INTELLIGENT DECISION SUPPORT SYSTEM FOR TOURISM PLANNING USING INTEGRATION MODEL OF K-MEANS CLUSTERING AND TOPSIS

[1]AHLIHI MASRURO, [2]FERRY WAHYU WIBOWO

[1,2]Informatics Engineering Department, STMIK AMIKOM Yogyakarta, Indonesia
E-mail: [1]ahlihi.m@amikom.ac.id, [2]ferry.w@amikom.ac.id

Abstract- A tourism issues are most competitive issue among countries, because the tourists aren't only international tourists, but also domestic tourists. To organize and forecast the visiting of the tourists using some criteria could implement intelligent decision support system (IDSS). The IDSS is the inter-discipliner major such as information systems, artificial intelligence, and decision science. In this paper the criteria which have been chosen to obtain optimization on the tourism planning is processed using two methods that are K-Means clustering and Technique for Order of Preference by Similarity to Ideal Solution abbreviated as TOPSIS method. For the validity of the measurement value is tested using partition coefficient (PC) and modified partition coefficient (MPC).

Keywords- Decision Support System, Intelligent, K-Means, TOPSIS, Tourism

I. INTRODUCTION

In this era, a tourism has competitive values among countries. The information services have many targets that are relevant and accurate, in order the tourists could be easier to determine and define tourism locations where the tourists will be visited in. Observing of tourists number that is still dominated by the domestic tourist, this issue needs decision support system for tourists to determine tourism locations and theirs information. Not many tourists have known the destination of tourism location, they just know from mouth-to-mouth and electronics medias. Information and communication technologies has been involved in the tourism industry, especially in the field of decision support system (DSS) 1. The decision support system could be used to forecast the tourism developing based on the desired tourists criteria concerning on the tourism location2. The DSS applications can give basic model to organization and people who manage tourism destinations in the concerning of the regulations and policies3. Decision support system method and data mining which use multi criteria decision making (MCDM) could be employed in the clustering or grouping of customer to gain effective on the advertising and minimalize usage of the resources 4. In other way, the MCDM in the real activity is very hard to define proper criteria as the tourists wish, so a supporting method. One of the purposed methods is TOPSIS, because of the proper valued criteria5.

Aim of this paper is to determine the tourism plan using intelligent decision support system for defining tourism location. In this case, the methods to support a decision of tourism plan deploy K-Means clustering algorithm and TOPSIS method which provide a data grouping and tourism location lists to the tourists. Basically, principle of K-Means clustering is to determine criteria values from alternative value groups. The value of this grouping is applied to define alternative list result which would be counted using TOPSIS method, so the data mining is implemented to support a decision in the system..

II. RELATED WORKS

2.1. Intelligent Decision Support System

The scope of intelligent decision support system (IDSS) is the collaborating major among artificial intelligence, decision science and information systems. This major has capability supporting data analysis, decision modeling, decision oriented, and the next planning orientation6. Many of methods have been improved and developed using multiple-parameter and intelligent technologies into intelligent systems. In some decision support system applications, some of them use criteria selections which have been established by system to be computed, in order giving information to the stake holders or who needs it. DSS itself is a computer-based system which is divided into three parts, i.e. a language system, a knowledge system and processing system. A language system is a way to have communication between the user and other components of the DSS, while a knowledge system is a knowledge repository which is used by the DSS, and a problem processing system is a connection among components which are required for the DSS7.

The intelligent decision support system is different from the conventional decision support system, because the function obtained by the decision support system including optimizing process and knowledge. The intelligent decision support system in this paper is applied to determine tourism planning using K-Means clustering algorithm and TOPSIS which is used to choose the tourism location objects, facilities and costs. In the implementation of K-Means clustering and TOPSIS methods, the available information from the K-Means clustering will be reprocessed by TOPSIS methods. The intelligent decision support system defines decision from the analyzing information and knowledge using the artificial intelligence, in that case it has an adaptive process as shown on figure 1.

