Indonesian Journal of Electrical Engineering and Computer Science
Vol. 24, No. 1, October 2021, pp. 464~472
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v24.i1.pp464-472

K-affinity propagation clustering algorithm for the classification of part-time workers using the internet

Novendri Isra Asriny[1], Muhammad Muhajir[2], Devi Andrian[3]
[1,2]Department of Statistics, Universitas Islam Indonesia, Yogyakarta, Indonesia
[3]Data Center and Employment Information, Indonesia Ministry of Labor, Indonesia

Article Info

Article history:

Received May 16, 2021
Revised Aug 3, 2021
Accepted Aug 7, 2021

Keywords:

Cluster
Internet
K-affinity propagation
Labor
Part-time workers

ABSTRACT

There has been a significant increase in the number of part-time workers in the last 3 years. Data collected from sakernas BPS showed that the number of part-time workers was 125,443,748 in the second period of 2016. This number rapidly increased in 2017, 2018 and 2019 in the same period, by 128,062,746, 131,005,641, and 133,560,880 workers. Based on the increase in the last 3 years, East Java province has the highest number of part-time workers that use the internet. This research aims to determine the number of part-time workers that use the internet by using the k-affinity propagation (K-AP) clustering. This method is used to produce the optimal number of cluster points (exemplar) is the affinity propagation (AP). Three clusters were used to determine the sum of the smallest value ratio. The result showed that clusters 1, 2, and 3 have 3, 23, and 5 members in Bondowoso, Jombang, and Surabaya districts.

Corresponding Author:

Muhammad Muhajir
Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia
Prof. Dr. H. Zanzawi Soejoeti Building, UII Integrated Campus, Yogyakarta, Indonesia
E-mail: mmuhajir@uii.ac.id

1. INTRODUCTION

Labor is the process by which a person renders services to the public in order to earn a living. An individual that carries out certain tasks and, in return receives wages, and other forms of compensation are called a worker [1]. Based on the manpower concept, the workforce is divided into two, namely labor and non-labor forces. The workforce comprises full-time and non-full-time workers aged 15 years and above, that do not have a permanent source of income. Meanwhile, students and people responsible for managing household affairs and other non-personal activities are excluded. Full-time workers work for more than 35 hours a week while non-full-time work for a lesser number of hours.

Non-full-time workers are classified into 2 categories, namely underemployed/semi-unemployed (freelance) and part-time workers. Semi-unemployed workers (freelance) are those that work for less than 35 hours a week and are willing to accept other job offers. Conversely, part-time workers are those that are not searching or willing to accept another job [2]. The population entering the labor force in Indonesia is shown in Table 1 [3]. According to Table 1, average increases of approximately 2.07% of the Indonesian population enter the labor force yearly. An important component of this arrangement is part-time work which is one of the threats to economic and social change. This is related to the increasing diversity of the workforce and its associated changes. Many disputes associated with growth are based on part-time employment, which positively impacts society as a whole. Until now, the stigma associated with part-time workers is considered unfavorable because the wages received are relatively small. However, it allows individuals to combine work with other activities such as e-commerce-based businesses, studying, or raising a family [4].

Table 1. Population entering the labor force in Indonesia

Year	The number of workers
2016	125,441,748
2017	128,062,746
2018	131,005,641
2019	133,560,880

According to Martin and Lale [5], [6], part-time work is increasingly relevant in many developed countries because it allows workers to play a dual role in the modern labor market. In "normal time," certain groups of employers offer alternative work arrangements, including full-time employment. Part-time jobs are usually on the increase during a recession because it prompts workers and companies to make certain adjustments to the new economic conditions. This takes several forms and differs among employees and establishments. Besides, part-time workers responded to these adjustments, especially during an economic downturn (recession). The cyclical increase is largely unintentional and is widespread across different segments of the labor market. Overall, based on the available evidence, the flexibility afforded by part-time work arrangements appears to be broadly positive. The trend of Indonesian part-time workers is shown in Table 2 [3].

