2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia

"TripBuddy" Travel Planner with Recommendation based on User's Browsing Behaviour

Merlinda Sumardi, Jufery, Frenky, Rini Wongso[*], Ferdinand Ariandy Luwinda

Computer Science Department, School of Computer Science, Bina Nusantara University, Jl. K.H. Syahdan No. 9 Kemanggisan, Jakarta 11480, Indonesia

Abstract

Tourism has been an irreplaceable part of economy growth of every country. By that factor, the authors were encouraged to build an application to serve detail information of tour destinations to provide easy preparation for travelling. The application, TripBuddy, was developed to learn user's behavior based on empiric data, which used to offer relevant destinations to certain user by using K-Means Clustering. TripBuddy is a web-based application which suggests optimal route with detail information of destinations, schedule, cost and duration to ease user's travel plan and it also gives recommendation based on user's browsing behavior

© 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 2nd International Conference on Computer Science and Computational Intelligence 2017.

Keywords: web application, tourism planning, k-means clustering, optimal route.

1. Introduction

Tourism, the combination of activities, services, and industry that offers travel experiences, transportation, accommodation, eat and drink, shopping, and any kind of hospitable service when someone is away from home [1], is one of the most important sectors and gives positive contribution to national economy. The most real contribution from this sector is the foreign exchange income. Additional foreign exchange from tourism sector has contributed to the strengthening of foreign exchange reserves. Strong foreign exchange reserves in turn will encourage the strengthening (appreciation) of Indonesian Rupiah against foreign currencies, particularly against the US Dollar [2]. The

* Corresponding author. Tel.: +62-21-5345830; fax: +62-21-5300244.
   E-mail address: rwongso@binus.edu

tourism sector also has the potential to encourage private sector growth and infrastructure development as well as increasing state revenues from taxes, particularly indirect taxes. In addition, the tourism sector is also boosting economic growth and employment [3].

In recent years, demand of tourism activities has greatly increased, whether it is to travel to foreign countries or explore the area in their own country. In Indonesia, for example, according to statistical reports of "Kementrian Pariwisata", in 2014 as indicated on http://www.kemenpar.go.id, the number of foreign tourist arrivals has reached 9.43 million increasing from the year 2008 and according to the statistical report of tourist profiles in 2014, a total of 251.2 million trips were taken by tourists.

The two variables that affect consumers' interest against travel packages are destination / tourism objects and travel package rates [4]. Not all schedules, rate, and destination are compatible with prospective tourists, whereas if they do not follow the package offered by travel agents, tourists are concern for not having actual and complete information in prior to or at the time of travel considering how rich the diversity the region / countries to be visited, especially those that have different local languages. One of the considerations of tourist in picking the travel package is the flexibility in managing the destination they would want to visit, the desired date, and how they easily get information about the travel destination without travel agent.

According to the survey of "VISA" in "Visa Global Travel Intentions Study 2015", tourists are likely to browse reviews and information of travel destination online before or at the time they do the travel. This is indicating that it has become a trend for people to prepare their travel by accessing online information.

In previous works, TripEneer was developed to personalized tour planning application. This application use five different ranking models, which are guides ranking, check-ins ranking, friends-base ranking, user-based ranking and hybrid ranking. TripEneer give recommendation place that match with user preferences according to those five models, and the users really like the system [5]. Another work was done by Ravi, a recommendation system based on social pertinent trust walker (SPTW) which is designed to discover interesting category of some users from the location based social network. As most of travel recommender systems lack the points of personalization, interactivity, and adaptivity, it is very important to hold specific information about users and their interests as a profile [6]. Another similar personal tour recommendation system was developed based on user interests and visit durations [7]. The result has shown that by using time-based user interest and personalized POI visit durations, the system have more accurate recommendation than using using frequency-based user interest and average visit durations.

