

# Mobile-based Recommendation System for the Tour Package Using the Hybrid Method

<https://doi.org/10.3991/ijim.v12i8.9483>

I Made Agus Wirawan<sup>(✉)</sup>, I Wayan Bayu Diarsa  
Universitas Pendidikan Ganesha, Bali, Indonesia  
imade.aguswirawan@undiksha.ac.id

**Abstract**—Although there has been a lot of research on traveler recommendations, there have been no studies that discuss the recommendations of tour packages from travel collections in the past. Therefore, in this study it is important to do a related study 1) Development of a mobile-based recommendation system using the Hybrid Method. 2) Accuracy testing of the system in providing tour package recommendations.

This study uses the CBR method to provide tour package recommendations based on past travel data. There are 4 stages of the process, such as: Retrieve, Reuse, Revision, and Retain. In this study, focused on the Retrieve stage using a hybrid method. The hybrid method on the mobile-based recommendation system is a combination of the methods of Naive Bayes, Bayes Theorem, and Dempster Shafer. Where Naive Bayes is used to calculating probabilities such as age and frequency of visits. The Bayes Theorem is used to calculate probabilities such as country, gender, and destination. To determine the mass value of a combination of evidence, then use the Dempster Shafer method. Based on the accuracy test, it was affirmed that the total accuracy of the tour package by the mobile-based recommendation is 95%. While the error rate of the system in providing tour packages is 5%.

**Keywords**—Tour package, Travelers, Recommendation system, Case Based Reasoning, Dempster Shafer Theory, Naïve Bayes, Bayes Theorem, Web Responsive, and k-Fold Cross-Validation

## 1 Introduction

Bali is a world tourist destination. Based on Trip Advisor in 2018, Bali was ranked fourth in the world as the 25 Most Popular Destinations in the World. As a world tourist destination, Bali is the largest foreign exchange producing province in Indonesia originating from tourism. Bali tourism combines natural beauty and cultural uniqueness in almost every village in Bali. Various efforts have been made by the Government through the Bali Office of Tourism and Culture to increase travelers visits, one of which is conducting market analysis. Market analysis activities are activities to analyze tourism promotion strategies and tourism trends [1]. However, the analysis process is carried out only at the description stage of the travelers that has been visited. Each traveler has characteristics that can be used to identify traveler interest. The characteristics of

travelers are a determining factor in the selection of tourist objects to be visited [2] [3] [4] [5] [6]. Although many studies have been conducted on the analysis and prediction of traveler visits, there has been no research on predictions of tourist destination packages based on the characteristics of travelers in the past [7] [8] [9]. Case-Based Reasoning (CBR) is a computer-based reasoning model that uses a knowledge-based approach to solve problems based on past experience. CBR can be used in providing recommendations on tourist objects to be visited according to the characteristics of travelers [10] [11] [12] [13]. In its application, the case data allows having imprecision and uncertainty information in producing an uncertain decision [14]. Dempster-Shafer theory allows representation of imprecision and uncertainty by applying the concept of belief and plausibility [14]. The belief value or basic belief assignment can be obtained from the probability of the Bayes Theorem method for the initial node [15] [16]. However, in determining the value of belief in continuous attributes it cannot use the Bayes Theorem method. To get the belief value in continuous attributes can use the gauss density formula on the Naïve Bayes method [17] [18] [19, p. 210]; [20]. Along with the development of the internet and mobile devices, the role of information technology in the tourism sector is a representative tool in the promotion and management of tourism objects [21] [22]. Along with access to information through various media such as computers, laptops, smartphones, or tablets, it has an impact on system development so that it can run on all different devices and platforms. This problem can be overcome by the responsive web. [23] [24].

Based on the problem and research that has been done before, it is deemed necessary to conduct research on 1) Development of a mobile-based recommendation system for tour packages using the hybrid method 2) Measuring the accuracy of the tour package recommendation system.

## 2 Background

There are several reviews of the literature used in this study: Travelers Segmentation Concepts; Case-Based Reasoning Method; Dempster Shafer Theory; Naïve Bayes; and Bayes Theorem.

### 2.1 Travelers Segmentation Concepts

Tourism marketing strategies can be done through on the traveler visits. An effective strategy in tourism marketing is to use market segmentation [25]. In determining the tourist attraction to be visited, travelers can be grouped into several segments according to the characteristics of travelers [26] Characteristics of each segmentation are [27] [28] [29]; [30]:

- a) **Geographical Segmentation.** In this segmentation, travelers are grouped by place or country. This segmentation is used to find out the market components (travelers) widely. The characteristics used in this segment are country of origin.
- b) **Socio-Demographic Segmentation.** Socio-demographic characteristics are reflections of "who wants what". Distribution based on these characteristics is most often

done for the sake of tourism analysis, planning, and marketing because it is very clear and relatively easy to group. Grouping of travelers based on socio-demographic characteristics is most evident in relation to their travel patterns. Gender and age are related to the choice of the type of tourism carried out. The characteristics used in this segment are age, gender, income, and occupation. [3].

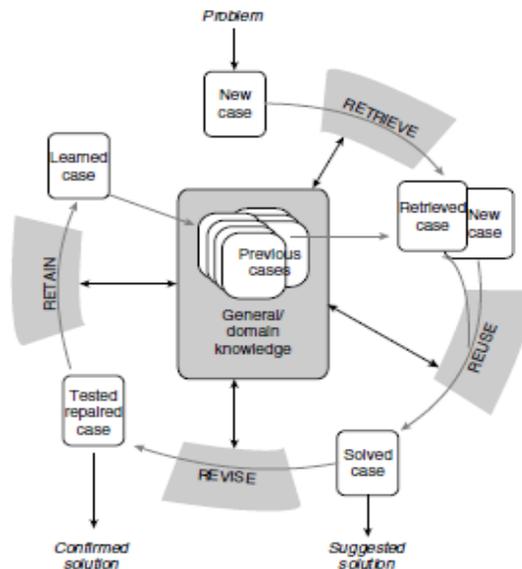
- c) **Segmentation of Behavior.** Because of the limitations of doing psychographic segmentation such as relatively high costs and time to conduct research, a new approach was developed based on behavioral segmentation. In this segmentation, travelers (markets) are grouped based on the frequency of visits.

