

Telecommunication Penetration Analysis in Indonesia by using K-Medoids Method

¹Purba Daru Kusuma and ²Walid Maulana Hadiansyah

¹Department of Computer Engineering, Faculty of Electrical Engineering,
Telkom University, Bandung, Indonesia

²Department of Telecommunication Engineering, Institute Teknologi Telkom Surabaya,
Faculty of Electrical Engineering, Surabaya, Indonesia

Abstract: Now a days, telecommunication grows fast both in speed and capacity, especially in Indonesia. In the other side, research about telecommunication implementation in Indonesia is not popular. Most researches in telecommunication focus on technology or business aspect. Based on this problem in this research, we analyze the statistic of telecommunication in Indonesia by using computational technology, especially, clustering method. We use K-medoids method rather than k-means method or agglomerative method as in our previous work. We analyze the statistic on computer usage, fixed line telecommunication, cellular telecommunication and internet access. Through this research, it is found that in fixed line telecommunication, disparity between provinces in Java island and provinces outside Java island is very wide. In the other side, in cellular telecommunication there is not any disparity between provinces in Java island and provinces outside Java island. In relation with average of net wage of the citizen in some regions there is positive relation between average of net wage and telecommunication penetration. Meanwhile, in other regions there is not any relation between average of net wage and telecommunication penetration.

Key words: Telecommunication, clustering, K-medoids, Indonesia, penetration, Java island

INTRODUCTION

Generally, researches in Information and Communication Technology (ICT) focus on the technology or business aspects. Many researches focus in speed increasing (Sattar and Matrawy, 2019), noise reduction (Chow *et al.*, 2007; Lee *et al.*, 2014) or data transmitting through many medium such as optical fibre (Chow *et al.*, 2007) or satellite (Sanctis *et al.*, 2015). Now a days, many studies in telecommunication focus on 5G technology (Ijaz *et al.*, 2016; Sattar and Matrawy, 2019; Nightingale *et al.*, 2018) or internet of things (Ijaz *et al.*, 2016; Sanctis *et al.*, 2015; Suryani *et al.*, 2017).

In business or industry aspects, many researches focus on the profitability or future trend in telecommunication industry such as Over the Top (OTT) service (Villagra and Cavalli, 2016; Datsika *et al.*, 2018; Hui, 2012; Farooq and Raju, 2019). In his research, Hui (2012) noted that to become a market winner, telecommunication industry must create innovation of its products and services to meet the diverse needs of its customers. Meanwhile, telecom companies meet Value Added Services (VAS) revenue reduction because of massive usage of OTT service such as: Whatsapp and Skype (Farooq and Raju, 2019).

This condition also occurs in Indonesia. Many researches about telecommunication in Indonesia that do not discuss about technology, usually discuss about

business aspect. Hidayat (2016) studied about the competition analysis of telecommunication industry in Indonesia, especially in consumer satisfaction aspect as switching barriers. In his research, he noted that besides competing in price, telecommunication companies should improve its customer satisfaction, technology and human resource and avoid switching barriers and at the same time attract other companies' customer (Hidayat, 2016). Ramadania *et al.* (2018) studied about parameters that affect customer loyalty in telecommunication provider. Rajasekar and Rae (2013) noted that in telecommunication industry in Oman, threat of entry and power of buyers has significant impact while power of suppliers has limited impact.

Meanwhile, researches or studies in telecommunication that focus on statistical analysis about the users are too few, especially in Indonesia. In the other side, Central Bureau of Statistics of Indonesia has already published the statistical reports of telecommunication in Indonesia. This report is published yearly. Unfortunately, as a report, most of statistical data are presented tabularly and lacks of deep and broad analysis.

Based on this problem in this research, we analyze the statistical telecommunication data of Indonesia by using computational method, especially, clustering process. The research purpose is to enrich and to give broader perspective about telecommunication in Indonesia. So, the value of this study is that it is one of

few papers that attempt to analyze non technical aspect of the telecommunication sector in Indonesia based on statistics data and by using computational technology.

In this research, we use K-medoids method in clustering process. K-medoids is a popular method in clustering method besides other popular clustering method such as k-means method. K-medoids method has been used broadly in many fields. Jung (2013) used K-medoid algorithm to construct hierarchical template tree for pedestrian contour detection. Wang *et al.* (2019) used K-medoid to clustering the data sequence. Ping *et al.* (2016) used K-medoid to cluster the breast cancer symptom.

This study is organized as follows. In the introduction, we explain the problem, gap analysis, research purpose and paper organization. In materials and methods, we explain source where statistics of telecommunication data in Indonesia that we use in this research and methods that we use to clustering this data. In the result analysis, we analyze the clustering result, especially, the cluster member and disparity among clusters. In discussion, we explore the research findings and relate the clustering result with economic, geographic and demographic data. In the conclusion, we conclude the work and propose future research potentials.

