K-Means Algorithm Implementation for Clustering of Foreign Tourists Visiting

Gita Muditha Kario and Endang Amalia

Abstract—The tourism sector plays an active role in economic growth for a country. Indonesia, which is one of the ASEAN states, shows that the role of tourism is one of the important sectors in the economy in Indonesia. However, the influence of the tourism sector has not been satisfactory for the government. The role of foreign tourist visits affects the economy in Indonesia by increasing foreign exchange for the country. In 2018, foreign exchange from the tourism sector continued to increase by 15.4 percent on an annual basis. However, it is unfortunate that Indonesia is still relatively small compared to other countries in the number of foreign tourist visits. The purpose of this study is to analyze the application of data mining in classifying the number of foreign tourist visits by Indonesia in ASEAN. The grouping is done by applying the K-Means clustering algorithm method. The data are grouped into 3 clusters, namely the high visit cluster (C1), the medium visit cluster (C2), and the low visit cluster (C3). So that the results obtained from the assessment of foreign tourist visits in ASEAN, namely, C1 namely Malaysia, C2 namely Singapore and Indonesia, and C3 namely the Philippines, Thailand, Vietnam, Myanmar / Burma, Brunei Darussalam, Cambodia, and Laos. The results of this study can be seen that Indonesia is in the medium visit grouping (C2). With this data, it can be a reference for the government to improve the tourism sector in visiting foreign tourists in Indonesia.

Keywords—Data mining, clustering, k-means, algorithm, foreign tourists, tourism

I. INTRODUCTION

Data mining, which is commonly referred to as knowledge discovery in databases (KDD), is an activity of collecting and using data that finds regularities and pattern relationships in large data sets. Output data mining is for future use in decision making. Clustering is one of the techniques in data mining. Clustering is a multivariate technique whose main objective is to classify objects based on their characteristics. Many methods can be applied for clustering involve the K-Means method, the LVQ (Learning Vector Quantization) method, FCM (Fuzzy C-Means), and so on.

The tourism sector has received significant attention from several countries in the world recently. The tourism sector influences economic growth of the country concerned. One of them is the countries in ASEAN with eastern customs, culture, societal friendliness, and natural beauty which are distinct advantages that can attract tourists to visit these countries. The application of data mining in the tourism sector can be a solution by analyzing large amounts of data. In the results of research launched by Twitter, Indonesia is included in the 10 countries most frequently visited by tourists using Twitter from the ASIA Pacific [1]. What about ASEAN, whether Indonesia is one of the countries with high foreign tourist visits. The application of data mining is intended to provide real solutions to the government to find out the highest amount of foreign tourist visits in ASEAN.

Indonesia, which is one of the ASEAN states, shows that in economy the role of tourism is one of the important sectors in Indonesia. However, the influence of the tourism sector has not been satisfactory for the government. According to the 2018 Ministry of Tourism's Performance report, the contribution of the tourism sector to the economy is still single digits. In 2018 the portion of tourism to Gross Domestic Product (GDP) was only 5.25 percent. In the same year, the realization of investment in the tourism sector reached US $ 1.6 billion or 80.43 percent of the target set by the government at that time, US $ 2 billion [2]. Nevertheless, the opportunity for the tourism sector to develop in Indonesia is still large, one of which is the influence of foreign tourist visits. The role of foreign tourist visits can affect the economy in Indonesia by increasing foreign exchange for the country. In 2018, foreign exchange from the tourism sector continued to increase, reaching Rp.229.5 trillion or an annual increase of 15.4 percent. However, it is unfortunate that Indonesia is still relatively small compared to other countries in the total of foreign tourist. In 2019, the Central Statistics Agency recorded the total of foreign tourist visits to Indonesia at 16.1 million, an increase of only 1.88 percent compared to 2018 [2]. As a result, the role of foreign tourists is very important so that the government continues to improve the achievement of the tourism sector. The application of data mining is intended to determine the position of Indonesian grouping in foreign tourist visits in ASEAN so that the government can also compare foreign tourist visits from various countries in ASEAN as one of the effective information to create an appropriate strategy to develop the tourism sector in Indonesia.

