ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 410

1st International Multidisciplinary Conference on Education, Technology, and Engineering (IMCETE 2019)

Clustering Tourist Destinations Based on Number of Visitors Using the K-Mean Method

1st Eka Mala Sari Rochman
Faculty of Engineering
University of Trunojoyo Madura
Indonesia Raya Telang Street PO BOX
02 Kamal-Bangkalan-Indonesia 69132
aery_r@yahoo.com

2nd Aeri Rachmad
Faculty of Engineering
University of Trunojoyo Madura
Indonesia Raya Telang Street PO BOX
02 Kamal-Bangkalan-Indonesia 69132

***Abstract*—The office door is the main access to enter the office in a company. Currently the office door generally still uses the key to be able to unlock the office door that has weak security. The authors will make the office door that integrated by a computer system that can unlock the office door using the registered faces in the system. The series of these tools use the camera to take a picture of the face and then sent it to the raspberry pi as the brain or the process tool to verification face, which will turn the face image into data information and make the decision to take action. After that solenoid will open after receiving the appropriate data from the raspberry pi. This tool becomes essential in helping employees to unlock the office door, so no need difficult to find the key to unlock the office door.**

Keywords: tourist, destinations, K-Mean Method

I. INTRODUCTION

Tourism is an inseparable part of human life, because traveling can eliminate the burden of mind due to very dense activities, especially regarding socio-economic activities which are seen as one of the prospective industries in the future [1]. The visit of local and foreign tourists to Indonesia will increase the country's foreign exchange and can improve the economy of the community in the tourist area. Bangkalan, a regency in the East Java region on Madura Island, has an attractive natural charm and diverse attractions [2]. Good nature tourism, beach tourism, religious tourism and culinary tourism [3]. But among the many attractions there are, there are still inequalities of tourists. This of course occurs due to many factors that can affect the interest of tourists both from the archipelago and foreign countries. To be able to equalize tourists, in this case a system can be classified into several groups according to data on the number of tourists who have visited. This system is expected to provide information quickly and precisely that can help in the field of tourism in Madura to optimize the development of tourist objects effectively and efficiently.

The case study takes place in Madura because there are still many attractions in Madura that are interesting and have high selling points but are still not very well known by the people. And the need for tourism publications about variations in the selection of attractions so that it can be better known by the wider community and can advance tourism objects in Madura. With this, the government can also quickly find out tourist attractions that are less attractive to tourists so that it is expected to improve the quality of tourist attractions in all aspects so that the spread of tourists can be generalized and the development of the population and economy throughout the Bangkalan region can be balanced. A solitive system that can help the development of these attractions is made using the K-Mean Clustering method.

Clustering will group data into several clusters (groups) based on the similarity of the characteristics of each data group [4]. In conducting this research, the author is inseparable from previous studies along with the methods used. Previous studies are a reference for writers to conduct and make this research. Where previous studies are comparative material to find out the advantages and disadvantages as well as to know and compare the methods used in previous studies. The author uses 4 previous studies relating to clustering. Research on the application of K-Means Clustering to classify tourist visits in Prov. DKI Jakarta [5]. The weakness of this study is that combining the number of tourists into one, namely the average tourist per year. Research on the application of K-Means to human infectious diseases [6]. The advantage of this research is that the features are used more than one. However, this research still does not test research results. Meanwhile, research on implementing Single Pass Clustering for web pages [7]. The advantage of this research is the existence of testing the research results using Precision and Recall.

Based on previous research, researchers will combine some of the advantages of the previous paper, namely the application of K-Means Clustering in tourists Bangkalan. The contribution in this research is to test the K-Means Clustering algorithm by using several features, namely in 2012-2018. After grouping, the results are tested using Precision and Recall. The contribution of this research is to group attractions based on the level of crowd visitors using the K-Means algorithm.

II. METHODOLOGY

This chapter will explain about data collection, clustering and k-mean algorithm.

A. Data Collection

Sample data of 20 records came from reports on the number of tourist visits to attractions in Bangkalan Regency. Then, the data is separated into two parts based on the type of tourists, namely Local Tourists and overseas Tourists.

B. Clusterring

Clustering Analysis is the process of dividing data in a set into several groups whose similarity of data in a group is greater than the similarity of the data with data in other groups [4]. The potential of clustering is that it can be used to find out structures in data that can be further used in a wide variety of applications such as classification, image processing, and pattern recognition [8].

