IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.14, No.2, April 2020, pp. 199~208
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.53028

Clustering followers of influencers accounts based on likes and comments on Instagram Platform

Puji Winar Cahyo*[1], Muhamad Habibi [2]
[1,2]Department of Informatics, FTTI UNJANI, Yogyakarta, Indonesia
e-mail: ***[1]pwcahyo@gmail.com**, [2]muhammadhabibi17@gmail.com

Abstrak

Promosi barang atau jasa saat ini dimudahkan dengan adanya penyebaran informasi melalui Instagram. Penyebaran informasi biasanya dilakukan oleh akun influencer ataupun akun promosi. Akun yang digunakan tentu memiliki banyak pengikut. Karena banyaknya data pengikut dalam akun itu, maka dapat dikelompokkan ke dalam karakter yang sama. Ini dilakukan untuk menentukan potensi promosi menggunakan akun media sosial. Penelitian ini menggunakan data dari 2 akun populer. Akun pertama adalah seorang seniman dengan nama pengguna ayutingting92. Akun kedua adalah infounjaya, akun promosi resmi dari Universitas Jenderal Achmad Yani Yogyakarta. Hasil pengelompokan dapat membagi data pengikut menjadi dua kelompok klaster dengan karakter interaksi yang berbeda. Perbedaan mendasar antara kedua kelompok terletak pada jumlah suka dan komentar. Hasil analisis akun infounjaya menunjukkan bahwa dari 4.906 pengikut, hanya 3.211 pengikut yang aktif terlibat dalam interaksi, 1.695 pengikut adalah pengikut pasif yang tidak suka atau tidak berkomentar tentang interaksi tersebut. Sementara dari analisis data kluster, hasil ayutingting92 menunjukkan bahwa dari 1 juta sample data pengikut, hanya 13.591 pengikut yang terlibat aktif dalam interaksi suka dan komentar, 986.409 adalah pengikut pasif.

Kata kunci*— Klaster, Instagram, Fuzzy, Media Sosial, Influencer*

Abstract

The promotion of goods or services is now facilitated by the dissemination of information through Instagram. Dissemination of information is usually done by influencers or promotional accounts. The account used certainly has a lot of followers. Because of the large amount of follower data in that account, it can be grouped into the same characters. This is done to determine the potential for promotion using social media accounts. This study uses data from 2 popular accounts. The first account is an artist with the username ayutingting92. The second account is Infounjaya, the official promotion account from Jenderal Achmad Yani University, Yogyakarta. The results of grouping can divide follower data into two cluster groups with different interactions. The basic difference between the two groups is the number of likes and comments. The infounjaya account analysis results showed that of 4,906 followers, only 3,211 followers were actively involved in the interaction, 1,695 followers were passive followers who did not like or did not comment on the interaction. Meanwhile, the results of the ayutingting92 follower cluster show that out of 1 million sample data followers, only 13,591 followers were actively involved in the interaction of likes and comments, 986,409 were passive followers.

Keywords— *Cluster, Instagram, Fuzzy, Social Media, Influencer*

Received December 30th, 2019; Revised April 25th, 2020; Accepted April 29th, 2020

1. INTRODUCTION

The promotion of goods and services through social media is an effective way to be applied in the current era, and it can be calculated from the minimum costs imposed and the management of the promotion process that does not spend time. Buzzer arises because of buzz marketing, and buzz marketing is an alternative way of advertising by utilizing influencers or trendsetters to spread information about a product [1] while the information disseminator is defined as a buzzer. Not only that, among social media fans, the name buzzer is taken from the behavior of social media users because they carry out word-of-mouth activities or informal communication is done to evaluate goods and services [2].

One type of buzzer that is on the Instagram platform is the endorsed account, which is generally done by many fans, including artists, artists, and the community. Endorse is taken from the word endorsement, as stated by [3] in [4] endorsement is a promotion strategy by conducting market analysis at the current time, utilizing cooperation with people who are followed by many fans in a certain period to get maximum profit. The endorsement is a promotional activity carried out by influencers, while to be an active influencer account, one of them can be determined by the total number of likes [5], the number of followers above one million unique followers [6] or positive public perceptions of account holders [7].

