Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X

# Application of Data Mining for Clustering of Foreign Tourist Visits Based on Arrival Entrance

**Handhy Nur Prabowo[1)], Resad Setyadi[2)], Wahyu Adi Prabowo[3)]**
[1)2)3)]Institut Teknologi Telkom Purwokerto,Indonesia
[1)]17103052@ittelkom-pwt.ac.id, [2)]resad@ittelkom-pwt.ac.id, [3)]wahyuadi@ittelkom-pwt.ac.id

**Submitted** : Nov 16, 2021 | **Accepted** : Jan 3, 2022 | **Published** : Jan 3, 2022

**Abstract:** Indonesia is a country with unique tourist destinations from each region. The tourism sector has an impact on the Indonesian economy which can encourage economic growth and increase the country's foreign exchange from foreign tourist visits. Tourism growth in Indonesia was disrupted due to the Covid-19 pandemic with the imposition of major social restrictions which resulted in a decrease in tourist visits and the paralysis of the tourism sector. Based on the problems described above, the authors are interested in conducting research in order to classify data on foreign tourist arrivals based on the entrance of foreign tourist arrivals. This research uses data mining method and K-Means Algorithm to form 5 clusters. The 5 clusters are divided into groups of tourist entrances which are categorized as very high (C1), high (C2), moderate (C3), low (C4) and very low (C5). In forming the 5 clusters, the researchers used Ms. Excel and Rapidminer 9.1 to process data. The results of this study obtained that the tourist entrance group was categorized as very high (C1) with 1 data, high (C2) with 1 data, moderate (C3) with 1 data, low (C4) with 1 data and very low (C5). ) that is with 21 data. This study aims to provide suggestions and future considerations to the Ministry of Tourism and Creative Economy of the Republic of Indonesia (Kemenparekraf) to carry out policies so that the Indonesian tourism sector can return to normal.

**Keywords:** Tourism; Covid-19; Indonesia; Data Mining; K-means; Clustering

## INTRODUCTION

Indonesia is an archipelagic country located on the equator with exotic natural and cultural wealth (Kadarisman, 2021). Indonesia's abundant and attractive natural and cultural wealth makes domestic and foreign tourists interested in traveling in Indonesia to relax and unwind (Kadarisman, 2021). The tourism sector has an impact on the Indonesian economy which can encourage economic growth and become an invisible export commodity because of its ability to bring in foreign exchange which is equivalent to the real sector of other commodities. (Rizki Munanda, 2019).

Tourism growth in Indonesia was disrupted due to the Covid-19 pandemic announced by the World Health Organization (WHO) on March 11, 2020 (Suprihatin, 2020). Global tourism is also paralyzed and has experienced a significant decline due to the Covid-19 pandemic, as a result of this many countries have implemented large-scale lockdowns and restrictions to stem the spread of the Covid-19 virus. (Suprihatin, 2020). In accordance with the Decree of the Head of the National Disaster Management Agency Number 9.A of 2020 concerning the determination of the status of certain emergency situations due to the Covid-19 disease outbreak in Indonesia, many people cannot carry out their daily routines as usual and many restrictions are placed between cities as a form of control. against the spread of Covid-19 (Ferdiansyah, Suganda, Novianti, & Khadijah, 2020). It is currently estimated that 75 million jobs in the tourism sector are experiencing shocks and the tourism industry is at risk of losing its turnover of more than US$2.1 trillion. (Zenker & Kock, 2020).

Based on the problems described above, the authors are interested in conducting research in order to classify data on foreign tourist visits based on the entrance. This research uses data mining method and K-Means Algorithm to form 5 clusters. The 5 clusters are divided into groups of tourist entrances which are categorized as very high (C1), high (C2), moderate (C3), low (C4) and very low (C5). In forming the 5 clusters, the researchers used Ms. Excel and Rapidminer 9.1 to process data. The results of this study aim to provide suggestions and future considerations to the Ministry of Tourism and Creative Economy of the Republic of Indonesia (Kemenparekraf) to carry out policies so that the Indonesian tourism sector can return to normal.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X

## LITERATURE REVIEW

### Related Research

Research conducted by (Kario & Amalia, 2021) explains the grouping of the number of foreign tourist visits in ASEAN using the K-Means Algorithm for clustering. The data is processed using RapidMiner by classifying tourist visits based on 3 clusters, namely C1 with high visits, C2 with moderate visits, C3 with low visits. The result of this research is to provide information that can be used as a reference in improving the tourism sector and a comparison to other countries.