## 2.2 Case Based Reasoning Method

Case-Based Reasoning (CBR) is a computer reasoning model that mimics the behavior of an expert and learns from the experience of previous cases [31]. This method can be used in various fields of economics, science, and tourism [12] [32] [31]. In many domains process, CBR has a nature in solving problems compared to rule-based systems [33]. Where a rule-based system requires sufficient time to acquire knowledge from experts. CBR does not require knowledge acquisition, it only requires a set of past experiences (case based). In the CBR there are four cycles, including Retrieve, Reuse, Revision, and Retain [34] [35].

1. **The retrieve process.** This stage is to matches the previous case that is the most relevant (similar) to the new case. The retrieval phase begins by describing the problem and determining the level of compatibility between the previous case and the new case. This section refers to aspects of identification, compatibility, search, selection, and execution. The retrieval stage is one of the most important stages in the CBR cycle, where the previous cases are taken (case-based) to solve the new case by measuring the similarity values of the old case.
2. **The reuse process.** Modeling knowledge and information from the results of the retrieval process. Modeling is based on the level of trust that is most relevant between the new case and the case data. The reuse process is producing a solution proposed for the new case.
3. **The revise process.** This process is carried out to review the resulting solution, by matching the real case. If the resulting solution is not suitable then the revision process can be made to adjust to the real case.
4. **The retain process.** This process is carried out to integrate/save of the solutions into case data. The stored solution can be used for the next case data.

From the four stages of the CBR, the following stages are described in Figure 1:



**Fig. 1.** Life cycle CBR  
Source (Snakar dan Simon, 2004)

In its application, the stages of reuse and revise are sometimes difficult to distinguish. The revision stage in CBR is still an open research because it is a complicated process in manipulating case solutions [36]. Although CBR can be used in various types of problem domains, there are a number of conditions that must be considered in applying the CBR, including: [36]; [37]

1. The CBR method can be applied if the problem domain is not fully understood or if success or failure factors from the solution cannot be explicitly modeled (for example, medical diagnosis or economic forecast).
2. The application of CBR method is a possible alternative to use if in the problem domain there are new experiences and data exceptions are often encountered.
3. If there are recurrent case data and a lack of similarity between new cases and old cases, it is better to build a domain model to get a solution.
4. One consideration in implementing a CBR method is the existence of significant benefits in the management of resources (such as system development time, company processing).
5. Ease in obtaining and recording the data needed from previous cases. The stored solutions can be adjusted in the future. These considerations are the basis for the selection of the CBR method.

Benefits of applying the CBR method [37]

1. Reducing tasks in the acquisition of knowledge.
2. Can handle errors in making solutions in the past.

3. Provide flexibility in knowledge modeling.
4. Can handle problems that have not been fully understood, defined, or modeled.
5. Allows making predictions of the solutions given.
6. Can do the learning process from time to time.
7. Reasoning can be done in the domain of knowledge in a small scope.
8. Reasoning can be done on incomplete data.
9. Avoid repetition of the steps taken in producing a solution.
10. Provide explained facilities.
11. Can be changed according to the purpose and domain of the case.
12. Reflects human reasoning

### 2.3 Dempster Shafer Theory

The Dempster-Shafer (DST) method was first introduced by Dempster, who experimented with the uncertainty model using a probability range rather than a single probability. Then in 1976 Shafer published the Dempster theory in a book called Mathematical Theory of Evident [14]. DST is a mathematical theory of evidence also called evidential reasoning that can handle information that is uncertain, and inaccurate. DST can be interpreted as a general form of probability theory, wherein probability theory, the evidence is associated with only one possible event, whereas in DST, the evidence is associated with some event (e.g. event set). Thus, the theory of evidence is more flexible than probability theory [38]. DST can be used in CBR [39].

**Frame of Discernment.** In Dempster-Shafer theory also known as Frame Of Discernment (FOD) or also called environment denoted by  $\Theta$  [14]:

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\} \quad (1)$$

where:

$\Theta$  = FOD or environment.

$\theta_1 \dots \theta_n$  = elements in the environment.

The environment/FOD is another term of the universe of discourse. FOD is a set of mutually exclusive objects. One way to determine the elements of an environment is to interpret a concept of questions and answers. For example, all the set of attractions in Buleleng:

$$\Theta = \{\text{Lovina Beach, Penimbangan Beach, Singaraja Museum}\}$$

If given a question "Which includes a marine tourism object is?", Then the possible answer from the  $\Theta$  subsets is:

$$\{\theta_1, \theta_2\} = \{\text{Lovina Beach, Penimbangan Beach}\}$$

Likewise with the question " Which are the marine tourism objects that are many foreign travelers?", Then the possible answer from subset  $\Theta$  is:

$$\{\theta_1\} = \{\text{Lovina Beach}\}$$

This subset is called a singleton set, because it has only one element on  $\Theta$ . Surely not all questions can be interpreted into the  $\Theta$  subset and are called the subset of empty  $\emptyset = \{\}$ . All subsets of the environment can produce a  $2^N$  number of subsets formed from  $\Theta$  to  $\emptyset$ , where  $N$  is the number of elements at  $\Theta$ . Here is an example of the patterns of all subset on the tourist object environment in Buleleng District  $\Theta = \{A, B, C\}$  can be seen in Figure 2:

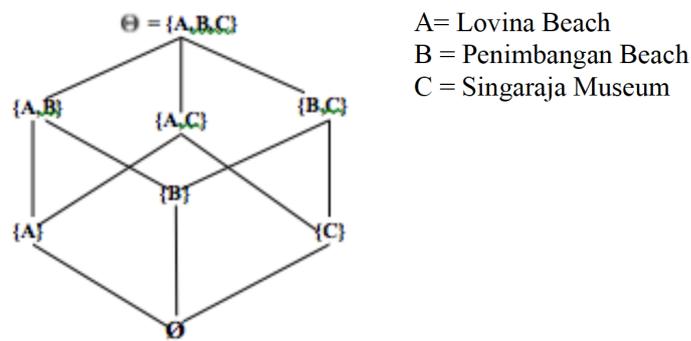


Fig. 2. Pattern evidence subsets  
Source [15]

Based on Figure 2, one of the patterns can be expressed into a subset of relational hierarchies between parent and child, for example:

$$\{\emptyset\} \subset \{A\} \subset \{A, B\} \subset \{A, B, C\}$$

The set  $\{A\}$  is a subset of  $\{A, B\}$  and the set  $\{A, B\}$  is a subset of  $\{A, B, C\}$ . It can be concluded.

$$X \subseteq Y$$

where all elements of  $X$  are elements of  $Y$ , so it can be expressed as follows:

$$X \subseteq Y \text{ means } \forall x \in X \rightarrow \forall x \in Y$$

In figure 2 the number of subsets formed is  $2^3 = 8$  subset. Collections of all these subsets define power sets or  $P(\Theta) = \{\emptyset, \{A\}, \{B\}, \{C\}, \{A, B\}, \{A, C\}, \{B, C\}, \{\}$ .