Generally, influencers do the activity of posting or promoting a product by including specific keywords or hashtags on the caption [8], to support accuracy in searching for popular posts on a platform [9]. Influencer account posts will undoubtedly generate interaction with followers of the account; the Instagram platform is realized with comments and likes [10]. From the amount of interaction data, cluster analysis can be performed to get a group of followers on an influencer account so that variations in followers of the buzzer account on the Instagram platform can be seen, which leads to passive or active followers. The results of the grouping of active followers can be used as a parameter for selecting influencers related to people's positive perceptions [7] towards account holders (for example, artists).

Based on the problems discussed earlier, we need a cluster model that can cluster followers of influencers on the Instagram platform. This study uses the Fuzzy C-Means algorithm in the clustering process. Research that has been done related to Fuzzy C-Means, among others, is the prediction of website pages that will be visited later by the user. This study aims to group user data based on the behavior of the pages visited [11]. In addition, other studies related to Fuzzy C-Means are used to cluster news sentiments using big data processing with 60 thousand data used. The data is divided into two, 30 thousand with positives labels, and 30 thousand negatives. The results of testing with 25 thousand data were divided into two, namely 12.5 thousand positive data and 12.5 thousand negative data showed an accuracy rate of 60.2% [12]. Therefore, the clustering of follower interactions is expected to help the public in selecting influencers of a product, based on positive perceptions that have been predetermined.

2. METHODS

The stages of the process in this study include data acquisition, preprocessing, and clustering using Fuzzy C-Means. The entire stage can be seen in Figure 1.

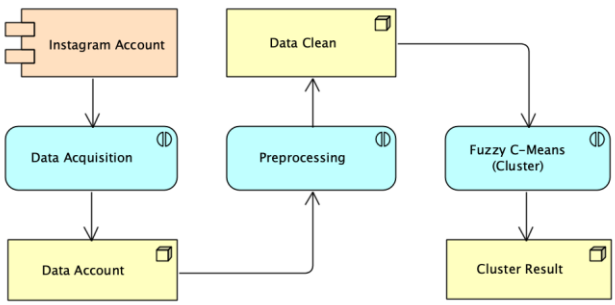

Figure 1 Research Stages

The first steps taken in this study are:

2. 1 Data Acquisition

Data availability on social media provides the flexibility of analysis to achieve specific goals. This analysis can be visualized into a chart or dashboard monitoring model [13]. For this reason, this study conducted a clustering of data on two Instagram accounts, which will be visualized in a simple cluster graph. The first account of the Instagram user analyzed was an artist and influencer, namely ayutingting92 (Indonesian artist) is the account with the most followers to date. The second account is taken from the official promotion account of Universitas Jenderal Achmad Yani Yogyakarta. Both influencer account data are based on the latest data received, 26 October 2019.

Ayutingting92 post data has an average of 33 thousand likes and an average of 300 comments per post. This study took a sampling of 1 million followers out of a total of 35 million followers. This is done because the retrieval process of making data that is quite large has limitations in terms of computing and time. Also, this study took data posting on the official promotion account of the University of Yogyakarta General Achmad Yani namely infounjaya, which has a total of 4906 followers. The number of posts owned by infounjaya account is 628 posts with an average of 300 likes and five comments.

This study uses two influencer accounts as data to be analyzed. The purpose of using these two accounts is to find out the comparison of interaction characteristics of followers with influencer accounts. The depth of the comment taken is only limited to the comments at the first level (not the replies to comments) for both accounts.

The second step taken in this research is:

2. 2 Preprocessing

Preprocessing is data which is prepared first before the data is done in the analysis phase [14]. Before the data is continued at the analysis stage, it is confirmed that the data is clean and uniform. At this stage, preprocessing is more focused on converting uniformity in the number of comments and likes made by each follower to the account posts that are followed. The uniformity of the data is realized by the numerical data that results from the conversion calculation counting the interaction of comments and likes on each follower. Data numbers of likes and comments interactions will be used as a cluster search uniformity feature of follower interactions with influencers accounts.

