IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.13, No.1, January 2019, pp. 1~10
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.24861

Cased Based Reasoning to Identify Cause Conflicts in Marriage

Arief Ichwani *[1], Suprapto[2]
[1]Institute of Technology Sumatera, Lampung Selatan, Indonesia
[2]Department of Computer Science and Electronics, UGM, Yogyakarta, Indonesia
e-mail**: *[1]arief.ichwani@if.itera.ac.id**, [2]sprapto@ugm.ac.id

Abstrak

Fungsi KUA dalam kegiatan agama Islam antara lain memberikan pelayanan dan bimbingan di bidang kepenghuluan dalam hal pelayanan nikah dan rujuk bagi umat beragama Islam, memberikan pelayanan dan bimbingan di bidang pengembangan keluarga sakinah, konsultasi konflik atau masalah rumah tangga, dan lain sebagainya.

Penelitian ini bertujuan untuk mengidentifikasi konflik dalam pernikahan dengan menerapkan algoritma Bayesian Model pada tahap penentuan kelompok data uji, yang selanjutnya akan dilakukan proses pencarian nilai similaritas tertinggi dengan menggunakan algoritma Nearest Neighbor. Adapun data sumber dan data uji yang digunakan terbagi menjadi dua kelompok, yaitu data nikah dan data riwayat konsultasi. Sedangkan kelompok konflik yang ingin diidentifikasi terbagi menjadi lima kelas, yaitu faktor pekerjaan, faktor usia, faktor pendidikan, faktor jumlah pernikahan, dan faktor status sosial.

Pengujian dilakukan dengan menggunakan data sebanyak 12 data yang terdiri dari 11 data kasus dan 1 data uji. Pada tahap penentuan kelompok konflik diperoleh bahwa data uji termasuk dalam kelompok satu yaitu F001 (faktor pekerjaan). Sehingga pada tahap mencari nilai similaritas digunakan hanya basis kasus yang berada pada kelas F001 yaitu KK001, KK003, dan KK008. KK001 memiliki nilai similaritas sebesar 0.476, KK003 sebesar 0.882, dan KK008 sebesar 0.142. Kasus dengan nilai similaritas paling tinggi akan disimpan sebagai basis kasus baru. Jika nilai similaritas yang didapat lebih kecil dari nilai threshold yaitu 0.8, maka solusi dari kasus tersebut akan direvisi oleh pakar. Hasil perhitungan tingkat akurasi dengan menggunakan 35 data uji baru, didapatkan nilai sebesar 82.86%.

Kata kunci*—case based reasoning, bayesian model, nearest neighbor*

Abstract

The function of KUA in the activities surrounding the religion of Islam, Including providing service and guidance in the area of present services in terms of marriage and reconcilement for Muslims, provide services and guidance in the field of development of Sakina, family consultation conflict or household problems, and so on. Integration between the computer and artificial intelligence into the post-wedding consulting services is one approach in Overcoming the limitations of the expert (religious instructor).

This research aims to identify conflict in marriage by applying Naive Bayes algorithm at the stage of Determining the groups of test the data (retrieve), then entered the stage of the search process of the highest similarity value by using the Nearest Neighbor algorithm (reuse). The data sources and the test of data used are divided into two groups items, namely marriage, and history of data consultation, While the group conflicts are identified will be divided into five classes items, namely an employment factor, the factor of age, educational factors, factors the number of weddings, and social status.

Received May 8th,2017; Revised August 16th, 2018; Accepted January 24th, 2019

Testing is performed by the use of 12 of data, consisting of 11 cases and 1 of data test data. At the stage of determination of group conflict acquired to test the data included in group one ie F001 (factor of the job), so at the stage of looking for value similarities used only the base case of the class F001 ie KK001, KK003, and KK008. The KK001 similarity has a value of 0476, KK003 of 0882, and KK008 of 0142. The case with most high similarity value will be stored as a base case. If the value of similarity Obtained less than the threshold value that is 0.8, then the solution of the case will be revised by experts. The results of the calculation accuracy, using 35 new tests of data that gets the value of 82.86%.

