Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 1, January 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11220
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Application of K-Means algorithm in visit grouping of Pagar Alam city tour

Desi Puspita[1)*], Sasmita[2)]
[1)2)] Sekolah Tinggi Teknologi Pagaralam
[1)]desiofira1@gmail.com, [2]sasmitha@gmail.com

Submitted : Nov 18, 2021 | **Accepted** : Dec 29, 2021 | **Published** : Jan 3, 2022

Abstract: The purpose of this study was to analyze the application of the k-means algorithm in classifying tourism visits to the city of Pagar Alam. The k-means algorithm in grouping tourist objects begins by determining the number of clusters to be formed, determining the centroid value of each cluster, calculating the distance between the data, and calculating the minimum object distance calculated. There are 10 tourism objects that are superior to the data from the Tourism Office of the City of Pagar Alam. The research data used is the number of tourist visitors during the COVID-19 pandemic, namely 2020. The data are grouped into 4 clusters, namely C1 = a high number of tourist visitors, C2 = moderate number of tourist visitors, C3 = a low number of tourist visitors, C4 = number of visitors travel is very low. the centroid values used are C1 = 92,494, Centroid C2 = 71,658, Centroid C3 = 26,981 and centroid C4 = 4,485. then we get the results of grouping C1=Green Paradise tourism, C2=Janang Orange Gardens,, C3=Curup Tujuh Kenangan, Curup Mangkok, Curup dew, Tegur Wangi Site, Pelang Kenidai Village, and C4= Lumai Site, Tebing Tinggi Site and Tanjung Aro Site . From the results of grouping for c4, it becomes a note for the government of the City of Pagar Alam in increasing the number of tourist visitors.

Keywords: *tourism, k-means, clustering, data mining*

INTRODUCTION

The development of technology and information is currently growing very rapidly, with technology it can help in carrying out all activities, especially in the world of work because with the development of the age of technology it becomes something that is not foreign to everyone, technology is an important role because technology is considered a medium of information. capable of managing and classifying tourist visit data quickly, easily, and accurately.

Based on a preliminary study through observation and interviews at the Pagar Alam City Tourism Office, that currently uses Ms.Word and Ms.Excel to process tourist visit data. End data from several tours will be recapitulated with Ms.Excel. The current grouping of tourist visit data has not been implemented because of the large amount of data that must be re-managed so that it has difficulty for official employees to classify the visit data. This makes researchers interested in applying data grouping with K-Means, which can later help find out tourists from which areas most visit tourist attractions in Pagar Alam City.

The K-Means algorithm is an algorithm that groups data in a cluster with the center point of the centroid that is closer to the data used. The purpose of the K-Means algorithm is to maximize the similarity of clustered data such as one cluster to another in determining the distance function. So that the maximum similarity of the data is obtained based on the distance between the data and the centroid point (Asroni & Adrian, 2015)

In a previous study conducted by (M.Hasyim Seregar, 2018) with the title "Clustering Sales of Building Tools Using the K-Means Method". In this study, it can be concluded that several things are needed to analyze the sales that occurred at the Adi Bangunan Store from July 1, 2017, to August 9, 2017. in demand. The application of the clustering method can determine the purchase of the needed stock of goods quickly, from the research conducted, it is known that the best-selling group of goods is 10 items so that the priority of purchasing the stock of goods is directed at these 10 items. The relationship with this research is that this study uses the K-Means method to group data, namely data on tourist visits from Pagar Alam City so that it becomes a reference for researchers. with the application of the K-Means Algorithm in classifying tourist

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 1, January 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11220
e-ISSN : 2541-2019
p-ISSN : 2541-044X

visits to Pagar Alam City, it can assist in managing tourist visit data based on existing tourist visits. This system will process the data using the K-Means Algorithm method and this method was chosen because it can group data precisely which separates the data into different groups.

LITERATURE REVIEW

The K-Means algorithm is included in the application of data mining clustering. Clustering is data that does not have a label/class so it is often called the unsupervised learning technique according to (He & Tan, 2012) in the book (Suntoro, 2019). The K-Means algorithm is an iterative clustering algorithm. The K-Means algorithm assigns cluster values (K) randomly, temporarily these values become the center of the cluster or commonly called the centroid, mean or "means". Then calculate the distance of each existing data to each centroid using the Euclidian formula until the closest distance of each data to the centroid is found. Classification of each data based on its proximity to the centroid. Do this step until the centroid value does not change (stable) (Retno Tri Vulandari, 2017).