Table 2. Trends of part-time workers in Indonesia

Year	The number of workers
2016	23,257,887
2017	24,674,737
2018	27,371,517
2019	28,405,787

Based on Table 2, the average increase in the population of part-time workers in Indonesia is 6.4%. However, this reflects the supply and demand factors in the labor market. The market research institute e-Marketer reported that the population of internet users in Indonesia ranked 6th globally, with 83.7 million people in 2014 [7]. According to APJII & Polling [8], in 2018, the total population was 264.16 million people, out of which 64.8% (171.17 million) were internet users. In 2017, the number of internet users and the total population were recorded as 143.26 million and 262 million. The contribution of internet users is dominated by 55.7%, 16.7%, 14.3%, and 13.5% of people from Java, West Java, Central Java, and East Java Province.

In 2019, the SAKERNAS data acquired for period 2 showed that the 5 most dominant provinces in terms of part-time workers that discharge their duties using the internet are East, Central and West Java as well as Aceh and North Sumatra with 2038, 1661, 1231, 860, and 764 workers respectively [3]. Based on this analysis, the province of East Java has the highest number of workers, hence it is used as the object of this research.

Therefore, to increase the availability of the part-time labor market and the workers' welfare, each district/city in East Java was grouped in accordance with related variables. These include using the internet as a means of communication at work, promotion activities, and the selling of goods and services through e-mails, social media, websites, and marketplace applications [9]. This aids the government, especially the East Java provincial manpower office, in enacting regulations related to the employment of part-time workers.

The districts and cities in East Java were grouped using the clustering affinity propagation (K-AP). This method was adopted to obtain the optimal number of exemplars and objects through affinity propagation (AP). One advantage of this approach is that the number of k need not be entered at the beginning, besides relatively small errors tend to occur when large datasets are used compared to other cluster methods [10]. The K-Affinity Propagation method is more stable than the K-Means, where the optimal cluster is obtained using C-Index, Davies Bouldin, and Connectivity [11], [12].

Based on the aforementioned description, this research aims to determine the grouping of part-time workers in East Java province using the "K-Affinity Propagation" method, also known as K-AP Clustering. This enabled policymaker to consider the adoption of certain steps or decisions related to the implementation of future innovations in the labor sector.

2. RESEARCH METHOD

2.1. Clustering method

The clustering method is a means of grouping objects with similar values and characteristics. It is commonly used to gain insight, statistical and image analyses, machine learning, data patterns, and retrieve information [13]. In general, the grouping involves groups with small distances between objects, the area density of the data space, and the interval or distribution of certain statistics. Therefore, this method serves as a reference for the optimization of multi-purpose problems [14].

Grouping is carried out based on the distance between objects. The shape of the cluster is only affected by the size of these distances. According to Maheswari [15] and Charrad *et al.* [16], this process is calculated using the Euclidean distance as follows.

Distance with Euclidean

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (1)$$

where
$D(x, y)$ = euclidiean distance between object,
xi, yi = object coordinate,
n = number of variables

2.2. Within sum-of-square (WSS) method

WSS is one of the methods used to evaluate intracluster variability. Generally, a cluster with a small WSS is more compact than one with a large square sum. Each observation is allocated to the nearest cluster with the distance calculated using the Cosine Similarity between the observation and cluster midpoint (centroid). Furthermore, each centroid is an average of the observations in each cluster. The formula for WSS is stated as follows [17], [18]:

$$\sum_{k=1}^{K}\sum_{i\in S_k}\sum_{j=1}^{p}\left(X_{ji} - \overline{X}_{kj}\right)^2 \quad (2)$$

information
S_k = Sample from the set-in cluster k
$\overline{X}_{kj}$= The variable from cluster j to that of k