With the object of foreign tourist visits in ASEAN to find out Indonesia's position, you can use the grouping method with the K-Means algorithm. This study have purpose is to hope that the results of grouping data using the K-Means algorithm can help the government to know that Indonesia is at high or low visit grouping. So that the government can increase the tourism sector in Indonesia to increase the number of foreign tourist visits which have an impact on the introduction of tourist objects in Indonesia, increase foreign exchange and the country's economy. With the above background, the authors wrote this research entitled Implementation of Data Mining on Foreign Tourist Visits Using the K-Means Clustering Algorithm.

II. THEORETICAL FRAMEWORK

A. Data Mining

Data mining can be defined as a process of finding new patterns with statistical methods, machine learning, database systems, and artificial intelligence. Data mining usually called knowledge discovery, which is taking data patterns to be processed, and then the output is very important information [3]. The purpose of data mining is to produce useful information from large data sources [4]. Data mining is a process of obtaining useful information from large database warehouses, using pattern recognition technology, these patterns are recognized by certain tools that can provide data analysis, as well as mathematical and statistical techniques [5]. Data mining can also be interpreted as extracting new information taken from large chunks of data that helps carry out analysis from reviewing data sets to find unknown relationships and compress data in new ways or methods that can be implemented and data that are useful for making decisions [6].

B. K-Means Clustering Algorithm

K-Means Clustering algorithm, K-means is included in partitioning clustering, where each data must be included in a certain cluster. K-means is a method derived from the simple idea of minimizing the double errors found in grouping problems [7]. K-Means functions to separate data into k separate subdivisions, where k is a positive integer number. So that the data obtained are grouped in the same cluster because they have the same character. The K-means algorithm classifies the data closest to the central cluster (centroid) [8]. The K-means algorithm is well known for its ability and convenience to cluster big data and outliers very quickly. The steps for the K-Means Clustering method are as follows [9], [10]:

1. Determine the value of c for the number of clusters to be formed
2. Choose the center of the initial cluster (centroid) of c.
3. Calculate the closest distance of each data to the centroid using the Euclidean distance formula. The following is the Euclidean Distance formula

$$d_{(x,y)} = \sqrt{(x_1 - S_1)^2 + (y_1 - t_1)^2} \quad \textbf{(Formula 1)}$$

   (x,y): the coordinates of the object
   (s,t): the coordinates of the object *centroid*
   d(x,y): Euclidean distance is the distance between point x and points y using mathematical calculations.
4. Grouping each data based on its closeness to the centroid (smallest distance).
5. Update value of the centroid. The new value of the centroid is obtained from the average total of clusters obtained by using the formula

$$\frac{x1 + x2 + \dots xn, \quad y1 + y2 + \dots yn}{n} \quad \textbf{(Formula 2)}$$

   x1 + x2 + … xn is the x coordinate point of the data incorporated in a centroid.
   y1 + y2 + … yn is the y coordinate point of the data incorporated in a centroid.

n is the number of data that is incorporated in a data center (centroid).

6. Repeating steps 3 to 5, until the components of each cluster have not changed or the result is the same as the previous one.

C. RapidMiner

RapidMiner is data processing software that is open (open source) using algorithmic principles and data mining, RapidMiner can use on all operating systems cause written by the java language [11]. RapidMiner can be used to integrate data with data mining, analyze text mining, and predictive analytics including operators for data preprocessing, input, output, and data visualization, so that they can make decisions for their users [12].

D. Research Methodology and Techniques

Methodology and Techniques Research are the processes of searching for knowledge in a systematic study to find solutions to a problem. Research systematically in the form of formulating problems that occur, then hypotheses, then collecting data or facts, then analyzing it so that it becomes a conclusion in the form of a solution to the problem. The purpose of the research methodology is to find answers and truths to existing problems with scientific application. It not only covers research methods but also considers logic in the context of the research being made and can explain the reasons for using the method or technique.