Whereas the third step carried out in this research is:

2. 3 Fuzzy C-Means Cluster

The follower account data that has gone through preprocessing is continued towards the clustering process. The clustering method is the learning process of scattered points (data), which are then put together into a homogeneous group based on the value of the distance between points to the center of the cluster [15].

At this stage, the clustering process uses the Fuzzy C-Means (FCM) method, where the nature of this algorithm changes the original discrete value {0,1} to a constant value [0,1]. Also, other algorithms that have similar characteristics with FCM are Fuzzy c-shells clustering (FCS) [16] and Mountain Method (MM) [17]. Where the algorithm can produce proper cluster spacing on two-dimensional and multi-dimensional features [18]. Following are the steps of the Fuzzy C-Means algorithm by Timothy [19] from Bezdek [20]:

1. Determine the number of clusters c $(2 \leq c \leq n)$ and the matrix value m' for the initial stage $U^{(0)}$, for each step labeled r where r = 0,1,2, ...

2. Start calculating the cluster center $\{V_i^r\}$ for each step, whereby using equation (1)

$$v_{ij} = \frac{\sum_{k=1}^{n} \mu_{ik}^{m'} . X_{ki}}{\sum_{k=1}^{n} \mu_{ik}^{m'}} \quad (1)$$

for *i* is the center of the cluster in the *i* feature and *j* is the *j* feature

3. Update the matrix partition for step *r*, $U^{(r)}$ following equation (2)

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}}\right)^{2/(m'-1)}\right]^{-1} \quad ; I_k = \emptyset \quad (2.a)$$

or $\mu_{ik}^{(r+1)} = 0 \text{ for } i \in I_k$ (2.b)

where $I_k = \{i | 2 \leq c \leq n; d_{ik}^{(r)} = 0\}$ (3)

and $\underset{\sim_k}{I} = \{1,2,\dots,c\} - I_k$ (4)

and $\sum_{i \in I_k} \mu_{ik}^{(r+1)} = 1$ (5)

4. If $\left\| \underset{\sim}{U}^{(r+1)} - \underset{\sim}{U}^{(r)} \right\| \leq \varepsilon_L$ then stop, if not then *r = r + 1* and return to the calculation stage 2.

In step 4, two fuzzy partition matrices will be compared in a row to achieve a reasonably good error rate accuracy ε_L. In step 3, some equations are shown, including equations (2) to equation (5). For equation (2.a) is run if the matrix partition set is still empty or still in stage r = 0, except if the result of d_{jk} is 0 then equation (2.b) will run with anticipation equation (3) and equation (4) by setting the partition membership value to 0 for each class. While equation (5) is to ensure that the sum of all columns on a fuzzy partition is not 0, the symbol $\underset{\sim}{U}$ is the total number of partitions. d_{ik} is the distance between the midpoint of the i-cluster with the k-data (where k is a member of m) using the euclidean distance method. The following equation (6) from Timothy [19] is used to find the value d_{ik}.

$$d_{ik} = d(x_k - v_i) ;$$

$$d(x_k - v_i) = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2\right]^{1/2} \quad (6)$$

3. RESULTS AND DISCUSSION

3.1 Cluster Results

The results of this study use the Instagram account data accountings92 and infounjaya that have been collected. Data clustering was performed based on membership values is exponential two, maximum iteration 1000, expected error achievement 0.005, and the number of

clusters is two. This results in a cluster distribution graph according to Figure 2 for ayutingting92 accounts and Figure 3 for infounjaya accounts.