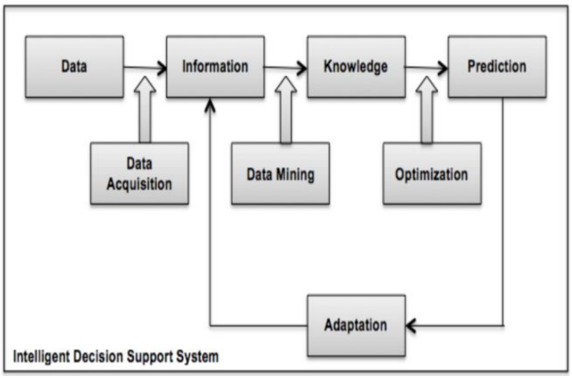

Fig.1. Diagram of Intelligent Decision Support System

2.2. K-Means Clustering Algorithm

Data Clustering is a process of grouping similar data objects or in other side, method of the ata mining which is categorized as unsupervised. Data mining is the mixing of the scope of the database, information retrieval, statistics, algorithm, and machine learning. The output clusters have to obtain minimum dissimilarity within the cluster and maximum dissimilarity with other clusters8. There are two kinds of the data clustering i.e. hierarchical data clustering and non-hierarchical data clustering. The K-Means is one of the non-hierarchical data clustering method which is partitioned into a form or more clusters that have same characteristics will be grouped into one cluster and different characteristics will be grouped into other clusters. The basic steps of K-Means clustering shown on figure 2.

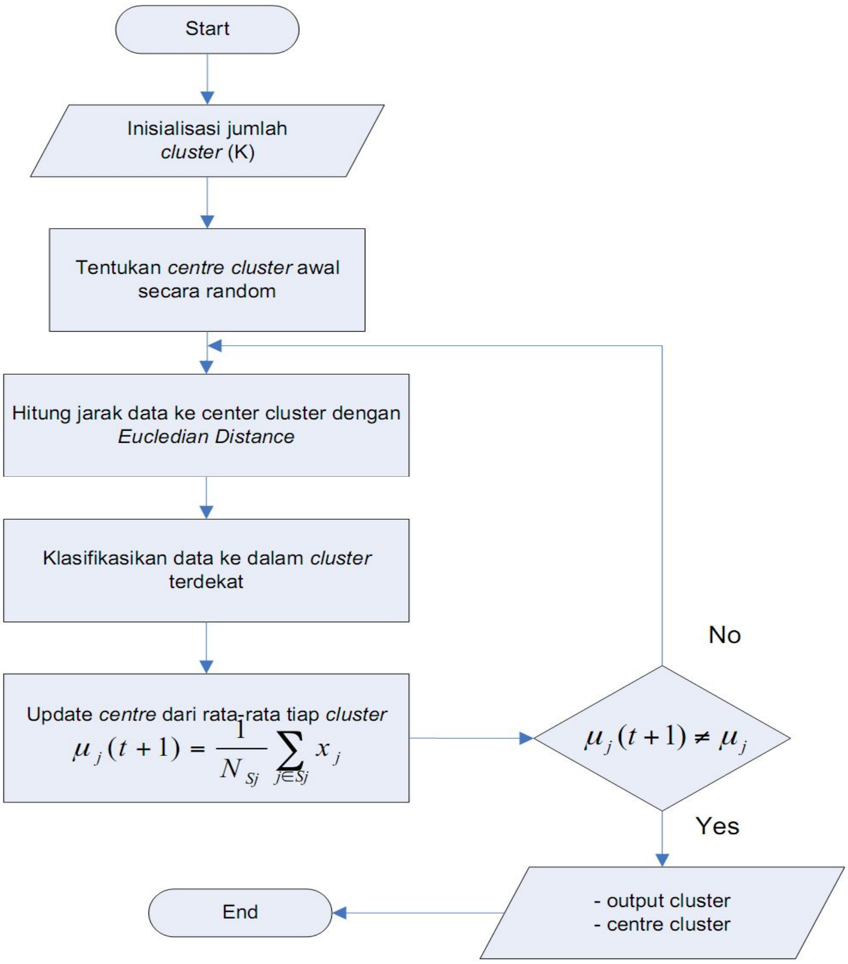

Fig.2. Basic Steps in K-Means Clustering

A Goal of the data clustering is to minimalize objective function set into clustering process. In the application is generated to minimalize variance data in the clustering and maximize among clusters. Some alternatives in the implementation of the K-Means are proposed computed theories, this is including of choosing distance space to count distance between data and centroid, re-allocated data method into each cluster, and used objective function.