**Mass Function and Ignorance.** In the DST the belief level of an event is called the degree of belief and represented as mass ( $m$ ) of the object or also called the Basic Probability Assignment (BPA) [40]. The basic belief assignment can be obtained from the prior probability for the initial node. In DST does not impose a belief to be given to ignorance or refutation of a hypothesis, but the mass value is given only to a subset in an environment that is believed as a degree of trust. An unsupported subset is considered the nonbelief ( $\Theta$ ). The nonbelief can be defined as a reserving judgment of disbelief

and additional belief of an evidence. For example, a set of attractions in Buleleng District  $\Theta = \{A, B, C\}$ . It is assumed that travelers coming from Country X have a belief to visit a tourist attraction ( $\{A, B\}$ ) of 0,7 or  $m(\{A, B\})=0,7$ . While the nonbelief of the tourist object are visited is  $m(\Theta) = 1 - 0,7 = 0,3$ . The determination of the value of  $m(\{A, B\}) = 0,7$  is derived from the probability of A and B traveler objects visited by travelers from Country X. In DST the mass value is given to a set of the subsets, whereas on probability theory is associated with a set of elements, for example, "I sure to visited Buleleng Regency, but I have no information about the tourist objects in Buleleng Regency it will be visited ". From the example the mass function value in DST can be obtained,  $m(\{A\}) = m(\{B\}) = m(\{C\}) = 0$  and the value  $m(\Theta) = 1$ , where  $\Theta = \{A, B, C\}$ . From this mass value can produce belief function,  $Bel(\{A\}) = Bel(\{B\}) = Bel(\{C\}) = 0$  and  $Bel(\Theta) = 1$ . But in probability theory, probability function (p) of each element is  $p(A) = 0.33$ ,  $p(B) = 0.33$ ,  $p(C) = 0.33$ , where  $S1 + S2 + S3 = 1$ . Each mass can be expressed as a function that maps the power set to the real number between intervals 0 to 1 [14]

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

and empty mass is defined:

$$m(\emptyset) = 0;$$

while the sum of the entire mass for each subset X of the power set is 1.

$$\sum_{X \in P(\Theta)} m(X) = 1 \tag{2}$$

Based on a sample of previously visited tourist attractions, it can be calculated the value of the power set [14].

$$\sum_{X \in P(\Theta)} m(X) = m(\{A, B\}) + m(\Theta) = 0,7 + 0,3 = 1$$

**Combining Evidence.** A combination of evidence is an important process in summarizing information from various sources. If there are several different sources (such as experts) in giving an assessment of a set of events, a combination of scores from each source is needed to summarize the information. To overcome this, it is necessary to use a rule known as Dempster's Rule of Combination on DST [14].

$$(m_1 \oplus m_2)(Z) = \sum_{X \cap Y = Z} m_1(X)m_2(Y) \tag{3}$$

In the process of combining the evidence allows for the occurrence of a conflict between the evidence. The conflict of evidence occurs when no elements are produced on the cross product between the evidence, for example, the combination between the evidence  $m(\{A\})$  and  $m(\{B\})$  which does not produce the element ( $\emptyset$ ). To overcome

this problem, we need to normalize the process by defining the evidential conflict (k) into Dempster's Rule of Combination in equation 3. So that it can be derived into the following equation 4 [14]:

$$(m_1 \oplus m_2)(Z) = \frac{\sum_{X \cap Y = Z} m_1(X)m_2(Y)}{1 - k} \tag{4}$$

where the evidential conflict (k) is formulated by:

$$k = \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y) \dots \tag{5}$$

So equation (5) substituted into equation (4) would be:  
where:

$$(m_1 \oplus m_2)(Z) = \frac{\sum_{X \cap Y = Z} m_1(X)m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y)} \tag{6}$$

- $(m_1 \oplus m_2)(Z)$  = mass function of evidence (Z), which is the result of a combination of evidence (X) and evidence (Y)
- $m_1(X)$  = mass function of evidence (X)
- $m_2(Y)$  = mass function of evidence (Y)
- k = mass evidential conflict
- $\oplus$  = operator direct sum

### 2.4 Theorem Bayes and Naïve Bayes

The Bayes Theorem and Naïve Bayes are one of the methods used in data mining that is based on the Bayes decision theory. The Bayes Theorem and Naïve Bayes have classification capabilities such as decision tree and neural network methods. This method can be used to predict the probability of membership in each class [41]. Here is the formula for the Bayes Theorem. [42]

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^n P(X_i|Y)}{P(X)} \tag{7}$$

The Bayes theorem is only able to process categorical type. The data with the continuous type is able to use the gauss density formula on the Naïve Bayes method. [6]:

$$P(X|Y) = \frac{1}{\sqrt{2\pi} \sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2} \tag{8}$$

Explanation:  
P(Y|X) = The probability of data with vector X in class Y.

$P(Y)$	=	Initial probability of class Y.
$\prod_{i=1}^q P(X_i Y)$	=	The Y class independent probability of all features in vector X.
$\mu$	=	Mean or average values of attributes with continuous features.
$\sigma$	=	The deviation standard

The Bayes Theorem and Naïve Bayes require a huge amount of data records to get good results. If the predictor category does not exist in the training data, then Bayes Theorem and Naïve Bayes assume that the new record with the predictor has a zero probability [20].

### 3 Implementation of the Mobile-Based Recommendation System

The traveler will input characteristics data as criteria for providing travel package recommendations. The Criteria which are inputted such as country, gender data, purpose, and frequent visits. The traveler is not required to input all of the criteria. The condition is called uncertainty. The method of Dempster Shafer can overcome the problem of uncertainty data. In this study, the criteria used in the tour package recommendations are country, gender, age, frequency, and purpose. To specify a mass value of the country, gender, and destination using the Bayes theorem, while age and frequency use Naive Bayes.

As an example, if the traveler named Bent (on the 11th data) wishes to get a tour package recommendation that matches his/her characteristics, then the system will provide travel package recommendations based on the same characteristics (country, gender, and age) of the 10 training data. Training data obtained from the distribution of questionnaires to travelers in some attractions. The questionnaire rubric containing the questions related to the tourist object (tour package) visited by travelers. The 1-10 traveler's data can be analyzed using a combination of Bayes Theorem, Naïve Bayes, and Dempster Shafer method, in order to provide the recommendation of the traveler attraction that will be visited by the 11th travelers. Here are 10 data tour packages visited by travelers (training data) can be seen in Table.1.