2.3. K-affinity propagation

K-affinity propagation (K-AP) modification of the AP method aims to produce an optimal number of exemplars. This new method identifies exemplars, thereby forming data point clusters. K is compared by several indices to determine its optimal form Jia *et al.* [19]. K-AP produces K clusters based on predetermined needs and parameters in terms of determining rules or controls in the message delivery process. Another advantage of this method is the belief in an object to serve as an exemplar which is automatically adapted by K-AP, while the AP is a parameter set by its users [20]. Besides, the overhead (memory usage during processing) computation of K-AP is insignificant compared to the AP. The algorithm of K-AP identifies exemplars by recursively sending real-valued messages between pairs of data points, this was inspired by AP [21]. The number of identified exemplars (clusters) is influenced by the input preferences values, although it also emerges from the message-passing procedure. The sum of *r(i, k)* and *a(i, k)* is used to determine whether or not the corresponding data point is a candidate exemplar k [22], [23]. After a data point has been selected, those placed closer to competing candidate exemplar k' are assigned to this cluster. K-AP generates k clusters by adding constraints in the process of swapping messages to limit its number while maintaining all AP clustering advantages [24], [25]. The algorithm of K-AP is stated as follows [26].

1. Input similarities matrix $s(i, k)$

$$\{s(i,k)_{i,k\in\{1,\dots,N\}i\neq k}, K\} \quad (4)$$

2. Initialize the availability $a(i, k)$, and confidence matrix $\eta^{out}(i)$

$$a(i,k) = 0 \quad (5)$$

$$\eta^{out}(i) = \min(s) \quad (6)$$

3. Renew responsibilities $r(i,k)$

$$r(i,k) = s(i,k) - \max\{\eta^{out}(i) + a(i,i)\}, \max_{k':k'\in\{i,k\}}\{a(i,k') + s(i,k')\} \tag{7}$$

4. Renewing self-responsibility $r(k,k)$

$$r(k,k) = \eta^{out}(i) - \max_{k':k'\neq k}\{a(k,k') + s(k,k')\} \tag{8}$$

5. Updating the availabilities matrix

$$a(i,k) \leftarrow \min\{0, r(k,k) + \Sigma_{i^{i}\notin\{i,k\}} max\{0, r(i',k)\}\} \tag{9}$$

6. Renewing self-availability $a(,k)$

$$a(k,k) = \Sigma_{i'\notin\{i,k\}}\max\{0, r(i',k)\} \tag{10}$$

7. Renew confidence $\eta^{in}(i)$

$$\eta^{in}(i) = a(i,i) - \max_{k':k'\neq i}\{a(i,k') + s(i,k')\} \tag{11}$$

$$\eta^{out}(i) = -R^{k}(\{\eta^{in}(j), j \neq i\}) \tag{12}$$

8. Combination of availability and responsibility $c(i,k)$ [27]

$$c(i,k) = argmax_{j}\{a(i,k) + r(i,k)\} \tag{13}$$

2.4. Standard deviation

The goodness of a cluster was determined using the standard deviation value [28], [29]. The (14) is the intra-cluster standard deviation equation,

$$S_w = K^{-1}\sum_{k=1}^{K} S_k \tag{14}$$

where S_k is the standard deviation for a variable k from cluster K. The equation of the inter-cluster standard deviation is stated as (15),

$$S_B = [(K-1)^{-1}\sum_{k=1}^{K}(\bar{X}_k - \bar{X})^2]^{\frac{1}{2}} \tag{15}$$

where $\bar{X}_k$ is the cluster average for a particular variable and $\bar{X}$ is the total average for all K clusters. The best method is the one with the smallest ratio value, after dividing Sw by Sb. However, assuming there are high homogeneity and heterogeneity values among members belonging to the same cluster, this simply means that it was appropriately formed [30].