In conducting a study, a research process is needed which consists of the steps required when carrying out research to be effective. In the research process, the following sequence of steps provides a useful procedural guide [13]:
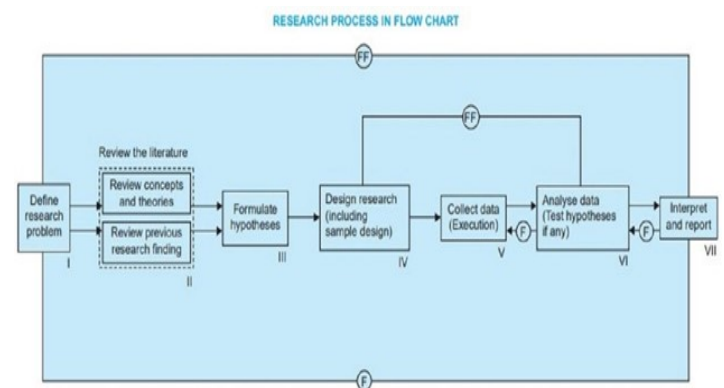

Fig. 1. Flowchart Research Process [13]

(1) Define Research Problem
    Define research problem is the initial stage that will be carried out in research, namely an overview of the problem that becomes a reference in research to be resolved.
(2) Review The Literature
    Research must be supported by a theory that underlies the research. A literature review can be done. Literature sources can be from anywhere such as journals, books, bibliographies, etc., depending on the problem being studied.
(3) Formulate Hypotheses
    Formulating a hypothesis is an initial assumption to be developed in research which is the main focus so that research can be more focused.

(4) Design Research
The design research is a structured concept plan that is carried out in research.

(5) Collect Data
Appropriate data is required at the data collection stage because several problems were found to be inadequate in data. For data collection, there are several ways including, primary data collection can be done through experiments or surveys.

(6) Analyse Data
Analyse data can be performed with several operations on existing data such as applying categories, coding, and drawing conclusions.

(7) Interpret And Report
If the researcher does not have a hypothesis, it will explain with findings based on several theories. That is known as interpretation. Finally, after the research has been carried out, the researcher makes a report about the research [13].

III. RESEARCH METHODS

In this study, using the research method C.R. Kothari. The flowchart of the research stages carried out exist in the following figure
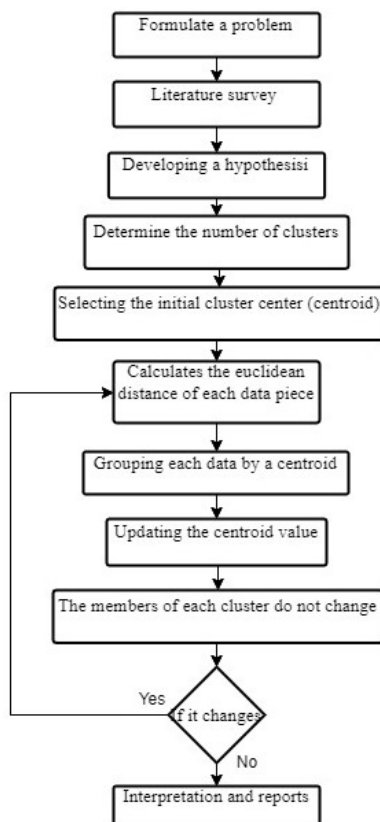

Fig. 2. Flowchart Research Stages

Based on Figure 2, the following is an explanation of the stages of the research

- The first stage in this research is to formulate a problem of research. The formulation of the problem needs to be done so that the research process is more focused so that it can provide instructions for data collection and search for the right method for solving the problem. The problem that occurs in this study is that foreign tourist visits in Indonesia are still relatively small compared to other countries, while foreign tourist visits have an important role in the economy of the country. The government needs to know the grouping of foreign tourist visits to be able to compare foreign tourist visits from various countries in ASEAN so that the government can create an appropriate strategy to increase the tourism sector in Indonesia in increasing the number of foreign tourist visits which have the impact to increase foreign exchange and the country's economy.

- The next stage to be carried out is a literature survey. At this stage, conducting a survey of literature related to research. The literature search is useful to avoid duplication of research implementation. At this stage, the writer uses 2 previous studies to become a comparison material to determine the advantages and disadvantages as well as to know and compare the methods used in previous studies.