MATERIALS AND METHODS

In this research, we use data from Book of Telecommunication Statistics in Indonesia 2017. This book is published by Central Bureau of Statistics of Indonesia. This book records statistic data and activity during 2017. This data is the newest telecommunication statistics data that has been published by this bureau. This book is published in November, 2018. Meanwhile, the statistical data of citizen's average of net wage is acquired from statistical year book of Indonesia 2018 that was also published by Central Bureau of Statistics of Indonesia (Anonymous, 2018 a, b).

The book content includes the ICT indicator, development of the cellular telephone and internet access. In this book, the scope of cellular communication is: signal strength, network coverage, wireless telecommunication, fixed wireless communication and cellular phone.

Because there are lots of telecommunication statistics data, we focus on several important data. Data that are processed is shown below. We also cluster the computer aspect because now a days, many people also connect to internet while they are working or studying by using their computer. In telecommunication platform, we also cluster the fixed line platform besides cellular phone. Although, cellular platform is more popular rather than fixed line platform, fixed line platform is still used as telecommunication media.

Telecommunication clustering list; process:

- Percentage of households that own computer
- Percentage of households that use computer
- Percentage of households that use fixed line telecommunication
- Percentage of households that use cellular telecommunication
- Average number of active cellular number that is owned by households
- Average number of household members that own cellular phone
- Percentage of citizen that own cellular phone
- Percentage of households that access internet in last 3 months

In this research, we use K-medoids method to process the telecommunication data by clustering the tabular data and analyzing the cluster characteristics. This method is used to enrich our work in clustering the statistics of Indonesia. In our previous works, we have used k-means method (Kusuma, 2019a, b) and agglomerative method. Agglomerative method uses deterministic approach. In the other side, k-means method uses both deterministic and stochastic approach. In k-means method, stochastic approach is used in determining initial centroid position. Meanwhile, deterministic approach is used in determining new centroid position. Agglomerative method uses hierarchical approach while k-means method uses non hierarchical approach.

We use K-medoids method to cluster the telecommunication data of Indonesia. Generally, K-medoids method is similar to k-means method. Both of they use non hierarchical approach. It means that the number of clusters does not change from the beginning of process to the end of process.

In its basic form, steps in k-means method are similar to steps in K-medoids method. Process in K-medoids clustering method is as follows. First, number of clusters (n_c) is determined. This process is depended on user intention in how many clusters that user wants to create. The next process is determining the initial centroid's position (p_c). In common method, initial centroid position is determined randomly in data set. If the set of data is represented in A, then the members are $\{a_1, a_2, a_3, a_4, \dots, a_n\}$ where n is the number of data members. So, in its basic formulae, the initial medoid's position is determined by using Eq. 1. Variable $p_{c,i}$ represents the position of medoidi. Meanwhile, p_a represents the position of the data member:

$$P_{c,i} = \text{random}(P_{a,1}, P_{a,n}) \quad (1)$$

In Eq. 1, it is shown that the initial medoid's position can be in at any member's position. It is also shown that the initial medoid's position follows uniform distribution.

The problem is that there is possibility where there is more than one medoid has same initial position. To solve this problem in this research, the initial medoid determination is improved by limiting the random range. This function is shown in Eq. 2:

$$P_{c,i} = \left\lfloor \frac{n}{n_c} \right\rfloor \cdot i + \text{rand}(0, n_c) \quad (2)$$

After initial medoids position is determined, the next process is allocating all nodes into clusters. Similar to k-means clustering method in K-medoids method, nodes will be allocated to cluster where its medoid is the closest one. This process is formalized by using Eq. 3. In Eq. 3, connection between node and medoid is represented as link (l). So, link can be defined as connection between node and its medoid. Variable $l_{i,j}$ means link between medoid i and node j. Meanwhile d is distance between the node and the medoid. We use Euclidean distance, so that, it is formalized by using Eq. 4:

$$l_{i,j} = \min(d(P_{c,i}, P_{a,j})) \wedge l \in L \quad (3)$$

$$d(P_c, P_a) = \|P_c - P_a\| \quad (4)$$

After all nodes have been linked to their medoid, the total links distance ($d_{tot,l}$) is calculated. It is sum of all links distance. This total links distance is formalized by using Eq. 5. Then, it becomes initial total links distance:

$$d_{tot,l} = \sum_{j=1}^n d(l_{i,j}) \quad (5)$$

The next process is repositioning the medoids position. In K-medoids method, the medoids repositioning process uses stochastic approach. The next medoid position is randomized among its members. This method is formalized by using Eq. 6. In Eq. 6, $C_{mem,c}(i,t)$ is set of cluster i's members at time t:

$$P_{c,i}(t+1) = \text{rand}(P_{a,1}, P_{a,n}) \wedge a_j \in C_{mem,c}(i,t) \quad (6)$$

After medoids are repositioned, links are recalculated. If the current total link distance is lower than previous total link distance then process continues to iterate. Else, iteration stops. This process is formalized by using Eq. 7:

$$\text{Action} = \begin{cases} \text{stop, } d_{tot,l}(t) \geq d_{tot,l}(t-1) \\ \text{continue, else} \end{cases} \quad (7)$$