Research conducted by (Herliyani Hasanah, Nugroho Arif Sudibyo, 2021) explains the grouping of foreign tourist visits based on nationality to Indonesia using the K-Means Algorithm using 2 clusters. C1 consists of countries that have a high level of visits and C2 countries that have a low level of visits to Indonesia. The results of this study become input data for tourism marketing strategies.

### Data Mining

Data mining known as pattern recognition is a data processing method used to obtain hidden patterns from the data to be processed. Data that is processed with data mining techniques will create a new scientific knowledge that comes from old data, the results obtained from processing the data can be used to determine decisions in the future. (Alkhairi & Windarto, 2019). Data mining is a process that can be used to identify and extract useful information and knowledge from various sciences in large databases with artificial intelligence, statistical techniques, mathematics, and machine learning. (Gustientiedina, Adiya, & Desnelita, 2019). Data mining, also known as knowledge discovery in database (KDD) can be used interchangeably by explaining the process of extracting hidden information (Toresa, 2020). Data Mining can be divided into four groups, namely prediction modeling, cluster analysis, association analysis and anomaly detection. (Fatmawati & Windarto, 2018).

### Clustering

Clustering is the process of dividing data in a set into several groups whose data similarity in one group is greater than the similarity of the data with data in other groups (Sari, Wanto, & Windarto, 2018). Clustering is a technique for grouping data based on the similarity of data characteristics (Aditya, Jovian, & Sari, 2020). In data mining there are two types of clustering methods used in data grouping, namely hierarchical clustering and non-hierarchical clustering. (Saragih, Sembiring, & Sayuthi, 2018).

### K-Means Algorithm

The K-means algorithm is one of the partitional algorithms, because K-Means is based on determining the initial number of groups by defining the initial centroid value. The K-means algorithm uses an iterative process to get a cluster database (Indraputra & Fitriana, 2020). The K-Means algorithm can be said to be a non-hierarchical clustering that groups data into one or more groups that have the same characteristics. (Darnita, Toyib, & Kurniawan, 2020). The K-Means algorithm starts with the formation of a cluster partition at the beginning then iteratively improves the cluster partition until there is no significant change in the cluster partition. (Sibuea & Safta, 2017). The steps for performing Clustering with the K-Means method are as follows (Rusdiansyah, Rasyid, & Sosrowidigdo, 2021):

1. Inputting or entering the number of clusters (k) in the data set that has been collected
2. Determine the center value (centroid)
3. Determination of the value of the centroid at the initial stage is carried out randomly, while in the iteration stage the formula is used as in the equation (1) the following:

$$\overline{V_{ij}} = \frac{1}{Ni}\sum_{k=0}^{Ni} X_{kj} \quad (1)$$

Information:
Vij = centroid the I-th cluster average for the j-variable
Ni = The number of members of the ith cluster
i, k = index of the cluster
j = the index of the variable
Xkj = the k-th data value of the j-th variable for the cluster

4. On each record, calculate the distance closest to the centroid. The centroid distance used is Euclidean Distance, with the formula as in the equation:

$$D_e \ \sqrt{(xi - si)^2} + \sqrt{(yi - ti)^2} \quad (2)$$

Information:
$D_e$ = Euclidean Distance

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X

i = The number of objects2
(x, y) = Object coordinates
(s, t) = Koordinat centroid

5. calculating the cluster center (centroid) from the data point and calculating the distance of each data to the nearest cluster center (centroid)
6. Recalculate each object that uses the new cluster center (centroid). If this process cluster center is not changed then the Clustering process is complete. And if there is still a change then it is repeated back to step 3 until the center of the cluster does not change.

**Rapidminer**

Rapidminer is a tool for analyzing data mining, text mining and predictive analysis. RapidMiner uses a variety of descriptive and predictive techniques to provide users with insights so they can make the best decisions. Rapidminer has approximately 500 data mining operators, including operators for input, output, data preprocessing and visualization. RapidMiner is written using the Java language so that it can work on all operating systems (Sibuea & Safta, 2017).