K-Means Clustering algorithm is commonly implemented in problems to understand consumer behaviors and to identify new product opportunities in the market. K-Means can also be used to summarize representation of objects in large number hence it is easier to describe the characteristics of each group or cluster [8]. In a study to determine the layout of the product in supermarkets, K-Means clustering algorithm is used for partitioning the data into clusters so that data with the same characteristics are grouped into the same cluster and data that have different characteristics are grouped into another cluster [9]. The result of clustering varies as number of cluster parameter changes hence main challenge of cluster analysis is that the number of clusters or the number of model parameters is seldom known, and it must be determined before clustering [10].

This study aims to develop an application that provides travel planning features with recommendation based on user behavior using K-Means Clustering algorithm. The application will be called as TripBuddy.

2. Methodology

Based on research done previously by Adisaputro [11], to obtain tourist information, tourists need to seek references from various sources that will take time. In addition, the communication between tourists who will be traveling travel together is also important in determining decision-making in planning and scheduling of the travel. The purpose of the study is to provide complete tourist information, planning and scheduling to ease user in designing their travel plan. Adaptive web-based application allows user to the application and not limited to certain operating system or device. They suggest that the application must provide get direction feature to the destination to give information of the travel routes that must be taken, adding register via email verification feature to prevent fraudulent users.

Hayati [12] define recommendation system as application in e-commerce website to suggest information and provide facilities for user in accordance to decision making. The system is considered as an illustration of user needs through

recommendation method by searching and recommending any item using rating based on the similarity of user information characteristics.

According to the previous researches mention above, we conduct surveys online and accept response from 323 people. We conclude the following requirements for the new app, which we called as TripBuddy:

- Recommendation of tourist destination.

  TripBuddy will provide destinations based on similar type of destination. When user is viewing certain destination, they will get recommendation of similar destination. TripBuddy will also give recommendation based on available nearest tourist destination. Other tourists' reviews will provide more information to user in deciding the travel destination. The other recommendation will be based on user's browsing experience. This will use K-Means as machine learning algorithm in the application. User will receive recommendation for some travel destinations that are likely to be visited by user or other users, hence user will get relevant suggestions.

- Detailed Searching

  To find travel destinations, each user has needs or interests that are different from other users. Therefore, the application will allow users to search for specific types of destinations that correspond to theirs. The filtering is based on attributes such as: price, rating, location, and category of tourist destinations.

- Travel Plan

  After providing information and recommendations on travel destinations, TripBuddy also provides feature that allows users to set the desired trip. In planning features, users can arrange destinations they would want to visit and the system will help in calculating the shortest distance so that the user can travel effectively. It will also calculate the total cost and duration of the trip, as well as the agenda for the trip. TripBuddy will also provide offline mode and this has become one of the benefits, so user can follow the travel agenda that had been prepared beforehand, without checking the information on the destinations on websites that require an internet connection.

Some existing similar applications that have been popular are being analyzed to make sure TripBuddy has the right features that meet with user needs. They are TripAdvisor, Agoda, Kaskus, Trivago, and Triip.me. The comparison result can be seen in Table 1 below.

Table 1. Similar Application Analysis

Features	TripAdvisor	Agoda	Kaskus	Trivago	Trip.me	TripBuddy
Destination Recommendation based on Browsing Experience						V
Tourist Destination Category	V				V	V
Detail Destination Searching	V		V	V		V
Destination Review	V	V	V	V	V	V
Travel Plan	V		V		V	V
Shortest Distance Route						V
Offline Plan						V
Features						V

The recommendation features that use K-Means works as follows. In the first stage, k-means will form a cluster, with the centroid (the center point of the cluster) randomly, while the number of clusters adjusted through the formula of rule of thumb. Object on this research are users who each has seven attributes of categories (in the example described in Table 2 below, we refer these categories as category 1 to 7). The implementation of categories' name may vary according to our preferences as TripBuddy's admin. The data is sorted by index with initial value or weight of 0. Weight increases if a user visits a page destination in accordance with the category of the destinations.