Keywords-*case-based reasoning, naive Bayes, nearest neighbor*

1. INTRODUCTION

Religious Affairs Office (KUA) is leading business units that carry out some tasks in the field of Islamic government. The role and function of KUA in the activities of the Islamic religion among others, provide services and guidance in the field of the marriage officer in terms of weddings and reconciliation for the people are Muslim, providing services and guidance in the development of harmonious family, provides consulting services to conflict or domestic problems, and others so.

Conflict is defined as a condition of an incompatibility between the values or goals to be achieved, both in individuals and in relations with others [1]. Conflicts can also occur in domestic life. The accumulation of problems which can not be resolved could lead to conflict. In addition, there are several factors that can affect the conflict in the household, such as 1) Incompatibility of the needs and expectations of each other, 2) Difficulties receive real differences (habits, needs, opinions, and values), 3) Problems finance (how to earn and spend), 4) problem child, 5) feelings of jealousy and have a spouse excessive that lack of freedom, 6) Failure to communicate, 7) couples not in line with initial interests and goals [2].

Along with the development of technology, it is possible that an expert in this study is an extension of Islam to document knowledge and experience in dealing with conflicts in marriage. This documentation can be contained in a system that uses artificial intelligence and knowledge base. In general, the settlement of the problems encountered can be done by looking at the experience of the cases that have gone before. It is the basis for this study using Case-Based Reasoning.

Case-Based Reasoning to solve the problems by emulate the human ability to use solutions from previous cases that have the highest similarity to the case is going on. The process of seeking closeness between the new cases against the former case can use a variety of methods. The method used in this research is to find the similarity case is Nearest Neighbor. This method is applied in the phase retrieval CBR cycle.

Retrieval is another step in finding information (usually in the form of documents) from unstructured nature, like text that can be extracted from a collection/collection of large data stored in the computer [3]. With today's computer capabilities, there is a need for a more towards the search process data. One way to avoid scanning linear with respect to each data record is with the labeling group on the entire document. One method that can be applied to the labeling process is Naive Bayes group.

Based on these descriptions, this study intends to develop a CBR system for the identification of conflict in marriage by implementing algorithms Naive Bayes and Nearest Neighbor algorithm. This research is expected to produce a system that can help identify the causes of conflict and recommend solutions to conflicts quickly and accurately.

Research domains on case-based reasoning have been done by various methods and various forms of cases, especially in the medical field. Research by applying CBR to support the diagnosis of skin and venereal diseases in humans. Global similarity calculation using the Nearest Neighbor method for the retrieval of cases, while the local similarity value of 1 if it has

the same symptoms and is 0 otherwise. Adaptation left entirely to the revised case expert. The diagnosis includes 36 class names in the skin and venereal diseases. The test results obtained by the accuracy of the system by 90% [4]. Next, research by proposing the use of CBR for the diagnosis of ENT diseases. The process of diagnosis is done by inserting a new case containing symptoms ENT diseases. Then the system will make the process of indexing using Backpropagation method for obtaining the index of new cases. Having obtained the index, the system performs the similarity calculation on the basis of new cases of the cases had the same index using the Cosine Coefficient. Cases with a similarity value below 0.8 are considered to have no solution and stored for later revision by an expert. Tests conducted on 111 new cases with the test results of 91.89% have a similarity above 0.8 and 8:11% by similarity under 0.8 [5]. Cases with a similarity value below 0.8 are considered to have no solution and stored for later revision by an expert. Tests conducted on 111 new cases with the test results of 91.89% have a similarity above 0.8 and 8:11% by similarity under 0.8 [5]. Cases with a similarity value below 0.8 are considered to have no solution and stored for later revision by an expert. Tests conducted on 111 new cases with the test results of 91.89% have a similarity above 0.8 and 8:11% by similarity under 0.8 [5].

2. METHODS

2.1 Description System

The system built is a system of identifying the causes of conflict in marriage. Results of the system output can be used as an insert to the KUA officials in providing advice to couples who consult domestic problems. CBR method used in this research is divided into two parts, the process of determining the group using the Naive Bayes algorithm and the determination process by applying a solution recommendation Nearest Neighbor algorithm.

