K-Means to Determine the e-commerce Sales Model in Indonesia

Ririn Restu Aria

Universitas Bina Sarana Informatika

ririn.rra@bsi.ac.id

Abstract

Sales of e-commerce in Indonesia are currently experiencing a pretty good increase by starting to switch sales that are done traditionally or offline to online sales using the use of internet media to facilitate trade transactions between sellers and buyers can be done anywhere and anytime. In this study using the K-means clustering method algorithm. The data source used in this study is based on sales using e-commerce based on provinces in Indonesia in 2018. The criteria used in the calculation are divided into 3 clusters, namely seller, reseller and dropship. Centroid data for cluster 1 = 68.30,33,58,4.95 namely Aceh, Riau, West Kalimantan, East Kalimantan, and Maluku. Centroid data for cluster 2 = 78.58,22,25.4,22 i.e. North Sumatra, West Sumatra, Jambi, Bangka Belitung Islands, D.K.I Jakarta, Central Java, East Java, Banten, Bali, Central Kalimantan, South Kalimantan, South Sulawesi and Papua and Centroid data for cluster 3 = 86.36,13.34,4.38 namely South Sumatra, Bengkulu, Lampung, Riau Islands, West Java, D.I Jogyakarta, West Nusa Tenggara, East Nusa Tenggara, North Kalimantan, Central Sulawesi, Gorontalo, West Sulawesi, North Maluku and West Papua.

*Keywords: e-commerce sales model, clustering, K-Means*

1. Introduction

The internet is one of the networks that currently makes it easy for users to carry out all activities that will be carried out without exception for the sellers to utilize the technology in marketing their products to consumers or buyers. In recent years the level of business through e-commerce has increased. Every e-commerce has its business process, for example, sales of physical goods and services, and so on [1]. E-commerce is a marketing medium that is used by sellers to be able to offer products that they want to sell to buyers. In the e-commerce business, there are several sales models, namely as a seller, a reseller and a drop shipper. During 2018 the sales model in e-commerce was dominated by sellers (80.81 percent), then followed by resellers (resellers) 19.41 percent and the smallest intermediary sellers with buyers (drop shipper) 4.78 percent. Cluster analysis is a multivariate technique whose main purpose is to group objects based on their characteristics [2]. The data used in this study is based on the 2019 e-commerce statistics document issued by the statistical center on the https://www.bps.go.id website page. In all e-commerce transaction activities that occur in Indonesia, researchers take the topic of sales models based on provinces in Indonesia. The results of calculations taken from the topic obtained by clustering by cluster into 3 namely seller.reseller and dropship.

2. Research and Methodology
2.1. E-commerce

Marketspace allows certain products or services to be sold through online transactions. Products that can be sold here are products that already have a name that is strong enough in the minds of customers and the product itself is easy to explain and distinguish between the two parties [3]. The benefits of implementing e-commerce a company can have an
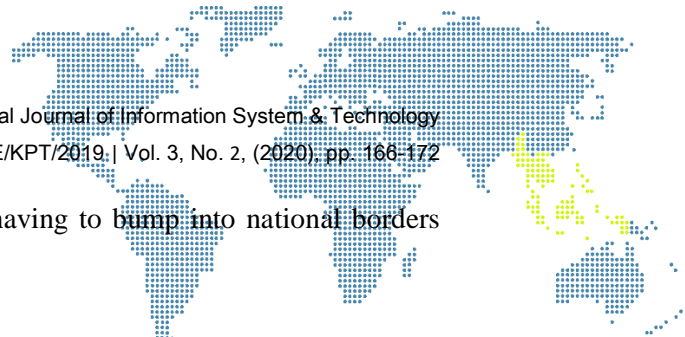
international market. Business can be run without having to bump into national borders with digital technology [4].

2.2. Algoritma K-Means

The k-means algorithm is an algorithm that partitions data into clusters so that similar data is on the same cluster and data that has inequality is on another cluster [5]. K-Means is one of the non-hierarchical (grouping) data grouping methods that can partition data into two or more groups. The method will partition the data into a group where the data with the same characteristics will be included in the same group while the data that has different characteristics will be grouped into other groups [6].