Previous studies related to k-means are as follows:

1. This research can be concluded that in order to assess the results of population health complaints by province, the Clustering K-Means method can be applied. The data is processed to obtain scores from the population with health complaints by province. The results obtained from this study can be input to the government, provinces that are of more concern to residents who have health complaints based on the clusters that have been carried out. (Rofiqo, Windarto, & Hartama, 2018)
2. In this study it can be concluded that data processing is carried out with the attributes of the initial stock and the stock being sold produces 3 product items that are most purchased by customers after knowing the level of sales of goods with what items are purchased by customers. It is expected that the company can determine stock items with any item to be added to the entry amount. ((Hutabarat & Sindar, 2019)
3. This research implements data mining with the K-Means Algorithm. The K-Means Clustering algorithm is used to classify people who are classified as entitled to receive social assistance and those who are not entitled to receive social assistance (Rusdiansyah, 2021)
4. The application of the K-Means method in grouping sales data at the Genta Corp Store can produce recommendations for goods that are in demand, not in demand, and quite in demand. So that the data becomes a reference for management to manage the stock of goods so that the store does not disappoint customers because the goods they want to buy are not available. (Indriani & Irfiani, 2019)

METHOD

Design Reseach

The steps are taken in the k-means clustering in Figure 1

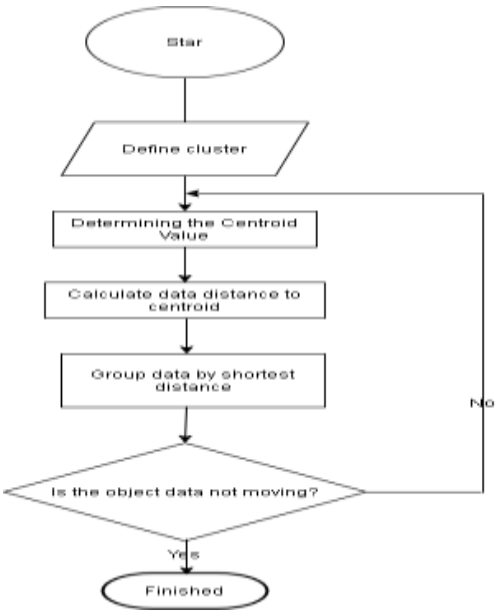

Figure 1. K-Means Clustering Flowchart

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 1, January 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11220
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In Figure 1. Flowchart of K-Means Clustering in general (Muliono and Sembiring 2019):

a. Determine the number of k as clusters to be formed. Determination of the number of clusters k is usually done by considering several factors, both theoretical and conceptual.
b. Determining the value of the centroid, to determine the initial centroid is carried out in a random form from several available objects as many as k clusters than to calculate the next i-th cluster centroid, using the following formula:

$$v = \frac{\sum_{i=1}^{n} xi}{n}; \qquad i = 1,2,3,\ldots n$$

Description :
v = Centroid in cluster
xi = object to-i
n = The number of objects that are members of the cluster

c. Calculating the distance from the centroid, calculating the distance between the object and the centroid which can be done using the euclidian distance mathematical formula:

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad ;i = 1,2,3,\ldots,n$$

Description:
xi = object x to-i
y = power y to-i
n = number of objects

d. Count each object to the nearest centroid
e. Perform the i-th iteration, then determine the position of the new centroid using the previous equation
f. Repeat the steps in point c if the position of the new centroid is not the same and the iteration or loop will stop if the ratio is not greater than the previous ratio value until the calculation results on each data converge.

Object of research

This research was conducted at the Pagar Alam City Tourism Office located on Jalan Laskar Wanita Mentarjo Gunung Gare, Pagar Alam City, South Sumatra Province.

Types of research

Based on the purpose and usefulness there are several types of research, in this study, the author conducted a qualitative type of research. Qualitative research is a research approach that uncovers certain social situations by describing and analyzing relevant data obtained from natural situations (Kristofora & Sujadi, 2017).

RESULT

Calculation of K-Means

At this stage, it will be explained about the use of the K-means algorithm in forming groups. The purpose of this stage is to prove that the k-means algorithm is able to provide the required information. The calculation is carried out with 10 sample data from the number of tourist visits to the leading tourist attraction in Pagar Alam City. The calculation starts from determining the initial centroid carried out in random form from several available objects as many as the number of clusters, then calculating the distance between the object and the centroid, Calculate each -each object to the nearest centroid using the Euclidian distance formula

Table 1. Cluster calculation

Tourist attraction	Cluster	Amount
Green Paradise	C1	92494
Kebun Jeruk Janang	C2	71658
Curup Tujuh Kenangan	C3	5422
Curup Mangkok	C3	6215
Curup embun	C3	5284
Situs Belumai	C4	1071

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 1, January 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11220
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Situs Tebing Tinggi	C4	2114
Situs Tegur Wangi	C3	3829
Situs Tanjung Aro	C4	1300
Desa Pelang kenidai	C3	6231

The clustering process in table 1 shows the results of clustering into four clusters so that this pattern will be the author's benchmark in applying the k-means algorithm in classifying tourist visits to Pagar Alam City.