The system was built consisting of three main process steps. The first stage begins by entering a problem/complaint is being experienced by married couples. This stage aims to get a consultation history data that is used as an input pair system. In addition, marriage data / marital personal use as input to the next process. The second stage is to find a group class of test data. At this stage, the test data (the result of the first step) is inserted into the system to do a comparison against a database of existing cases in the database. Probability / highest probability value of the comparison results are then used as label group conflict of test data.

The third stage is entered in the retrieval process. This process aims to obtain a solution of a previous case that has the highest similarity value of the test data. At this stage, the calculation results of the second phase in the form of a label group used as the basis for calculating the similarity of cases. Case similarity calculation is applied only to cases that have the same group label to the test data. The process of calculating the similarity value case is divided into two steps, namely: 1) Finding the value of local similarity to the features of type symbolic and numeric-type features, 2) for the global similarity value by multiplying each similarity of each attribute with a value of weight.

Threshold (Limit value) is applied in this study of 0.8. The threshold value aims to maintain system performance. If the global similarity value is smaller than the threshold value, then the repair process will be conducted by an expert solution. Conversely, if the value of the resulting global similarity is more than the threshold value, then the solution of the previous case can be used as a reference in providing advice/solutions conflict.

2.2 Representation Case

CBR important component in the system is in the form of data storage previous cases into storage. The cases are stored in the base case contains at least two components, namely the problem which is the real situation when cases and solutions to these problems. Selection domain representation depending on the needs and duties as well as the structure of the case data is available. Two common views on how a case represented, the functional information and the ease of obtaining such information.

In general, a case is represented in the form of a collection of features that characterize and solutions for dealing with such cases. These features are stored are parameters to obtain a solution. These features are obtained from the acquisition of knowledge, such as an interview with an expert. Type and method of evaluation features are presented in Table 1.

Table 1 Names and assessment features

Name Features	Value Features
Husband:	
Age	in terms of years
Occupation	Equal = 1 and No = 0
Education	Equal = 1 and No = 0
Number of Marriage	1, 2, 3, etc ..
Wife:	
Age	in terms of years
Occupation	Equal = 1 and No = 0
Education	Equal = 1 and No = 0
Number of Marriage	1, 2, 3, etc ..
Factors problem: 1. M001 2. M002 3. M003 Etc.	 1. The lack of economic 2. Intimidation and physical violence 3. Often quarrel Etc.
Identify the cause of the conflict: 1. F001 2. F002 3. F003 4. F004 5. F005	 1. Job factor 2. Age factor 3. Educational factors 4. Number of marriages factor 5. Social status factor

After the case represented in some form, then the data will be stored in case the database is indexed. Index here in the form of labeling/group of each record. Thus, in the case retrieval process can be done more quickly without having to perform a scan of each record in the database.

2.3 Retrieval

CBR is based on the hypothesis that a solution to the case that never happened before can help solve new cases that occur on the condition that there is a similarity between them [6]. Measurement is done by comparing the similarity of the features of the new cases with similar features that exist in the base case. A case is said to be identical to other cases if it has a similarity value equal to one, whereas if the value of similarity is less than one, then it can be said that such cases are similar.

The increasing number of case data stored in the base case, the longer the time required for scanning/matching features between the test data with the base case. To overcome this, indexing performed on every record the data contained in the database. In this study, indexing is implemented by labeling group conflicts that exist in the base case. Labeling groups of the conflict on the base of cases performed by extension Islam based on cases that have been dealt with previously.

2.3.1 Naive Bayes Algorithm Functions

Naive Bayes algorithm used in the process of determining the group (label data) to the Algorithm test data is very efficient because the reasoning process is done by utilizing the existing input with a relatively faster way to find a group or classification of data which is most similar to the classes predetermined. Bayes expressed as a hypothesis called Apriori Maximum Probability Hypothesis (HMAP). The formula used in the algorithm described in equation (1) [7].