2.3. Data Mining

Data mining, often also called knowledge discovery in database (KDD), is an activity that includes the collection, use of historical data to determine regularity, patterns or relationships in large data sets.

To complete the grouping with the data mining method as follows:
(a) Data collection stage,
(b) Data processing stage,
(c) Clustering Stage and
(d) Analysis Phase [7]

The steps in the method are explained further as follows:

a. Data collection stage

The data used in this study is based on provinces that use sales models in e-commerce. The source of the data was obtained from the 2019 e-commerce statistical document published by the central statistical body, which can be accessed from https://bps.go.id. This study divided into 3 clusters, namely seller cluster, reseller cluster and dropship cluster.

b. Data processing stage

For the data processing stage, it is obtained from sales transactions using e-commerce in 2018 based on provinces and then grouped to get the clustering calculation process.

c. Clustering Stage

Cluster analysis is a method used to divide data sets into groups based on predetermined similarities [8] [7].

d. Analysis Phase

In this stage, the sales model data analysis in e-commerce is analyzed. The data obtained is then processed using the weight calculation of each index into 3 clusters, namely seller, reseller and dropship by province.

3. Results and Discussion

In conducting clustering, the data obtained will be calculated in advance based on sales transactions in e-commerce based on provinces and sales models.

Table 1. Percentage of E-Commerce Businesses by Year of Starting the E-Commerce Busines and Sales Model

Year of Business Start E-commerce	Seller	Reseller	Drop shipper
< 2010	89,89	14,44	3.25
2010-2016	81,57	19,22	5,29
2017-2018	80,03	19,70	4,51

Source: https://bps.go.id
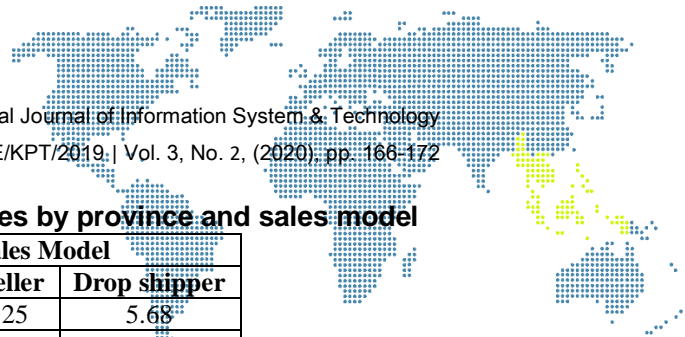
Table 2. Percentage of e-commerce enterprises by province and sales model

| Province	Sales Model		
	Seller	Reseller	Drop shipper
Aceh	76.70	31.25	5.68
Sumatera Utara	85.24	16.99	5.85
Sumatera Barat	68.90	30.71	4.33
Riau	77.25	22.75	6.88
Jambi	79.75	21.10	1.27
Sumatera Selatan	79.59	20.12	1.46
Bengkulu	87.80	14.23	2.85
Lampung	86.02	15.04	4.42
Kep. Bangka Belitung	85.65	15.28	3.24
Kepulauan Riau	79.59	24.90	5.31
DKI Jakarta	83.66	7.90	9.52
Jawa Barat	78.73	20.54	5.63
Jawa Tengah	79.07	21.59	5.63
D.I. Yogyakarta	90.78	9.64	3.63
Jawa Timur	82.86	17.00	4.55
Banten	68.07	31.93	4.83
Bali	86.25	16.11	4.32
Nusa Tenggara Barat	74.62	26.30	4.28
Nusa Tenggara Timur	67.42	40.15	4.55
Kalimantan Barat	79.93	21.53	4.74
Kalimantan Tengah	88.32	12.41	3.65
Kalimantan Selatan	85.34	14.92	2.62
Kalimantan Timur	82.14	18.83	2.60
Kalimantan Utara	65.17	33.71	8.99
Sulawesi Utara	86.90	9.61	11.35
Sulawesi Tengah	78.11	23.40	3.02
Sulawesi Selatan	78.95	17.84	5.85
Sulawesi Tenggara	74.71	24.71	5.88
Gorontalo	63.00	36.00	3.00
Sulawesi Barat	86.39	12.43	1.78
Maluku	82.61	15.42	3.56
Maluku Utara	68.85	31.15	3.28
Papua Barat	91.26	9.71	0
Papua	79.07	25.58	2.33