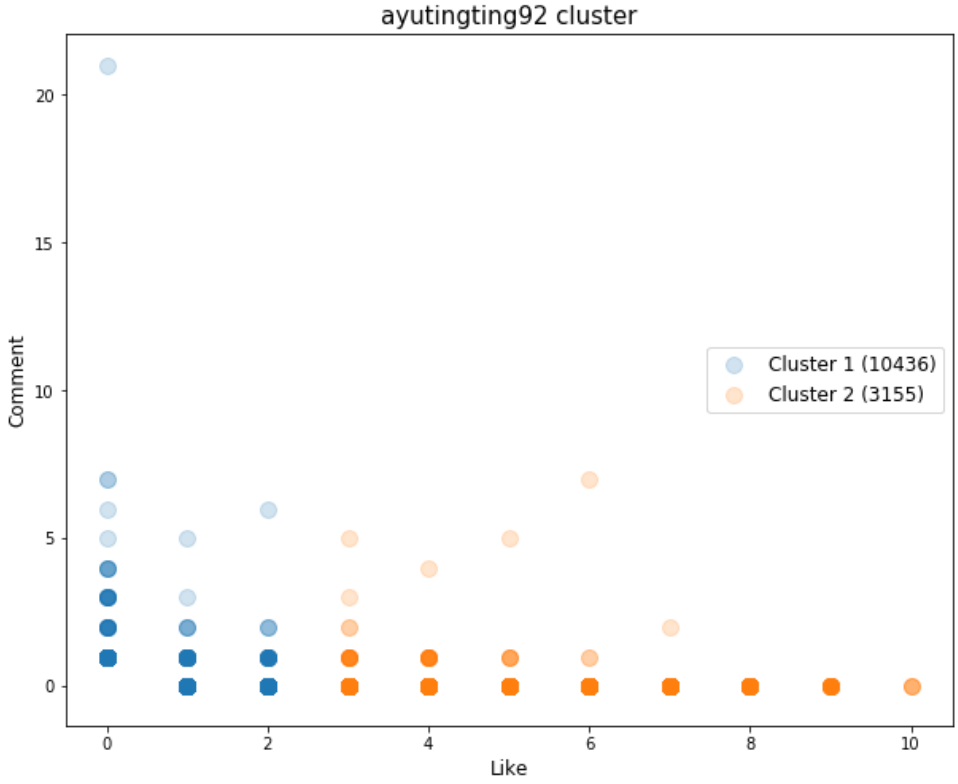

Figure 2 ayutingting92 cluster

Based on the graph shown in Figure 2, it can be seen that from 1 million follower data, only 13,591 followers who had interactions like and commented on the last ten posts. The followers were divided into 2 cluster groups, with 10,436 followers in the first cluster and 3,155 followers in the second cluster. The analysis in the first cluster shown in blue shows a group of followers who have a low amount of interest in the interaction of likes and comments. While the second cluster shown in orange is a group of followers who have a high interest in the interaction of likes compared to comments, details of the top ten interactions number account followers of ayutingting92 can be seen in Table 1.

Table 1 Top ten number of interaction followers of ayutingting92

No	Cluster 1 (Top Ten)			Cluster 2 (Top Ten)		
	Like	Comment	Total Account	Like	Comment	Total Account
1	1	0	7430	3	0	1435
2	2	0	2758	4	0	780
3	0	1	141	5	0	432
4	1	1	41	6	0	247
5	2	1	18	7	0	125
6	0	2	11	8	0	60
7	0	3	10	9	0	25
8	0	4	4	4	1	12
9	1	2	3	3	1	11
10	2	2	2	5	1	4

Table 1 shows that of 10,436 followers grouped in the first cluster, 7430 accounts have the same number of interactions 1 likes and 0 comments. While for the second cluster of 3,155 followers who have been grouped, there is the highest number of followers, namely 1435 accounts interacting 3 likes and 0 comments. The two cluster account interactions are based on the last ten posts made by ayutingting92. As for the infounjaya account clustering, see Figure 3.

