

COMPARATIVE ANALYSIS EXPERT SYSTEM OF CERTAINTY FACTOR METHOD WITH DEMPSTER SHAFER METHOD BY USING OPEN DECISION MAKER TOOL FOR TODDLER AT VACCINE HOUSE

Eka Fitri Rahayu

Universitas Gunadarma, ekafitri@staff.gunadarma.ac.id

ABSTRACT

Expert system is a form progress of Information and Communications Technology. Development of an expert system can be performed in variety areas, including health. In its development required a result can be trusted by the user. Therefore it required the value of belief for any decision issued by the expert system.

In expert system, there are several uncertainty methods can be used namely Certainty Factor, Dempster Shafer, Naïve Bayes, Fuzzy Logic, etc. Therefore, in this study will be performed the comparison between Certainty Factor and Dempster Shafer method using data infectious diseases in toddler. Both of methods have different ways of working in the calculation, but both are equally give results to support a decision.

In the comparative analysis of both methods will be used a decision support system tool namely Open Decision Maker with Analytical Hierarchy Process approach. From the comparison result of eight cases has been obtained that the Certainty Factor method have more higher belief values as big as 82.54% than Dempster Shafer method as big as 17.46%, but significantly the value of belief between the two was not differ greatly.

Key Words: Certainty Factor, Dempster Shafer, Expert System, Open Decision Maker.

INTRODUCTION

Artificial Intelligence is one part development progress of Information and Communication Technology (ICT). In some fields, artificial intelligence has been widely applied to support and enable people to making a decision. One example is the development of expert systems in various fields, including health. Expert systems are computer based systems that use knowledge, fact and reasoning technique in solving a problem that usually can only be solved by an expert in the field. Representation of the expert system is done based on the facts, rules, and approaches in the form of reasoning, one of them is Forward Chaining approach (trace forward). In addition to using a technique of reasoning, would be much better if the expert system is supported by the

certainty values of a hypothesis. In obtaining the certainty value, in expert system there are several methods used to calculate the value of the certainty of the results such as: Certainty Factor method, Naïve Bayes, Dempster Shafer and Fuzzy.

In the development of expert systems, researchers must be proficient in considering and choosing what methods will be used to obtain the best results of an expert system that has been built. This is necessary so that the expert system can provide accurate results to the user so as to provide confidence and allows users to take a decision to resolve the problem. In this case, it will be made a comparison between the two methods of expert system method there are Certainty Factor and Dempster Shafer. Both methods have a different way of

working, the certainty factor method easier in the process of calculating the level of trust, and the Dempster Shafer methods have to go through a fairly complex calculation. Thus it would be a comparison between two methods is the complexity of calculations affects the results of the high level of trust. Comparison of the two methods will be made by means of manual calculations and the use open decision maker tool. The data used to perform a comparative analysis is data contained in the application of web-based expert system for diagnosing infectious diseases in toddlers. Each of these approaches has a way of working or different calculation processes, but has a same goal of providing the results accuracy of a hypothesis. The results of both approaches can be analyzed the compared with each other, so that researchers can consider which method is better for use in building an expert system.

RESEARCH METHOD

The research data is the data will be used to support the way of research process, which data of this study has two types used namely the primary data and secondary data.

Primary Data

Primary data is the research data obtained directly from primary sources or do not through the mediator. Primary data can be subject opinion individually or group. In this study, the primary data source is obtained directly in the field of answers a pediatrician regarding infectious disease in toddlers.

Secondary Data

Secondary data is data or information obtained from other parties or indirectly. Secondary data can be a good record which has been published or not. Secondary data source to support this research was obtained from

the literature on the subject related to the study.