Based on its formalization, the K-medoids clustering algorithm is developed. This algorithm is shown in Fig. 1. This algorithm then is implemented into

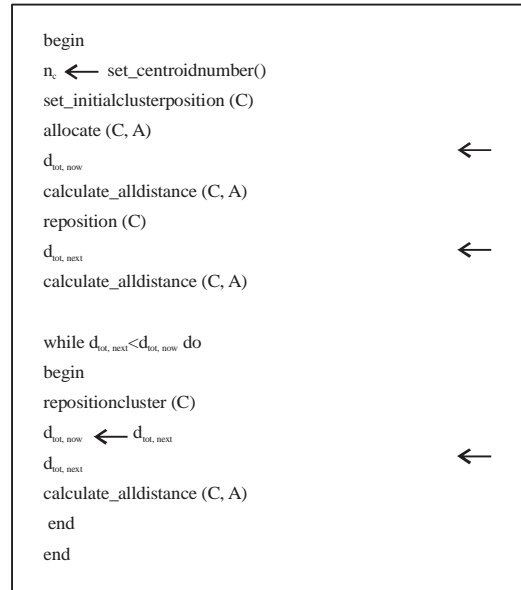


Fig. 1: K-medoids clustering algorithm

telecommunication clustering application. This application is developed by using PHP language, so that, it is a web based application. Initially, tabular statistics data is provided in pdf format. So, before it can be processed this tabular data must be entered into database. In this research, we use MySQL database. Then, the application collects data from database to be processed. The output of the clustering process is clustered data that contains information about the number of provinces, the average score and the provinces list in every cluster. In this research, there are five clusters or medoids in every clustering work.

RESULTS AND DISCUSSION

In this section, we analyze the result of the clustering process. The analysis focuses on disparity among clusters. Gap or distance in average score between clusters is also measured. Besides, provinces distributions also analyzed.

The first analysis is on the clustering result of the percentage of households that own computer. This percentage comes from number of households that own computer divided with the total households in the provinces. This parameter is important because now a days, computer is one of tools to access internet, especially, for working and studying. As penetration of mobile internet increases, accessing internet through computer increases too. The clustering result of percentage of households that own computer is shown in Table 1.

In Table 1, it is shown that the percentage of households that own computer in Indonesia is low. The

Table 1: Clustering result of percentage of households that own computer

Cluster No.	No. of provinces	Average percentage	Provinces list
1	5	12.89	North Sumatera, Lampung, West Nusa Tenggara, East Nusa Tenggara, Papua
2	7	16.85	Jambi, South Sumatera, Central Java, East Java, West Kalimantan, Gorontalo, West Sulawesi
3	6	18.78	Aceh, Bengkulu, West Java, Central Sulawesi, Maluku, North Maluku
4	9	21.07	West Sumatera, Riau, Bangka Belitung, Banten, Central Kalimantan, South Kalimantan, North Sulawesi, South-East Sulawesi, West Papua
5	7	29.96	Riau island, DKI Jakarta, DI Yogyakarta, Bali, East Kalimantan, North Kalimantan, South Sulawesi

Table 2: Clustering result of percentage of households that use computer

Cluster No.	No. of provinces	Average percentage	Provinces list
1	9	13.64	Aceh, South Sumatera, Lampung, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, West Sulawesi, North Maluku, Papua
2	12	17.45	Jambi, Bengkulu, Bangka Belitung, Central Java, East Java, Central Kalimantan, South Kalimantan, Central Sulawesi, South-East Sulawesi, Gorontalo, Maluku, West Papua
3	4	19.76	North Sumatera, Riau, West Java, North Sulawesi
4	5	21.70	West Sumatera, Banten, Bali, North Kalimantan, South Sulawesi
5	4	28.83	Riau island, DKI Jakarta, DI Yogyakarta, East Kalimantan

average percentage in the fifth cluster is <50%. Meanwhile, disparity among clusters in percentage of households that own computer is not wide. The number of provinces in a cluster ranges from 5-9 provinces. Meanwhile, the average percentage of the first cluster is only 43% of the average percentage of the fifth cluster.

In province distribution, there is not any difference between provinces in Java island and provinces outside Java island. Provinces in Java island are distributed from the second cluster to the fifth cluster. All provinces in the first clusters are from Sumatera island, Nusa Tenggara region and Papua island. Provinces in Papua island are distributed in the first cluster and the fourth cluster. Provinces in Sumatera island are distributed in all clusters.

The second analysis is analysis on clustering result of the percentage of households that use computer. This percentage comes from number of households that use computer compared with the total households in the provinces. This parameter is also measured based on assumption that not all households that own computer use computer and in the other side, not all households that use computer own computer. The clustering result of percentage of households that use computer is shown in Table 2.

Table 2 shows that there is similarity between percentage of households that use computer and percentage of households that own computer as it is shown in Table 1. Generally, percentage of households that use computer is also low as it is below 50%. The average percentage ranges from higher than 10% to lower than 30%. Three clusters have average percentage lower than 20% while two clusters have average percentage higher than 20%. The average percentage of the first cluster is 47.31% of the average percentage of the fifth cluster.