3.1. Determination of Centroid

In the application of the K-means algorithm, the midpoint or centroid value of the data obtained provided that the desired clusterization is 3, Cluster determination is divided into three parts namely cluster seller (C1), cluster reseller (C2) and drop shipper (C3) based on 3 If the province is taken as preliminary data, the midpoint value or centroid also has 3 points. These point values can be seen in the following Table 3:

Table 3. Preliminary Centroid Data

Atribut	Seller	Reseller	Drop shipper
C1	76.70	31.25	5.68
C2	85.65	15.28	3.24
C3	90.78	9.64	3.63

3.2. The calculation results

Furthermore, the K-Means classification algorithm will be used to group existing data. Existing data will be grouped into 3 groups. The steps of grouping data are as follows:
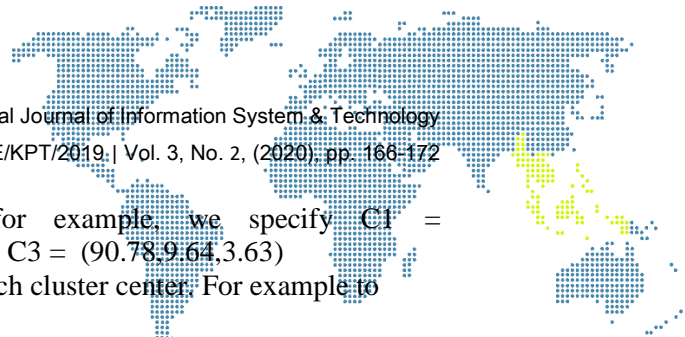
a) Determine the cluster center randomly, for example, we specify C1 = (76.70,31.25,5.68); C2 = (85.65,15.28,3.24); dan C3 = (90.78,9.64,3.63)

b) Calculate the distance of each existing data to each cluster center. For example to calculate

   1) the distance of the first student data with the center of the first cluster is:

$$d_{11} = \sqrt{((76.70 - 76.70)^2 + (31.25 - 31.25)^2 + (5.68 - 5.68)^2)} = 0.00$$

   2) Distance of first student data with the center of the second cluster:

$$d_{12} = \sqrt{((76.70 - 85.65)^2 + (31.25 - 15.28)^2 + (5.68 - 3.24)^2)} = 18.47$$

   3) Distance of first student data with the center of the third cluster:

$$d_{13} = \sqrt{((76.70 - 90.78)^2 + (31.25 - 9.64)^2 + (5.68 - 3.63)^2)} = 25.87$$

Complete calculation results in Table 4:

Table 4. Results of calculating cluster center distance 1

Province	Sales Model			C1	C2	C3	Distance
	Seller	Reseller	Drop shipper				Shortest
Aceh	76.70	31.25	5.68	0.00	18.47	25.87	0.00
Sumatera Utara	85.24	16.99	5.85	16.62	3.15	9.47	3.15
Sumatera Barat	68.90	30.71	4.33	7.93	22.80	30.38	7.93
Riau	77.25	22.75	6.88	8.60	11.82	19.12	8.60
Jambi	79.75	21.10	1.27	11.48	8.52	16.08	8.52
Sumatera Selatan	79.59	20.12	1.46	12.25	7.96	15.48	7.96
Bengkulu	87.80	14.23	2.85	20.52	2.42	5.53	2.42
Lampung	86.02	15.04	4.42	18.74	1.26	7.24	1.26
DKI Jakarta	83.66	7.90	9.52	24.67	9.89	9.40	9.40
Jawa Barat	78.73	20.54	5.63	10.90	9.01	16.37	9.01
Jawa Tengah	79.07	21.59	5.63	9.95	9.42	16.85	9.42
D.I. Yogyakarta	90.78	9.64	3.63	25.87	7.63	0.00	0.00
Jawa Timur	82.86	17.00	4.55	15.57	3.53	10.85	3.53
Banten	68.07	31.93	4.83	8.70	24.27	31.84	8.70
Bali	86.25	16.11	4.32	17.95	1.49	7.93	1.49
Nusa Tenggara Barat	74.62	26.30	4.28	5.55	15.63	23.22	5.55
Nusa Tenggara Timur	67.42	40.15	4.55	12.91	30.86	38.44	12.91
Kalimantan Barat	79.93	21.53	4.74	10.29	8.60	16.13	8.60
Kalimantan Tengah	88.32	12.41	3.65	22.23	3.94	3.70	3.70
Kalimantan Selatan	85.34	14.92	2.62	18.73	0.78	7.65	0.78
Kalimantan Timur	82.14	18.83	2.60	13.90	5.03	12.66	5.03
Kalimantan Utara	65.17	33.71	8.99	12.25	28.15	35.55	12.25
Sulawesi Utara	86.90	9.61	11.35	24.59	9.97	8.64	8.64
Sulawesi Tengah	78.11	23.40	3.02	8.41	11.08	18.71	8.41
Sulawesi Selatan	78.95	17.84	5.85	13.60	7.63	14.56	7.63
Sulawesi Tenggara	74.71	24.71	5.88	6.84	14.68	22.15	6.84
Gorontalo	63.00	36.00	3.00	14.75	30.70	38.30	14.75
Sulawesi Barat	86.39	12.43	1.78	21.52	3.29	5.52	3.29
Maluku	82.61	15.42	3.56	17.03	3.06	10.01	3.06
Maluku Utara	68.85	31.15	3.28	8.21	23.11	30.72	8.21
Papua Barat	91.26	9.71	0	26.61	8.54	3.66	3.66
Papua	79.07	25.58	2.33	7.00	12.26	19.82	7.00

Data will be a member of a cluster that has the smallest distance from the center of the cluster. For example for the first data, the smallest distance is obtained in the third cluster, so the first data will be a member of the third cluster. Likewise, for the second data, the smallest distance is in the third cluster, then the data will enter the third cluster. The complete cluster position can be seen in Table 5
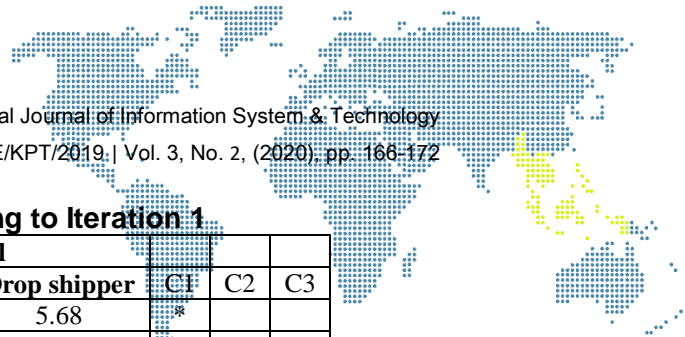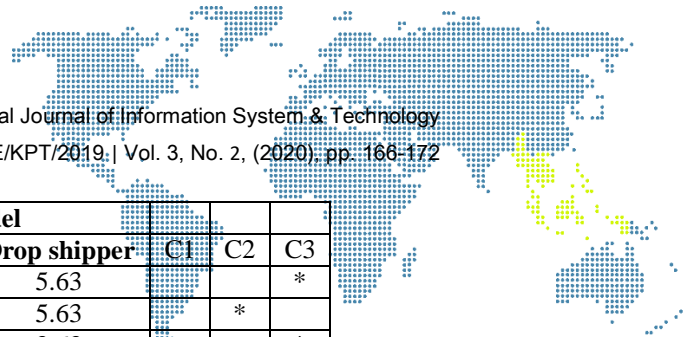
Table 5. Clustering according to Iteration 1