The computation clustering of K-Means method have some steps i.e.

1. Define many K-cluster which is desired to form

2. Generate random value for initial cluster centre called as centroid as many of k.

3. Calculate distance every input data to each centroid using Eucledian distance until finding nearest distance from each data with centroid. The equation of the Eucledian distance shown on equation 1.

$$\varepsilon(\alpha_i, \beta_j) = \sqrt{(\alpha_i - \beta_j)^2} \quad (1)$$

where ε is the Euclidean distance, αi is the last centroid and βj is the initial centroid.

4. Clasify each data related to the centroid (smallest distance).

5. Update centroid value which is the new centroid value obtained from clusters average due to the equation 2.

$$\beta_j(n+1) = \frac{1}{N_{Sj}} \sum_{j \in Sj} \alpha_j \quad (2)$$

where βj(n+1) is the new centroid at the iteration n+1, Nsj is the number of data at the cluster of Sj.

6. Do step 2 to 5 until the member of each cluster there is no change.

7. If the step 6 reached, then the average value of the cluster centre βj at the end iteration will be employed as parameter for radial basic function in the hidden layer.

The value of measurement validity is tested by partition coefficient (PC) and modified partition coefficient (MPC). The PC is method of measuring cluster numbers when the data have overlap. The index of PC is shown on equation 3.

$$PC(\varsigma) = \frac{1}{\eta} \sum_i^{\varsigma} = 1 \sum_j^{\eta} = 1\left(\tau_{ij}^2\right) \quad (3)$$

Where ς is the cluster number and η is the data number τij is the data member degree of j-th at the cluster of i-th. PC(ς) is the index value of the PC at the cluster of ς-th, the value of PC is in the range of 1/ς < PC(ς) < 1. Generally, cluster number that is optimal determined from biggest PC value number of max2<ς<η-1PC(ς). The partition coefficient is rarely getting different to various cluster number of ς. The modified partition coefficient (MPC) is capable to reduce different various cluster number that is shown on equation 4.

$$MPC(\varsigma) = 1 - \frac{\varsigma}{\varsigma - 1}\left(1 - PC(\varsigma)\right) \quad (4)$$

MPC(ς) is index value of MPC at cluster of ς-th. The MPC value is at the range of 0 < PC(ς) < 1. In general, cluster number that is optimal determined from biggest MPC value number of max2<ς<η-1PC(ς).

2.3. TOPSIS Method

The rule of the short path has a consequence on the implementation of the scheduling and the solution of the problem of the long path. One of these methods is The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). TOPSIS is method to find solution from the owned data which perform the result based on the values of the nearest data between both desired value and longest value of the desired value. In general, TOPSIS procedure has some steps, i.e.

1. Making normalized decision matrices;
2. Making decision matrices with normalized weight;
3. Making positive ideal solution and negative ideal solution matrices;
4. Defining distance between alternative every value with positive ideal solution and negative ideal solution matrices;
5. Defining preference values for each alternative.

TOPSIS method needs work rating for every alternative criteria Ai at every criteria Cj which is normalized as shown on equation 5.

$$\kappa_{ij} = \frac{\mu_{ij}}{\sqrt{\sum_{i=1}^{m} \mu_{ij}^2}} \quad (5)$$

where i=1,2,...,m and j=1,2,...,n. Meanwhile κij is a normalized matrices of [i] and [j], thus μij is a decision matrices of [i] dan [j]. Positive ideal solution of A+ and negative ideal solution of A- could be determined by normalized weight rating of [i] and [j] or it symbolized as ξij which the symbol of ξij is the multiplication betwen weight vector of [i] from the analytical hierarchy process (AHP) that symbolized as ωi and normalized matrices of κij, where i=1,2,...,m and j=1,2,...,n. The positive ideal solution of A+ is set of maximal of ξij if j is the profit atributes, so the value of the A+ could be written as ξ1+, ξ2+, ..., ξn+ and the negative ideal solution of A- is set of minimize of ξij if j is the cost attributes, so the value of the A- could be written as ξ1-, ξ2-, ..., ξn-. The distance between both alternative criteria Ai and positive ideal solution A+ can be written as shown on equation 6.