Meanwhile, there is difference in the distribution of the number of provinces among clusters between percentage of households that own computer and

percentage of households that use computer. The minimum number of provinces is 4 provinces and the maximum number of provinces is 12 provinces. So in percentage of households that own computer aspect, provinces are distributed more equally. In the other side in aspect of percentage of households that use computer, provinces are more concentrated in the second cluster.

In provinces distribution, there is little difference between percentage of households that own computer and percentage of households that use computer. In the fifth cluster there are four provinces. Two provinces are in Java island while the others are outside Java island. Provinces in Java island are distributed from the second cluster to the fifth cluster. In the first cluster, there are provinces from Sumatera, Kalimantan, Sulawesi, Little Sunda, Maluku and Papua regions.

The third analysis is on the clustering result of the percentage of households that use fixed line telecommunication. Now a days in Indonesia, although, wireless telecommunication is more popular than the fixed line telecommunication for some people or entities they still use fixed line telecommunication. They are offices, formal institutions or families that have installed fixed line telecommunication before wireless telecommunication era. For certain people, formal institution is perceived more credible if it has fixed line telecommunication. The clustering result of percentage of households that use fixed line telecommunication is shown in Table 3.

Table 3 shows that percentage of households that use fixed line telecommunication in Indonesia is very low. The lowest average percentage is only 0.97% while the highest average percentage is only 14.58%. Besides, disparity among cluster in percentage of households that use fixed line telecommunication is very wide. The average percentage of the first cluster is only 6.65% of the fifth cluster. The average percentage of the fourth cluster is only 26.4% of the average percentage of the fifth cluster.

Table 3: Clustering result of percentage of households that use fixed line telecommunication

Cluster No.	No. of provinces	Average percentage	Provinces list
1	9	0.97	Aceh, Lampung, West Nusa Tenggara, East Nusa Tenggara, Central Kalimantan, Central Sulawesi, Gorontalo, West Sulawesi, Papua
2	5	1.38	Bangka Belitung, West Kalimantan, South-East Sulawesi, North Maluku, West Papua
3	8	1.83	North Sumatera, Riau, Jambi, Bengkulu, South Kalimantan, North Sulawesi, South Sulawesi, Maluku
4	11	3.85	West Sumatera, South Sumatera, Riau island, West Java, Central Java, DI Yogyakarta, East Java, Banten, Bali, East Kalimantan, North Kalimantan
5	1	14.58	DKI Jakarta

Table 4: Clustering result of percentage of households that use cellular telecommunication

Cluster No.	No. of provinces	Average percentage	Provinces list
1	7	76.91	West Nusa Tenggara, East Nusa Tenggara, West Sulawesi, Maluku, North Maluku, West Papua, Papua
2	5	86.38	Aceh, Central Java, East Java, West Kalimantan, Central Sulawesi
3	6	88.58	West Sumatera, Bengkulu, Lampung, West Java, DI Yogyakarta, Banten
4	3	89.87	Bali, North Sulawesi, Gorontalo
5	13	92.99	North Sumatera, Riau, Jambi, South Sumatera, Bangka Belitung, Riau island, DKI Jakarta, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, South Sulawesi, South-East Sulawesi

Table 5: Clustering result of average number of active cellular number that owned by household

Cluster No.	No. of provinces	Average number	Provinces list
1	3	2.14	West Nusa Tenggara, East Nusa Tenggara, West Sulawesi
2	8	2.42	Aceh, North Sumatera, Jambi, South Sumatera, Bengkulu, Lampung, Central Sulawesi, North Maluku
3	13	2.60	West Sumatera, Riau, Bangka Belitung, Central Java, East Java, West Kalimantan, Central Kalimantan, South Kalimantan, North Sulawesi, South-East Sulawesi, Gorontalo, Maluku, Papua
4	6	2.78	Riau island, West Java, DI Yogyakarta, Bali, South Sulawesi, West Papua
5	4	3.03	DKI Jakarta, Banten, East Kalimantan, North Kalimantan

In provinces distribution, only DKI Jakarta is in the fifth cluster. Other provinces are distributed in other clusters. In this result, it is shown that there is disparity between provinces in Java island and provinces outside Java island. Besides DKI Jakarta, provinces in Java island are in the fourth cluster. Provinces outside Java island are distributed in the first cluster to the fourth cluster. In fourth cluster, there are three provinces from Sumatera region and two provinces from Kalimantan region. Provinces in Maluku and Papua regions are distributed in the first cluster and second cluster.

The fourth analysis is analysis on the result of clustering process of the percentage of households that use cellular telecommunication. This parameter is important to be analyzed because penetration of the cellular telecommunication is very massive, especially in Indonesia. The clustering result of percentage of households that use cellular telecommunication is shown in Table 4.