## METHOD

Research method is the process of gathering information with the aim of improving, modifying or developing an investigation or group of investigations (Nana & Elin, 2018). The following is the method used by the author in this research:

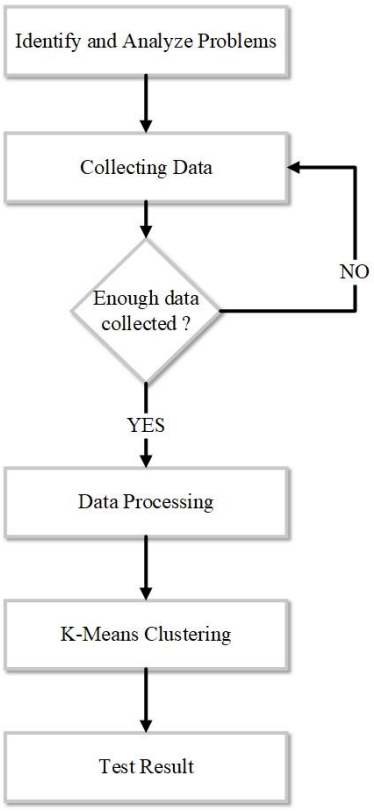



Figure 1. Research Method

**Identify and Analyze Problem**

At this stage the researchers identified the existing problems to be investigated. This stage is also carried out to determine the object of research and find solutions to overcome these problems.

**Collecting Data**

After identifying the problem, the researcher collects data sourced from the Central Statistics Agency (BPS) until the data collected is sufficient to continue the next process. The following is a data set of foreign tourist visits based on the arrival of the entrance until 3rd quarter 2021:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1. Data of Foreign Tourist Visits Based on Entrance Arrival

| Pintu Masuk | Bulan | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Januari | Februari | Maret | April | Mei | Juni | Juli | Agustus | September |
| Ngurah Rai, Bali | 2 | 12 | 3 | 9 | 8 | 1 | 0 | 0 | 0 |
| Soekarno-Hatta, Banten | 1238 | 5958 | 10188 | 15471 | 12775 | 13448 | 5466 | 1071 | 4081 |
| Juanda, Jatim | 20 | 69 | 52 | 106 | 160 | 181 | 52 | 10 | 2 |
| Kualanamu, Sumut | 11 | 17 | 45 | 22 | 50 | 38 | 28 | 4 | 3 |
| Husein Sastranegara, Jabar | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| YIA, DIY | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| BIL, NTB | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 26 | 0 |
| Sam Ratulangi, Sulut | 433 | 1027 | 2513 | 2685 | 1015 | 2435 | 695 | 5 | 593 |
| Minangkabau, Sumbar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S. Syarif Kasim II, Riau | 1 | 3 | 8 | 1 | 0 | 88 | 0 | 0 | 0 |
| S. Iskandar Muda, NAD | 0 | 0 | 0 | 1 | 4 | 3 | 1 | 0 | 0 |
| Ahmad Yani, Jateng | 0 | 0 | 2 | 2 | 1 | 3 | 0 | 0 | 0 |
| Supadio, Kalbar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sultan Hasanuddin, Sulsel | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sultan M. Badaruddin II, Sumsel | 0 | 0 | 0 | 0 | 0 | 0 | 804 | 477 | 0 |
| Batam, Kep. Riau | 260 | 157 | 343 | 191 | 245 | 250 | 153 | 149 | 129 |
| Tj. Uban, Kep. Riau | 0 | 0 | 81 | 10 | 20 | 20 | 0 | 0 | 0 |
| Tj. Pinang, Kep. Riau | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tj. Balai Karimun, Kep. Riau | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Tj. Benoa, Bali | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Jayapura, Papua | 2 | 2 | 0 | 3 | 0 | 0 | 0 | 0 | 3 |
| Atambua, NTT | 4 | 26 | 0 | 3 | 13 | 152 | 12 | 12 | 73 |
| Entikong, Kalbar | 6 | 11 | 14 | 2 | 0 | 0 | 1 | 0 | 0 |
| Aruk, Kalbar | 3 | 2 | 13 | 1 | 1 | 0 | 1 | 0 | 0 |
| Nanga Badau, Kalbar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Data Processing**

The data set that has been obtained before processing is carried out by cleaning the data and selecting attributes before testing the results using the K-means Algorithm.

**K-Means Clustering**

The processed data is then clustered using the K-means Algorithm manually using Ms. Excel.