$$\delta_i^+ = \sqrt{\sum_{i=1}^{n} (\xi_i^+ - \xi_{ij})^2} \quad (6)$$

where δi+ is the alternative distance of Ai with the positive ideal solution. In addition, the distance between both alternative criteria Ai and negative ideal solution A- can be written as shown on equation 7.

$$\delta_i^- = \sqrt{\sum_{j=1}^{n} (\xi_{ij} - \xi_i^-)^2} \quad (7)$$

where δi- is the alternative distance of of Ai with the negative ideal solution. So the preference value for each alternative of can be written as shown on equation 8.

$$\psi_i = \frac{\delta_i^-}{\delta_i^- + \delta_i^+} \quad (8)$$

where Ψi is the nearest distance for each alternative at ideal solutions. If the value of Ψi is bigger, so the alternative value of Ai is more chosen.

III. RESULTS AND DISCUSSION

The determining of tourism location planning employing K-Means clustering and TOPSIS shown on figure 3.

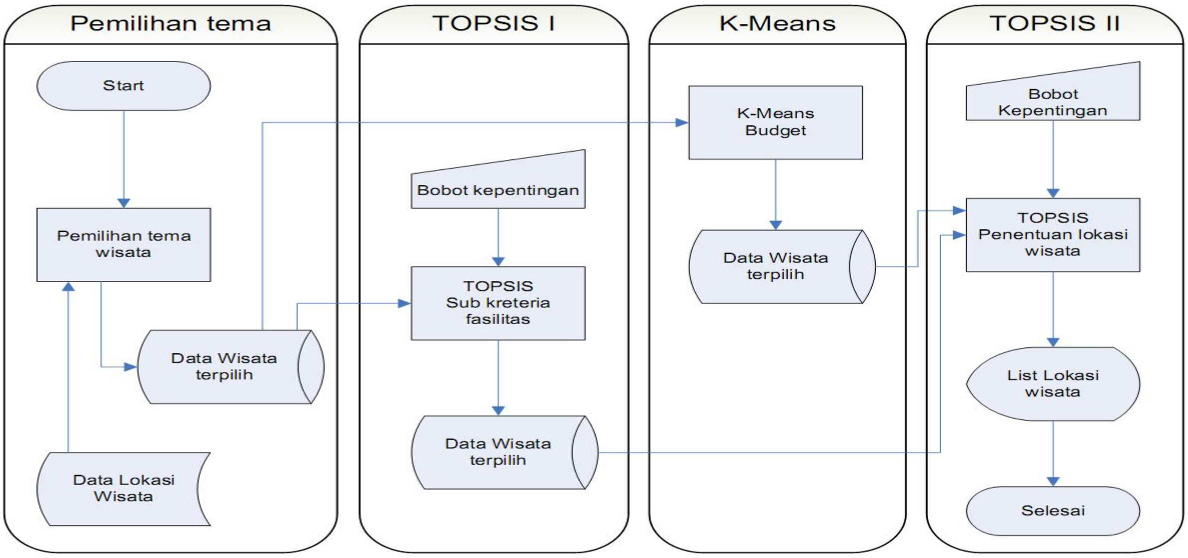

Fig.3. A Chart of Determining Tourism Planning Using K-Means Clustering and TOPSIS Method

The K-Means clustering provides values which will be reprocessed by TOPSIS. This process started from the choosing criteria until it will be emerged information of the tourism location. From the result, it used to define tourism location terms, thus the user should define an available sub-criteriaweight degree. The resulting prediction is reprocessedagain within calculating budget values using K-Means to get the budget clustering. After obtained facility and budget values in the previous result, it used to compute the next result using new weights.So the IDSS will choose smallest budget and biggest facility values. Pseudo-code of this decision support system for tourism planning using integration model of K-Means clustering and TOPSIS written bellow,