Province	Sales Model			C1	C2	C3
	Seller	Reseller	Drop shipper			
Aceh	76.70	31.25	5.68	*		
Sumatera Utara	85.24	16.99	5.85		*	
Sumatera Barat	68.90	30.71	4.33	*		
Riau	77.25	22.75	6.88	*		
Jambi	79.75	21.10	1.27		*	
Sumatera Selatan	79.59	20.12	1.46		*	
Bengkulu	87.80	14.23	2.85		*	
Lampung	86.02	15.04	4.42		*	
Kep. Bangka Belitung	85.65	15.28	3.24		*	
Kepulauan Riau	79.59	24.90	5.31	*		
DKI Jakarta	83.66	7.90	9.52			*
Jawa Barat	78.73	20.54	5.63		*	
Jawa Tengah	79.07	21.59	5.63		*	
D.I. Yogyakarta	90.78	9.64	3.63			*
Jawa Timur	82.86	17.00	4.55		*	
Banten	68.07	31.93	4.83	*		
Bali	86.25	16.11	4.32		*	
Nusa Tenggara Barat	74.62	26.30	4.28	*		
Nusa Tenggara Timur	67.42	40.15	4.55	*		
Kalimantan Barat	79.93	21.53	4.74		*	
Kalimantan Tengah	88.32	12.41	3.65			*
Kalimantan Selatan	85.34	14.92	2.62		*	
Kalimantan Timur	82.14	18.83	2.60		*	
Kalimantan Utara	65.17	33.71	8.99	*		
Sulawesi Utara	86.90	9.61	11.35			*
Sulawesi Tengah	78.11	23.40	3.02	*		
Sulawesi Selatan	78.95	17.84	5.85		*	
Sulawesi Tenggara	74.71	24.71	5.88	*		
Gorontalo	63.00	36.00	3.00	*		
Sulawesi Barat	86.39	12.43	1.78		*	
Maluku	82.61	15.42	3.56		*	
Maluku Utara	68.85	31.15	3.28	*		
Papua Barat	91.26	9.71	0			*
Papua	79.07	25.58	2.33	*		

From the results of the calculation of the cluster in iteration 1, the calculation of the cluster for the next iteration must be resumed until the calculation result does not change the value. The calculation will finish clustering iterated to 8 which is shown as table 6

Table 6. Clustering according to Iteration 8

Province	Sales Model			C1	C2	C3
	Seller	Reseller	Drop shipper			
Aceh	76.70	31.25	5.68	*		
Sumatera Utara	85.24	16.99	5.85		*	
Sumatera Barat	68.90	30.71	4.33		*	
Riau	77.25	22.75	6.88	*		
Jambi	79.75	21.10	1.27		*	
Sumatera Selatan	79.59	20.12	1.46			*
Bengkulu	87.80	14.23	2.85			*
Lampung	86.02	15.04	4.42			*
Kep. Bangka Belitung	85.65	15.28	3.24		*	
Kepulauan Riau	79.59	24.90	5.31			*
DKI Jakarta	83.66	7.90	9.52		*	

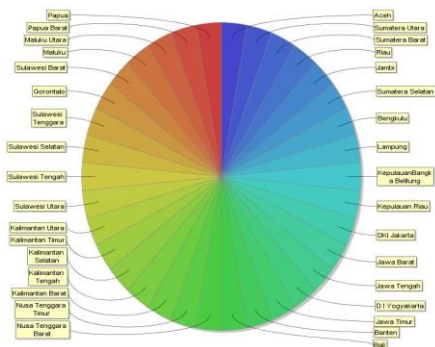| Province	Sales Model			C1	C2	C3
	Seller	Reseller	Drop shipper			
Jawa Barat	78.73	20.54	5.63			*
Jawa Tengah	79.07	21.59	5.63		*	
D.I. Yogyakarta	90.78	9.64	3.63			*
Jawa Timur	82.86	17.00	4.55		*	
Banten	68.07	31.93	4.83		*	
Bali	86.25	16.11	4.32		*	
Nusa Tenggara Barat	74.62	26.30	4.28			*
Nusa Tenggara Timur	67.42	40.15	4.55			*
Kalimantan Barat	79.93	21.53	4.74	*		
Kalimantan Tengah	88.32	12.41	3.65		*	
Kalimantan Selatan	85.34	14.92	2.62		*	
Kalimantan Timur	82.14	18.83	2.60	*		
Kalimantan Utara	65.17	33.71	8.99			*
Sulawesi Utara	86.90	9.61	11.35			*
Sulawesi Tengah	78.11	23.40	3.02			*
Sulawesi Selatan	78.95	17.84	5.85		*	
Sulawesi Tenggara	74.71	24.71	5.88			*
Gorontalo	63.00	36.00	3.00			*
Sulawesi Barat	86.39	12.43	1.78			*
Maluku	82.61	15.42	3.56	*		
Maluku Utara	68.85	31.15	3.28			*
Papua Barat	91.26	9.71	0			*
Papua	79.07	25.58	2.33		*	