#### 3.1 Retrieve process

The first step, the traveler input data criteria of the country, gender, and age on the system. For the frequency and purpose criteria is assumed not entered by the traveler – see Figure 3 (according to the data in table 1).

After the criteria entered, the next step clicks the process button. Figure 4 shows the flowchart of the retrieve process.

**Table 1.** The visitor data.

No.	Name	Country	Gender	Purpose	Age	Frequency	Tour packages
1	Erick	France	M	Vacation	32	1	W1,W3,W4
2	Jhon Lee	China	M	Research	25	1	W1,W2,W3, W4
3	Sofia	US	F	Vacation	21	2	W1, W2,W3
4	Andi	Indonesia	M	Vacation	24	4	W1,W3, W4
5	Merry	France	F	Vacation	35	2	W1,W3, W4
6	Kevin	US	M	Visiting Relative	49	1	W1,W2,W3, W4
7	Lee Young	China	M	Research	26	4	W1, W2,W3
8	Coi Dae	China	F	Vacation	31	4	W1,W2,W3, W4
9	Ahmad	Indonesia	M	Visiting Relative	38	2	W1, W2,W3, W4
10	West	France	F	Vacation	26	5	W1, W2,W3, W4
11	<b>Bent</b>	<b>France</b>	<b>M</b>	<b>(null)</b>	<b>33</b>	<b>(null)</b>	<b>?</b>

**Fig. 3.** Retrieve process in the system.

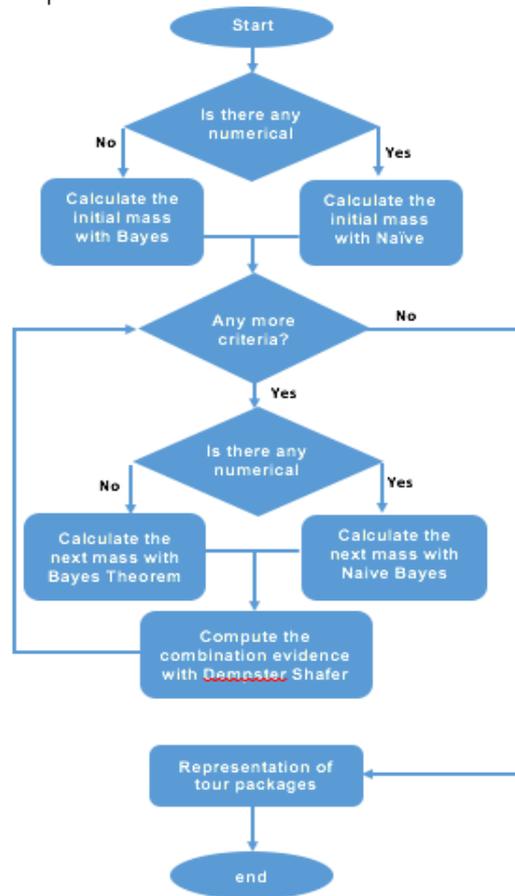


Fig. 4. Retrieve flowchart

- d) based on the flowchart in Figure 4, can be explained:
- e) The first step is to check the value of numerical or category criteria. If the value is a category, then the next step calculates the value of the initial mass function using the Bayes Theorem formula, if the value is numerical, then the next step calculates the value of the initial mass function using the Naïve Bayes Classifier formula.
- f) To determine the value of the next mass by using the Bayes theorem (if the category attribute) or using naïve Bayes (if the criteria are continuous). But if there are no next criteria, then present the recommendation of the tour package.
- g) The results of the initial mass value and the next mass values combined using the Dempster Shafer method.
- h) If there is a new criterion, then the process of determining the mass value. But if there are no new criteria, then present the recommendation of the tour package based on the mass value of the results of a combination of the Dempster Shafer method.

The following are manual calculations of the retrieve process using Bayes Theorem, Naive Bayes, and Dempster Shafer methods.

- i. Calculate the mass function of the "France" using the Bayes Theorem formula.

$$m_1(\{W1, W3, W4\}) = \frac{2}{3} = 0,66666667$$

$$m_1(\{W1, W2, W3, W4\}) = \frac{1}{3} = 0,33333333$$

$$m_1(\Theta) = 1 - (0,66666667 + 0,33333333) = 0$$

The highest value is the tour package W1, W3, and W4. If traveler input gender criteria, then proceed to stage 2.

- ii. Calculate the mass function of the "male" criteria by using the Bayes Theorem formula.

$$m_2(\{W1, W3, W4\}) = \frac{2}{6} = 0,33333333$$

$$m_2(\{W1, W2, W3, W4\}) = \frac{3}{6} = 0,5$$

$$m_2(\{W1, W2, W3\}) = \frac{1}{6} = 0,16666667$$

$$m_2(\Theta) = 1 - (0,33333333 + 0,5 + 0,16666667) = 0$$

- iii. Conduct a combined evidence test. Since  $m_1(\Theta) = 0$  and  $m_2(\Theta) = 0$ , it is not combined.

**Table 2.** Evidence calculations ( $m_1 \oplus m_2$ ).

$m_1 \oplus m_2$	$m_2(\{W1, W3, W4\}) = 0,333$	$m_2(\{W1, W2, W3, W4\}) = 0,5$	$m_2(\{W1, W2, W3\}) = 0,167$
$m_1(\{W1, W3, W4\}) = 0,667$	$(\{W1, W3, W4\}) = 0,222$	$(\{W1, W3, W4\}) = 0,333$	$(\{W1, W3\}) = 0,111$
$m_1(\{W1, W2, W3, W4\}) = 0,333$	$(\{W1, W3, W4\}) = 0,111$	$(\{W1, W2, W3, W4\}) = 0,167$	$(\{W1, W2, W3\}) = 0,056$

- iv. The result of the combined evidence process, calculated by the formula 6.

$$m_3(\{W1, W3, W4\}) = \frac{0,222 + 0,333 + 0,111}{1 - 0} = 0,666$$

$$m_3(\{W1, W2, W3, W4\}) = \frac{0,166}{1 - 0} = 0,167$$

$$m_3(\{W1, W3\}) = \frac{0,111}{1 - 0} = 0,111$$

$$m_3(\{W1, W2, W3\}) = \frac{0,056}{1 - 0} = 0,056$$

The highest value is **W1, W3, and W4** as the recommendation of tour package based on the characteristic of country and gender of 0,666.

