Smart Tourisms: Decision Support Systems In The Strategic Location Of Inn/Hotel For Travelers Using The Naive Bayes Algorithm

Husain¹, Muhammad Zarlis², Mochamad Wahyudi³, Heroe Santoso¹, Hengki Tamando Sihotang⁴, Nurul Fitriani⁵
¹Faculty of Computer Science, Universitas Bumigora, Mataram, Indonesia
²Faculty of Computer Science and Information Technology, Universitas Sumatera Utara Medan, Indonesia
³Teknologi Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia
⁴Teknik Informatika, STMIK Pelita Nusantara, Medan, Sumatera Utara, Indonesia
⁵SMKN 3 Mataram, Indonesia

Husain@universitasbumigora.ac.id

Abstract: This research makes the system in decision making for tourists who will visit an area. In making decision decision systems using the Bayes algorithm. Bayes's theorem was put forward by a 1763 British Presbyterian minister named Thomas Bayes. In this study Bayes's theorem is used to calculate the probability of occurrence of an event based on the influence obtained from the observations. In this paper, the Naive Bayes algorithm is used to support the Decision Making System (DSS) to determine the best and strategic location in determining hotel accommodations, especially for tourists who are associated with having mediocre funds in traveling to Lombok Island. In this study applied a number of criteria and data sets obtained from questionnaires from some visiting tourists. There are 3 variables including Hotel Prices, Distance, and the presence or absence of transportation from tourist destinations, from the three variables, to the training dataset. So that in the future there will be questions in the form of data testing to answer the decision whether the hotel can be take or not by looking at the highest probability value.

Keywords: Tourism, DSS, Naive Bayes, Probability, training data, data testing, dataset

1. Introduction

The role of tourism in increasing state revenues is very important especially economic development. Today it cannot be denied that tourism activities have become a lifestyle for most people in this world to visit tourist destinations in all corners of the world [1][2]. The advancement of information technology is one aspect that makes tourism progress in a region inseparable from the role of ICT (Information and Computer Technology) as well as making a system to help smooth travelers such as research conducted in this paper to create a system to help prospective tourists who will visit an area to get convenience in determining hotels/inns according to the criteria. There is a tendency that when someone is planning a tourist visit, some of the tourists will visit, of course, looking for interesting lodging places, close to tourist sites, easy transportation and of course affordable lodging costs. To overcome these problems the researchers load a decision-making system in terms of determining the location of suitable accommodation and going to select by prospective tourists based on the criteria applied. Decision Support System and Intelligent Systems were first revealed by Michael S. Scott Morton around the 1970s with the term Management Decision Systems [3][1]. define DSS as an interactive system, which helps decision makers through the use of data and decision models to solve problems that are semi-structured and unstructured [4][2]. Decision Making System in helping tourists to choose the location of the accommodation is very suitable with the existing criteria. This criterion was obtained from several
speakers so that it was made into training data. The following are some questions for resource persons to be used as training data:

The following is the criteria for finding lodging:

1.) Hotel prices
   a. Expensive
   b. Is being
   c. Cheap

2.) Location / distance
   a. Far
   b. Is being
   c. Close

3.) There Is No Transportation
   a. Not fortunate
   b. There is a transportation

4.) Hotel facilities
   a. Well
   b. Enough
   c. Less

From the above criteria, 10 criteria are made which refer to the 3 questions above. As in the picture below:

| Data | Hotel Price | Distance to Destination | Transportation | Hotels Facilities | selected |
|------|-------------|-------------------------|----------------|------------------|----------|
| 1    | Cheap       | Close                   | No             | Less             | Yes      |
| 2    | Is being    | Close                   | No             | Enough           | Yes      |
| 3    | expensive   | Close                   | No             | Well             | Yes      |
| 4    | expensive   | Far                     | No             | Enough           | No       |
| 5    | expensive   | Is being                | No             | Enough           | No       |
| 6    | Is being    | Far                     | Yes            | Well             | No       |
| 7    | Cheap       | Far                     | Yes            | Less             | No       |
| 8    | Cheap       | Is being                | No             | Enough           | Yes      |
| 9    | expensive   | Far                     | Yes            | Well             | No       |
| 10   | Is being    | Is being                | Yes            | Well             | Yes      |