2.5. Methodology

Secondary data was obtained from the Indonesia Ministry of Labor website, namely www.linda.kemnaker.go.id. This study adopted the 2019 data on the number of part-time workers in East Java. The research variables include Gender, Level of Education, Health Insurance, Wage Payment System, Using the Internet at Work, Training, Activities with the Most Time, Taking Care of Household, Job Status in Main Job, Contract Agreement, School Participation, and Marital Status, Trade Union Member, Age, Wage, and Number of Household Members. The K-Affinity propagation clustering method was adopted. The RStudio software was used to group districts in East Java province based on the number of part-time workers that used the internet.

In this study, the first step is to determine the topic, then determine the problems found, formulate the problem, search for study literature, preprocess the data, then make an overview, then determine the number of clusters with WSS and expert recommendations, perform clustering with the K-AP algorithm. Finally, Evaluation of cluster goodness to determine the best number of clusters. For the schematic in the form of images can be seen in Figure 1.

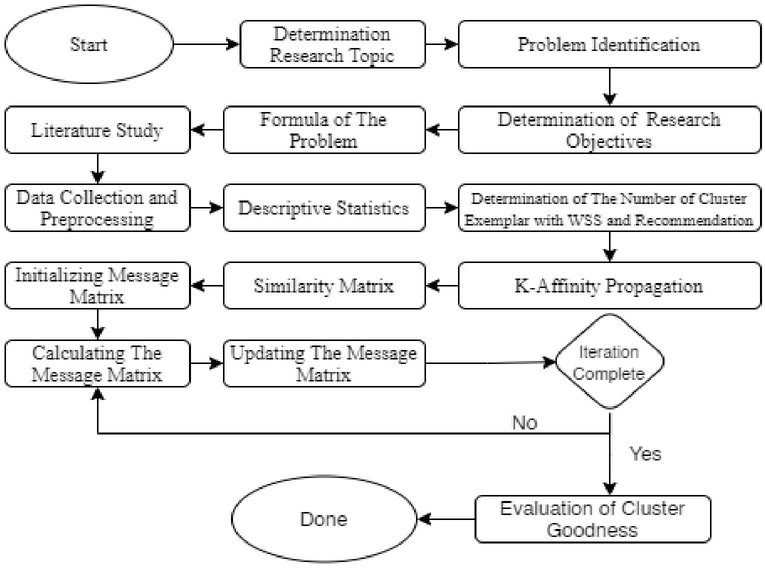

Figure 1. Research stages using k-affinity propagation method

3. RESULTS AND DISCUSSION

3.1. Overview of part-time workers in East Java

Based on the trend in the past 3 years, part-time workers using the internet to execute their main job have increased. Figure 2 shows the provinces with the highest number of part-time workers that use the internet to carry out their main job in 2019. Based on Figure 2, it is evident that they are mostly found on Java Island. The most dominating 3 provinces, are East, Central, and West Java Provinces with 2038, 1661, and 1231 workers. The other provinces have less than 1,000 part-time workers that use the internet to carry out their main job. In [32] reported that there were 20.20 million workers in East Java as of February 2019 working in various sectors, from the agricultural, forestry, and fisheries, services, and trade, to the manufacturing industry. Based on the East Java employment report, in February 2018, the number of part-time workers in the province increased from 12.52 million in 2017 to 13.55 million in 2018.

Figure 2. Distribution of part-time workers

3.2. Cluster number validation

The cluster validity test or recommendations of experts in the manpower field is used to determine the number of clusters to be used. The results of the cluster validity test are shown in Figure 3. Based on Figure 3, it is evident that from the WSS validation test, the most optimal number of clusters is 4. This is smaller than the other WSS values. Experts in the field of labor recommended the use of 3 clusters. Therefore, 3 and 4 clusters are the ideal numbers used for grouping part-time workers using the internet to discharge their main jobs. Afterward, cluster profiling was performed.