- v. When travelers input the age criteria, then the next step is to calculate the mass function using the gauss density formula (formula 8) of Naïve Bayes Classifier.

- a. Determine the mean and standard deviation on age criteria.

$$\begin{aligned} \mu_{W1,W3,W4} &= \frac{32 + 24 + 35}{3} = 30,33 \\ \mu_{W1,W2,W3,W4} &= \frac{25 + 49 + 31 + 38 + 26}{5} = 33,8 \\ \mu_{W1,W2,W3} &= \frac{21 + 26}{2} = 23,5 \\ \sigma^2_{W1,W3,W4} &= \frac{(32 - 30,33)^2 + (24 - 30,33)^2 + (35 - 30,33)^2}{3 - 1} = 32,33 \\ \sigma_{W1,W3,W4} &= \sqrt{32,33} = 5,69 \\ \sigma^2_{W1,W2,W3,W4} &= \frac{(25 - 33,8)^2 + (49 - 33,8)^2 + (31 - 33,8)^2 + (38 - 33,8)^2 + (26 - 33,8)^2}{5 - 1} \\ &= 98,7 \\ \sigma_{W1,W2,W4} &= \sqrt{98,7} = 9,935 \\ \sigma^2_{W1,W2,W3} &= \frac{(21 - 23,5)^2 + (26 - 23,5)^2}{2 - 1} = 12,5 \\ \sigma_{W1,W2,W3} &= \sqrt{12,5} = 3,536 \end{aligned}$$

- vi. Calculate the value of gauss density.

From the new data on the age criterion = 33 years, then calculated using the gauss density formula to determine the probability value of the age criterion:

$$\begin{aligned} P(\text{Age} = 33|W1, W3, W4) &= \frac{1}{\sqrt{2\pi} \cdot 5,69} \exp \frac{-(33-30,33)^2}{2 \times 32,33} = 0,063 \\ P(\text{Age} = 33|W1, W2, W3, W4) &= \frac{1}{\sqrt{2\pi} \cdot 9,935} \exp \frac{-(33-33,8)^2}{2 \times 98,7} = 0,040 \\ P(\text{Age} = 33|W1, W2, W3) &= \frac{1}{\sqrt{2\pi} \cdot 3,536} \exp \frac{-(33-23,5)^2}{2 \times 12,5} = 0,003 \end{aligned}$$

The next step is to normalize the probability value of the age criterion to determine the final probability/mass ( $m_4$ ).

$$\begin{aligned} m_4(W1, W3, W4) &= \frac{0,063}{0,063 + 0,040 + 0,003} = 0,594 \\ m_4(W1, W2, W3, W4) &= \frac{0,040}{0,063 + 0,040 + 0,003} = 0,377 \\ m_4(W1, W2, W3) &= \frac{0,025}{0,063 + 0,040 + 0,003} = 0,028 \end{aligned}$$

- vii. Conduct a combination of evidence between  $m_3 \oplus m_4$ .  
viii. The result of the combined evidence process, calculated by the formula 6.

$$\begin{aligned} m_5(\{W1, W3, W4\}) &= \frac{0,396+0,099+0,251}{1 - 0} = 0,746 \\ m_5(\{W1, W3\}) &= \frac{0,019 + 0,066 + 0,033 + 0,003 + 0,042}{1 - 0} = 0,163 \end{aligned}$$

$$m_5(\{W1, W2, W3, W4\}) = \frac{0,063}{1 - 0} = 0,063$$

$$m_5(\{W1, W2, W3\}) = \frac{0,005 + 0,002 + 0,021}{1 - 0} = 0,028$$

Based on the calculation of the last evidence, it turns out the greatest value is in  $m_5(\{W1, W3, W4\})$ . o the result of the recommendation of tour package for traveler to-11 is to Penglipuran, Pure Kehen and Trunyan equal to 0,746 or 74,6%.

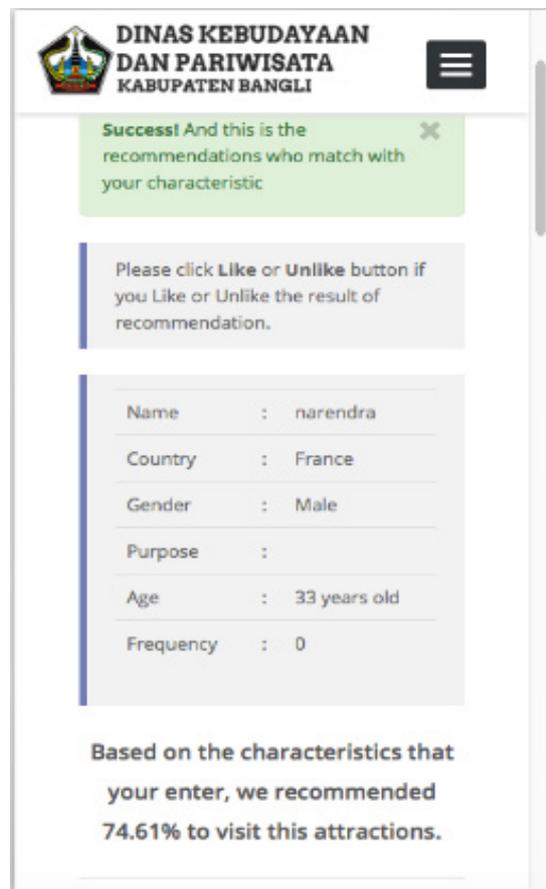


Fig. 5. The result of the retrieve process.

### 3.2 Reuse process

Tour packages (Penglipuran, Pura Kehen, and Trunyan) are expressed in the recommendation system. Recommended tour packages, will display pictures, descriptions, and coordinates of the location of the tourist attraction.