Pseudo-code of the DSS for Tourism Planning

```
input(k)
for i <-- 1 to k do
   centroid[i] <-- random(n)
endfor
   done <-- false
   max_iteration <-- 0
while (done = false or max_iteration <= 100) do
   for i <-- 1 to n do
      for j <-- 1 to k do
           min[j] <-- a[i] - centroid[k]
           square[i] <-- min[j]*min[j]
           dist[i,j] <-- sqrt(square[i])
      endfor
   endfor
   for i <-- 1 to N do
      dist <-- dist[i,1]
      cluster <-- 1
      for j <-- 2 to k do
         if dist[i,j] < dist then
            dist <-- dist[i,j]
            cluster <-- j
         endif
      endfor
      member[i].value <-- data[i]
      if member[i].cluster <-- cluster = i then
         c <-- member[j].nilai
         result[i,j] <-- c
      endif
   endfor
   for 1 <-- 1 to k do
      sum <-- 0
      count <-- 0
      for j <-- 1 to n do
         sum <-- sum + hasi[i,j]
         if hasil[i,j] = 0 then
            count <-- count + 1
         endif
      endfor
      centroid[i] <-- count / (j-1-count)
   endfor
      temp <-- 0
      temp1 <-- 0
      for i <-- 1 to k do
         temp <-- temp + centroid[i]
         temp1 <-- temp1 + centroid[i]
      endfor
   if temp - temp1 then
      done <-- true
   else
      for j <-- 1 to k do
         centre[j] <-- centroid[j]
      endfor
   endif
   max_iter <-- max_iter + 1
endwhile
for i <-- 1 to n do
   write(member[i].value,' ',member[i].cluster)
endfor
```

Some criteria which are implemented to define parameterized facilities written bellow.

1. Tourism locations which have parameterized facilities

I. Available

a. Proper rest area numbers
b. Good facilities
c. Easier facility access
d. Good information center

e. Organized and wide parking area

II. Less available

a. Rest area number is improper within tourism location wide ratio

b. Less good facility condition

c. Easier facility access

d. Less information center

e. Unorganized parking area

III. Unavailable

a. There is no rest area

b. Bad facility condition

c. Difficult facility access

d. There is no information center

e. Organized and wide parking area

2. Budget is divided into 5 groups clustering done by K-Means. In the determining cost is calculated by the cost transportation from the airport until tourism location plus tourism location tickets. This cost chosen within concerning on the average cost from the tour guide agents.

3. *Tourism location terms used some parameters such as,*

I. Nature

II. Beach

III. Mountain

IV. History

In the determining of tourism location terms is just done by selection based on the term input as shown on figure 4.

Fig.4. Defining Terms of Tourism Location

From the obtained data, the sub-facility of each location will be applied to weight operating. So the result is a facility criteria. The TOPSIS method is applied to determine value not to determine ranking of the tourism location as shown on figure 5.

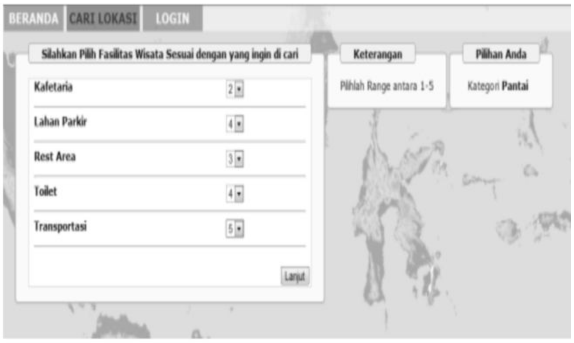

Fig.5. Resulting Facility Criteria

Then the final result is done by weighting on the facility and budget criteria. The first weighting process is resulted budget clustering and other weighting process is resulted facility clustering. The both results will be ranked by system as shown on figure 6.