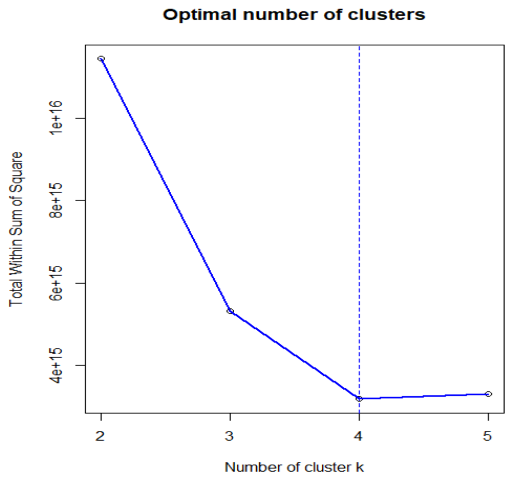

Figure 3. The optimum number of clusters using WCSS

3.3. Cluster results based on the k-ap algorithm

Based on Table 3, the use of the K-AP algorithm shows that cluster 1 is exemplar of the Kediri district or city with 3 members. Cluster 2 has an exemplar, namely the Ponorogo regency or city with 23 members. Conversely, cluster 3 has an exemplar, namely the Jember district or city with 5 members. The profiling obtained using the K-AP algorithm with 3 clusters can be known the characteristics of cluster 1 are poor, while those of 2 and 3 are very dominant and dominant.

Table 3. The analysis of K-AP using 3 clusters

Cluster	Exemplar	Objects	Characteristic of Cluster
1	Bondowoso	Bangkalan, Bondowoso, and Lumajang	All the lowest variables compared to the cluster other
2	Jombang	Banyuwangi, Blitar, Bojonegoro, Gresik, Jember, Jombang, Kediri, Lamongan, Madiun, Mojokerto, Nganjuk, Ngawi, Pacitan, Pamekasan, Pasuruan, Ponorogo, Probolinggo, Sampang, Situbondo, Sumenep, Trenggalek, Tuban, dan Tulungagung	Sex, namely the female gender, is very dominant. At the educational level, high school graduates are very dominant. This is followed by health insurance, the wage payment system dominated by wholesale and weekly packages. Furthermore, using the internet in executing the main job is very dominant, particularly as a means of communication. The school participation status, namely having been schooling and dropping out of school, is very dominant, followed by training, activities such as work, managing the household, and job status in terms of not trying to be assisted by temporary labor or female workers. In terms of the agricultural sector, fixed and non-permanent contract agreements are very dominant. The characteristics of trade union members such as age, namely those above 40 years, wages, and workers earning relatively Rp. 1,500,000, and the number of family members as many as 4 people are very dominant.
3	Surabaya	Batu, Magetan, Malang, Sidoarjo, dan Surabaya	Dominant gender characteristics, namely sex, the female gender, educational level such as high school graduates, health insurance, wage payment system particularly wholesale and weekly packages, using the internet, as a means of communication, school participation status, namely having been schooling and dropping out of school, trainings, activities particularly work, household management, work status, namely not engaging in businesses that requires the assistance of temporary labor or family worker are dominant. In terms of the agricultural sector, the fixed and the non-fixed contract agreements are dominant. The High labor union members' characteristics are dominated by age, namely those over 40 years, wage namely workers earning relatively Rp. 1,500,000, and the number of household members is mostly 4 people.

Based on Table 4, the K-AP algorithm was used to show that clusters 1, 2, 3, and 4 have exemplars with 14, 9, 7, and 1 members in Probolinggo, Madiun, Ponorogo, and Surabaya district or city. The profiling using the K-AP algorithm with 4 clusters can be known the characteristics of cluster 1 are very dominant, while that of 2, 3, and 4 are dominant, poor, and very poor, respectively.