Table 2. Example of Dataset of Category Attributes for 8 users

User	Category 1 [0]	Category 2 [1]	Category 3 [2]	Category 4 [3]	Category 5 [4]	Category 6 [5]	Category 7 [6]
1	18	2	12	0	25	13	10
2	4	0	9	0	15	0	12
3	7	0	8	0	1	0	20
4	0	2	0	11	0	11	0
5	1	5	0	9	0	5	15
6	0	0	0	21	0	9	0
7	8	1	0	9	6	0	11
8	8	0	2	11	0	0	4

After that, algorithm will insert data into nearest centroid (after the calculation of data distance to the centroid). Then, data will form new centroid if the average data value is different from that centroid. This step will repeat until the average data value of old and new centroid is the same. From the clustering result, we get user data based on cluster. The centroid value will be used as destination recommendation for user inside that cluster, by sorting the destination based on the category that has the highest value of weight to the lowest.

Table 3. Example of of Centroid Cluster Result for k = 2

User	Category 1 [0]	Category 2 [1]	Category 3 [2]	Category 4 [3]	Category 5 [4]	Category 6 [5]	Category 7 [6]
1	5.5	1	5.25	8	10	8.25	5.5
2	6	1.5	2.5	7.25	1.75	1.25	12.5

Table 4. Example of Dataset Result on 2 Cluster

User	Cluster 0	Cluster 1
1	▒	
2	▒	
3		▒
4	▒	
5		▒
6	▒	
7		▒
8		▒

2.1 K-Means Clustering

Machine learning is the technique used to convert data into information through learning of data repeatedly [13]. One of the implementation of machine learning is clustering. Han & Kamber [14] defines clustering as process to group data into class or cluster, so all the objects inside a cluster will have the highest similarity between one and another but has no resemblance with objects inside other clusters. Cluster analysis is the activity to identify the density of the area in object space and therefore we can find the overall distribution patterns and interesting correlation between data attributes.

Jung, Kang, and Heo [15] mentioned that the two frequently used clustering algorithm are KMeans and Expectation Maximization, whereas when both algorithms are compared in terms of accuracy and speed in classifying a set of object under investigation, produces satisfying accuracy, wiith an accuracy of K-Means that is higher but requires a longer time than expectation mazimization. K-Means is an unsupervised learning algorithm which is simple and has a high enough accuracy in accordance to the size of the object, so that the algorithm is more scalable and efficient for object processing in large number in classifying the characteristics of the object. In addition, K-Means algorithm is not influenced by the order of objects [8]. In the implementation of K-Means clustering, the data that can be processed in the calculation is numerical data in the form of numbers. While the non-numeric data can also be processed, but must be translated or represented by code to simplify the calculation of distance or the similarity of characteristics of each object.

Sebayang [9] explains the way the K-Means algorithm works. It first randomly selects n number of data as centroid. Then, distance of data and centroid is calculated by using Euclidean Distance. The formula of Euclidean distance can be described as:

$$d(x,y) = (\sum_{j=1}^{p}(x_j - y_j)^2)^{\frac{1}{2}}$$ (1)

, where d is the distance, j represents attribute values, p is the dimension of data, and x,y represents the position.

Next, the data will be placed inside the nearest cluster, which is calculated from the midpoint of the cluster. The centroid will be determined when all data have already been placed inside the nearest cluster by calculating the data average in the cluster.

The process of determining the centroid and placement of data in the cluster is repeated until the value of centroid is convergent (the centroid of all the cluster does not change anymore). Here is how the k-means algorithm works:

- Step 1 : Define k as number of cluster to be created
- Step 2: Randomly define initial k centroid (center point of cluster)
- Step 3: Calculate distance of each data to each centroid
- Step 4: Each data selects the nearest centroid
- Step 5: Define new position of centroid by calculating the average value of data which are in the same centroid
- Step 6: Repeat step 3 if the position of new and old centroid is different.