Table 1. Literature survey

No	Author	Journal Contents
1	Agus Perdana Windarto	In his research, he conducted a grouping of rice import data by using the method of K-Means clustering so that it could assess the rice imports of the leading countries [14].
2	Andy Sapta dan Fitri Larasati Sibuea	In his research, students grouped data applying the K-means clustering method to determine a pattern assessment from the criteria determined by the school to produce data on outstanding students [15].

- The next stage is to develop a hypothesis, that is, in the grouping of data on foreign tourist visits there are significant differences between one grouping to another.

- The next stage is the clustering stage applying the K-Means algorithm, then data analysis is carried out with the Rapid Miner application. In the first stage, namely decide the total of cluster from secondary data obtained from the document of foreign tourist visits per month according to nationality in 2020 produced by the Ministry of Law and Human Rights (Directorate General of Immigration) through the website. https://www.bps.go.id. The data will then be processed by grouping foreign tourist visits according to nationality in ASEAN into 3 clusters, namely the high visit cluster, medium visit cluster, and low visit cluster.

- After determining the cluster then selecting the initial cluster center (centroid) from the initial data.

- Then calculate Euclidean distance from each data to each centroid with the formula of Euclidean Distance.

- Next, grouping each data by a centroid.

- Then update the value of the centroid of the corresponding cluster mean. The value of the new centroid is needed for further data grouping, where the results of grouping members of each cluster do not change or the results are the same as the previous one.

- If the calculation results are not the same then perform a loop of the steps to calculate the euclidean distance.

However, if the results are the same, the calculation is complete.

- In the final stage, the results of data processing carried out through the testing process will be used in the concluding stage of the results as the conclusion of the study.

IV. RESULTS AND DISCUSSION

This section, clustering will be formed, the data obtained will be calculated first using the K-Means algorithm. Sample data as much as 10 records from the document of Foreign Tourist Visits per month by Nationality in 2020 in ASEAN. The data is then accumulated. The sum exist in table 2.

Table 2. Data of Foreign Tourist Visits per month (ASEAN) in 2020

Nationality	Data of Foreign Tourist Visits per month (ASEAN) in 2020											
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
Malaysia	206532	164372	113848	61527	66385	62013	58053	57863	53373	45325	43285	46264
Brunei Darussalam	1219	923	444	8	17	17	4	8	13	26	9	9
Philippines	17174	13487	6041	1427	1819	1535	806	676	714	1003	1275	1342
Singapore	138625	84669	39751	2075	1335	1132	1169	1328	1405	1614	1930	2432
Thailand	7349	7463	2679	325	498	359	315	290	318	404	376	380
Vietnam	9152	6691	1499	265	304	238	181	190	162	175	169	180
Laos	307	396	36	-	-	-	2	2	-	1	-	2
Kamboja	1136	788	193	2	5	5	13	17	14	15	5	6
Myanmar/Burma	3992	1910	1309	783	736	664	502	485	442	502	490	539
Indonesia	55075	46769	26399	1588	2801	2527	2715	2625	3140	2309	1542	1309

Table 3. Accumulated Data Foreign Tourist Visits

No	Nationality	Amount
1	Malaysia	978.840
2	Brunei Darussalam	2.697
3	Philippine	47.299
4	Singapore	277.465
5	Thailand	20.756
6	Vietnam	19.206
7	Laos	746
8	Kamboja	2.199
9	Myanmar/Burma	12.354
10	Indonesia	148.799

After summarizing the data, the value of all foreign tourist visits according to nationality (ASEAN) in 2020 will be obtained. Then the data will be made into clusters, so first it will be determined as follows:

1. Decide the total of cluster, there are 3 clusters
2. Decide the random cluster center. The cluster center value 1 (C1) = 978,840, the cluster center value 2 (C2) = 47,299, and the cluster center value 3 (C3) = 746.
3. Decide the cluster value of each data. At this stage, the K-Means algorithm is used to classify the data

into 3 cluster, namely the high visit cluster (C1), the medium visit cluster (C2), and the low visit cluster (C3). In the application of the K-means algorithm, the midpoint or centroid is obtained. Euclidean Distance Space is used to count the distance between data and centoid. One of the equations that can be used is Euclidean Distance Space. The clustering process of each processed data is taken from the closest distance. For example, to calculate the distance of the first instance from the center of the first cluster is:

$$d_{(x,y)} = \sqrt{(x_1 - S_1)^2 + (y_1 - t_1)^2}$$

$$d_{1,1} = \sqrt{(978.840 - 978.840)^2} = 0$$
$$d_{1,2} = \sqrt{(47.299 - 978.840)^2} = 931.541$$
$$d_{1,3} = \sqrt{(746 - 978.840)^2} = 978.094$$

The calculation continues on each data. From the results of the above calculations, the value of cluster 1, cluster 2, and cluster 3 in iteration 1 has been grouped. The calculation results of alliteration 1 calculations exist in Table 4.

Table 4. Calculations Iteration 1

Initials	Nationality	Amount	Cluster				
			C1	C2	C3	Distance	Result
1	Malaysia	978.840	0	931.541	978.094	0	C1
2	Singapore	277.465	701.375	230.166	276.719	230.166	C2
3	Indonesia	148.799	830.041	101.500	148.053	101.500	C2
4	Philippines	47.299	931.541	0	46.553	0	C2
5	Thailand	20.756	958.084	26.543	20.010	20.010	C3
6	Vietnam	19.206	959.634	28.093	18.460	18.460	C3
7	Myanmar/Burma	12.354	966.486	34.945	11.608	11.608	C3
8	Brunei Darussalam	2.697	976.143	44.602	1.951	1.951	C3
9	Kamboja	2.199	976.641	45.100	1.453	1.453	C3

| 10 | Laos | 746 | 978.094 | 46.553 | 0 | 0 | C3 |

In table 3, in this process, K-means will continue iterating so that the last iteration data is the same as the previous iteration data grouping. The graph of iteration 1 grouping exist in the following figure:
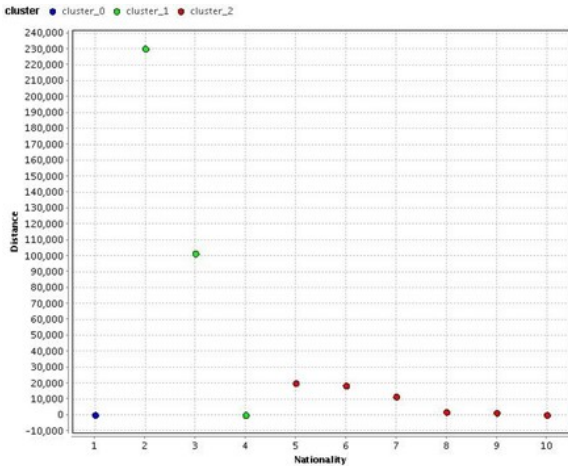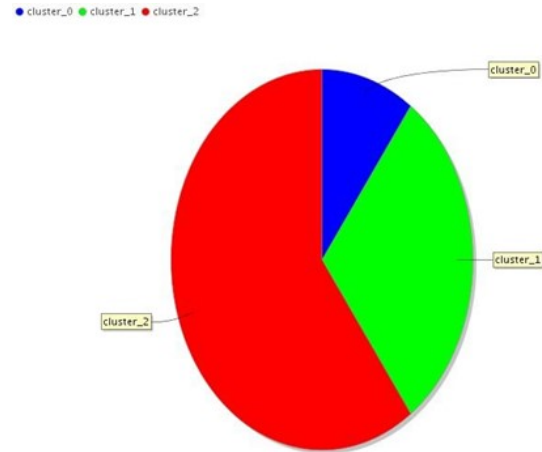

Fig. 3. Clustering Iteration Data 1

Fig. 4. Graph Iteration Data 1

In iteration 1, the input value data cluster is obtained which exist in table 4. In iteration 2, the centroid value search will be carried out for iteration 2 with the results of grouping in iteration 1. The results exist in Table 5:

Table 5. Centroid Iteration Data 2

Attribute	Cluster 1 (C1)	Cluster 2 (C2)	Cluster 3 (C3)
Amount of visit	978.840	157.854	9.660

After the centroid has been obtained, then the next process is to find the closest distance to each data. Process the data that grouped, the search for the closet or shortest distance in iteration 2, and the calculation of the results of the clustering data exist in the following table:

Table 6. Calculations Iteration 2

Initials	Nationality	Amount	Cluster				
			C1	C2	C3	Distance	Result
1	Malaysia	978.840	0	820.986	969.180	0	C1
2	Singapore	277.465	701.375	119.611	267.805	119.611	C2
3	Indonesia	148.799	830.041	9.055	139.139	9.055	C2
4	Philippines	47.299	931.541	110.555	37.639	37.639	C3
5	Thailand	20.756	958.084	137.098	11.096	11.096	C3
6	Vietnam	19.206	959.634	138.648	9.546	9.546	C3
7	Myanmar/Burma	12.354	966.486	145.500	2.694	2.694	C3
8	Brunei Darussalam	2.697	976.143	155.157	6.963	6.963	C3
9	Kamboja	2.199	976.641	155.655	7.461	7.461	C3
10	Laos	746	978.094	157.108	8.914	8.914	C3

Table 6 shows the results of the grouping iteration 2 has different results from iteration 1. The process will continue in the next iteration. The resulting graph of iteration 2 exist in the following figure:
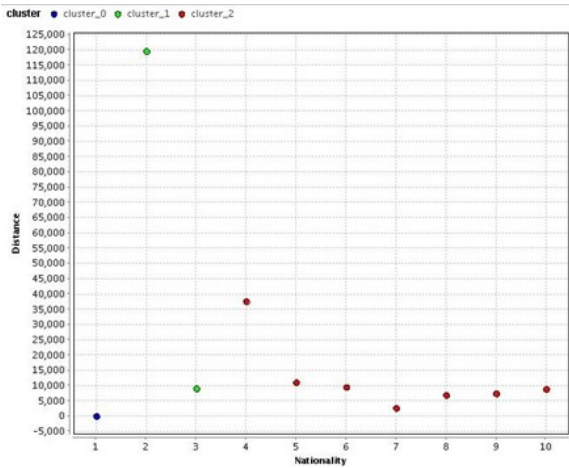
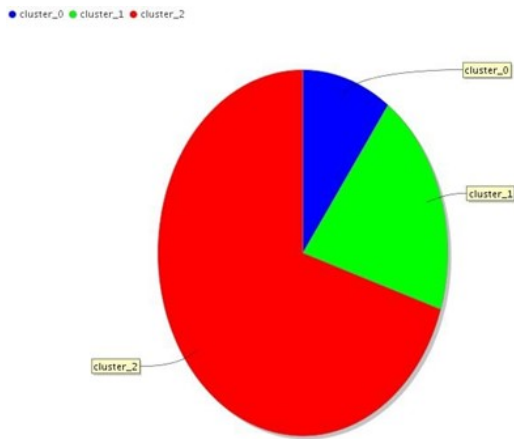Fig. 5. Clustering Iteration Data 2

Fig. 6. Graph Iteration Data 2

From Figure 5, the data on foreign tourist visits by nationality (ASEAN), iteration groupings 2 obtained 3 cluster. The high visit cluster (C1) is Malaysia, the medium visit cluster (C2) is Singapore and Indonesia and the low visit cluster (C3) is the Philippines, Thailand, Vietnam, Myanmar / Burma, Brunei Darussalam, Cambodia, and Laos. The results of iteration 2 show dissimilar results from the results of iteration 1. Then the process will continue to iteration 3 in the same way, determining the new centroid from iteration 2 and finding the closest distance from the value of the centroid in iteration 3.