**Test Result**

To get better results, a clustering process is carried out using rapidminer 9.1 to get more detailed test results.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X

## RESULT

This study forms 5 clusters of 25 data on foreign tourist visits based on the entrance to be processed using the K-means algorithm. The calculation starts by randomly determining the centroid value in the first iteration and then calculating the closest distance to get the cluster value because the k-means method allocates data into clusters from data that has the closest distance to the center point of each cluster by calculating the distance of each data using the Euclidean formula. distance. The next step is to do the second iteration by recalculating the closest distance to the centroid value that has been determined from the average value in the first iteration to get the cluster value in the second iteration. This research stops at the second iteration because there is no more data moving from each cluster. For more details, the steps in clustering in the application of the K-means algorithm are as follows:

### Determine the centroid value

Calculations on the K-means algorithm begin by determining the value of the center point or centroid at random, here are the randomly selected centroid values:

Table 1. Centroid Values

| Pintu Masuk Kedatangan | Januari | Februari | Maret | April | Mei | Juni | Juli | Agustus | September |
|---|---|---|---|---|---|---|---|---|---|
| Soekarno-Hatta, Banten | 1238 | 5958 | 10188 | 15471 | 12775 | 13448 | 5466 | 1071 | 4081 |
| Sam Ratulangi, Sulut | 433 | 1027 | 2513 | 2685 | 1015 | 2435 | 695 | 5 | 593 |
| Batam, Kep. Riau | 260 | 157 | 343 | 191 | 245 | 250 | 153 | 149 | 129 |
| Sultan M. Badaruddin II, Sumsel | 0 | 0 | 0 | 0 | 0 | 0 | 804 | 477 | 0 |
| Nanga Badau, Kalbar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### Calculate the distance of each data using the Euclidean distance formula

After determining the 5 center points or centroids, the next step is to calculate the distance of each data using the Euclidean formula and the following is the result of the calculation:

Table 2. The Result of The Calculation on Iteration 1

| Pintu Masuk | Januari | Februari | Maret | ... | September | C1 | C2 | C3 | C4 | C5 | Shortest Distance |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ngurah Rai, Bali | 2 | 12 | 3 | ... | 0 | 27770,26 | 4738,161 | 645,5494 | 935,0123 | 17,4069 | 17,40689519 |
| Soekarno-Hatta, Banten | 1238 | 5958 | 10188 | ... | 4081 | 0 | 23302,34 | 27219,39 | 27621,89 | 27783,19 | 0 |
| Juanda, Jatim | 20 | 69 | 52 | ... | 2 | 27517,49 | 4513,357 | 463,5591 | 927,9326 | 283,3196 | 283,3196075 |
| Kualanamu, Sumut | 11 | 17 | 45 | ... | 3 | 27702,84 | 4674,463 | 578,326 | 912,5662 | 87,58995 | 87,58995376 |
| Husein Sastranegara, Jabar | 0 | 0 | 0 | ... | 0 | 27782,71 | 4749,304 | 656,2439 | 934,8508 | 1 | 1 |
| YIA, DIY | 0 | 0 | 0 | ... | 0 | 27783,19 | 4749,817 | 656,6239 | 934,8503 | 0 | 0 |
| BIL, NTB | 1 | 0 | 2 | ... | 0 | 27781,42 | 4748,712 | 649,7661 | 921,8579 | 26,09598 | 26,0959767 |
| Sam Ratulangi, Sulut | 433 | 1027 | 2513 | ... | 593 | 23302,34 | 0 | 4196,747 | 4723,598 | 4749,817 | 0 |
| Minangkabau, Sumbar | 0 | 0 | 0 | ... | 0 | 27783,19 | 4749,817 | 656,6239 | 934,8503 | 0 | 0 |
| S. Syarif Kasim II, Riau | 1 | 3 | 8 | ... | 0 | 27736,52 | 4699,725 | 622,609 | 939,0229 | 88,42511 | 88,42510956 |
| S. Iskandar Muda, NAD | 0 | 0 | 0 | ... | 0 | 27779,15 | 4746,714 | 653,4784 | 934,0043 | 5,196152 | 5,196152423 |
| Ahmad Yani, Jateng | 0 | 0 | 2 | ... | 0 | 27779,43 | 4745,877 | 653,4883 | 934,8599 | 4,242641 | 4,242640687 |
| Supadio, Kalbar | 0 | 0 | 0 | ... | 0 | 27783,19 | 4749,817 | 656,6239 | 934,8503 | 0 | 0 |
| Sultan Hasanuddin, Sulsel | 0 | 0 | 0 | ... | 0 | 27783,19 | 4749,817 | 656,6239 | 934,8503 | 0 | 0 |
| Sultan M. Badaruddin II, Sumsel | 0 | 0 | 0 | ... | 0 | 27621,89 | 4723,598 | 957,5646 | 0 | 934,8503 | 0 |
| Batam, Kep. Riau | 260 | 157 | 343 | ... | 129 | 27219,39 | 4196,747 | 0 | 957,5646 | 656,6239 | 0 |