Table 4 shows that penetration of cellular telecommunication in Indonesia is very high. Besides, disparity among clusters in percentage of households that use cellular telecommunication is narrow. The lowest average percentage is 76.91% while the highest average percentage is 92.99%. One cluster has average percentage <80%. Three clusters have average percentage from 80-90%. One cluster has average percentage higher than 90%. The average percentage of the first cluster is 82.71% of the average percentage of the fifth cluster. So, all clusters have average percentage higher than 50%. It can

be said that most households in Indonesia use cellular telecommunication. This condition is very different than condition in percentage of households that use fixed line telecommunication as it is shown in Table 3.

Table 4 also shows that the most populous cluster is the fifth cluster. The fifth cluster contains more than ten provinces while other clusters contain <10 clusters each.

Table 4 shows that there is not any difference between provinces in Java island and provinces outside Java island. Provinces in Java island are distributed in the second, third and fifth clusters. Unfortunately, all provinces in Maluku and Papua regions are in the first cluster. In the fifth cluster, there are provinces from Java, Sumatera, Kalimantan and Sulawesi regions. Besides West Kalimantan, all provinces in Kalimantan region are in the fifth cluster.

The fifth analysis is analysis on the clustering result of the average number of active cellular numbers that owned by household. It is common in Indonesia that one person owns more than one active cellular number. For example, a person owns two active cellular numbers. One number is for business purpose and another one is for private purpose. Meanwhile, a household usually contains four members: father, mother and two children. The clustering result of average number of active cellular number that owned by household is shown in Table 5.

Table 5 shows that in average, a household that accesses cellular phone has at least two active cellular numbers. Disparity among clusters in aspect of average

Table 6: Clustering result of average number of household members that own cellular phone

Cluster No.	No. of provinces	Average number of members	Provinces list
1	1	1.84	West Nusa Tenggara
2	8	2.20	North Sumatera, South Sumatera, Lampung, Central Java, East Java, East Nusa Tenggara, Gorontalo, West Sulawesi
3	2	2.29	Bengkulu, Central Sulawesi
4	8	2.32	Aceh, Jambi, Bangka Belitung, West Java, DI Yogyakarta, West Kalimantan, South Kalimantan, South-East Sulawesi
5	15	2.56	West Sumatera, Riau, Riau island, DKI Jakarta, Banten, Bali, Central Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, South Sulawesi, Maluku, North Maluku, West Papua, Papua

Table 7: Clustering result of percentage of citizen that own cellular phone

Cluster No.	Number of provinces	Average percentage	Provinces list
1	3	41.35	East Nusa Tenggara, West Sulawesi, Papua
2	11	54.07	Aceh, North Sumatera, South Sumatera, Lampung, West Nusa Tenggara, West Kalimantan, Central Sulawesi, South-East Sulawesi, Gorontalo, Maluku, North Maluku
3	8	59.44	West Sumatera, Jambi, Bengkulu, Central Java, East Java, Banten, South Sulawesi, West Papua
4	8	64.33	Riau, Bangka Belitung, West Java, DI Yogyakarta, Bali, Central Kalimantan, South Kalimantan, North Sulawesi
5	4	73.32	Riau island, DKI Jakarta, East Kalimantan, North Kalimantan

number of active cellular phone number that owned by household is not wide. The lowest average percentage is 2.14% while the highest average percentage is 3.03%. The average percentage of the first cluster is 70.62% of the fifth cluster.

Distribution of the number of provinces among clusters follows normal distribution. The third cluster contains the highest one with 13 provinces. The number of provinces in the second cluster is close to the number of provinces in the fourth cluster and they are lower than the number of provinces in the third cluster. The number of provinces in the first cluster is almost equal to the number of provinces in the fifth cluster and they are lower than the number of provinces in the second cluster or in the fourth cluster.

In provinces distribution, there are little disparity between provinces in Java island and provinces outside Java island. Provinces in Java island are distributed from the third cluster to the fifth cluster. In the other side, provinces outside Java island are distributed in all clusters. In the fifth cluster, two provinces are in Java island while two other provinces are in Kalimantan island. All provinces in Nusa Tenggara region are in the first cluster. The interesting part is provinces in Papua region are in the third cluster and the fourth cluster.

The sixth analysis is analysis on the clustering result of the average number of household members that own cellular phone. This clustering process is to enrich result shown in the fifth cluster. If there are two household members that own cellular phone, it is assumed that they are husband and wife. The clustering result of average number of household members that own cellular phone is shown in Table 6.

Table 6 shows that most of average number of household members that own cellular phone ranges from 2 to 3 members. Only a province with average number of household members is <2 members. Disparity among

provinces in average number of household member aspect is narrow. The average number of members of the first cluster is 71.87% of the average number of members of the fifth cluster.

Basically, there is not any difference between provinces in Java island and provinces outside Java island. Provinces in Java island are in the second, fourth cluster and fifth cluster. Provinces in Sumatera are distributed from the second to the fifth cluster. Provinces in Kalimantan region are in the fourth and fifth cluster. Provinces in Sulawesi are distributed in the second to fifth cluster. All provinces in Maluku and Papua region are in the fifth cluster.