Data Collection Method

Data collection techniques to support the process of this research are as follows:

1. Interview

At this stage the researcher conducted interviews with pediatrician, dr. Piprim B. Yanuarso, Sp.A (K). At this stage the researcher asked several questions related to the research needs such as symptoms data and infectious diseases in toddlers with the certainty value of some symptoms in an illness. Informants can give value with a range 0 to 1 for each symptom for every disease. Scoring is based on the theory contained in the Certainty Factor and Dempster Shafer method which states if given value is 0 (zero), it indicates the absence of symptoms (evidence); if given value is 1, it indicates the presence of certainty.

2. Literature Study

Data collection through literature study aims to explore as much information as possible about the objects and elements involved in this study. At this stage the literature study conducted by searching for information through e-books, articles, and journals.

Research Data Analysis

The data used in comparison of two methods is the value assigned by expert for each symptoms data in a disease. This value will be used as input data in the calculation process. Then, in the calculation process for each method will produce an output data will be used as support in making decision. In this case the way to generate an output of the certainty factor method is different with dempster shafer method.

In this study contained 25 symptoms and 8 diseases data. List of symptoms that will be input for each method are shown in Table 1.

Table 1.
Decision Table

No	Symptoms	A	B	C	D	E	F	G	H
1	Spleen enlarges slightly	x							
2	Child has a rash (redness of skin) is flat, especially on the chest and abdomen and sometimes spreads to the face, arms and Legs	x							
3	Can swelling of lymph nodes in the back of the head, neck next to the side and behind the ears	x							
4	Despite the high fever, but the child remained conscious and active	x							
5	When body temperature started to increase, 5-10% of patients experienced febrile seizures (seizures due to high fever)	x							
6	Fever	x	x		x			x	x
7	Respiratory disorders		x						
8	After passing 2 - 4 days, the rash spreads to the line of the body, arms and legs		x						
9	Rash (like a punch) on both cheeks		x						
10	There is a rash of small blisters containing pus and scab around the face, hands, head			x					
11	After one or two days, appeared spots - red color and become blisters filled with water				x				
12	Sometimes breathing sounds when breathing or after coughing					x			
13	Flu/Cold					x	x		x
14	Persistent cough discontinuous and can choke or vomit					x			
15	Glands at the back of the neck will swell						x		
16	Speckled rash that appears within one or two days - first in the face, then the rest of the body						x		
17	Difficult to chew and swallow							x	
18	Headache							x	
19	Dry mouth							x	

Tabel lanjutan 1

20	The gland swollen and tender under the ears and under the chin	x
21	White patches in the mouth (Koplik spots)	x
22	The rash appears on the third or fourth day	x
23	The spots will be flushed and more and more, but not itchy.	x
24	Coughing hard	x
25	Watery eyes, inflamed, red	x

Description : A: Roseola Infantum, B : Red Check Syndrome (Parvovirus B19), C : Impetigo, D: Chickenpox, E: Whooping Cough, F: Rubella (German Measles), G: Mumps, H: Measles (Rubeola, Measles 9 Days).

Tabel 2.
Range Value

No	Symptoms	Variable	Value
1	Spleen enlarges slightly		
2	Child has a rash (redness of skin) is flat, especially on the chest and abdomen and sometimes spreads to the face, arms and legs		
3	Can swelling of lymph nodes in the back of the head, neck next to the side and behind the ears		
4	Despite the high fever, but the child remained conscious and active	yes	0.1
5	When body temperature started to increase, 5- 10% of patients experienced febrile seizures (seizures due to high fever)		
6	Fever		
7	Respiratory disorders		
8	After passing 2 - 4 days, the rash spreads to the line of the body, arms and legs		
9	Rash (like a punch) on both cheeks		
10	There is a rash of small blisters containing pus and scab around the face, hands, head		
11	After one or two days, appeared spots - red color and become blisters filled with water		
12	Sometimes breathing sounds when breathing or after coughing	No	0
13	Flu/Cold		
14	Persistent cough discontinuous and can choke or vomit		
15	Glands at the back of the neck will swell		
16	Speckled rash that appears within one or two days - first in the face, then the rest of the body		
17	Difficult to chew and swallow		
18	Headache		
19	Dry mouth		