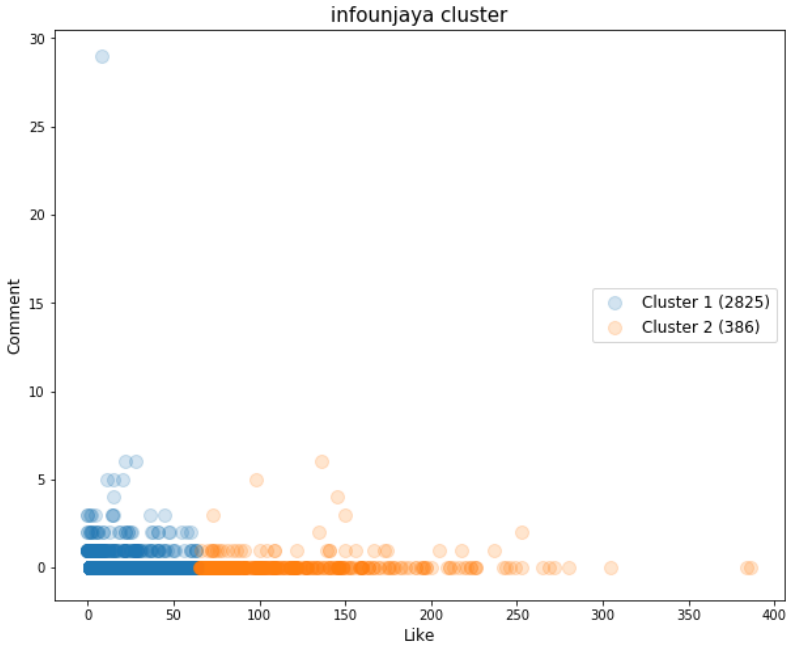

Figure 3 infounjaya cluster

From Figure 3, it can be seen that the total number of infounjaya followers who are actively interacting is 3211 followers out of 4,906 total followers. The amount is based on likes and comments from all posts made by infounjaya, a total of 3211 accounts grouped into 2 clusters. The first cluster with 2825 accounts is shown in blue, the second cluster with 386 accounts is shown in orange. Figure 3 shows the interaction of comments is not much done by the followers. The first cluster and the second cluster almost have the same interaction characteristics in the number of comments. The distinguishing characteristic is the number of likes, where the first cluster is in the range of groups of likes with a small number, while the second cluster is in the range of groups of likes with a high number. Details of the top ten number interactions of infounjaya account followers can be seen in Table 2.

Table 2 Top ten number of interaction followers of infounjaya

No	Cluster 1 (Top Ten)			Cluster 2 (Top Ten)		
	Like	Comment	Total Account	Like	Comment	Total Account
1	1	0	400	66	0	10
2	2	0	236	75	0	8
3	3	0	186	68	0	8
4	4	0	145	78	0	7
5	5	0	132	74	0	7
6	6	0	110	80	0	7
7	7	0	95	81	0	7
8	8	0	79	99	0	7
9	10	0	74	71	0	7
10	11	0	64	83	0	7

Table 2 shows that out of 2825 followers in the first cluster had the most interaction similarities, namely the number of interactions 1 likes and 0 comments with 400 followers of the account. As for the second group, the highest number of interaction similarities was in 66 likes and 0 comments with 10 followers. Both interactions in the cluster are based on 628 posts made by the infounjaya account. From the top ten interactions in the two clusters, there are comment interactions with the number 0. If it is added up, it generates 1596 account followers from 3211 active followers. It can be intended that almost half of the followers belonging to an active account without interacting comments on posts made by infounjaya account.

3.2 Cluster Evaluation

Validation of the total number of cluster divisions is necessary. Due to the validation, it can be seen from the optimal number of cluster division that will be applied to a large data set [21]. The optimal estimate for the distribution of the number of clusters in Fuzzy C-Means can be determined by calculating the value of the fuzzy partition coefficient (PC) for each cluster. Then the results of each fuzzy partition coefficient value will be compared. The closer the fuzzy partition coefficient value to 1, the more optimal the cluster division becomes [22]. We use cluster validation from Choubin et al to calculate the fuzzy partition coefficient on each cluster by equation (7) [23].

$$PC = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{nc} u_{ij}^2 \quad (7)$$

PC values are based on the range [1 / nc, 1], where nc is the number of clusters. u_ij is a fuzzy partition on the cluster. Whereas N is the number of data sets carried out by the cluster.