The seventh analysis is analysis on the clustering result of the percentage of citizen that own cellular phone. This clustering process is to enrich result shown in the fifth cluster. The clustering result of percentage of citizen that own cellular phone is shown in Table 7.

Table 7 shows that percentage of citizen that own cellular phone is lower than percentage of households that use cellular telecommunication but still much higher than percentage of households that use fixed line telecommunication. The highest average percentage in Table 7 is still lower than the lowest average percentage in Table 4. The average percentage among cluster ranges from 41.35-73.32%. Meanwhile, disparity in aspect of percentage of citizen that own cellular phone is little bit wide. The average percentage of the first cluster is 56.39 of the average percentage of the fifth cluster.

In number of provinces aspect, the distribution follows Poisson distribution. There is concentration in the second cluster. The number of provinces in the first cluster is the lowest among all clusters and it is little bit lower than the number of provinces in the fifth cluster. Meanwhile, the number of provinces in the third cluster is equal to the number of provinces in the fourth cluster.

Table 8: Clustering result of percentage of households that access internet in last 3 months

Cluster No.	No. of provinces	Average percentage	Provinces list
1	2	31.76	East Nusa Tenggara, Papua
2	12	46.21	Aceh, South Sumatera, Bengkulu, Lampung, West Nusa Tenggara, West Kalimantan, Central Sulawesi, South-East Sulawesi, West Sulawesi, Maluku, North Maluku, West Papua
3	4	52.77	North Sumatera, Jambi, Central Kalimantan, Gorontalo
4	8	56.91	West Sumatera, Riau, Bangka Belitung, Central Java, East Java, South Kalimantan, North Sulawesi, South Sulawesi
5	8	69.84	Riau island, DKI Jakarta, West Java, DI Yogyakarta, Banten, Bali, East Kalimantan, North Kalimantan

In provinces distribution, there is gap between provinces in Java island and provinces outside Java island. Nevertheless, this gap is not significant. Meanwhile, percentage of citizen that uses cellular telecommunication in Kalimantan region is very high. Four of five provinces in Kalimantan are in the fourth and fifth clusters. Provinces in Java island are distributed from the third cluster to the fifth cluster. The fifth cluster contains one province from Sumatera region, one province from Java region and two provinces from Kalimantan region. Three provinces in Java island are in the third cluster and two provinces in Java island are in the fourth cluster. Generally, provinces in Sulawesi region are distributed in the middle and low clusters. All provinces in Maluku are in the second cluster.

The eighth analysis is analysis on the clustering result of the percentage of households that access internet in last 3 months. This parameter is important to find connection between cellular telecommunication and computer. The clustering result of percentage of households that access internet in last 3 months is shown in Table 8.

Table 8 shows that generally, percentage of households that access internet in last 3 months is higher than percentage of households that use computer and lower than percentage of households that use cellular telecommunication. This fact strengthens argument that now a days, besides by using computer, internet is also accessed through cellular telecommunication or mobile phone. In the other side, not all of cellular phones are used to access internet. The percentage of households that access internet in last 3 months among clusters ranges from 31.76-69.84%. Meanwhile, disparity among clusters is wide enough. Average percentage of the first cluster is 45.48% of average percentage of the fifth cluster. In number of provinces aspect, provinces are concentrated in the second cluster. Nevertheless, the distribution does not follow Poisson distribution.

In provinces distribution, provinces in Java island tend to have higher percentage rather than provinces outside Java island. In Java island, two provinces are in the fourth cluster and four provinces are in the fifth cluster. Two provinces in Kalimantan are in the fifth cluster while other three provinces are distributed from the second cluster to the fourth cluster. All provinces in

Maluku are in the second cluster. Provinces in Sumatera region are distributed into from the second cluster to the fourth cluster.

Compared among clustering process, it is shown that cellular telecommunication dominates telecommunication platform in Indonesia. Its penetration is much higher than fixed line telecommunication. For telecommunication operator that provides both fixed line and cellular telecommunication, cellular telecommunication cannibalizes the fixed line one (Dustin, 2017). Fixed line telecommunication is very concentrated in Java island, especially in DKI Jakarta. It can be understood because population density in Java island is much higher than outside Java island. In the other side, investment in fixed line telecommunication is much more expensive rather than cellular telecommunication. So, fixed line telecommunication deployment is feasible only in high density area. It makes disparity among clusters in fixed line telecommunication is much wider than disparity among clusters in cellular telecommunication.

In the other side, cellular telecommunication penetration in Indonesia is very high. Now a days, most of households in Indonesia use cellular telecommunication. Meanwhile, disparity in accessing cellular telecommunication among provinces is narrow. Besides its mobility and its agility in deploying cellular telecommunication, competition in cellular telecommunication among companies in Java island is much harder than outside Java island. It is shown that through Telkomsel, Telkom as the national leader in cellular subscriber market share holds 78% of market outside Java and in contrast holds only 23% of Java market (Dustin, 2017).