Table lanjutan 2

20	The gland swollen and tender under the ears and under the chin
21	White patches in mouth
22	The rash appears on the third or fourth day
23	The spots will be flushed and more and more, but not itch
24	Coughing hard
25	Watery eyes, inflamed, red

Tabel 3.
Description Output of Certainty Factor Method

Case	Formula	Result
1	$CF(A) > CF(B) > CF(D) > CF(H)$	A
2	$CF(B) > CF(A) > CF(D) > CF(G) > CF(H)$	B
3	$CF(E) > CF(F) > CF(H)$	E
4	$CF(H) > CF(F) > CF(E)$	H

Tabel 4.
Description Output of Dempster Shafer Method

Case	Formula	Result
1	$m(A) > m(ABDHG) > m(\Theta)$	A
2	$m(F) > m(F,H,E) > m(\Theta)$	F
3	$m(D) > m(A,D,H,B,G) > m(\Theta)$	D
4	$m(E) > m(E,H,F) > m(\Theta)$	E

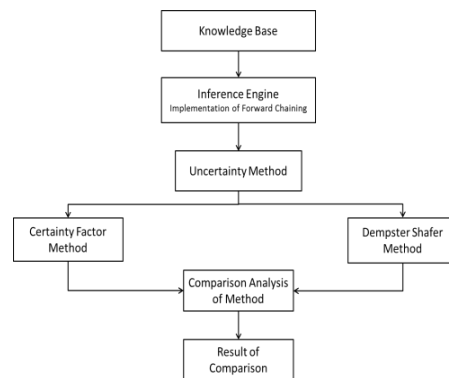


Figure 1 Methodology of Comparative Analysis for Certainty Factor and Dempster Shafer Method

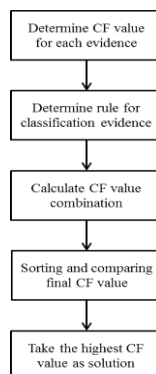


Figure 2 Steps of Certainty Factor Method Calculation

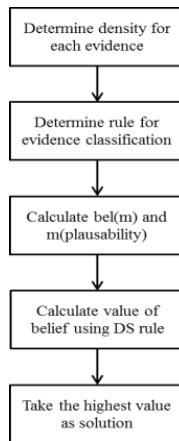


Figure 3 Steps of Dempster Shafer Method Calculation

Range of Input Data

Each data that will be input in the calculation each method has a range more than 0 to 1. However, if data is not used as an input, then, regarded as a variable No and the value is 0.

Output of Certainty Factor Method

Output data of certainty factor method obtained through calculation result CF value every symptom selected then conducted a comparative analysis of the certainty factor final value from several possible diseases which can occur. The highest CF value will be used as the final decision on the method.

For example symptoms were selected for case 1 is symptom no 1, 2, 3, 4, 5, 6, and 7 then the possibility of disease suffered is disease A, B, D, H. Because the symptoms selected most is symptom of disease A (no 1, 2, 3, 4, 5, 6), then the possible CF value on disease A larger than disease B, D, H which only have symptom no 7.

In Table 3 shown some examples the output data acquisition of certainty factor method.

Output of Dempster Shafer Method

Output data of Dempster Shafer method obtained through calculation results density values every symptom selected then conducted a comparative

analysis the mass function final value of several possible diseases which can occur. The highest mass function value will be used as final decision on the method.

For example symptoms were selected for case 1 is symptom no 1, 2, 3, 4, 5, 6, and 7 then the possibility of disease suffered is disease A, B, D, H and G. Because the symptoms selected most is symptom of disease A (no 1, 2, 3, 4, 5, 6), then the possible mass function value on disease A higher than disease B, D, H, G which only have symptom no 7.

Table 4 shows some examples the output data acquisition of dempster shafer method.