Rule of thumb is the simple method that can be used for determining the number of clusters k. It is simple and can be applied to all types of data sets (Kodinariya & Makwana, 2013). The formula can be seen below, with n as the number of objects or data.

$$k \approx \sqrt{n/2} k \approx \sqrt{n/2}$$ (2)

3. Result and Discussion

TripBuddy is developed on website platform and can be accessed through Browser. It also uses Google Maps API in accordance to the map and routing display of the travel plan. The architecture is described in the following Figure 1:
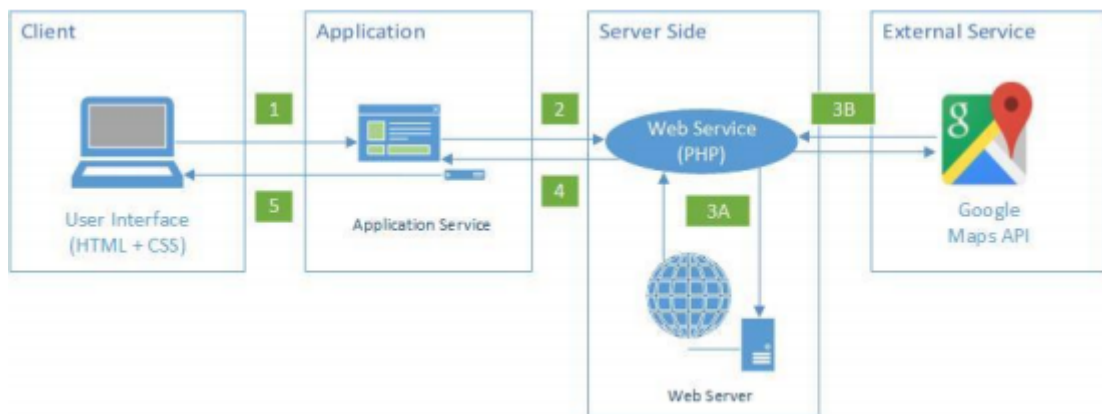

Figure 1. TripBuddy Architecture

TripBuddy offers destination that the tourists may like, which consists of travel destination recommendation based on browsing experience (user behavior). New user or member will be asked to explore some destinations before TripBuddy can offer the recommendation.
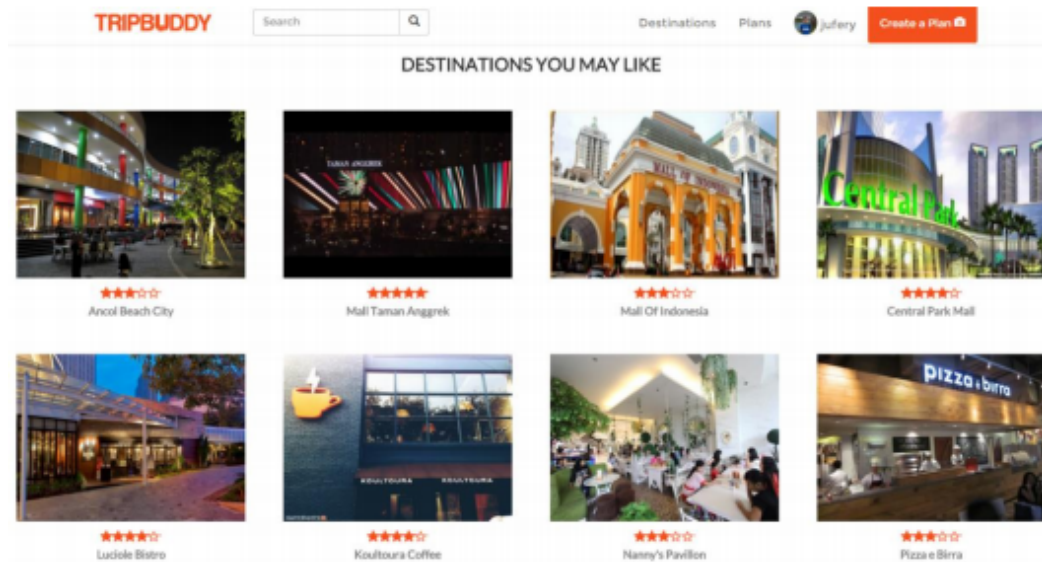
Figure 2. TripBuddy Recommendation Page

Destinations that are displayed are the result of K-Means clustering which is performed on the user data when they access the website. The example of TripBuddy Recommendation Page is shown in Figure 2 above. Some users that are grouped in the same cluster will have similar display of recommendation. This feature is built based on the recommendation from previous works, which suggest a recommender feature to have the points of personalization and adaptivity. Hence, TripBuddy will give recommendation based on user browsing behavior, with the adaptivity to recalculate the most appropriate destinations according to the number of users (n) that influenced the value of k in the rule of thumb of K-means clustering. TripBuddy also offers filter based on the available category, price as the admission ticket price or estimation of cost for the destination, destination's rating based on the rate given by user, and location, which is based on province of country.