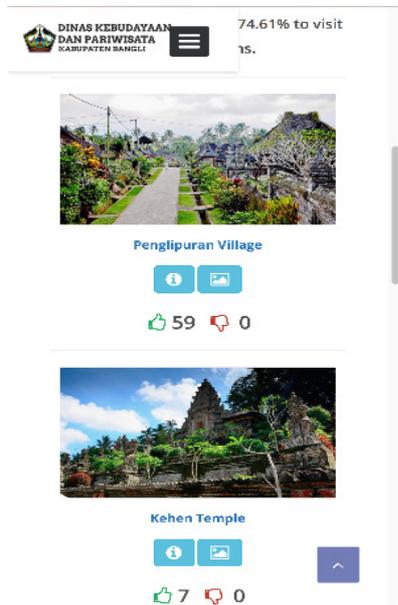


Fig. 6. The results of the tour package

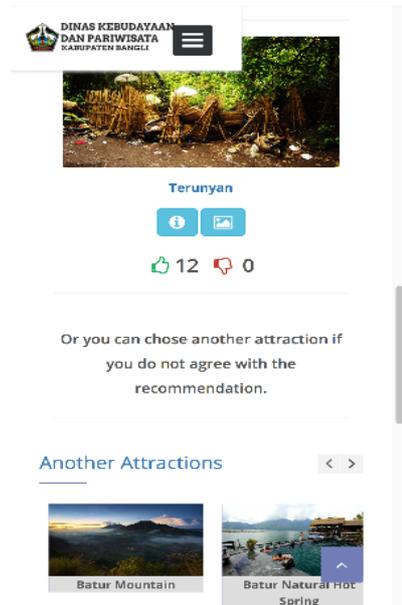


Fig. 7. The results of the tour package

When the detail button is clicked on each tourist attraction, it will display the description of traveler information, address, coordinates, and event.



Fig. 8. Description of the tour object

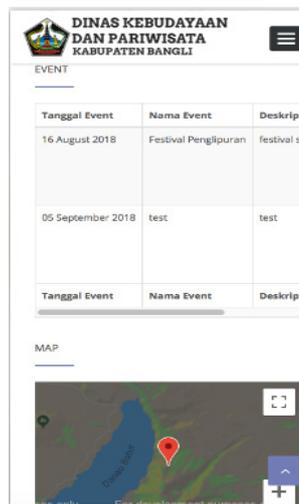


Fig. 9. Map and event of the tour object

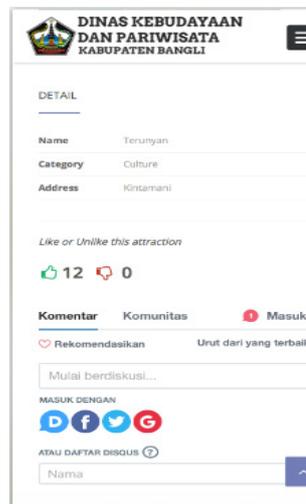


Fig. 10. Feedback of the tour object

### 3.3 Revise process

At this stage, the system displays the attractions (in addition to the recommended attractions) and sorted by the highest rating. The revision process is done by clicking LIKE button on detail tourist object.

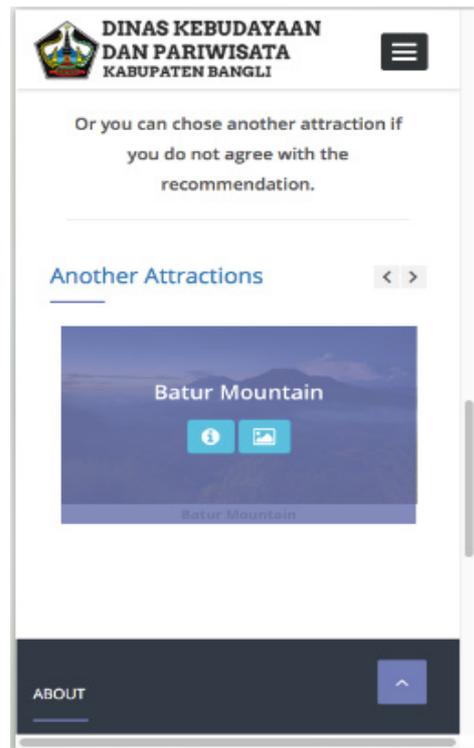


Fig. 11. The process of revise.

### 3.4 Retain process

At this stage, the recommended tourist attraction and favored by travelers will be stored in the database. The preferred tourist object will be used in the training data for the next recommendation case.

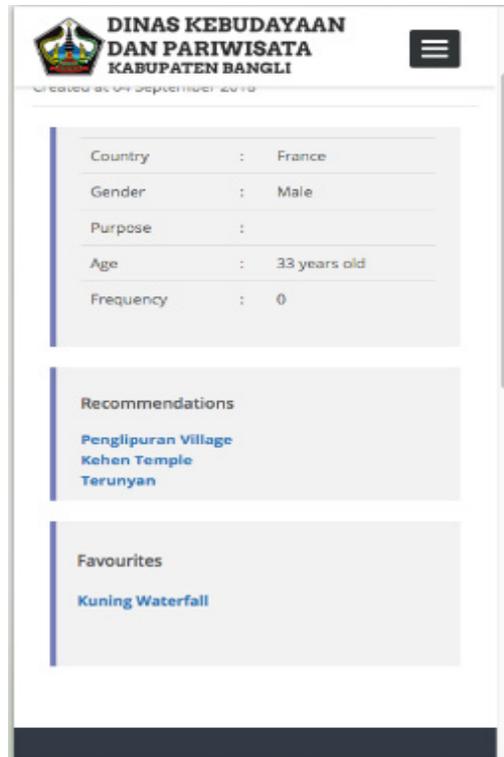


Fig. 12. The process of retain

### 3.5 System accuracy testing

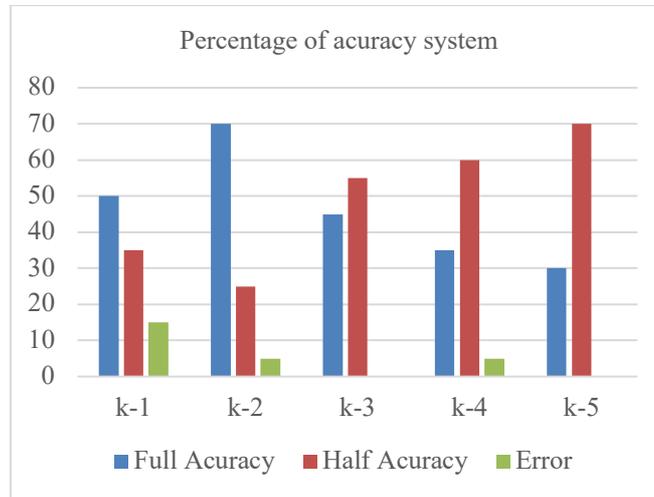
To measure the accuracy level of the system in providing tour package recommendations, need to be tested accuracy. The accuracy testing by involving large data sets can use k-Fold Cross-Validation. In the system accuracy test is divided into three types of accuracy such as full accuracy, half accuracy, and error. The amount of data is used in the test is 100 data and divided into 5 groups (k-Fold). Each k-Fold consists of 20 data testing and 80 training data. The testing mechanism for each group can be seen in table 4.

