 1–7, 2015.