In the process of resolving decision-making, determining the best location of Hotels / Hotels using the Naive Bayes method, because as expressed by A. Yunita et al. Naive bayes is a method that applies a statistical approach to induction instructions on classification problems [3].
2. Literature Review

2.1 Tourism

Tourism is a sector that has taken an important role in the economic development of the nations of the world. Increasing progress and prosperity have made tourism a central part of human needs or lifestyles, and moved people to get to know nature and culture in other countries [5]. So that indirectly, human movement will affect the mutually sustainable economic chain into a service industry that contributes to the world economy, the economy of nations, to the improvement of economic welfare at the level of local communities. Regions that have advanced tourism can not be separated from the system of application of technology in supporting the implementation of tourism trips in an area such as a tourist search system, hotels etc [6] [7].

2.2 Data Mining

Data mining is the process of finding relationships in data unknown to the user and presenting them in a way that can be understood so that the relationship can be the basis of decision making [8]. Data Mining is a series of processes to explore added value in the form of information that is not known manually from a database by extracting patterns from data with the aim of manipulating data into more valuable information obtained by extracting and recognizing important patterns or draw from data contained in the database [9]. Data Mining is also known by other names such as: Knowledge discovery (mining) in databases (KDD), extraction of knowledge (knowledge extraction) Analysis of data / patterns and business intelligence (business intelligence) and is an important tool for manipulating data for presenting information as needed user with the aim to assist in analyzing the collection of behavioral observations [9]. In general the definition of data-mining can be interpreted as follows:

a. The process of finding interesting patterns of stored data in large numbers.
b. Extraction of useful or interesting information (non-trivial, implicit, before the potential usefulness is known), pattern or knowledge of the data stored in large amounts.
c. Exploration of analysis automatically or semi-automatically to large amounts of data to find meaningful patterns and rules.

2.3 Decision Making System

According to (Raymond Mc Leod, Jr., 1995: 348) Decision support system is a system of producing information aimed at a particular problem that must be solved by the manager and can assist managers in decision making. Decision support systems are an integral part of the totality of the overall organizational system. An organizational system includes physical systems, decision systems and information systems [8][6]. To produce good decisions in a decision support system, it needs to be supported by information and quality facts including:

a. Accessibility
   This attribute is related to the ease of getting information, information will be more meaningful to the user if the information is easily obtained, because it will be related to the activity of the value of the information.

b. Completeness
   This attribute is related to the completeness of the content of information, in this case the content does not only concern volume but also conforms to the expectations of the user so often these completeness is difficult to quantify.

c. Accuracy
   This attribute relates to the level of possible errors in the processing of large amounts of data. Two types of errors that often occur are related to calculations.
d. Accuracy
This attribute is related to the compatibility between the information produced with the user's needs. As with completeness, accuracy is also very difficult to quantify.

e. Punctuality
The quality of information is also very much determined by the exact time of delivery and actualization. For example, information relating to daily planning will very useful if delivered every two days.

f. Clarity
This attribute is related to the form or format of information delivery. For a leader, information presented in the form of graphics, histograms, or images will usually be more meaningful than information in the form of long words.

g. Flexibility
This attribute is related to the level of adaptation of the information produced to the needs of various decisions to be taken and to a group

2.4 Naïve Bayes
Naïve Bayes algorithm is one of the algorithms found in the classification technique [4]. Naïve Bayes is a classification with probability methods and statistics. The Bayes theorem predicts opportunities in the future based on previous experience [5].

The following are the naive bayes algorithms and theorems as follows:

\[
P(H | X) = \frac{P(X | H)P(H)}{P(X)}
\]

- **X** = Data with unknown classes
- **H** = X data hypothesis is a class Specific
- **P (H | X)** = The probability of the hypothesis H is based on condition x (posteriori prob.)
- **P (H)** = Probability of hypothesis H (prior prob.)
- **P (X | H)** = Probability X based on the condition
- **P (X)** = Probability of X

In classifying a dataset using Naive Bayes describes the process before starting experiments using the Bayes theorem to the output to be produced[10].