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{\sum_{i=1}^{n} P(X|C_i)P(C_i)} \quad (1)$$

where *P (Ci | X)* is the posterior probabilities conditional hypothesis C_i occur if given evidence *X, P (X | C_i)* is the conditional probability of hypothesis *X* occurs if given evidence C_i and *P (C_i)* is the probability of prior hypotheses C_i occur regardless any evidence. While C_i is the recommendation results and *X* is an input parameter. So the solution to be fulfilled according to equation (2).$X \in C_i$.

$$P(X_n|C_i)P(C_i) = \max(P(X|Ci)\ P(Ci)) \quad (2)$$

2.3.2 Nearest Neighbor Algorithm Functions

Nearest Neighbor is one of the machine learning algorithms are relatively simple to determine the classification of an object. This algorithm included in the category of supervised learning which requires a number of reference data has the class attribute. The workings of this algorithm are to compare each of the attributes of the target case of the attributes contained in the base case, then the comparison is calculated using a similarity function. There are two types of similarity used in the Nearest Neighbor algorithm, namely:

1. Local similarity

   The local similarity is defined as the level of similarity contained in the data/features. Local similarity can be divided into two, namely numerical (discrete) and symbolic (continuous). Discrete data is data whose values are natural numbers, while the continuous data is data whose value is in a certain interval. In this study, the data included in this kind of symbolic are education, jobs, and a history of consultation (problem/complaint). As for features such as age and income are included in numeric types.

   a. For numeric data types with a range of data, normalization can be done with the aim of scaling process values to fall in a certain range shown in equation (3) [8].

      $$f(s, t) = 1 - \frac{|s-t|}{f_{max}-f_{min}} \quad (3)$$

      where *s,t* is the value you want to compare features, *s* is the source case and *t* is the target case, while f_{max} is the highest value used features and fmin is the lowest value used features.

   b. For symbolic data type is shown in equation (4).

      $$f(s,t) = \begin{cases} 1 & if\ s = t \\ 0 & other \end{cases}, where\ s, t \in \{true, false\} \quad (4)$$

2. Global similarity

   The global similarity is defined as the level of similarity contained in the object or contained within the case. The global similarity is used to calculate the similarity between the new case by case stored in the base case. There are many ways to measure the distance between the proximity of cases, such as Euclidean Distance Manhattan Distance, and Nearest Neighbor. Nearest Neighbor has several advantages, namely resilience training data

that have a lot of noise and effectively if it has data training in large numbers. Calculating the value of similarity (similarity cosine) between document test with all the training documents can be used equation (5) [8].

$$Sim(S,T) = \frac{\sum_{i=1}^{n} f(S_i,T_i)*w_i}{\sum_{i=1}^{n} w_i} \quad (5)$$

where *Sim (S, T)* is the similarity between the case of *S* and *T*, *n* is the number of attributes that are used, S_i is the i-th attribute the source case, T_i is the i-th attributes that exist in the target case, w_i is the attribute weights to-i; where the weight is determined by experts, and $f(S_i, T_i)$ is a function of local similarity.

Differences in weight/value of interest on each feature (data consulting marriage and history), then the formula in equation (5) needs to be improved. These improvements are used to deal with the value of the trust and the handling of the case or the target of a new problem. These improvements carried out in accordance with the equation (6) [9].

$$Sim(S,T) = \left(\frac{\sum_{i=1}^{n} f(S_i,T_i)*w_i}{\sum_{i=1}^{n} w_i}\right) * \frac{J(S_i,T_i)}{J(T_i)} * P(S) \quad (6)$$

where $P(S)$ is the level of confidence or belief of experts to a case in source case, $J(S_i, T_i)$ is the number of features contained in the target case that appear in the source case, $J(T_i)$ is the number of features contained in the target case.

2.4 Flow Model System

The model is built based on the system flow analysis system needs. Models built workflow system consists of several parts with each function as illustrated in Figure 1.