Table 7. Centroid Iteration Data 3

Attribute	Cluster 1 (C1)	Cluster 2 (C2)	Cluster 3(C3)
Amount of visit	978.840	213.132	15.037

After the centroid has been obtained, then the next process is to find the closest distance to each data. Process the data that grouped, the search for the closet or shortest distance in iteration 2, and the calculation of the results of the clustering data exist in the following table:

Table 8. Calculations Iteration 3

| Initials	Nationality	Amount	Cluster				
			C1	C2	C3	Distance	Result
1	Malaysia	978.840	0	765.708	963.803	0	C1
2	Singapore	277.465	701.375	64.333	262.428	64.333	C2
3	Indonesia	148.799	830.041	64.333	133.762	64.333	C2
4	Philippines	47.299	931.541	165.833	32.262	32.262	C3
5	Thailand	20.756	958.084	192.376	5.719	5.719	C3
6	Vietnam	19.206	959.634	193.926	4.169	4.169	C3
7	Myanmar/Burma	12.354	966.486	200.778	2.683	2.683	C3
8	Brunei Darussalam	2.697	976.143	210.435	12.340	12.340	C3
9	Kamboja	2.199	976.641	210.933	12.838	12.838	C3
10	Laos	746	978.094	212.386	14.291	14.291	C3
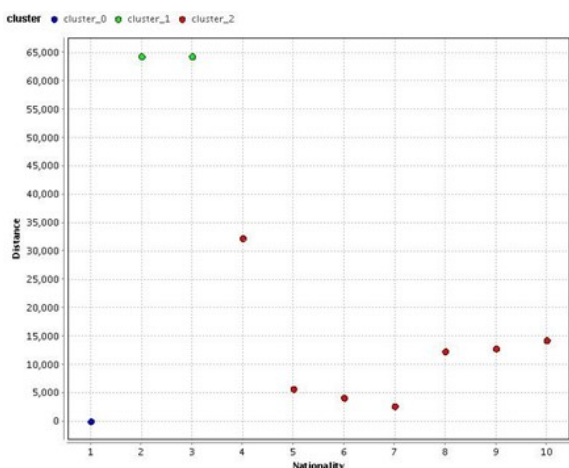
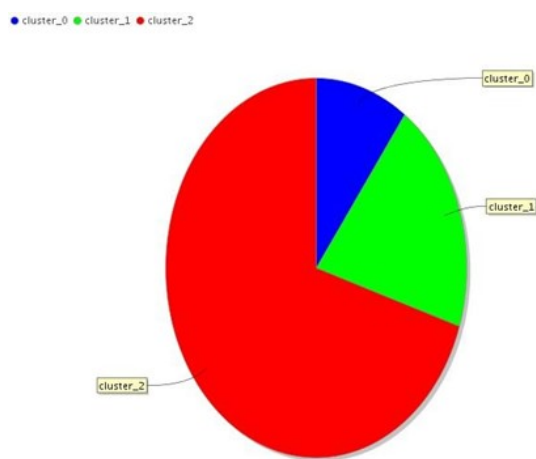Fig. 7. Clustering Iteration Data 3

Fig. 8. Graph Iteration Data 3

In the 3rd iteration, the data grouping was carried out in 3 clusters with iteration 2, the same results were obtained. From 10 data on foreign tourist visits by nationality (ASEAN), 1 cluster of high visits (C1), namely Malaysia, 2 medium visit clusters (C2) namely Singapore and Indonesia, and 7 clusters of low visits (C3) the Philippines, Thailand, Vietnam, Myanmar / Burma, Brunei Darussalam, Cambodia, and Laos.

V. CONCLUSION

To assess the grouping of the number of Foreign Tourist Visits in ASEAN, it can be applying the K-Means clustering algorithm. The data will be processed to obtain the value of foreign tourist visits in Indonesia. The data is processed using RapidMiner by conducting K-Means which are grouped into 3 clusters, namely the high visit cluster (C1), the medium visit cluster (C2), and the low visit cluster (C3). Data centroid C1 = 978,840, data centroid C2 = 213.132, data centroid C3 = 15.037 So that results obtained from an assessment of foreign tourist visits in ASEAN, namely, 1 high visit cluster (C1), namely Malaysia, 2 medium visit clusters (C2) namely Singapore and Indonesia, and 7 low visit clusters (C3) namely Philippines, Thailand, Vietnam, Myanmar / Burma, Brunei Darussalam, Cambodia, and Laos. The results of this study can be seen that Indonesia is in the medium visit grouping (C2). With this data, it can be a source of information for the government as a reference