Member can also create travel plan by filling in the available form which requires travel dates, destination and duration for visiting each destination. Routing map will be generated based on the nearest distance of each destination and member may change the order of destination
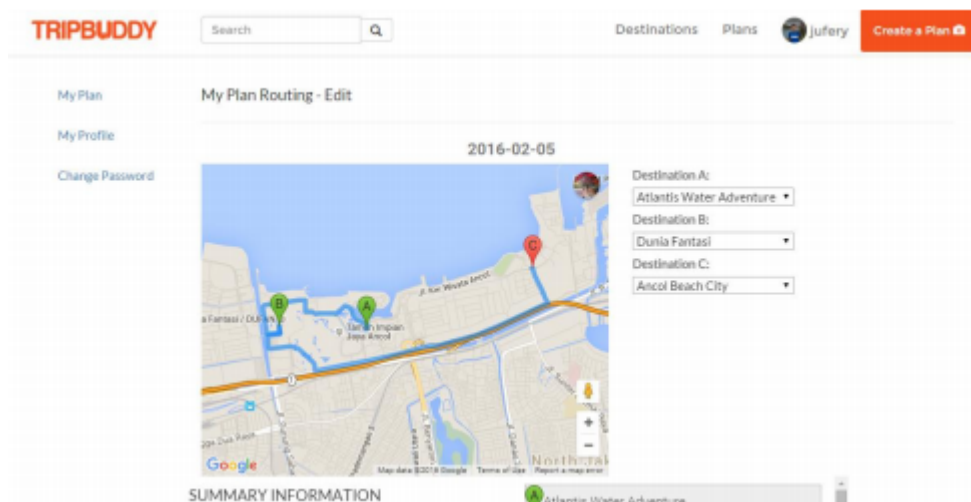

Figure 3. Routing Map

TripBuddy also offers offline mode, where user can save their travel plan into device, so they can use during their travel without internet access.
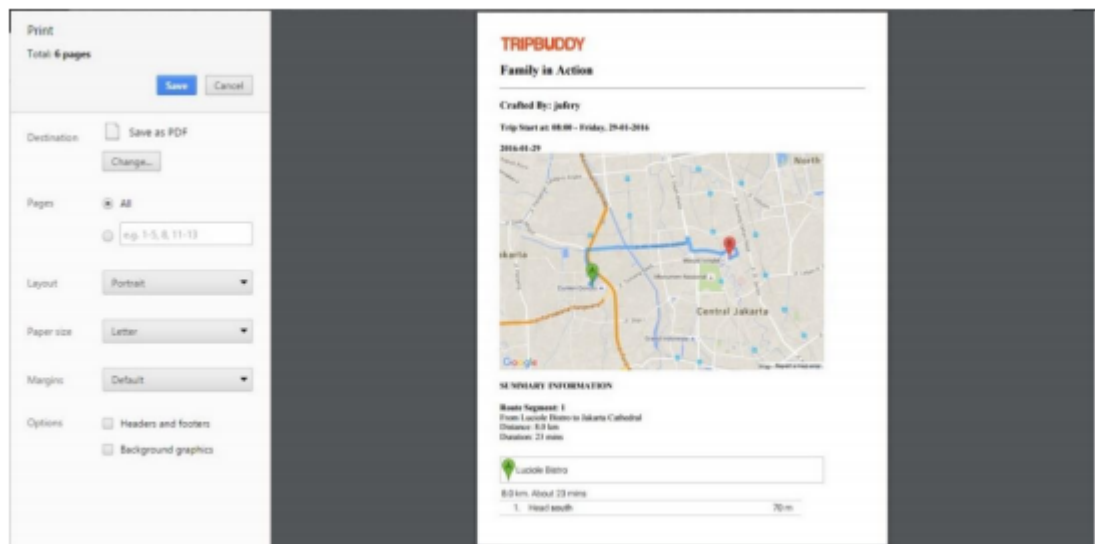