Next, each test phase will calculate the percentage of accuracy by using the formula in equation 9 and equation 10.

$$\text{Full / Half Accuracy} = \frac{\text{The number of recommendations accordingly}}{\text{Total of recommendations.}} \times 100\% \quad (9)$$

$$\text{error} = \frac{\text{Number of incorrect recommendations}}{\text{Total of recommendations.}} \times 100\% \quad .(10)$$

The value of the final accuracy is obtained from the overall average number of accuracy results in each test.



**Fig. 13.**The result of k-Fold Cross-Validation

Based on each value of the full accuracy, half accuracy, and error on each iteration, we can find the average value of the total accuracy of the system. The value of the full accuracy and the half accuracy can determine the total value of accuracy. The value of the error rate is 5% and the total accuracy of 95%. So it can be concluded that this system has been very good in providing a recommendation tourist attraction for the travelers.

#### 4 Conclusions and Future Work

The development of a tourist package recommendation system using the hybrid method (a combination of the methods of Bayes Theorem, Naive Bayes and Dempster Shafer) can accommodate tour packages. Where Naive Bayes is used to calculating the probability of continuous criteria such as age and frequency of visits. The Bayes Theorem is used to calculating the probability of categories such as country, gender, and purpose of visit. To determine the mass value of a combination of evidence using the Dempster Shafer method. The k-Fold Cross Validation method is used to measure the accuracy of recommended system tour packages. Based on the accuracy test, it was affirmed that the total accuracy of the tour package by the mobile-based recommendation is 95%. While the error rate of the system in providing tour packages is 5%. For further research, the system must be developed in a comprehensive manner, so that all relevant parties (Tourism Office, Tourism Awareness Group, and Travelers) can manage tourism objects in an integrated manner. In addition, to find out the experience of users in using the system, further studies are needed regarding user experience in the tour package using the mobile-based recommendation system.

## 5 References

- [1] J. S. Atmojo and I. J. Dewi, Implementasi dan Implikasi Kelembagaan Pemasaran Pariwisata yang Bertanggung Jawab, I ed., Jakarta: Pinus Book Publisher untuk Kementerian Pariwisata dan Kebudayaan Republik Indonesia, 2011.
- [2] Y. K. Sung, Chang, K. C and Sung, Y. F., "Market Segmentation of International Tourists Based on Motivation to Travel: A Case Study of Taiwan," *Asia Pacific Journal of Tourism Research*, vol. 21, no. 8, pp. 862-8820, 2015. <https://doi.org/10.1080/10941665.2015.1080175>
- [3] M. Luo, Feng, R and Cai, L. A., "Information search behavior and tourist characteristics: The internet vis-à-vis other information sources," *Journal of Travel & Tourism Marketing*, vol. 17, no. 2-3, pp. 15-25, 2004. [https://doi.org/10.1300/J073v17n02\\_02](https://doi.org/10.1300/J073v17n02_02)
- [4] Boo, S and Busser, J. A., "The Hierarchical Influence of Visitor Characteristics on Tourism Destination Images," *Journal of Travel & Tourism Marketing*, vol. 19, no. 4, pp. 55-67, 2005. [https://doi.org/10.1300/J073v19n04\\_05](https://doi.org/10.1300/J073v19n04_05)
- [5] C. V. Vuuren and E. Slabbert, "Travel Motivation and Behavior of Tourists to A South African Resort," in *International Conference on Tourism & Management Studies*, Portugal, 2011.
- [6] D. SIRISACK, S. XAYAVONG, S. PHONGSAVATH and N. VONGSANGA, "The Characteristics and Motivations of Foreign Tourists Who Visit Luang Prabang Province, Lao PDR," *International Journal of Business and Social Science*, vol. 5, no. 9, pp. 262-275, 2014.
- [7] S. Boo and J. A. Busser, "The Hierarchical Influence of Visitor Characteristics on Tourism Destination Images," *Journal of Travel & Tourism Marketing*, vol. 19, no. 4, pp. 55-67, 2005. [https://doi.org/10.1300/J073v19n04\\_05](https://doi.org/10.1300/J073v19n04_05)
- [8] M. Braunhofer and F. Ricci, "Selective contextual information acquisition in travel recommender systems," *Information Technology & Tourism*, vol. 17, no. 1, pp. 5-29, 2017. <https://doi.org/10.1007/s40558-017-0075-6>
- [9] W. Wornl, A. Hefele and D. Herzog, "Recommending a sequence of interesting places for tourist trips," *Information Technology & Tourism*, vol. 17, no. 1, pp. 31-54., 2017. <https://doi.org/10.1007/s40558-017-0076-5>
- [10] A. A. Niknafs, M. E. Shiri and M. M. Javidi , "A Case-Based Reasoning Approach in E-Tourism: Tour Itinerary Planning," in *1Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03)*, 2003. <https://doi.org/10.1109/DEXA.2003.1232122>
- [11] C. Lemnaru, M. Dobrin, M. Florea and R. Potolea, "Designing a Travel Recommendation System using Case-Based Reasoning and Domain Ontology," in *Intelligent Computer Communication and Processing (ICCP)*, Cluj-Napoca, Romania, 2012 .
- [12] P. Chanvarasuth and L. Boongasame, "Hybridizing Principles of the ELECTRE III Method with Case-Based Reasoning for a Travel Advisory System: Case Study of Thailand," *Asia Pacific Journal of Tourism Research*, vol. 20, no. 5, pp. 585-598, 2015. <https://doi.org/10.1080/10941665.2014.908229>
- [13] J. Giarratano and G. Riley, *Expert Systems Principles and Programming Fourth Edition*, 1 ed., Canada: Thomson Course Technology, 2005.
- [14] I. M. A. Wirawan and S. Hartati, "Mobile-based CBR System using Dempster-Shafer Modification Rule for Tourist Spots Recommendations in Buleleng Regency," *International Journal of Computer Applications*, vol. 83, no. 7, pp. 9-12, 2013.
- [15] N. W. N. Prasistayanti, D. G. H. Divayana and I. M. A. Wirawan, "Sistem Pendukung Keputusan Diagnosa Penyakit Gangguan Jiwa Dengan Metode Dempster-Shafer," *Jurnal Nasional Pendidikan Teknik Informatika*, vol. 6, no. 3, pp. 218-236, 2017.