Stages that must conducted before making a comparative analysis is the calculation stage for every method with the same variable or input. In this stage will be shown how calculate the confidence value for every method which will be used as output by the system. Each of these methods has different calculations technique, but do not close possibility that the results will be issued not far adrift with each other. Figure 1 shows the stages in the process of doing a comparison between Certainty Factor and Dempster Shafer method.

Implementation of Uncertainty Method

The next step is implement method or approach used to calculate values of facts contained in the knowledge base. In this study, will used two methods namely Certainty Factor and Dempster Shafer method. Each method has a different calculation. The following are the steps in the process of calculation to be performed.

The Steps of Certainty Factor Method Calculation

In performing calculation using Certainty Factor there are several stages. The flow calculation CF Value for symptoms selected as follows :

The Steps of Dempster Shafer Method Calculation

In performing calculations using Dempster Shafer there are several stages. The flow of calculation CF value for symptoms selected as follows.

Comparison Analysis of Method

The next step after implement calculation method is do a comparison for every results issued from each methods. Then, the results will be analyzed to determine which method is more accurate and better used. The

comparison is not only done manually by looking at the end result of every calculation, but researcher will compare the results using the Decision Support System applications. It can be a support for the results of comparative analysis method.

Result of Comparison

The results of comparison method will be made in form table containing results of calculation for every case in every method, weighting alternative, weighting criteria, alternative ranking, alternative Main Criteria Matrix which is result of Open Decision Maker application.

RESULT AND DISCUSSION Calculation Techniques with Certainty Factor Method

The first step to calculate the trust value by using Certainty Factor is initializing symptoms contained in the knowledge base. These symptoms will be used as input followed by classifying the symptoms and determine the rules are suitable for the chosen symptoms to obtain required results. Initialize the symptoms will be named using symptoms code. List of initialization symptoms contained in the knowledge base can be seen in table 5:

**Table 5.
Symptom Initialization**

Code	Symptoms
A001	Spleen enlarges slightly
A002	Child has a rash (redness of skin) is flat, especially on the chest and abdomen and sometimes spreads to the face, arms and legs
A003	Can swelling of lymph nodes in the back of the head, neck next to the side and behind the ears
A004	Despite the high fever, but the child remained conscious and Active
A005	When body temperature started to increase, 5-10% of patients experienced febrile seizures (seizures due to high fever)
A006	Fever

Tabel lanjutan 5

B001	Respiratory disorders
B002	After passing 2 - 4 days, the rash spreads to the line of the body, arms and legs
B003	Rash (like a punch) on both cheeks
C001	There is a rash of small blisters containing pus and scab around the face, hands, head
D001	After one or two days, appeared spots - red color and become blisters filled with water
E001	Sometimes breathing sounds when breathing or after coughing
G003	Headache
G004	Difficult to chew and swallow
H002	White patches in the mouth (Koplik spots)
H003	The rash appears on the third or fourth day
H004	The spots will be flushed and more and more, but not itchy.
H005	Coughing hard
H007	Watery eyes, inflamed, red

Table 6.
CF Value of Symptoms

Code	CF Value	Code	CF Value
A001	0.1	E003	0.8
A002	0.7	F001	0.7
A003	0.6	F002	0.8
A004	0.5	G001	0.6
A005	0.4	G002	0.8
A006	0.8	G003	0.5
B001	0.5	G004	0.8
B002	0.4	H002	0.5
B003	0.1	H003	0.8
C001	0.7	H004	0.7
D001	0.8	H005	0.7
E001	0.6	H007	0.8
E002	0.7		

The next step is determining CF value for every symptom in every disease. This value is determined by an expert or a pediatrician. Table 4.2 show the CF values which given by experts to give a solution After determining the CF values for every symptom, the next process is calculate the values with CF combination equation as follow.

Calculation for the selected symptoms are classified based on rule that has been created, which each rule will be calculated the certainty value so that will be obtained the highest CF

value of comparison results would be the solution.