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Tj. Uban, Kep. Riau | 0 | 0 | 81 | ... | 0 | 27729,12 | 4687,154 | 599,5248 | 938,8323 | 86,37708 | 86,37708029 |
| Tj. Pinang, Kep. Riau | 0 | 0 | 0 | ... | 0 | 27783,19 | 4749,817 | 656,6239 | 934,8503 | 0 | 0 |
| Tj. Balai Karimun, Kep. Riau | 0 | 15 | 0 | ... | 0 | 27779,98 | 4746,596 | 653,1998 | 934,9706 | 15 | 15 |
| Tj. Benoa, Bali | 8 | 0 | 0 | ... | 0 | 27782,84 | 4749,094 | 653,4975 | 934,8845 | 8 | 8 |
| Jayapura, Papua | 2 | 2 | 0 | ... | 3 | 27780,56 | 4747,134 | 653,906 | 934,8642 | 5,09902 | 5,099019514 |
| Atambua, NTT | 4 | 26 | 0 | ... | 73 | 27683,02 | 4652,673 | 583,8698 | 934,2334 | 172,0203 | 172,0203476 |
| Entikong, Kalbar | 6 | 11 | 14 | ... | 0 | 27774,13 | 4738,231 | 643,6342 | 934,1815 | 18,92089 | 18,92088793 |
| Aruk, Kalbar | 3 | 2 | 13 | ... | 0 | 27776,65 | 4741,32 | 647,3453 | 934,0889 | 13,60147 | 13,60147051 |
| Nanga Badau, Kalbar | 0 | 0 | 0 | ... | 0 | 27783,19 | 4749,817 | 656,6239 | 934,8503 | 0 | 0 |

Based on the data in iteration 1, the results of clustering are as follows:

Table 3. Result of Iteration 1 Clustering

| Cluster | Data |
|---|---|
| C1 | 1 |
| C2 | 1 |
| C3 | 1 |
| C4 | 1 |
| C5 | 21 |

From the cluster mapping data that has been obtained, new centroids are generated as follows:

Table 4. New Centroid Value

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **C1** | 1238 | 5958 | 10188 | 15471 | 12775 | 13448 | 5466 | 1071 | 4081 |
| **C2** | 433 | 1027 | 2513 | 2685 | 1015 | 2435 | 695 | 5 | 593 |
| **C3** | 260 | 157 | 343 | 191 | 245 | 250 | 153 | 149 | 129 |
| **C4** | 0 | 0 | 0 | 0 | 0 | 0 | 804 | 477 | 0 |
| **C5** | 2,76190476 | 7,4761905 | 10,47619 | 7,619047619 | 12,238095 | 23,190476 | 4,5238095 | 2,4761905 | 3,8571429 |

**Calculating 2nd Iteration with new centroid**

The next step is to calculate the closest distance again based on the new centroid, following the results of the calculation using the new centroid:

Table 5. Iteration 2

| C1 | C2 | C3 | C4 | C5 | JARAK TERDEKAT |
|---|---|---|---|---|---|
| 27770,25634 | 4738,160614 | 645,5493784 | 935,0122994 | 25,11324953 | 25,11324953 |
| 0 | 23302,33716 | 27219,38594 | 27621,88558 | 27754,97922 | 0 |
| 27517,49331 | 4513,356955 | 463,5590577 | 927,9326484 | 254,0414323 | 254,0414323 |
| 27702,83744 | 4674,462643 | 578,3260326 | 912,566162 | 61,29253871 | 61,29253871 |
| 27782,70658 | 4749,30437 | 656,2438571 | 934,8507902 | 30,24819502 | 30,24819502 |
| 27783,1906 | 4749,816944 | 656,6239411 | 934,8502554 | 30,98924494 | 30,98924494 |
| 27781,42259 | 4748,711825 | 649,7661118 | 921,8579066 | 38,27732099 | 38,27732099 |
| 23302,33716 | 0 | 4196,747074 | 4723,597781 | 4722,471955 | 0 |
| 27783,1906 | 4749,816944 | 656,6239411 | 934,8502554 | 30,98924494 | 30,98924494 |
| | | | | | |
| 27736,51883 | 4699,724886 | 622,6090266 | 939,0228964 | 66,81810609 | 66,81810609 |
| 27779,14587 | 4746,714443 | 653,4783853 | 934,0042826 | 26,94444844 | 26,94444844 |
| 27779,43167 | 4745,876842 | 653,488332 | 934,8598825 | 27,24586761 | 27,24586761 |
| 27783,1906 | 4749,816944 | 656,6239411 | 934,8502554 | 30,98924494 | 30,98924494 |