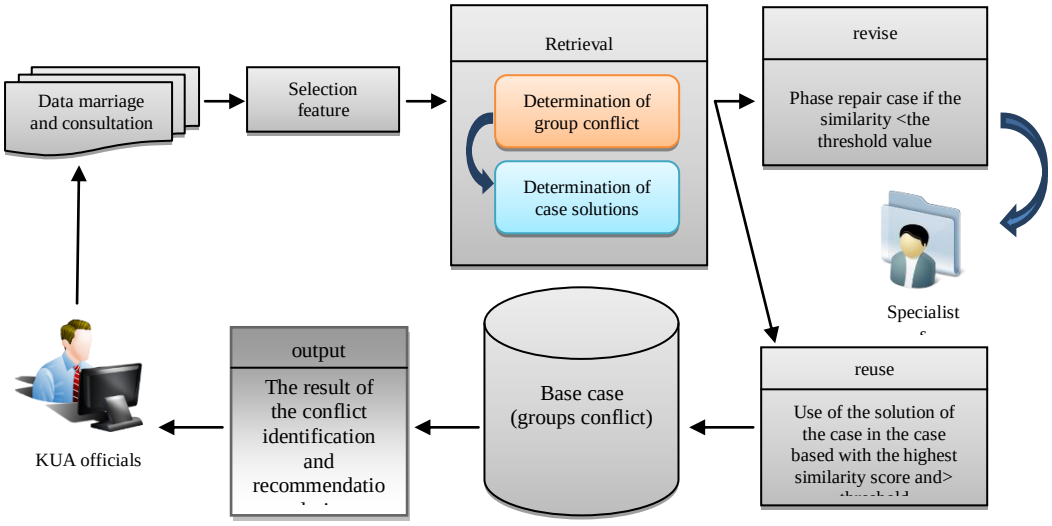

Figure 1 Model grooves CBR system for identifying the causes of marital conflict

2.5 Database Design

Database designed in the form of the model database diagram as shown in Figure 2.

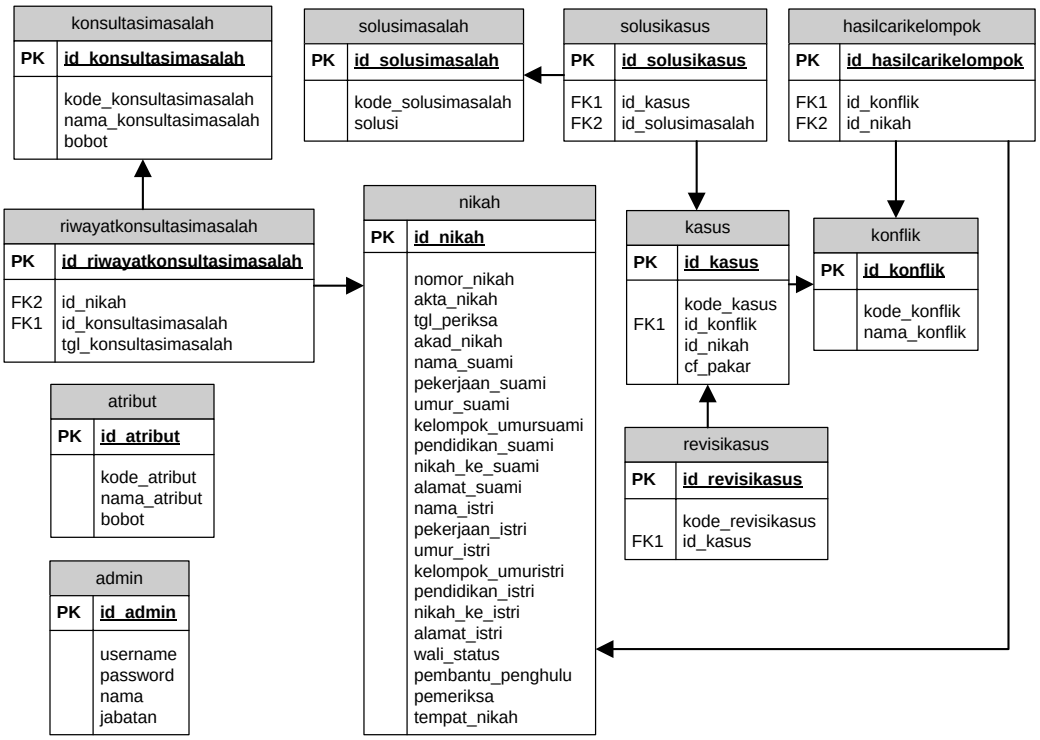

Figure 2 Design of the model database diagram

Database diagram model on Figure 2 illustrates the relationship between tables. Relationships between tables are described in the following description:

1. Relationships that occur from the table of marriage and relationship issues consultancy binary form that is M: N. So that in this relation there are new tables (associative) in the form of consulting history table problem.
2. The relation between the marriage table and the conflict is M: N, where each record of marriage can have more than one conflict. Results of the second relation table raise new associative table form table results search results group.
3. Relations that occur between tables solutions to problems and cases in the form of a binary relation that is M: N. So that in this relation there are new tables (associative) in the form of case solutions table.
4. Relations that occur between the tables of cases and revision cases was 1: 1.
5. Relations that occur between the tables of cases and conflicts are 1: M, where each case may have the number of conflicts is more than one.

2.6 Data and Methods of Testing Systems

The data used in the research is a marriage of data (personal data) husband and wife and consulting historical data post-wedding couple. Data obtained from SIMKAH KUA marriage Sukabumi Bandar Lampung. The data used is a time series data from 2014 through 2015.

In evaluating the performance of a system required a systematic way. The confusion matrix is the right way to analyze the system's accuracy in recognizing the tuple of a different class. Accuracy is the degree of proximity measurement of the quantity of the actual value. Analysis of the test results performed by using four parameters: true positive (TP), false positive (FP), true negative (TN) and false negatives (FN). TP and TN provide that information correctly diagnose system while the FP and FN indicate that the wrong diagnosis system. Table 2 illustrates how the system is tested using the confusion matrix.

Table 2 Testing of the system with the confusion matrix

Value Classification	Actual Value	
	TRUE	***FALSE***
TRUE	TP (True Positive) *Correct result*	FP (False Negative) *Unexpected result*
FALSE	FN (False Negative) *missing result*	TN (True Negative) *The correct absence of a result*

Table 2 is an evaluation of the classification results by using two classes, TRUE and FALSE. Subsequently obtained the value of precision, recall, and accuracy. To calculate the value of precision and recall can be used equation (7), (8) [10].

$$Precision = \left(\frac{TP}{TP + FP}\right) * 100\% \quad (7)$$

$$Recall = \left(\frac{TP}{TP + FN}\right) * 100\% \quad (8)$$

The truth value of the results of the identification system can be done by calculating the degree of accuracy by using equation (9).

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) * 100\% \quad (9)$$

3. RESULTS AND DISCUSSION

The data used as a basis in the case of testing the accuracy of the system is as much as 100 data. The data consist of data and history data consulting marital problem married couples. These data have been given the label group by extension causes of the conflict in accordance with the understanding and the results of consultations couples. Table 3 presents the distribution of the database of the case is used as a knowledge base system.

Table 3 Distribution of base case

Conflict Group	Amount of Data
F001 - Factors jobs	27
F002 - The age factor	14
F003 - education Factor	25
F004 - Factors number of marriages	13
Factors F005 social status	21

Five classes that exist in this system has features formers. Features formers obtained from interviews with experts (extension of Islam) and reference sources used in the study.

Accuracy testing is done using 35 new data is not yet known group conflict by the system. Testing is done by comparing the class identification results with an actual class system that has been determined by extension Islam. Data accuracy testing results are presented as the Tabel 4 system.

Table 4 Results of testing with 35 test data

Actual Classroom	Amount of Data	Conflict Identification				
		F001	F002	F003	F004	F005
F001	14	13		1		
F002	5		4		1	
F003	8	1		6		1
F004	4		1		3	
F005	4	1				3

Furthermore, the test results obtained from Table 4 can be used to calculate the value of precision and recall of every class. Precision and recall value can be found using equation (7) and (8). Results calculated from the value of precision and recall are shown in Table 5.