Based on the above calculation using the combination equation in Certainty Factor method, final results are obtained for H disease is 0.99989. Whereas for the CF value on disease A, B, D and G is 0.8, because there is only one selected evidence on the disease that is evidence A006. So that the final result is disease H with CF value 0.99989.

Calculation Technique with Dempster Shafer Method

In the process of calculating of the trust value by using Dempster Shafer, the stage is initializing the symptoms that are contained in knowledge base. These symptoms will be used as input then performed symptoms classification based on diseases which have a relationship with these symptoms.

By use of the Dempster rule then obtained value for m13 as follows.

Based on the calculation above can be compared to the value of m for

every possible disease namely $m_{13}(H) > m_{13}(A,B,D,G,H) > m_{13}(E,F,H) > m_{13}(\theta)$. The m value for disease H is the highest, it is 0.999208.

Comparative Analysis

Calculations for each method performed by testing eight cases according to the number of diseases data in previous studies. Based on the calculations have been done to compare the methods of Certainty Factor and Dempster Shafer of the obtained results as shown in Table 7.

$$CF(R1,R2) = CF(R1) + CF(R2) - [(CF(R1) \times CF(R2))] \dots\dots\dots (1)$$

Or

$$CF(R1,R2) = CF(R1) + CF(R2) * [1-CF(R1)] \dots\dots\dots (2)$$

$$CF(H1) = CF(A006) + CF(H002) * [1-CF(H001)] = 0,8 + 0,5 * [1 - 0,8] = 0,9 \dots\dots\dots (2)$$

$$CF(H2) = CF(H003) + CF(H1) * [1-CF(H003)] = 0,8 + 0,9 * [1 - 0,8] = 0,98 \dots\dots\dots (2)$$

$$CF(H3) = CF(H004) + CF(H2) * [1-CF(H004)] = 0,7 + 0,98 * [1 - 0,7] = 0,994 \dots\dots\dots (2)$$

$$CF(H4) = CF(H005) + CF(H3) * [1-CF(H005)] = 0,7 + 0,994 * [1 - 0,7] = 0,9982 \dots\dots\dots (2)$$

$$CF(H5) = CF(E006) + CF(H4) * [1-CF(E006)] = 0,7 + 0,9982 * [1 - 0,7] = 0,99946 \dots\dots\dots (2)$$

$$CF(H7) = CF(H007) + CF(H4) * [1-CF(H007)] = 0,8 + 0,99946 * [1 - 0,8] = 0,99989 \dots\dots\dots (2)$$

$$M13(H) = 0.796832 + 0.001728 + 0.001008 + 0.000432 / (1-0) = 0.999208 \dots\dots\dots (5)$$

$$M13(E,F,H) = 0.000252 / (1-0) = 0.000252 \dots\dots\dots (5)$$

$$M13(A,B,D,G,H) = 0.000432 / (1-0) = 0.000432 \dots\dots\dots (5)$$

$$M13(\theta) = 0.000108 / (1-0) = 0.000108 \dots\dots\dots (5)$$

**Tabel 7.
Result of Calculation**

Criteria	CF (%)	DS (%)
A	99.22	96.11
B	94.6	77.5
C	60	60
D	96	80
E	97.6	92
F	98.2	94
G	99.84	98.4
H	99.98	99.92
AVG	93.16	87.3225

Tabel 8.
Weighting Criteria

Criteria	Quarrel of Criteria	Weight	Criteria	Quarrel of Criteria	Weight
A	3.11	13.99	7	E	5.6
B	17.1			C	0
C	0	3.11	2	F	4.2
A	3.11			C	0
D	16	12.89	7	G	1.44
A	3.11			C	0
E	5.6	2.49	2	H	0.06
A	3.11			C	0
F	4.2	1.09	1	E	5.6
A	3.11			D	16
G	1.44	1.67	1	F	4.2
A	3.11			D	16
H	0.06	3.05	2	G	1.44
A	3.11			D	16
C	0	17.1	9	H	0.06
B	17.1			D	16
D	16	1.1	1	F	4.2
B	17.1			E	5.6
E	5.6	11.5	6	G	1.44
B	17.1			E	5.6
F	4.2	12.9	7	H	0.06
B	17.1			E	5.6
G	1.44	15.66	8	G	1.44
B	17.1			F	4.2
H	0.06	17.04	9	H	0.06
B	17.1			F	4.2
D	16	16	9	H	0.06