Table 4. The analysis of K-AP using 4 clusters

Cluster	Exemplar	Objects	Characteristic of cluster
1	Probolinggo	Blitar, Jember, Jombang, Kediri, Lamongan, Madiun, Mojokerto, Pamekasan, Pasuruan, Ponorogo, Probolinggo, Situbondo, Tuban, Tulungagung	Highly dominated by the sex characteristics, namely the female gender, followed by the educational level, particularly high school graduates. In addition, the health insurance, the wage payment system, particularly the wholesale and weekly packages, using the internet as a means of communication, school participation status, namely having been schooling and dropping out of school, training, activities such as work, household management, job status, namely not needing the assistance of non-permanent workers. Fixed and non-permanent contract agreements dominate the agricultural sector. In terms of trade union members, age is the very dominant characteristic, namely those over 40 years, wage particularly those earning Rp. 500,000, and the number of household members as many 4 people are very dominant.
2	Malang	Bangkalan, Banyuwangi, Bojonegoro, Bondowoso, Gresik, Lumajang, Magetan, Malang, dan Trenggalek	Dominant gender characteristics, namely females, followed by the educational level, particularly high school graduates, health insurance, wage payment system, especially the wholesale and weekly packages. It also uses the internet as a means of communication, high school participation status, namely having been schooling and dropping out of school, trainings, activities namely work, household management, work status, such as not engaging in businesses that require temporary labor or family worker. In the agricultural sector, fixed and non-fixed contract agreements are dominant. In terms of labor union members, age is a dominant characteristic, namely those over 40 years, wage particularly those earning Rp. 2,000,000, and the number of household members is 4 people.
3	Batu	Batu, Nganjuk, Ngawi, Pacitan, Sampang, Sidoarjo, dan Sumenep	Less gender characteristics, namely female, poor educational level especially high school graduates, health insurance, including poor wage payment system namely the wholesale and weekly packages. Subsequently, using the internet as a means of communication, poor school participation status, namely having been schooling and dropping out of school, poor training, and work. Furthermore, the household management, and work status are poor, particularly not engaging in businesses, which require the assistance of temporary or family workers. In the agricultural sector, the fixed and non-fixed contract agreements are poor. The characteristics of the union members are poor. This includes the age characteristics, which are extremely dominant, namely those over 40 years, they earn a poor wage of relatively Rp. 1,500,000, and the number of household members is 4 people.
4	Surabaya	Surabaya	The least characteristics of all variables compared to the others

3.4. Evaluation of cluster goodness

Standard deviation was used to determine the ideal number of clusters, as shown in Table 5. Based on the results obtained, exemplar 3, have the smallest value of 13.45. Therefore, the smaller the standard deviation value, the greater the similarity of the objects. It was discovered that the K-AP method is the best procedure for grouping part-time workers that use the internet to execute their main job. This is also determined the Standard Deviation of K-Means and K-medoids. Based on Figure 4, comparison of three varying clustering methods resulted in a significantly different standard deviation value. Therefore, the K-AP method was used to achieve a smallest standard deviation value compared to the K-Means and K-medoids.

Table 5. The evaluation of cluster goodness

The Number of Clusters	SD
3	13,43
4	22,76

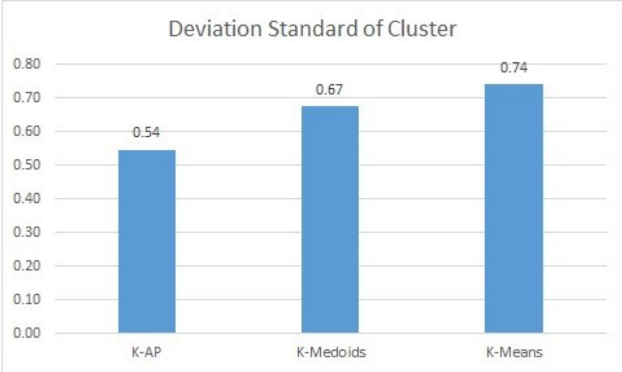

Figure 4. Standard deviation of clustering method

4. CONCLUSION

Based on SAKERNAS data, part-time workers that use the internet to carry out their jobs are mostly in Java Island, and mostly dominating are 3 cities, namely East, Central, and West Java Provinces with 2,038, 1661, and 1232 workers respectively. There were 20.20 million workers in East Java as of February 2019 working in several sectors. The best cluster of data is 3, with exemplars of cluster 1. The characteristics obtained are clusters 1, 2, and 3 are poor, very dominant, and dominant, respectively. Cluster 3 has high gender characteristics, namely females. The characteristics of using the internet as a means of communication are high. In terms of agriculture, the characteristics of the fixed and the non-permanent contracts are high. The Rp. 1,500,000 wage is high and household members, which comprises of 4 family members.