The evaluation was carried out on the Instagram account ayutingting92 and infounjaya six times. In the fuzzy partition coefficient calculation experiment, the exponential membership value is set two, the maximum iteration is 1,000 times, and the expected error achievement limit is 0.005, resulting in the evaluation results shown in Figure 4 for ayutingting92 and Figure 5 for infounjaya accounts.

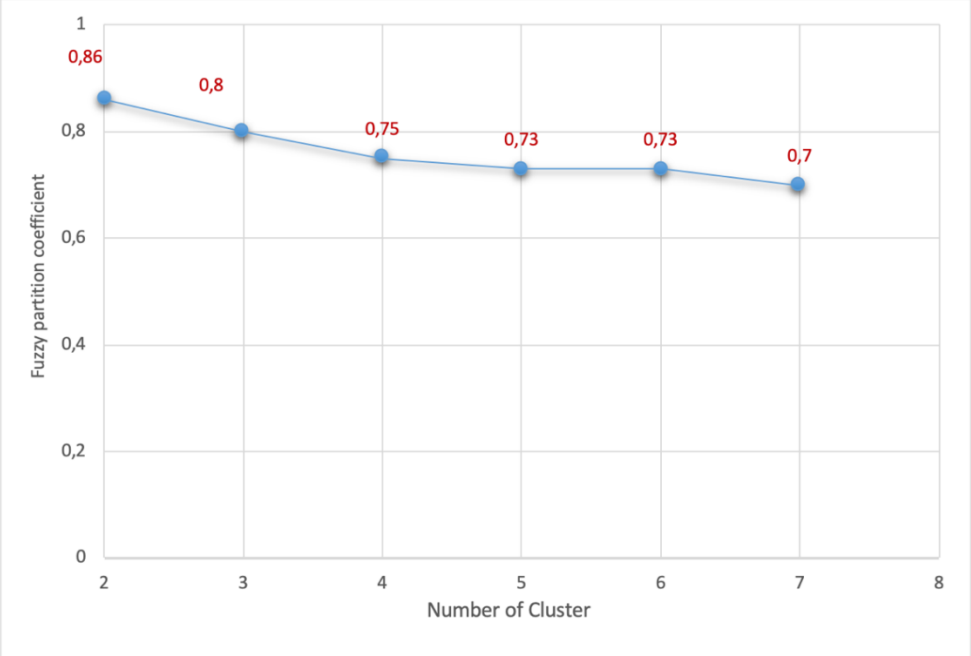

Figure 4 PC ayutingting92

Based on Figure 4, it can be seen that in the first experiment using 2 clusters produces a fuzzy partition coefficient value of 0.86. The second experiment with 3 clusters resulted in a coefficient value of 0.8 continued until the 6th experiment of 7 clusters, the fuzzy partition coefficient value decreased to 0.7.