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X

| | | | | | |
|---|---|---|---|---|---|
| 27783,1906 | 4749,816944 | 656,6239411 | 934,8502554 | 30,98924494 | 30,98924494 |
| 27621,88558 | 4723,597781 | 957,5646192 | 0 | 930,198171 | 0 |
| 27219,38594 | 4196,747074 | 0 | 957,5646192 | 630,4964055 | 0 |
| 27729,12449 | 4687,154361 | 599,5248118 | 938,8322534 | 71,79638781 | 71,79638781 |
| 27783,1906 | 4749,816944 | 656,6239411 | 934,8502554 | 30,98924494 | 30,98924494 |
| 27779,97777 | 4746,596254 | 653,1998163 | 934,9705878 | 31,0008597 | 31,0008597 |
| 27782,83528 | 4749,094335 | 653,4975134 | 934,8844848 | 31,30720847 | 31,30720847 |
| 27780,56173 | 4747,133767 | 653,9059565 | 934,8641613 | 29,60617 | 29,60617 |
| 27683,02131 | 4652,672995 | 583,8698485 | 934,2333756 | 148,3094781 | 148,3094781 |
| 27774,12512 | 4738,23121 | 643,6342129 | 934,1814599 | 28,06893838 | 28,06893838 |
| 27776,65014 | 4741,319648 | 647,3453483 | 934,0888609 | 27,88761915 | 27,88761915 |
| 27783,1906 | 4749,816944 | 656,6239411 | 934,8502554 | 30,98924494 | 30,98924494 |

Based on the data in iteration 1, the results of clustering are as follows:

Table 6. Result of iteration 2 clustering

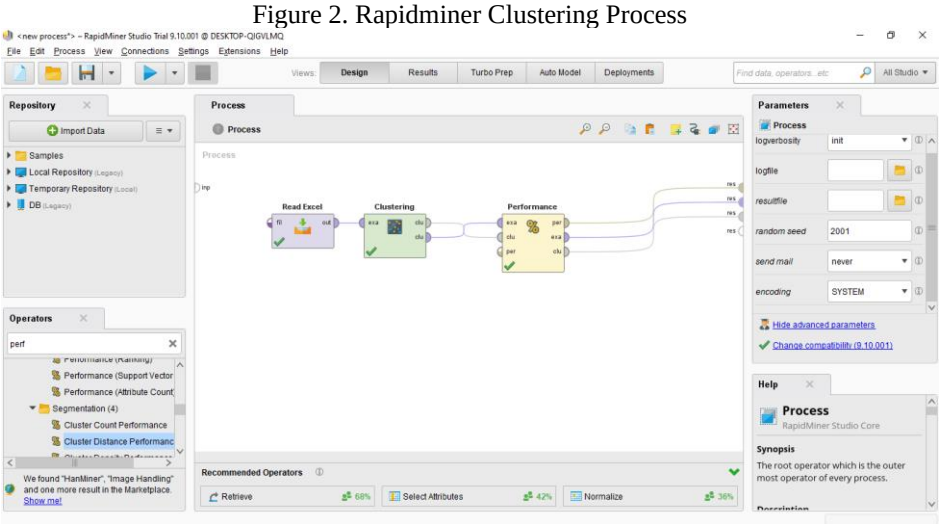
| Cluster | Data |
|---|---|
| C1 | 1 |
| C2 | 1 |
| C3 | 1 |
| C4 | 1 |
| C5 | 21 |

Based on the results of clustering from tables 4 and 6 there is no change and the process ends until the second iteration.