Comparative analysis will be performed using the Open Decision Maker tool with AHP approach (Analytical Hierarchy Process). The use of these tool aims to prove and strengthen the comparison of manual calculations.

The next step is giving weight to each pair of criteria. It is used to calculate a rating value for each criterion of both methods. The means used to give weight is to compare the difference value between criteria of two alternatives. Scale is used to perform the weighting of criteria based on the table 8.

Table 8 reveals the rating result for alternative that has been calculated. Alternative Certainty Factor (CF) is superior than alternative Dempster Shafer (DS), which is the value for the alternative CF at 82.54% while the value of alternative DS is equal to 17.46%.

CONCLUSION AND SUGGESTION

The result of a comparative analysis between the two methods shows that the Certainty Factor method is superior in generating a higher level of belief in more cases than Dempster Shafer method. This is shown by the results of data processing confidence

level of both methods using Open Decision Maker application. Certainty Factor method obtain value as big as 82.54%, while Dempster Shafer method obtain value as big as 17.46% from processing of eight cases. However, the level of belief obtained by each method for each hypothesis has not much difference significantly. In the calculation of the Dempster Shafer method, resulting hypotheses are more varied, it is because the entire evidence selected will be combined with each other. While Certainty Factor method, the calculation is done by sorting or grouping directly into the rules of evidences has been made so that the combination only limited evidence contained in any rules.

In this study, the comparisons were performed only using eight cases from one expert to obtain data on the level of belief. The result of study is limited manual calculations and has not been implemented into the system. Therefore, further research is expected so that the analysis of comparison in determining the uncertainty method performed with more cases, the data come from another expert, and also can be tested by using statistical tests. Then, both methods can be implemented into a system so that can be utilized by the community in diagnosing infectious diseases in toddlers.

In this study, the comparisons were performed only using eight cases from one expert to obtain data on the level of belief. The result of study is limited manual calculations and has not been implemented into the system. Therefore, further research is expected so that the analysis of comparison in determining the uncertainty method performed with more cases, the data come from another expert, and also can be tested by using statistical tests. Then, both methods can be implemented into a system so that can be utilized by the community in

diagnosing infectious diseases in toddlers.

REFERENCES

- Army, W. (2018). Sistem pakar diagnosis penyakit menular dengan metode forward chaining Dan certainty factor. *Jurnal Sains dan Informatika*, 4(2), 171.
- Leman, D. (2018). Expert system diagnose tuberculosis using Bayes theorem method and Shafer Dempster method. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*.
- Liebowitz, J. (2019). *The handbook of applied expert systems*. CRC Press.
- Minardi, J., & Suyatno, S. (2016). Sistem pakar untuk diagnosa penyakit kehamilan menggunakan metode Dempster-Shafer Dan decision tree. *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 7(1), 83.
- Panggabean, E. K. (2018). Comparative analysis of Dempster Shafer method with certainty factor method for diagnose stroke diseases. *International Journal of Artificial Intelligence Research*, 2(1), 32.
- SIHOTANG, H. T. (2019). Sistem pakar mendiagnosa penyakit kolesterol pada remaja dengan metode certainty factor (CF) berbasis web.
- Widyaningsih, M., & Gunadi, R. (2017). Dempster Shafer Untuk Sistem Diagnosa Gejala Penyakit Kulit Pada Kucing. *Jurnal SAINTEKOM*, 7(1), 81.