ACKNOWLEDGEMENTS

The authors are grateful to all related institutions, especially the Directorate of Academic Development, for funding supports, the Faculty of Mathematics and Natural Sciences, Islamic University of Indonesia for providing image processing facilities, coding or programming supports, and encouragement. The authors are also grateful to their colleagues for helping in completing and processing the data.

REFERENCES

[1] Indonesia government, Indonesia constitution Number 13 of 2003 on Manpower, 2003.
[2] Minister of Manpower Decree, Indonesia constituion Number 206 of 2017 on Labor development index measurement guidelines, 2017.
[3] Statistics Indonesia. Labor Force Situation in Indonesia August 2019, 2020.
[4] P. Schoukens and A. Barrio, “The changing concept of work: When does typical work become atypical?,” *European Labour Law Journal*, vol. 8, no. 4, pp. 306-332, 2017, doi: 10.1177/2031952517743871.
[5] D. B.-Martins, “Why does part-time employment increase in recessions?,” *IZA World Labor*, pp. 397-397, 2017, doi: 10.15185/izawol.397.
[6] D. B.-Martins and E. Lalé, “The ins and outs of involuntary part-time employment,” *Labour Econ.*, vol. 67, no. 101940, 2020, doi: 10.1016/j.labeco.2020.101940.
[7] D. Dzahaban and M. R. Shihab, "Customer social experience as antecedents of social commerce: Insights from Kaskus," *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS),* 2016, pp. 251-256, doi: 10.1109/ICACSIS.2016.7872766.
[8] A. Gumelar, M. I. Nasution, I. F. Oesman, F. Ramadini, and M. Irfan, “Technology mobile banking on customer satiscation”, J. Phys. Conf. Ser., vol. 15, no. 7, 2020, doi:10.1088/1742-6596/1477/7/072020.
[9] C. L. Adkins, V. Farmville, S. A. Premeaux, and L. Rock, “The use of communication technology to manage work-home boundaries,” *Journal of Behavioral and Applied Management*, vol. 15, no. e, pp. 82-100, 2014, doi: 10.21818/001c.17939.
[10] X. Zhang, C. Furtlehner, C. Germain-Renaud and M. Sebag, "Data Stream Clustering With Affinity Propagation," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 7, pp. 1644-1656, July 2014, doi: 10.1109/TKDE.2013.146.
[11] M. Muhajir and N. N. Sari, “K-Affinity Propagation (K-AP) and K-Means Clustering for Classification of Earthquakes in Indonesia,” *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, 2018, pp. 6-10, doi: 10.1109/SAIN.2018.8673344.
[12] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, “Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty,” *Expert Systems with Applications*, vol. 37, no. 7, pp. 5259-5264, 2010, doi: 10.1016/j.eswa.2009.12.070.

[13] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, "An overview of clustering methods," *Intelligent Data Analysis*, vol. 11, no. 6, pp. 583-605, 2007, doi: 10.3233/IDA-2007-11602.

[14] S. S. Ghuman, "Clustering Techniques-A Review," *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, no. 5, pp. 524-530, 2016.

[15] K. Maheswari, "Finding Best Possible Number of Clusters using K-Means Algorithm," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1S4, pp. 533-538, 2019, doi: 10.35940/ijeat.a1119.1291s419.

[16] M. Charrad, N. Ghazzali, and A. N. Boiteau, "NbClust: an R package for determining the relevant number of clusters in a data set," *J. Stat. Softw.*, vol. 61, no. 6, pp. 1-36, 2014, doi: 10.18637/jss.v061.i06.