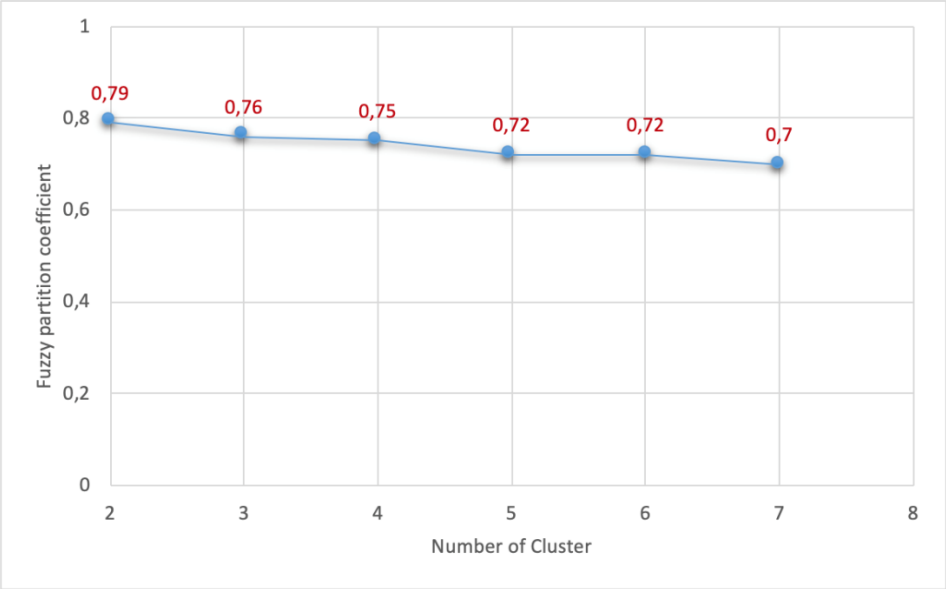

Figure 5 PC infounjaya

As shown in Figure 5, the fuzzy partition coefficient value in the first attempt reached 0.79 in the data sharing of 2 clusters. In the second experiment decreased to 0.76 in 3 clusters, decreased to 0.7 in the 6th experiment with a total of 7 clusters.

Based on experiments, the fuzzy partition coefficient values that have been done on both accounts can be seen that the most optimal number of clusters to be applied is 2 clusters. The optimal fuzzy partition coefficient value on account ayutingting92 is 0.86, while for infounjaya accounts is 0.79. Meanwhile, the fuzzy partition coefficient value decreases with the increasing number of clusters being tested.

5. CONCLUSIONS

This research succeeded in creating a model for grouping followers based on the number of interactions on influencer accounts on the Instagram platform. The optimal number of ayutingting92 clusters and infounjaya followers is two groups. Based on the results of research that has been done, the basic difference between the two clusters is on the number of likes and comments. A first cluster is an account group with a low interaction between the number of likes compared to comments, while the second cluster has a high interest in the interaction of likes compared to comments. Although the top ten data of the two clusters, there are interaction data with 0 comments.

The results of grouping analysis of infounjaya account showed that of 4,906 followers, only 3,211 followers were actively involved in the interaction, while the rest were passive followers without interactions. Analysis of cluster data sampling ayutingting92 shows that out of 1 million followers, only 13,591 followers actively interact with likes and comments, while 986,409 followers are passive followers. The results of the research from the two accounts can find out how many active followers and passive followers based on the interaction of likes and comments.

ACKNOWLEDGEMENTS

The data source used in this study comes from median-analytics.com (social media analysis platform), which is owned by the center of study and data analytic services at Jenderal Achmad Yani University, Yogyakarta. For this reason, the research team would like to thank you because with the help of data openly, providing data in the form of influencer accounts and follower interaction from Instagram platform so that this research can be completed and published.

REFERENCES

[1] M. Irwansyah, "Kajian Humas Digital: Transformasi Dan Kontribusi Industri 4.0 Pada Stratejik Kehumasan," *J. Teknol. Inf. dan Komun.*, vol. 7, no. 1, pp. 27–36, 2018.

[2] E. Sivadas and R. P. Jindal, "Alternative measures of satisfaction and word of mouth," *J. Serv. Mark.*, vol. 31, no. 2, pp. 119–130, 2017.

[3] P. Katri, "Celebrity Endorsement of Meta-Analysis?," *West. J. Nurs. Res.*, vol. 31, no. 4, pp. 435–436, 2009.

[4] R. Ahmed, S. Seedani, M. Ahuja, and S. Paryani, "Impact of Celebrity Endorsement on Consumer Buying Behavior," *SSRN Electron. J.* ·, 2015.

[5] P. L. Breves, N. Liebers, M. Abt, and A. Kunze, "The Perceived Fit between Instagram Influencers and the Endorsed Brand," *J. Advert. Res.*, vol. 59, no. 4, p. 440 LP-454, Dec. 2019.