In the cluster analysis process the method used to divide the data into a subset of data based on similarities or similarities that have been determined previously [9]. So cluster analysis in general can be said that:

- Data contained in one cluster has a high degree of similarity, and
- Data contained in a different cluster has a low level of similarity.

C. K-Means

K-means is one of the non-hierarchical (grouped) data grouping methods that try to partition existing data into two or more groups [8]. This method of partitioning data into groups so that data with the same characteristics are included in the same group and data with different characteristics are grouped into other groups. The purpose of grouping this data is to minimize the objective functions arranged in the grouping process, which generally seeks to minimize variations within a group and maximize variation between groups.

K-means algorithm basically performs 2 processes, namely the process of detecting the location of the cluster center and the process of finding members from each cluster. The clustering process begins by identifying the data to be clustered, X_{ij} (i = 1, ..., n; j = 1, ..., m) where n is the amount of data to be clustered and m is the number of variables. At the beginning of the iteration, the center of each cluster is set freely (arbitrarily), C_{kj} (k = 1, ..., k; j = 1, ..., m). Then calculated the network of each data with each cluster center. The basic process of the k-means algorithm can be seen below:

1) *Determine the number of clusters to be formed and set the center of the cluster k. 1.*

By determining the number of clusters as much as 3. Researchers use 3 clusters because researchers want to group the number of visitors based on 3 categories, namely quiet, moderate, and crowded..

2) *Using euclidean distances then count each data to the center of the cluster*

$$\mathrm{d}(\mathbf{i},\mathbf{k}) = \sqrt{\sum_{i}^{m}(C_i j - C_k j)^2}. \quad (1)$$

Specifies the cluster center randomly. Before determining the cluster center, we determine a limit on each cluster to determine the cluster center by finding the maximum and minimum values. Then determine the limit by dividing 3 parts in the minimum and maximum range.

3) *Group the data into clusters with the shortest distance by the equation*

$$\min \sum_{k}^{i} -a_{ik}- = \sqrt{\sum_{i}^{m}(C_i j - C_k j)^2} \quad (2)$$

4) *Calculate the center of the new cluster using the equation*

$$C_{kj} = \frac{\sum_k^i x_{ij}}{p} \quad (3)$$

With X_{ij} Cluster to k, p = number of k-cluster members

Determine the cluster value of each data. In this case must determine which cluster value is closest to the data, it is necessary to calculate the distance of each data with the center point of each cluster. At this stage Distance Space is used to calculate the distance between data and centroid. One of the equations that can be used is Euclidean Distance Space because the result that can be obtained is the shortest distance between the two points calculated. The formula used is (1) From the results of the above calculation, the values of cluster 1, cluster 2 and cluster 3 are generated in iteration 1. Repeat steps two to four so that there is no more data moving to other clusters [10].

III. Results and Discussion

From the 20 input value data, it can be seen, C1 consists of 1 leading tourist attraction namely Cow Race, C2 consists of 3 superior attractions namely Konang Natural Fire, Mother's Tomb, Syeichona Cholil Tomb, C3 consists of 16 leading tourist objects namely Mount Geger , Kolla Langgundih, Tomb of Sultan Abdul Kadirun, Lighthouse, Bangkalan Museum, Maneron Beach, Rongkang Beach, Siring Kemuning Beach, City Recreation Park, Jaddih Kapur Hill, Pelalangan, Mangrove Forest, Sunan Cendana Tomb, Bidadari Dhurjan Waterfall, Kec. Waterfall. Galis, and Anyar Beach. Shown by figure 1.

Meanwhile, based on Nusantara Tourists, it can be seen that, C1 consists of 2 leading tourist objects namely the Tomb of Sultan Abdul Kadirun, City Recreation Park, C2 consists of 16 leading tourist objects namely Konang Natural Fire, Mount Geger, Cow Race, Kolla Langgundih, Lighthouse, Bangkalan Museum, Maneron Beach, Rongkang Beach, Siring Kemuning Beach, Kapur Jaddih Hill, Pelalangan, Mangrove Forest Kec. Ten, Tomb of Sunan Cendana, Bidadari Dhurjan Waterfall, Kec. Galis, and Anyar Beach, C3 consists of 2 leading tourist objects namely the Tomb of Mother's Tears, and the Tomb of Syeichona Cholil. Shown by figure 2.