BERANDA | CARI LOKASI | LOGIN

Hasil Rekomendasi Wisata Sesuai dengan yang di cari

No	Nama Lokasi	Alamat	Tarif	Fasilitas	Matrik
1.	Wediombo	Wediombo	Rp 250.000,00	Kurang Memadai	1
2.	Ngobaran	Ngobaran	Rp 750.000,00	Kurang Memadai	0.72891251819387
3.	Ngrenehan	Ngrenehan	Rp 900.000,00	Kurang Memadai	0.66678645265488
4.	Trisik	Trisik	Rp 1.540.000,00	Kurang Memadai	0.66678645265488
5.	Sepanjang	Sepanjang	Rp 1.130.000,00	Kurang Memadai	0.66678645265488
6.	Pok Tunggal	Pok Tunggal	Rp 1.500.000,00	Kurang Memadai	0.66678645265488
7.	Congot	Congot	Rp 600.000,00	Kurang Memadai	0.58312088668961
8.	Sadeng	Sadeng	Rp 960.000,00	Kurang Memadai	0.55075164284348
9.	Siung	Siung	Rp 350.000,00	Memadai	0.52176278294205
10.	Sundak	Sundak	Rp 435.000,00	Memadai	0.52176278294205
11.	Indrayanti	Indrayanti	Rp 560.000,00	Memadai	0.47294932220327
12.	Glagah	Glagah	Rp 1.250.000,00	Memadai	0.38570007648896
13.	Parangtritis	Parangtritis	Rp 670.000,00	Memadai	0.10473350337864
14.	Depok	Depok	Rp 1.050.000,00	Memadai	0
15.	Parangkusumo	Parangkusumo	Rp 1.850.000,00	Memadai	0

Tutup

Fig.6. Ranked Result

CONCLUSIONS

Intelligent decision support system for tourism plan using integration model of K-Means clustering and TOPSIS has major conclusions are as follows:

1. K-Means algorithm was just employed to define budget cost, because budget is the continued data in this case.

2. K-Means algorithm and TOPSIS have criteria which is converted the data, unless the data is the continued data.

ACKNOWLEDGMENTS

We thank STMIK AMIKOM Yogyakarta which gives us chance to present our research.

REFERENCES

[1]. S. P. Singh, J. Sharma, and P. Singh, "A Web-Based Tourist Decision Support System for Agra City", International Journal of Instrumentation, Control & Automation (IJICA), Vol. 1, Issue 1, pp. 51-54, 2011.

[2]. A. Patelis, C. Petropoulos, K. Nikolopoulos, B. Lin, and V. Assimakopoulos, "Tourism Planning Decision Support Within An E-Government Framework", International journal of Electronic Government, Vol. 2, No. 2, pp. 134-143, 2005. DOI: 10.1504/EG.2005.007091

[3]. R. Baggio and L. Caporarello, "Decision Support Systems in a Tourism Destination: Literature Survey and Model Building", Italian Chapter of Association for Information Systems (itAIS 2005), 1-2 December 2005.

[4]. A. H. Azadnia, P. Ghadimi, M. Molani-Aghdam, "A Hybrid Model of Data Mining and MCDM Methods for Estimating Customer Lifetime Value", Proceedings of the 41st International Conference on Computers & Industrial Engineering, pp. 44-49, 2011.

[5]. M. N. Mokhtarian and A. Hadi-Vencheh ,"A New Fuzzy TOPSIS Method Based on Left and Right Scores: An

Application For Determining An Industrial Zone For Dairy Products Factory", Applied Soft Computing, Vol. 12, Issue 8, pp. 2496-2505, 2012. DOI: 10.1016j.asoc.2012.03.042

[6]. E. Seniwati and F. W. Wibowo, "Comparison of Nutritional Status Data Calculation between K-Nearest Neighbour and Bayesian Algorithms", Proceedings of 5th International Seminar on Industrial Engineering and Management (5th ISIEM), 2012.

[7]. A. B. El Seddawy, T. Sultan, and A. Khedr, "Applying Classification Technique Using DID3 Algorithm to Improve Decision Support System Under Uncertain Situations", International Journal of Modern Engineering Research, Vol 3, Issue 4, pp. 2139-2146, 2013.

[8]. D. D. Nimbalkar and P. Shah, "Data Mining Using RFM Analysis", International Journal of Scientific & Engineering Research, Vol. 4, Issue 12, pp. 940-943, 2013.

★★★