In cellular telecommunication penetration, provinces in Java island and provinces outside Java island meet different problem. With more populous area with high telecommunication traffic, capacity is the main concern in deploying cellular telecommunication infrastructure in Java island. In the other side with less populous area and heavy terrain, connectivity among terrestrial infrastructure is the main concern in deploying cellular telecommunication infrastructure outside Java island.

This condition becomes reason why provinces in Papua and Maluku regions are in the lowest cluster in cellular telecommunication penetration. Papua and

West Papua are provinces with the lowest people density in Indonesia. Population density in West Papua is 9 persons km⁻² and population density in Papua is 10 persons km⁻². Meanwhile, population density in DKI Jakarta is 15,624 persons km⁻². With mountainous terrain, deploying cellular telecommunication infrastructure in Papua is difficult. Besides mountainous, based on data in Indonesia Statistical Yearbook 2018 there are 4,108 islands in West Papua. It makes Papua becomes province with highest number of islands among provinces in Indonesia. It also makes deploying cellular telecommunication infrastructure in West Papua is very expensive. This condition also occurs in Maluku and North Maluku. There are 1,286 islands in Maluku and 856 islands in North Maluku.

By comparing result in computer, cellular telecommunication and internet access, it is shown that now a days, people tend to access internet while they are using computer. This condition is triggered by the easiness of accessing internet through WiFi or connecting computer to cellular phone. As the percentage of internet access is higher than percentage of computer usage, accessing internet is not only by using computer but also by using cellular phone too. It is because cellular phone offers better mobility and flexibility rather than computer. In the other side, percentage of cellular telecommunication usage is higher than percentage of internet access. It means that not all of cellular phone is used to access internet.

In Java region, DKI Jakarta usually leads in all aspects among provinces. This position is usually followed by Banten, DI Yogyakarta and West Java. Meanwhile, Central Java and East Java usually become the lowest in telecommunication penetration in most aspects. There is relation between telecommunication penetration and average net wage of formal employee. The average of net wage in DKI Jakarta is the highest among provinces in Java island and its gap is wide. The second highest is Banten and then followed by West Java. East Java becomes the lowest provinces which its employee receive the lowest average of net wage.

In Sumatera region, Riau island becomes leader in telecommunication penetration among provinces. Then, it is usually followed by Riau and Bangka Belitung. After that, the position of other provinces is usually dynamic based on the parameter. Relating to the average of net wage data in Sumatera region, telecommunication penetration tends to have positive relation with average of net wage, although in some parameters, the relation is not always positive. Riau and Bangka Belitung are two highest provinces in Sumatera in average of net wage. Then, they are followed by West Sumatera and North Sumatera. Other provinces have average of net wage lower than 2 million rupiah per month. South Sumatera becomes province with the lowest average of net wage.

In Kalimantan region, it is shown that East Kalimantan is very dominant among provinces in all

aspects. Meanwhile, North Kalimantan becomes the second best. In some parameters, North Kalimantan is in the same cluster with East Kalimantan while in other aspects, North Kalimantan in different cluster with East Kalimantan. West Kalimantan is usually in the lowest among provinces. In Kalimantan, there is relation between telecommunication penetration and average of net wage. In Kalimantan, East Kalimantan is province with the highest average of net wage and the gap with other provinces is significant. Meanwhile, North Kalimantan is province with the second highest average of net wage. The third position is Central Kalimantan. Disparity in average of net wage between Central Kalimantan and North Kalimantan is very low. West Kalimantan and South Kalimantan are provinces with the lowest average of net wage.

In Sulawesi region, North Sulawesi and South Sulawesi are provinces that telecommunication penetration is the highest. Meanwhile, West Sulawesi is usually in the lowest among provinces. Position of other provinces is usually dynamic. In the other side in average of net wage aspect, West Sulawesi is the highest one. In the other side, South Sulawesi is province with the second lowest in average of net wage aspect.

In Little Sunda region, Bali leads the telecommunication penetration in most aspects. This condition is rationale because Bali is the most important tourism destination, not only among in Little Sunda region but also among all provinces in Indonesia. In many aspects, disparity between Bali and other provinces in Little Sunda region is wide. In relation with average of net wage, Bali is also the province with the highest among other provinces.

CONCLUSION

In this research, we have clustered and analyzed the telecommunication penetration in Indonesia. In this research, it is shown that cellular telecommunication penetrates more significant rather than fixed line telecommunication. In accessing internet when they are operating computer they access internet too. Meanwhile, people access internet not only by using computer but also by using cellular phone. In cellular telecommunication, disparity among provinces in Indonesia is not significant. In the other side in fixed line telecommunication, disparity between provinces in Java island and provinces outside Java island is very wide. In relation with average of net wage in some regions there are positive relation between average of net wage and telecommunication penetration. Meanwhile in other regions, there is not any relation between average of net wage and telecommunication penetration.