- [16] P. YILDIRIM and D. BIRANT, "Naive Bayes Classifier for Continuous Variables using Novel Method (NBC4D) and Distributions," in In Innovations in Intelligent Systems and Applications (INISTA) Proceedings, Turkey, 2014.
- [17] R. . R. Bouckaert, "Naive bayes classifiers that perform well with continuous variables.," in In Australasian Joint Conference on Artificial Intelligence , Berlin, Heidelberg., 2004. [https://doi.org/10.1007/978-3-540-30549-1\\_106](https://doi.org/10.1007/978-3-540-30549-1_106)
- [18] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," Citec Journal, Vol. 2, No. 3, Mei 2015 – Juli 2015, 2015.
- [19] B. A. Muktamar, "Implementasi Data Mining dengan Naive Bayes Classifier untuk Mendukung Strategi Pemasaran di Bagian Humas STMIK AMIKOM Yogyakarta," Naskah Publikasi, p. 3, 2013.
- [20] D. Buhalis and R. Law, "Progress in Information Technology and Tourism Management: 20 years on and 10 years after Internet. The state of e-Tourism research," Tourism Management, vol. 29 , no. 4, p. 2008, 609-623.
- [21] Y. Park and U. Gretzel, "Success factors for destination marketing web sites: a qualitative meta-analysis," Journal of Travel Research, vol. 46, pp. 46-63, 2007. <https://doi.org/10.1177/0047287507302381>
- [22] Lestari, D. M., Hardianto, D and Hidayanto, A. N., "Analysis of user experience quality on responsive web design from its informative perspective," Int. J. Softw. Eng. Appl, vol. 8, no. 5, pp. 53-62, 2014.
- [23] Glassman, N. R and Shen, P, "One site fits all: responsive web design," Journal of Electronic Resources in Medical Libraries, vol. 11, no. 2, pp. 78-90, 2014. <https://doi.org/10.1080/15424065.2014.908347>
- [24] T. F. Srihadi, Hartoyo, D. Sukandar and A. W. Soehadi, "Segmentation of the tourism market for Jakarta: Classification of foreign visitors' lifestyle typologies," Tourism Management Perspectives, pp. 32-39, 2016.
- [25] P. A. Weaver, K. W. McCleary, J. Han and P. E. Blosser, "Identifying leisure travel market segments based on preference for novelty," Journal of Travel & Tourism Marketing, vol. 26, no. 5-6, pp. 568-584, 2009. <https://doi.org/10.1080/10548400903163129>
- [26] A. C. Tynan and J. Drayton, "Market segmentation," Journal of Marketing Management, vol. 2, no. 3, pp. 301-335, 2010. <https://doi.org/10.1080/0267257X.1987.9964020>
- [27] A. Tkaczynski and S. R. Rundle-Thiele, "Event segmentation: A review and research agenda," Tourism Management, vol. 32, pp. 426-434, 2011. <https://doi.org/10.1016/j.tourman.2010.03.010>
- [28] A. Beerli and J. D. Martin, "Tourists' characteristics and the perceived image of tourist destinations: a quantitative analysis—a case study of Lanzarote, Spain," Tourism Management, vol. 25, no. 5, pp. 623-636, 2004. <https://doi.org/10.1016/j.tourman.2003.06.004>
- [29] I. G. Pitana and P. G. Gayatri, Sosiologi Pariwisata: Kajian sosiologis terhadap struktur, sistem, dan dampak – dampak pariwisata, I ed., Daerah Istimewa Yogyakarta, Daerah Istimewa Yogyakarta: ANDI Yogyakarta, 2005.
- [30] P. Zhang, A. Essaid, C. Zanni-Merk and D. Cavallucci, "Case-based Reasoning for Knowledge Capitalization in Inventive Design Using Latent Semantic Analysis," in International Conference on Knowledge Based and Intelligent Information and Engineering Systems, Marseille, France, 2017. <https://doi.org/10.1016/j.procs.2017.08.004>
- [31] I. M. AHMED, M. ALFONSE, M. AREF and A.-B. M. SALEM, "Reasoning Techniques for Diabetics Expert Systems," in Procedia Computer Science, Netherlands, 2015.

- [32] I. Watson, Applying case-based reasoning: techniques for enterprise systems, 1st Edition ed., San Francisco: Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1998, 1998, p. 289.
- [33] D. Guessoum, M. Miraoui and C. Tadi, "Contextual Case-Based Reasoning Applied to a Mobile Device," International Journal of Pervasive Computing and Communications, vol. 13, no. 3, pp. 282-299, 2017. <https://doi.org/10.1108/IJPC-11-2016-0056>
- [34] J. . L. Kolodner, "An Introduction to Case-Based Reasoning," Artificial Intelligence Review, vol. 6, no. 1, pp. 3-34, 1992. <https://doi.org/10.1007/BF00155578>
- [35] K. Sankar and C. Simon, Foundation of Soft Case-Based Reasoning, I ed., Hobokon, New Jersey: A John Wiley & Sons, Inc, 2004.
- [36] S. Chowdhury, "Fusing probability density function into Dempster-Shafer theory of evidence for the evaluation of water treatment plant," Environmental monitoring and assessment, vol. 185, no. 5, pp. 3917-3929, 2013. <https://doi.org/10.1007/s10661-012-2840-5>
- [37] S. Petrovic, N. Mishra and S. Sundar, "A novel case based reasoning approach to radiotherapy planning," Expert Systems with Applications, vol. 38, no. 9, pp. 10759-10769, 2011. <https://doi.org/10.1016/j.eswa.2011.01.109>
- [38] C. Simon and P. Weber, "Bayesian Networks Implementation of the Dempster Shafer Theory to Model Reliability Uncertainty," in First Int. Conference on Availability, Reliability and Security (ARES), 2006. <https://doi.org/10.1109/ARES.2006.38>
- [39] Kusriani, Emha and T. Luthfi, Algoritma Data Mining, Yogyakarta: Andi Yogyakarta, 2009.
- [40] I. M. A. Wirawan, Metode Penalaran dalam Kecerdasan Buatan, 1 ed., Depok, Jawa Barat: RajaGrafindo Persada, 2017, p. 196.

## 6 Authors

**I Made Agus Wirawan** is a lecturer at the Education of Informatics Engineering, Universitas Pendidikan Ganesha, Indonesia. His research focus is in the field of artificial intelligence and mobile applications.

**I Wayan Bayu Diarsa** is a student majoring in informatics engineering. Currently he is continuing his studies in computer science at the Universitas Pendidikan Ganesha.

Article submitted 04 September 2018. Resubmitted 05 November 2018. Final acceptance 06 November 2018. Final version published as submitted by the authors.