**Calculation using rapidminer 9.1**

Rapidminer 9.1 is then used for clustering to describe the results and visualize the data. The process begins by entering the data set and selecting the K-means algorithm on the rapidminer operator, the following is the process:

Figure 2. Rapidminer Clustering Process




After all processes have been carried out, the following results are obtained:

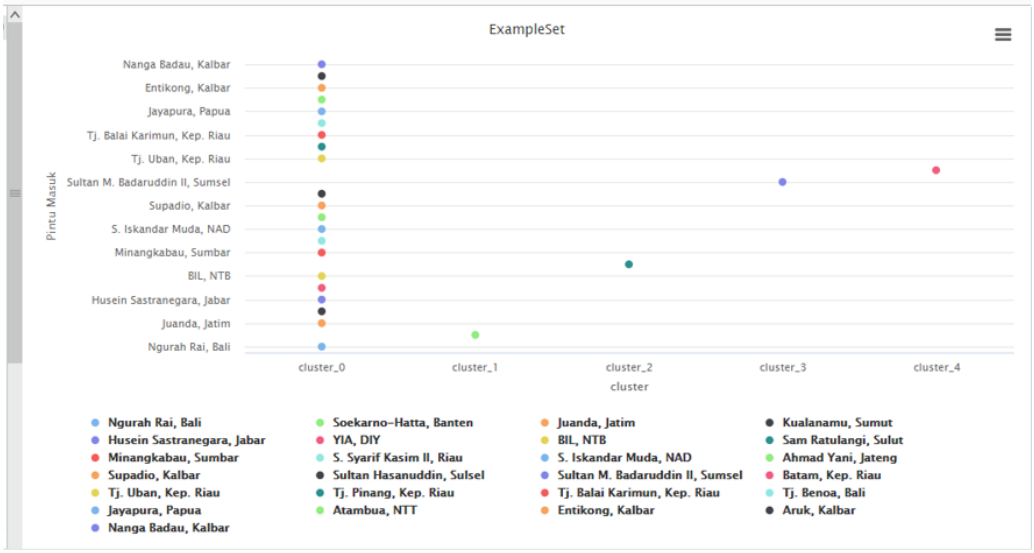
Figure 3. Clustering Result

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 1, Januari 2022
DOI : https://doi.org/10.33395/sinkron.v7i1.11217
e-ISSN : 2541-2019
p-ISSN : 2541-044X


The next step is to visualize the data, the visualization can be seen in figure 3 below:

Figure 4. Visualisation Result




## DISCUSSIONS

In this study, the K-means algorithm is used to classify categories of foreign tourist visits based on air, sea and land arrivals. This study was conducted to find out which entry gates have the highest to lowest foreign tourist arrivals during the COVID-19 pandemic in the 3rd quarter of 2021. In the clustering process, the existing datasets are grouped into 5 clusters with the provisions of C1 with a very high category, namely the incoming arrivals category, C2 high category, C3 medium, C4 low and C5 very low. The clustering process in the K-means algorithm has a weakness, namely determining the initial center of the cluster or centroid. The result of the cluster formed from the k-means algorithm is very dependent on the initial value of the specified cluster center, this makes it very difficult to get a unique initial centroid result. (Hablum, Khairan, & Rosihan, 2019).

## CONCLUSION

Based on the results of this study, it can be concluded that in the research conducted using the datamining method and the K-Means Algorithm to perform clustering using a manual process and reinforced using rapidminer 9.1 to describe the data in more detail and of course to visualize the data. The results obtained in the clustering process in forming 5 clusters by grouping the data on foreign tourist visits based on the entrance, the group of tourist entrances is categorized as very high (C1) with 1 data, high (C2) with 1 data, moderate (C3) with 1 data, low (C4) with 1 data and very low (C5) with 21 data. The results of this study aim to provide an overview of the number of foreign tourist visits based on entrances in the 3rd quarter of 2021 so that it can be compared with the previous year's data whether there was an increase or decrease as a result of the Covid-19 pandemic. Furthermore, this research provides suggestions and future considerations to the Ministry of Tourism and Creative Economy of

*name of corresponding author

the Republic of Indonesia (Kemenparekraf) to be able to make decisions based on the results of this study to establish specific policies to restore the tourism sector.

## REFERENCES

Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, *4*(1), 51. https://doi.org/10.30865/mib.v4i1.1784

Alkhairi, P., & Windarto, A. P. (2019). Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Seminar Nasional Teknologi Komputer & Sains*, 762–767.