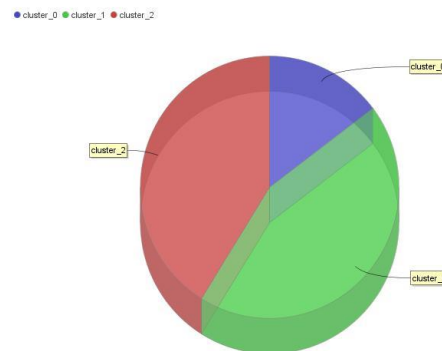**Figure 1. results of iteration cluster**

Figure 2. results of iteration cluster 8

4. Conclusion

To evaluate the sales model by province, the K-Means clustering method can be applied. Data is processed to obtain cluster values from the sales model by province. The data is processed using Ms. Excel to determine the centroid value in 3 clusters namely cluster seller, cluster reseller and cluster drop shipper. For provincial data, the data in cluster 1 seller is Aceh, Riau, West Kalimantan, East Kalimantan, and Maluku. Cluster 2 Resellers are North Sumatra, West Sumatra, Jambi, Bangka Belitung Islands, D.K.I Jakarta, Central Java, East Java, Banten, Bali, Central Kalimantan, South Kalimantan, South Sulawesi and Papua. Cluster 3 Drop shipper is South Sumatra, Bengkulu, Lampung, Riau Islands, West Java, D.I Jogyakarta, West Nusa Tenggara, East Nusa Tenggara, North Kalimantan, Central Sulawesi, Gorontalo, West Sulawesi, North Maluku, and West Papua.

References

[1]   G. L. Pritalia, "Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce," *Indones. J. Inf. Syst.*, vol. Vol.1, no. No.1, pp. 47–56, 2018.

[2]   A. P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering," *J. Teknol. Inf. Tecno.com*, vol. vol.16, no. no.4, pp. 348–357, 2017.

[3]   D. Achjari, "Potensi Manfaat Dan Problem Di E-Commerce," *J. Ekon. dan Bisnis Indones.*, vol. Vol. 15, no. No.3, pp. 388–395, 2000.

[4]   R. Shabur Miftah Maulana, Heru Susilo, "Implementasi E-Commerce Sebagai Media Penjualan Online (Studi Kasus Pada Toko Pastbrik Kota Malang)," *J. Adm. Bisnis*, vol. Vol.29, no. No.1, pp. 1–9, 2015.

[5]   F. P. Jaroji, Danuri, Putra, "K-Means Untuk Menentukan Calon Penerima Beasiswa Bidik Misi Di Polbeng," *J. INOVTEK POLBENG - SERI Inform.*, vol. vol.1, no. no.1, pp. 87–94, 2016.

[6]   Y. D. Gustientiedinaa, M.Hasmil adiya, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *J. Nas. Teknol. dan Sist. Inf.*, vol. Vol.5, no. No.1, pp. 17–24, 2019.

[7]   D. H. Sadewo, Mhd Gading, Agus Perdana Windarto, "Penerapan Datamining Pada Populasi Daging Ayam Ras Pedaging Di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering," *J. Nas. Inform. dan Teknol. Jar.*, vol. vol.2, no. no.1, pp. 60–67, 2017.

[8]   D. Wyatt, J. C, Spiegelhalter, "Field Trials of Medical Decision-Aids: PotentialProblems and Solutions," in *Proc. 15th Symposium on ComputerApplications in Medical Care*, 1991, vol. vol.1, no. ed.2.

Authors

1st Author
Ririn Restu Aria
Lecturer of Universitas Bina Sarana Informatika, Indonesia