The stages are as follows:

a. Determine which cafeteria will be used in the form of 3 questions such as Price, Distance and Existence of No Transportation.
b. After being fed, the filling is done by the correspondent so that it is made into a table
c. From the criteria that have been prepared, then put into a table as many as 10 criteria to be combined so that if the correspondent chooses one of the attributes with the answer then the answer is yes or no.
d. Next, make another table for the appearance of each value for four times the attribute of each criterion (C1-C4) to find the probability value
e. Calculating the Likelihood of Yes and Likelihood No.
f. Next calculate the probability value by normalizing the likelihood value Yes and likelihood Not to get the value 1
g. The final result of the normalization calculation is obtained with the highest value, then the answer is YES (Take) and vice versa if the answer is low eating No (Not taken).
2.5 System Analysis

From the two formulas between Likelihood YES and Likelihood Not compared to the results. If Probability Value YES $\geq$ Probability Value No, the Hotel is selected. Conversely if the value of Probability is Yes $\leq$ Probability value No then it is not taken. The following is the Flowchart process for determining the best location to choose a hotel that is suitable for use by Travelers.

Flowchart calculation process for system decision making

![Flowchart](image)

Figure 1 Flowchart calculation process

The explanation of the flowchart above is started by determining the value i, if the variable i, is not the same as the number of criteria then the criteria i will be calculated until the number of criteria is
fulfilled, then the likelihood calculation process is done and the likelihood is not, the value of the likelihood result is yes and likelihood no. Next is the calculation of probability values, if the probability value Y is greater than the probability value No then the inn / hotel will be chosen But if the probability value Yes is lower or greater the probability value is not then the inn / hotel is not taken.

3. Experiment
3.1. Dataset
In this study, a dataset of 10 students was used to find lodging for tourists who visited Lombok. From the 10 criteria there are 4 variables including Hotel Prices, Distance of tourist places, There is no transportation and Take or not the inn.

Table 2: Data Training and Testing

| data | Hotel Price | Distance to destination | Transportation | Hotel Facilities | Selected |
|------|-------------|-------------------------|---------------|-----------------|----------|
| 1    | Cheap       | Close                   | No            | Less            | Yes      |
| 2    | Is Being    | Close                   | No            | Enough          | Yes      |
| 3    | Expensive   | Close                   | No            | Well            | Yes      |
| 4    | Expensive   | Far                     | No            | Enough          | No       |
| 5    | Expensive   | Is being                | No            | Enough          | No       |
| 6    | Is being    | Far                     | Yes           | Less            | No       |
| 7    | Cheap       | Far                     | Yes           | Well            | No       |
| 8    | Cheap       | Is being                | No            | Enough          | Yes      |
| 9    | Expensive   | Far                     | Yes           | Well            | No       |
| 10   | Is being    | Is being                | Yes           | Well            | Yes      |
| 11   | cheap       | Close                   | Not           | Enough          | ?        |

From the table above there are 10 criteria called Data Training, then at number 11 called testing data. The question is based on the Testing data whether later tourists will take or have Yes or No.

Table 3: The first step is to calculate the probability of C1 (price).

| Hotel Price | Number of selected events | Probabilities |
|-------------|---------------------------|---------------|
|             | Yes | No | Yes | No |
| Cheap       | 3   | 1  | 0.5 | 0.25 |
| Is Being    | 2   | 1  | 0.3 | 0.25 |
| Expensive   | 1   | 2  | 0.2 | 0.5 |
| Total       | 6   | 4  | 1   | 1   |
Table 4: The second step calculates the probability of C2 (distance)

| Distance | Number of selected events | Probabilities |
|----------|---------------------------|---------------|
|          | Yes | No  | Yes | No  |
| Close    | 3   | 1   | 0.5 | 0.3 |
| Is Being | 2   | 1   | 0.3 | 0.3 |
| Far      | 1   | 2   | 0.2 | 0.5 |
| Total    | 6   | 4   | 1   | 1   |

Table 5: Steps when calculating the probability of C3 (There is a Claim for Angkot to a Tourist Destination)

| Distance | Number of selected events | Probabilities |
|----------|---------------------------|---------------|
|          | Yes | No  | Yes | No  |
| Yes      | 3   | 2   | 0.5 | 0.5 |
| No       | 3   | 2   | 0.5 | 0.5 |
| Total    | 6   | 4   | 1   | 1   |

Table 6: Probability of occurrence of C4 (Hotel Facilities)

| Distance | Number of selected events | Probabilities |
|----------|---------------------------|---------------|
|          | Yes | No  | Yes | No  |
| Less     | 1   | 2   | 0.2 | 0.5 |
| Enough   | 2   | 1   | 0.3 | 0.25|
| Well     | 3   | 1   | 0.5 | 0.25|
| Total    | 6   | 4   | 1   | 1   |

Table 7: The fourth step calculates the probability of C4 (Selected yes or no)

| Transportation | Number of Selected Events | Probabilities |
|----------------|---------------------------|---------------|
|                | Yes | No  | Yes | No  |
| Total          | 5   | 5   | 0.5 | 0.5 |

After calculating the likelihood, it is possible that the likelihood is possible. Not from C1-C4 it matches the Testing data number 11.