for improving the tourism sector so that it can maintain and even increase the number of foreign tourist visits. The government can also compare with other countries so that it can be equal to countries in high clusters, even more than other countries in foreign tourist visits which have an impact on the introduction of tourist objects in Indonesia and increase in foreign exchange. In the future for similar research using the K-Means clustering method, it can be compared with other clustering methods to get maximum modeling results. In processing the data, the clustering can give weight to the criteria in the grouping process to get more accurate results. Future research can focus on Indonesia so that the government can find out which areas need to be improved in foreign tourist visits.

REFERENCES

[1] A. M. M. P. Senja, "Alasan Utama Turis Asing Berwisata ke Indonesia," 26 March 2019. [Online]. Available: https://travel.kompas.com/read/2019/03/26/171100327/alasan-utama-turis-asing-berwisata-ke-indonesia.. [Accessed March 2021].

[2] I. CNN, "Menghitung Kontribusi Sektor Pariwisata Bagi Ekonomi RI," 26 February 2020. [Online]. Available: https://www.cnnindonesia.com/ekonomi/20200226121314-532-478265/menghitung-kontribusi-sektor-pariwisata-bagi-ekonomi-ri. [Accessed March 2021].

[3] Suyanto, Data Mining Untuk Klasifikasi dan Klasterisasi Data, Bandung: Penerbit Informatika, 2017, p. 10.

[4] R. Purohit and D. Bhargava, "An Illustration to Secured Way of Data Mining Using Privacy Preserving Data Mining," *Journal of Statistics and Management Systems,* p. 637, 2017. doi:https://doi.org/10.1080/09720510.2017.1395183.

[5] E. Sikumbang, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *Jurnal Teknik Komputer,* p. 156, 2018.

[6] E. Prasetyo, Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab, Bandung: Andi Offset, 2014, p. 15.

[7] S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method to Determine Nutritional Status," *Jurnal Biometrika dan Kependudukan,* p. 63, 2020. doi:10.20473/jbk.v9i1.2020.

[8] Asroni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Jurnal Ilmiah Semesta Teknika,* p. 78, 2015.

[9] W. Azis and D. Atmajaya, "WS Azis and D Atmajaya. Pengelompokan Minat Baca Mahasiswa Menggunakan Metode K-Means," *ILKOM Jurnal Ilmiah,* pp. 89-90, 2016.

[10] S. Handoko, Fauziah and E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Teknologi dan Rekayasa,* pp. 80-81, 2020. doi:https://doi.org/10.35760/tr.2020.v25i1.2677.

[11] S. Haryati, A. Sudarsono and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *Jurnal Media Infotama,* p. 133, 2015.

[12] B. Rahmat, A. Gafar, N. Fajriani, U. Ramdani, F. Uyun, P. Yuwanda and N. Ransi, "Implementasi K-Means Clustering Pada RapidMiner

Untuk Analisis Daerah Rawan Kecelakaan," *Seminar Nasional Riset Kuantitatif Terapan,* p. 60, 2017.

[13] C. Kothari, Garg and Gaurav, Research Methodology: Methods and Techniques [Fourth multi color edition], New Age International, 2019, pp. 1-11.

[14] A. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *International Journal Of Artificial Intelligence Research.,* 2017. doi:10.29099/ijair.v1i2.17.

[15] F. Sibuea and A. Sapta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi),* p. 88, 2017.

BIOGRAPHY OF AUTHORS

Gita Muditha Kario Information System, Faculty of Technique, Widyatama University, Cikutra Street No.204A Sukapada, Bandung, 40125, Indonesia
gita.muditha@widyatama.ac.id

Endang Amalia Information System, Faculty of Technique, Widyatama University, Cikutra Street No.204A Sukapada, Bandung, 40125, Indonesia
endang.amalia@widyatama.ac.id