[17] W. Hardi, W. A. Kusuma, and S. Basuki, "Clustering topic groups of documents using K-Means algorithm: Australian Embassy Jakarta media releases 2006-2016," *Berk. Ilmu Perpust. dan Inf.*, vol. 15, no. 2, pp. 226-238, 2019, doi: 10.22146/bip.36451.

[18] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the Number of Clusters in a Data Set Via the Gap Statistics," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 63, no. 2, pp. 411-423, 2002, doi: 10.1111/1467-9868.00293.

[19] H. Jia, L. Wang, H. Song, Q. Mao, and S. Ding, "A K-AP Clustering Algorithm Based on Manifold Similarity Measure," *International Conference on Intelligent Information Processing (eds) Intelligent Information Processing IX*, 2018, pp. 20-29, vol. 538, doi: 10.1007/978-3-030-00828-4_3.

[20] A. F. Moiane, and Á. M. L. Machado, "Evaluation of the clustering performance of affinity propagation algorithm considering the influence of preference parameter and damping factor," *Boletim de Ciências Geodésicas*, vol. 24, no. 4, pp. 426-441, 2018, doi: 10.1590/s1982-21702018000400027.

[21] A. M. Serdah and W. M. Ashour, "Clustering LargeScale Data Based on Modified Affinity Propagation Algorithm," *Journal of Artificial Intelligence & Soft Computing Research*, vol. 6, no. 1, pp. 23-33, 2016, doi: 10.1515/jaiscr-2016-0003.

[22] L. Wang *et al.*, "An Improved Integrated Clustering Learning Strategy Based on Three-Stage Affinity Propagation Algorithm with Density Peak Optimization Theory," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/6666619.

[23] J. Meng, H. Hao, and Y. Luan, "Classifier ensemble selection based on affinity propagation clustering," *J Biomed Inform*, vol. 60, pp. 234-242, 2016, doi: 10.1016/j.jbi.2016.02.010.

[24] H. Jia, S. Ding, L. Meng, and S. Fan, "A density-adaptive affinity propagation clustering algorithm based on spectral dimension reduction," *Neural Comput & Applic*, vol. 25, pp. 1557-1567, 2014, doi: 10.1007/s00521-014-1628-7.

[25] X. Zhao and W. X. Xu, "An extended affinity propagation clustering method based on different data density types," *Comput. Intell. Neurosci.*, pp. 1-8, 2015, doi: 10.1155/2015/828057.

[26] X. Zhang, W. Wang, K. Nørvåg, and M. Sebag, "K-AP: Generating specified K clusters by efficient Affinity Propagation," Proc. - IEEE Int. Conf. Data Mining, ICDM, no. December, pp. 1187–1192, 2010, DOI: 10.1109/ICDM.2010.107.

[27] M. J. Bunkers, J. R. Miller, and A. T. DeGaetano, "Definition of climate regions in the northern plains using an objective cluster modification technique," *Journal of Climate*, vol. 9, no. 1, pp. 130-146, 1996, doi: 10.1175/1520-0442(1996)009<0130:DOCRIT>2.0.CO;2.

[28] C. J. Veenman, M. J. T. Reinders and E. Backer, "A Maximum Variance Cluster Algorithm," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1273-1280, Sept. 2002, doi: 10.1109/TPAMI.2002.1033218.

[29] F. A. Sebayang, M. S. Lydia, and B. B. Nasution, "Optimization on Purity K-Means Using Variant Distance Measure," *2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT)*, 2020, pp. 143-147, doi: 10.1109/MECnIT48290.2020.9166\600.

[30] A. R. Barakbah and K. Arai, "Identifying Moving Variance to Make Automatic Clustering for Normal Data Set," *in Proceedings of the IECI Japan Workshop 2004 (IJW 2004),* 2004, pp. 26-30.