[6] S. Kim, S. Yoo, J. Han, and M. Gerla, "How Are Social Influencers Connected in Instagram ?," *SocInfo 2017*, vol. 2, pp. 257–264, 2017.

[7] M. De Veirman, V. Cauberghe, and L. Hudders, "Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude," *Int. J. Advert.*, vol. 36, no. 5, pp. 798–828, Sep. 2017.

[8] C. Abidin, "Visibility labour: Engaging with Influencers' fashion brands and #OOTD advertorial campaigns on Instagram," *Media Int. Aust.*, vol. 161, no. 1, pp. 86–100, 2016.

[9] M. Habibi and P. W. Cahyo, "Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 4, pp. 399–408, 2019.

[10] I. Sen, A. Aggarwal, S. Mian, S. Singh, P. Kumaraguru, and A. Datta, "Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram," *WebSci'18 10th ACM Conf. Web*, pp. 205–209, 2018.

[11] R. Katarya and O. P. Verma, "An effective web page recommender system with fuzzy c-mean clustering," pp. 21481–21496, 2017.

[12] V. N. Phu, N. Duy, D. Vo, T. Ngoc, T. Vo, T. Ngoc, and T. A. Nguyen, "Fuzzy C-means for english sentiment classification in a distributed system," *Appl. Intell.*, 2016.

[13] P. W. Cahyo and E. Winarko, "Model Monitoring Sebaran Penyakit Demam Berdarah di Indonesia Berdasarkan Analisis Pesan Twitter," Universitas Gadjah Mada Yogyakarta, 2017.

[14] M. Habibi and Sumarsono, "Implementation of Cosine Similarity in an automatic classifier for comments," vol. 3, no. 2, pp. 110–118, 2018.

[15] P. W. Cahyo, "Klasterisasi Tipe Pembelajar Sebagai Parameter Evaluasi Kualitas Pendidikan Di Perguruan Tinggi," *Teknomatika*, vol. 11, no. 1, pp. 49–55, 2018.

[16] T. Wang, "A Flexible Possibilistic C-Template Shell Clustering Method with Adjustable Degree of Deformation," in *Fuzzy Systems (FUZZ-IEEE)*, 2016, pp. 1516–1522.

[17] K. P. Sinaga, J. Hsieh, J. B. M. Benjamin, and M.-S. Yang, "Modified Relational Mountain Clustering Method," in *Computing Sciences and Engineering (ICCSE)*, 2018, pp. 690–701.

[18] R. Winkler, F. Klawonn, and R. Kruse, "Fuzzy C-Means in High Dimensional Spaces

Fuzzy c-means in high dimensional spaces," *Int. J. Fuzzy Syst. Appl.*, vol. 11, no. 1, 2010.

[19] Timothy J. Ross, *Fuzzy Logic With Engineering Applications*, Third Edit. United Kingdom: John Wiley & Sons Ltd, 2010.

[20] J. C. Bezdek, "FCM : THE FUZZY c-MEANS CLUSTERING ALGORITHM," vol. 10, no. 2, pp. 191–203, 1984.

[21] M. Ren, P. Liu, Z. Wang, and J. Yi, "A Self-Adaptive Fuzzy ? -Means Algorithm for Determining the Optimal Number of Clusters," *Comput. Intell. Neurosci.*, vol. 2016, no. 1, 2016.

[22] Z. Hu, Y. Y. Bodyanskiy, and O. K. Tyshchenko, "A Cascade Deep Neuro-Fuzzy System for High- Dimensional Online Possibilistic Fuzzy Clustering," in *2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT)*, 2016, pp. 119–122.

[23] B. Choubin, K. Solaimani, H. M. Roshan, and A. Malekian, "Watershed classification by remote sensing indices : A fuzzy c-means clustering approach," *J. Mt. Sci.*, vol. 14, pp. 2053–2063, 2017.