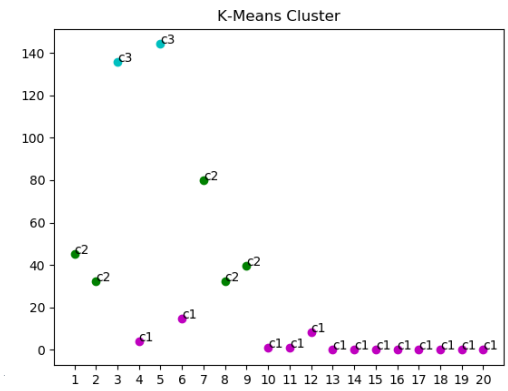

Table 1. Results of Precision Recall Testing in the local tourist

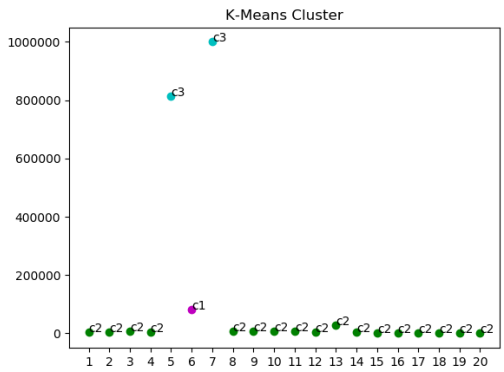

Figure 2. Clustering Final Results Overseas tourists Local tourists

In the precision testing and recall the results obtained are as in Table 1 and Table 2.

TABLE 1. RESULTS OF PRECISION RECALL TESTING IN THE OVERSEAS TOURIST

	c1	c2	c3
Presisi	1.0	0.0	1.0
Recall	0.0555555555556	0.0	1.0

	c1	c2	c3
Presisi	1.0	0.333333333333	1.0
Recall	0.941176470588	1.0	0.5

From Table 1, we can see that the value of prescion and recall in C2 is 0. In addition, the value of recall in C1 is also very small, which is 0.05. This shows that the results obtained are not accurate. Meanwhile, in Table 10, the values shown are relatively large, even though the C2 precision is small, which is 0.33. The inaccurate results are likely to be caused by dynamically taking upper boundary values in determining the initial cluster center without considering expert opinions.

IV. CONCLUSIONS

The conclusions obtained from the results of trials in this study include:

- The results of K-Means mention grouping based on Nusantara Tourists, C1 consists of 2 leading tourism objects, C2 consists of 16 leading tourism objects, and C3 consists of 2 featured tourism objects. While based on Foreign Tourists, C1 consists of 16 flagship attractions, C2 consists of 2 flagship attractions, and C3 consists of 1 flagship tourist attraction.
- Precision and Recall results, some small values are seen. In the archipelago tourist data, the value of precision and recall on C2 is 0%. While in foreign tourist data, the precision value on C2 is 0%, but the recall value is 100%. This is very influential on the accuracy of the data.

REFERENCES

[1] Dhani Eko Setyo Purnomo, “Sistem pendukung keputusan untuk pemilihan obyek wisata di surakarta menggunakan metode fuzzy tahani,” UNIVERSITAS STIKUBANK (UNISBANK) SEMARANG, 2013.

[2] Anamisa, Devie Rosa, Faikul Umam, and Aeri Rachmad. "Sistem Informasi Pencarian Lokasi Wisata di Kabupaten Jember Berbasis Multimedia." Prosiding SNATIKA 4 (2017): 32-36.

[3] Anamisa, Devie Rosa, Eka Mala Sari Rochman, and Aeri Rachmad. "Analysis of the Result of the Ant Colony System Adaptive on Tourism Object." JMEST, 2019.

[4] A. P. Windarto, “Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method,” *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017.

[5] L. Maulida, “Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov . Dki Jakarta Dengan K-Means,” *JISKa*, vol. 2, no. 3, pp. 167–174, 2018.

[6] A. Bastian *et al.*, “Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka),” *Jsi.Cs.Ui.Ac.Id*, vol. 14, no. 1, pp. 26–32, 2018.

[7] H. Februariyanti and E. Zuliarso, “Algoritma Single Pass Clustering untuk Klastering Halaman Web,” *Pros. Semin. Nas. Komput. dan Elektro*, pp. 1–8, 2012.

[8] O. Ong, Johan Oscar, “Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university,” *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20, 2013.

[9] A. K. Wardhani, “Implementasi Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Kajen Pekalongan,” *J. Transform.*, vol. 14, no. 1, pp. 30–37, 2016.

[10] M. G. Sadewo, A. P. Windarto, and D. Hartama, “Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 2, no. 1, pp. 60–67, 2017.