Table 8: calculates the likelihood of possible YES =

| Hotel Price | Distance to destination | Transportation | multiplied | Results |
|-------------|-------------------------|----------------|------------|---------|
| Is Being    | Is Being                | No             | 1/2        | 0.5     |
Table 9: Calculates the Likelihood of possible No =

| Hotel Price | Distance to destination | Transportation | multiplied | Results |
|-------------|-------------------------|---------------|------------|---------|
| Is Being    | Is Being               | No            | ½          | 0.5     |
| 0.3         | 0.3                    | 0.5           | 0.3        | 0.5     |

4. Results
From the results of the likelihood calculation by looking at the value of the results obtained at a higher likelihood value, so the location of the hotel is likely to be taken. But to get results more accurately calculate the probability value.

\[
\text{Probabilitas}_{\text{YES}} = \frac{0.021}{0.021 + 0.004} = 0.842
\]

\[
\text{Probabilitas}_{\text{NO}} = \frac{0.04}{0.004 + 0.021} = 0.158
\]

5. Conclusions
a. From what has been discussed above, we understand how the Naive Bayes algorithm works to solve a specific problem in decision making.

b. The results of the experiment from this journal resulted in the Naive Bayes algorithm for decision making in determining the best location for lodging / hotel after calculating the Likelihood value and the probability value obtained greater than Yes = 0.889, compared the probability value not = 0.111.

c. Price = Medium, Distance = Moderate, Public transportation = No, then the location is very possible to take.

Bibliography
[1] S. Chookaew, O. Chanin, J. Charatarawat, P. Sriprasert, and S. Nimpaya, “Increasing Halal Tourism Potential at Andaman Gulf in Thailand for Muslim Country,” J. Econ. Bus. Manag., vol. 3, no. 7, pp. 739–741, 2015.
[2] Y. Li, C. Hu, C. Huang, and L. Duan, “The concept of smart tourism in the context of tourism information services,” Tour. Manag., vol. 58, pp. 293–300, 2017.
[3] R. Riedel, V. Wiers, and J. C. Fransoo, “Modelling dynamics in decision support systems,” Behav. Inf. Technol., vol. 31, no. 9, pp. 927–941, 2012.
[4] H. T. Sihotang, M. Zarlis, S. Efendi, D. Jollyta, and Husain, “Evaluation of Maturity Level of Information and Communication Technology (ICT) Governance with CobIT 5.0 Case Study: STMIK Pelita Nusantara Medan,” J. Phys. Conf. Ser., vol. 1255, no. 1, pp. 0–6, 2019.
[5] M. Masoud, K. Zarkesh, N. Almasi, and F. Taghizadeh, “Ecotourism Land Capability Evaluation Using Spatial Multi Criteria Evaluation,” Res. J. Appl. Sci. Eng. Technol. 3(7) 693-700, 2011 ISSN 2040-7467 © Maxwell Sci. Organ. 2011, vol. 3, no. 7, pp. 693–700, 2011.
[6] D. J. T. and R. Sharpley, Tourism and Development in the Developing World, Second Edi. California: Disaster and Development, 2017.
[7] K. Boes, D. Buhalis, and A. Inversini, “Conceptualising Smart Tourism Destination Dimensions,” 2015.
[8] T.-P. L. C. Efraim Turban, Jay E. Aronson, “Decision Support Systems and Intelligent Systems,” Get. Res. Find. into Pract. Second Ed., pp. 123–137, 2007.

[9] D. J. Hand, “Principles of Data Mining (CONFERENCE PAPER),” Drug Saf., vol. 30, no. 7, pp. 621–622, 2007.

[10] P. Berchialla, F. Foltran, and D. Gregori, “Naïve Bayes classifiers with feature selection to predict hospitalization and complications due to objects swallowing and ingestion among European children,” Saf. Sci., vol. 51, no. 1, pp. 1–5, 2013.