Figure 4. Offline Mode

4. Conclusion

According to the surveys conducted to 32 people near Bina Nusantara University after they tested the application, we can conclude that the tourist can get more information about the travel destination in accordance to their interest through TripBuddy as it offers many recommendation features. The recommendation features have been able to give related suggestion according to user's browsing behavior in TripBuddy (100% respondents agree that the feature is very helpful). Search destination or travel plans are easier because of the available platform to make adjustment or customization in the detailed search offered by application. Tourists may arrange their own travel plans more easily, and get more information on itineraries that can be used. They can prepare an effective journey even if they travel to a new place because the application is equipped with the most optimal settings so it does not need guidance from a tour guide.

We plan to develop the application by adding service integration feature with lodging and local transportation to provide a complete service and routing system that implements GPS (Global Positioning System) to guide journey in visual.

References

1. Cooper C, Wanhill S, Fletcher J, Gilbert D, Fyall A. Tourism: Principles and Practice. 3rd ed.: Prentice Hall; 2006.
2. Nizar MA. PENGARUH JUMLAH TURIS DAN DEVISA PARIWISATA TERHADAP. Jurnal Kepariwisataan Indonesia. 2012; II: p. 495-512.
3. Nizar MA. PENGARUH PARIWISATA TERHADAP PERTUMBUHAN EKONOMI DI INDONESIA. Jurnal Kepariwisataan Indonesia. 2011 June; 6(2).
4. Prihatno. Analisa Perilaku Konsumen Terhadap Jasa Paket Perjalanan Wisata di Yogyakarta. Media Wisata. 2010: p. 37-46.
5. Yerva SR, Grosan F, Tandrau A, Aberer K. TripEneer: User-based Travel Plan Recommendation Application. In AAAI Publications; 2013. p. 750-751.

6. Ravi L, Vairavasundaram S. A Collaborative Location Based Travel Recommendation System through Enhanced Rating Prediction for the Group of Users. Journal Computational Intelligence and Neuroscience. 2016 March; 2016(7).

7. Lim KH, Chan J, Leckie C, Karunasekera S. Personalized tour recommendation based on user interests and points of interest visit durations. In IJCAI'15 Proceedings of the 24th; 2015; Buenos Aires. p. 1778-1784.

8. Ediyanto , Mara MN, Satyahadewi N. PENGKLASIFIKASIAN KARAKTERISTIK DENGAN METODE K-MEANS CLUSTER ANALYSIS. Bimaster. 2013: p. 133-136.

9. SEBAYANG SA. SISTEM MARKET BASKET UNTUK MENENTUKAN TATA. Medan: UNIVERSITAS SUMATERA UTARA, Teknologi Informasi; 2012.

10. Kodinariya Tm, Makwana PR. Review on Determining Number of Cluster in K-Means Clustering. International Jurnal of Advance Research in Computer Science and Management Studies. 2013 January; I(6): p. 90-95.

11. A KNAR, Liemunandar K, Lintang UPL. Layanan Perencanaan dan Penjadwalan Pariwisata Masyarakat Indonesia dengan Konsep Jejaring Sosial Berbasis Web. Thesis S1. Jakarta: Bina Nusantara University, Computer Science; 2012.

12. HAYATI N. METODE HYBRID (CONTENT DAN COLLABORATIVE BASED) NEAREST. Thesis. Sumatera Utara University, Teknologi Informasi; 2012.

13. Harrington P. Machine Learning in Action. 1st ed. New York: Mining Publication; 2012.

14. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. 3rd ed. New York: Morgan Kaufmann; 2011.

15. Jung YG, Kang MS, Heo J. Clustering performance comparison using K-means and expectation maximization algorithms. Biotechnology & Biotechnological Equipment. 2014; XXVIII: p. 44-48.