Import Data Connection With K-Means Method

Tools to perform data simulation using RapidMiner software. RapidMiner is one of the processing of data mining such as clustering with the K-means method. After opening the RapidMiner software, make a connection between the imported data using the k-means method then drag the data and add the k-means operator to the process sheet then connect the data with k-means clustering and cluster performance distance to find out the output by clicking the play or run button.

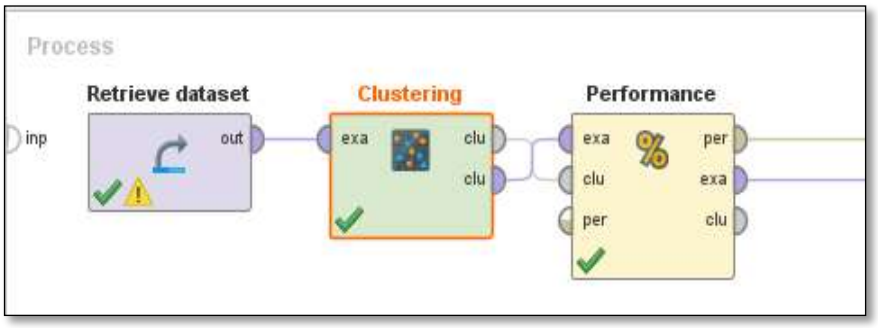

Figure 2. Display Connectivity

In Figure 2 it can be explained where the K-Means operator, the Cluster Distance operator with Performance crosses the first line with the second line and vice versa. After that it is reconnected with the output to be processed. When Run is clicked, it will generate a cluster if the data is successfully imported. After getting the cluster, the next step is to search for the centroid value of the cluster results. The results of the centroid value or the midpoint value can be seen in table 2 as follows:

Table 2. Data centroid

Atribute	C1	C2	C3	C4
Number of Visits	92.494	71.658	26.981	4.485

Analysis of data from data grouping carried out on 4 clusters with 10 tourist attraction data values can be seen in C1 the leading tourist attraction, namely Green Paradise tourism, C2 the leading tourist attraction, namely Kebun Jeruk Janang,, C3 consists of 5 leading tourist objects, namely Curup Tujuh Kenangan, Curup Mangkok, Curup dew, Tegur Wangi Site, Pelang Kenidai Village, and C4 consist of 3 leading tourism objects, namely the yetai site, the Tebing Tinggi site and the Tanjung Aro site.

CONCLUSION

Based on the results and discussion of the research, it can be concluded that the results of the data analysis that has been carried out by researchers with the application of the k-means algorithm where the variables used are the number of visitors who are grouped into 4 clusters, namely the high cluster (C1) where the number of tourist visitors is high, the cluster is medium (C2) where the number of tourist visitors is moderate, the cluster is low (C3) where the number of tourist visitors is low and the cluster is very low (C4) where the number of tourist visitors is very low. Where the results of the value of centroid C1 = 92,494, Centroid C2 = 71,658, Centroid C3 = 26,981 and centroid C4 = 4,485.

REFERENCES

Asroni, & Adrian, R. (2015). Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface . *Jurnal Ilmiah Semesta Teknika Vol.18, No.1*, 76-82.

Elita Amrina, I. K. (2019). Perancangan Sistem Informasi Pemasaran Biduk Wisata Pulau Berbasis Web Mobile. *Jurnal Optimasi Sistem Industri Vol.18 No.2*, 142-152.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Hutabarat, S. M., & Sindar, A. (2019). Data Mining Suku Cadang Sepeda Motor Menggunkan Algoritma K-Means. *Jurnal Nasional Komputasi dan Teknologi Informasi, Vol.2, No.2*, 126-132.

Indriani, F., & Irfiani, E. (2019). Clustering Data Penjualan Pada Toko Perlengkapan Outdoor Menggunkan Metode K-Means. *Jurnal Informatika Volume.7 No.2*, 109-114.

Kristofora, M., & Sujadi, A. (2017). Analisis Kesalahan Dalam Menyelesaikan Masalah Matematika Dengan Menggunakan Langka Polya Siswa Kelas VV SMP. *PRISMA*, 11.

M.Hasyim Seregar, S. (2018). Klasterisasi Penjualan Alat-Alat Bngunan Menggunakan Metode K-Means. *Jurnal Teknologi Dan Open Source*, 83-89.

Retno Tri Vulandari, S. M. (2017). *Data Miining Teori dan Aplikasi Rapidminer.* Yogyakarta: GAVA MEDIA.

Rofiqo, N., Windarto, A. P., & Hartama, D. (2018). Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means. *KOMIK(Konferensi Nasional Teknologi Iformasi dan Komputer*, 216-223.

Rusdiansyah, H. S. (2021). Data Mining using K-means method for feasibility selection of Non-cash food Assistance recipients in the Era of Covid-19. *Jurnal dan Penelitian Teknik Informatika*.

Suntoro, J. (2019). *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP.* Jakarta.

*name of corresponding author