Table 5 Results of testing precision and recall value each class

Conflict Group	Precision	Recall
F001 - Factors jobs	92.86%	86.67%
F002 - The age factor	80.00%	80.00%
F003 - education Factor	75.00%	85.71%
F004 - Factors number of marriages	75.00%	75.00%
F005 - social status factor	75.00%	75.00%

After testing the precision and recall value of each class is done, then the next calculation of the overall average grade/group conflict based on the value of precision, recall, and accuracy. The accuracy of calculation according to the equation.

Table 6 System performance test results with threshold 0.8

PERFORMANCE		
Precision	***Recall***	***Accuracy***
79.57%	80.48%	82.86%

Testing is done by using a threshold of 0.8. Giving the threshold value is based on the results of a literature review has been done. Precision is the degree of accuracy of the information requested by the user with the answers given by the system, while the recall is the level of success of the system in rediscovering information.

The test for the CBR system by applying the confusion matrix may use a threshold value that varies from 0.6 to 1. The higher the required threshold value will maintain the accuracy of the system in identifying the cause of the conflict. This is because if the result of the similarity of cases reached 1, then the new cases that were tested had the absolute similarity of the cases were in the base case.

4. CONCLUSION

Based on the results of research and testing has been done, it was concluded that:

1. This research resulted in the CBR system for the identification of the causes of conflict in marriage by applying the algorithm Naive Bayes and Nearest Neighbor algorithm.
2. Naive Bayes is used for the determination of group conflict on test data (new cases). It aims to speed up the scanning process in conducting retrieval.
3. Nearest Neighbor is used to find a solution to the conflict in the retrieval stage. Search on conflict solution of the test data is applied only to cases with the same label conflict group.
4. The test results of 35 new test data indicate that the threshold limit value of 0.8, the performance of the system has an accuracy of 82.86%.

5. SUGGESTIONS

Some things that can be recommended for further research are:
1. The process of scoring or weighting can be done by applying a weighting algorithm, such as methods SWING, Simple Additive Weighting (SAW), and the Analytic Hierarchy Process (AHP).
2. Necessary to add more features in identifying the conflict to get more accurate results.

REFERENCES

[1] Gradianti, TA & Suprapti., V., 2014, Marital Conflict Resolution Style on a pair of Dual earner, *Jurnal Psikologi Education and Development*, Vol 3, No 3, December 2014.

[2] Yanuarti, D & Sriningsih., 2012, Marital Adjustment on Conflict at Work Spouse, *http://sriningsih.mercubuana-yogya.ac.id/2012/01*, Accessed on January 15, 2017.

[3] Manning, CD, Raghavan, P & Chutze, H., 2008, *Introduction to Information Retrieval*, Cambridge University Press, New York.

[4] Abdiansyah & Hartati, S., 2008, Case-Based Reasoning to Support Skin and Venereal Disease Diagnosis in Humans, *Thesis*, Graduate Program in Computer Science, Univ. Gadjah Mada, Yogyakarta.

[5] Rismawan, T & Hartati, S., 2012, Case-Based Reasoning for Disease Diagnosis ENT, *Jurnal IJCCS*, No2, Vol6, pp67 ~ 78.

[6] Wahyudi, E & Hartati, S., 2017, Case-Based Reasoning for diagnosis Heart Disease, *Journal IJCCS*, No. 1, Vol 11, pp 1 ~ 10.

[7] Kusrini & Lutfi ET 2009, *Data Mining Algorithms,* Andi Offset, Yogyakarta.

[8] Adawiyah, R., 2016, Case-Based Reasoning for the Diagnosis of Disease Due to Dengue virus type, *Thesis*, Graduate Program in Computer Science, Univ. Gadjah Mada, Yogyakarta.

[9] Mancasari, UA, 2012, Expert System Using Case-Based Reasoning for the Diagnosis of Neurological Disease in Children, *Thesis*, S1 Computer Science, Univ. Gadjah Mada, Yogyakarta.

[10] Kustiyahningsih, Y., 2013, the Feasibility Application Medium Business Lending Using ID3 Algorithm, *Proceedings of the Opportunities and Challenges in Computational Intelligence, Big Data, and Cybernetics*, Yogyakarta, October 18th, 2014.