This research is a part of analysis of Informatic and Communication Technology (ICT) penetration and development in Indonesia. Besides this research, there are many other aspects that are challenging to be explored for

example in content and digital behaviour. Meanwhile, exploring and analyzing about technology platforms that are used in ICT in Indonesia and impact of ICT implementation to economic and social development in Indonesia is also challenging.

ACKNOWLEDGEMENTS

We thank to Telkom University for funding this research and Institute Teknologi Telkom Surabaya for the collaboration in this research.

REFERENCES

- Anonymous, 2018. Statistical yearbook of Indonesia 2018. Central Bureau of Statistics, Jakarta, Indonesia. <https://seadelt.net/Documents/?ID=329>
- Anonymous, 2018. Telecommunication statistics in Indonesia 2017. Central Bureau of Statistic, Jakarta, Indonesia.
- Chow, C.W., G. Talli and P.D. Townsend, 2007. Rayleigh noise reduction in 10-Gb/s DWDM-PONs by wavelength detuning and phase-modulation- induced spectral broadening. *IEEE. Photonics Technol. Lett.*, 19: 423-425.
- Datsika, E., A. Antonopoulos, D. Yuan and C. Verikoukis, 2018. Matching theory for over-the-top service provision in 5g networks. *IEEE. Trans. Wirel. Commun.*, 17: 5452-5464.
- Dustin, G., 2017. Telecommunication: Coming of age. *Mirae Asset Sekuritas Indonesia*, South Jakarta, Indonesia. https://www.miraeasset.co.id/files/bbs/01202/8980_1.pdf
- Farooq, M. and V. Raju, 2019.. Impact of Over-the-Top (OTT) services on the telecom companies in the era of transformative marketing. *Global J. Flexible Syst. Manage.*, 20: 177-188.
- Hidayat, S., 2016. The competition analysis of telecommunications industry in Indonesia: Switching barriers and consumer satisfaction. *Int. J. Econ. Manage. Sci.*, Vol. 6, No. 1. 10.4172/2162-6359.1000392.
- Hui, M., 2012. The research on development trend of telecommunications industry. *IERI. Procedia*, 3: 232-236.
- Ijaz, A., L. Zhang, M. Grau, A. Mohamed and S. Vural *et al.*, 2016. Enabling massive IoT in 5G and beyond systems: PHY radio frame design considerations. *IEEE. Access*, 4: 3322-3339.
- Jung, H.G., 2013. Medoid selection from sub-tree leaf nodes for K-medoid clustering-based hierarchical template tree construction. *Electron. Lett.*, 49: 108-109.
- Kusuma, P.D., 2019. Strategic ports activities analysis in Indonesia by using agglomerative Clustering method. *J. Eng. Applied Sci.*, 14: 8733-8742.
- Kusuma, P.D., 2019. Woman participation analysis in public universities in Indonesia by using K-means method. *J. Eng. Appl. Sci.*, 14: 5690-5698.
- Lee, K., J. Cho and Y. Park, 2014. Channel prediction-based noise reduction algorithm for dual-microphone mobile phones. *IEEE. Trans. Consum. Electron.*, 60: 393-401.
- Nightingale, J., P. Salva-Garcia, J.M.A. Calero and Q. Wang, 2018. 5G- QoE: QoE modelling for ultra-HD video streaming in 5G networks. *IEEE. Trans. Broadcast.*, 64: 621-634.
- Ping, Q., C.C. Yang, S.A. Marshall, N.E. Avis and E.H. Ip, 2016.. Breast cancer symptom clusters derived from social media and research study data using improved K-medoid clustering. *IEEE. Trans. Comput. Soc. Syst.*, 3: 63-74.
- Rajasekar, J. and M.A. Rae, 2013. An analysis of the telecommunication industry in the sultanate of Oman using Michael Porter's Competitive Strategy model. *Competitiveness Rev.*, 23: 234-259.
- Ramadania, M. Theresia and I. Sadalia, 2018.. The determinants of customer loyalty for telecommunication provider. *J. Res. Bus. Econ. Manage.*, 10: 1918-1931.
- Sanctis, M.D., E. Cianca, G. Araniti, I. Bisio and R. Prasad, 2015. Satellite communications supporting internet of remote things. *IEEE. Internet Things J.*, 3: 113-123.
- Sattar, D. and A. Matrawy, 2019. Optimal slice allocation in 5G core networks. *IEEE. Networking Lett.*, 1: 48-51.
- Suryani, V., S. Sulistyono and W. Widyanan, 2017.. Internet of Things (IoT) framework for granting trust among objects. *J. Inf. Process. Syst.*, 13: 1613-1627.
- Villagra, D.R. and A.R. Cavalli, 2016. Analysis and influence of economical decisions on the quality of experience of OTT services. *IEEE. Latin Am. Trans.*, 14: 2773-2776.
- Wang, T., Q. Li, D.J. Bucci, Y. Liang, B. Chen and P.K. Varshney, 2019. K-medoids clustering of data sequences with composite distributions. *IEEE. Trans. Signal Process.*, 67: 2093-2106.