Darnita, Y., Toyib, R., & Kurniawan, Y. (2020). Penerapan Metode K-Means Clustering Pada Aplikasi Android Pada Tanaman Obat Herbal. *Pseudocode*, *7*(2), 105–114. https://doi.org/10.33369/pseudocode.7.2.18-27

Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi. *Computer Engineering, Science and System Journal*, *3*(2), 173. https://doi.org/10.24114/cess.v3i2.9661

Ferdiansyah, H., Suganda, D., Novianti, E., & Khadijah, U. L. (2020). PENGELOLAAN MITIGASI KRISIS PARIWISATA AKIBAT PANDEMI COVID-19 DALAM MENGHADAPI FASE NEW NORMAL (Studi Kasus Di Desa Wisata Nglanggeran Yogyakarta). *Open Journal Systems*, *15*(3), 4133–4135.

Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 17–24. https://doi.org/10.25077/teknosi.v5i1.2019.17-24

Hablum, R., Khairan, A., & Rosihan, R. (2019). Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means. *JIKO (Jurnal Informatika Dan Komputer)*, *2*(1), 26–33. https://doi.org/10.33387/jiko.v2i1.1053

Herliyani Hasanah, Nugroho Arif Sudibyo, R. M. G. (2021). Data Mining Using K-Means Clustering Algorithm for Grouping Countries of Origin of Foreign Tourist. *Nusantara Science and …*, *2021*, 88–94. Retrieved from http://www.nstproceeding.com/index.php/nuscientech/article/view/455

Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, *10*(3), 3.

Kadarisman, A. (2021). Government public relations dalam pengembangan pariwisata masa pandemi COVID-19 di Geopark Ciletuh. *PRofesi Humas Jurnal Ilmiah Ilmu Hubungan Masyarakat*, *5*(2), 270. https://doi.org/10.24198/prh.v5i2.29800

Kario, G. M., & Amalia, E. (2021). K-Means Algorithm Implementation for Clustering of Foreign Tourists Visiting. *International Journal of Open Information Technologies*, *9*(6), 20–27.

Nana, D., & Elin, H. (2018). Memilih Metode Penelitian Yang Tepat: Bagi Penelitian Bidang Ilmu Manajemen. *Jurnal Ilmu Manajemen*, *5*(1), 288. Retrieved from https://jurnal.unigal.ac.id/index.php/ekonologi/article/view/1359

Rizki Munanda, S. A. (2019). PENGARUH KUNJUNGAN WISATAWAN MANCANEGARA, RATA-RATA PENGELUARAN DAN TINGKAT HUNIAN HOTEL TERHADAP PENDAPATAN INDONESIA PADA SEKTOR PARIWISATA. *Kajian Ekonomi Dan Pembangunan*, *1*, 37–48.

Rusdiansyah, R., Rasyid, H. Al, & Sosrowidigdo, S. (2021). Implementation of address recording management using the K-Means clustering classification algorithm in Kebayoran District, DKI Jakarta. *SinkrOn*, *5*(2), 184–191. https://doi.org/10.33395/sinkron.v5i2.10855

Saragih, A. T. R., Sembiring, A. S., & Sayuthi, M. (2018). Penerapan Metode Clustering K-Means untuk Proses Seleksi Calon Peserta Lomba MTQ. *Pelita Informatika*, *17*(April), 117–122. Retrieved from https://ejurnal.stmik-budidarma.ac.id/index.php/pelita/article/download/776/704

Sari, R. W., Wanto, A., & Windarto, A. P. (2018). Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *2*(1), 224–230. https://doi.org/10.30865/komik.v2i1.930

Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring. *Jurteksi*, *4*(1), 85–92. https://doi.org/10.33330/jurteksi.v4i1.28

Suprihatin, W. (2020). Analisis Perilaku Konsumen Wisatawan Era Pandemi Covid-19 ( Studi Kasus Pariwisata di Nusa Tenggara Barat ). *Jurnal Bestari*, *19*(1), 56–66.

Toresa, D. (2020). Implementasi K-Means Terhadap Penyebaran Penyakit Tbc Di Riau Menggunakan Rapid Miner. *JUTIM (Jurnal Teknik Informatika Musirawas)*, *5*(1), 35–42. https://doi.org/10.32767/jutim.v5i1.809

Zenker, S., & Kock, F. (2020). The coronavirus pandemic – A critical discussion of a tourism research agenda. *Tourism Management*, *81*(May). https://doi.org/10.1016/j.tourman.2020.104164

*name of corresponding author