Comparison of Multiple Distances On Panel Data in Non-Hierarchical Clustering Method

Adella Sari Cahyani Sugiono*, Indahwati*, I Made Sumertajaya*

*Department of Statistics, IPB University

DOI: 10.29322/IJSRP.10.07.2020.p10320
http://dx.doi.org/10.29322/IJSRP.10.07.2020.p10320

Abstract- Cluster analysis is a multivariate analysis that classifies objects based on their characteristics. Clustering analysis is generally used in cross-section data, that typically taken one point in time and unlike panel data that are taken at multi-times of several objects. This study explores methods for clustering analysis of panel data via distance measures. The objective of this research is to compare the Manhattan Distance, Euclidean Distance, Maximum Distance, Frechet Distance and Dynamic Time Warping (DTW) Distance for Clustering Analysis of Panel Data. The best of the distance measure was implemented empirically for Clustering of Indonesian's Province that based on the Human Development Index (HDI), from 2010-2019. Results show that the Manhattan, Euclidean and Maximum provide distances with optimum performances, when the generated data between the clusters are not overlapping. However, when there were overlaps between clusters, the Manhattan distance was the most appropriate.

Index Terms : *Clustering panel data, Manhattan Distance, Euclidean Distance, Maximum Distance, Frechet Distance, DTW Distance*

I. INTRODUCTION

Cluster analysis is a multivariate analysis that aims to classify objects based on their characteristics (Matjik 2011). Methods for clustering analysis are generally divided into two types: non-hierarchical and hierarchical. The development of the clustering method can also be classified according to the size of the data, namely fewer objects (cluster 1 stage) usually using a hierarchical method and many objects (cluster 2 stages) that use non-hierarchical methods. In this research, the clustering analysis method used is the non-hierarchical via the K-means method.

The cluster method can also be distinguished based on the characteristics of the data. Clustering analysis is generally used in cross-section data. Unlike panel data, cross-section data pose main weakness of only taken at one point in time and thus ignoring the data trend overtime. This study uses a panel data that allow combination of time series data and cross-section data taken in several times on an object. Because the structure of the panel data is different from the cross-section data, the choice of distance measure must accommodate the panel data structure, that is dynamic following time of some object. The most basic of cluster analysis is "distance", generally measured using the Euclidean distance which is not appropriate for the panel data.

Several reports on clustering analysis with panel data employ several distance measures. Zheng et al (2014) used an “Euclidean distance timed and spaced”. Moreover, Genolini et al (2015), used the R packages to implement k-means designed to work specifically on trajectories (kml) or on joint trajectories (kml3d). They used the “Minkowski Distance” to calculate the distance for cluster analysis in which Manhattan distance was assigned when the rank of the distance formula is 1 and Euclidean distance was assigned when the rank is 2. The distance becomes maximum when it is heading to the infinite rankings ($p \rightarrow \infty$). Later, Genolini et al (2016) assessed the longitudinal data hordes based on its data form and utilized the Frechet distance and the Dynamic Time Warping (DTW). The literature used different distance measure for cluster analysis on longitudinal and panel data. The used clustering methods were not based on a model, but rather the raw data. To our best knowledge, no report so far that compare various distance measure to cluster panel data.

The objective of this study is to compare the Manhattan, Euclidean, Maximum, Frechet and Dynamic Time Warping (DTW) Distance for Clustering Analysis of Panel Data. The best of the distance measure was implemented empirically on real data of Province of Indonesia based on the Human Development Index (HDI) variables, from 2010-2019. The data used for distance comparison were panel data with time of observation during ten years and the number of observation objects were 34 (number of provinces in Indonesia). The used variables were Life Expectancy, School Length Expectation, Average School Length, and Percentage Expenditures.

II. MATERIALS AND METHODS

2.1 Clustering Analysis in Panel Data

Clustering Analysis in Panel Data is different from cluster analysis in cross-section data, in this clustering the algorithm that used must be able to support the structure of panel data. K-Means method can be used in panel data clustering by modifying the size of the distance adjusted to the panel data structure. However, there are several approaches that can be used for clustering panel data. According to Liao (2005) there are three categories of approaches in panel data cluster analysis. First the raw data-based approach, how it works by calculating the distance between objects in a certain period of time using raw data. The results obtained are used for the clustering analysis. Frequently used distances are distances based on correlation values. Second, the approach based on features, by eliminating the influence of

noise data and reducing the dimensions of the data, after which the distance calculation is done and the clustering process is carried out. Third, the model-based approach that has been formed from raw data. In this research, the approach used for panel data clustering is raw data based approach to determine the distance in the clustering analysis process. Clusters analyzing with the K-Means method using raw data from generated data, with the following stages:

a. Partitioning the observation object into k-cluster randomly.
b. Calculating the cluster center of each k-cluster.
c. Calculating the observation object's distance to the cluster center by using several distance equation as follows: Manhattan Distance, Euclidean Distance, Maximum Distance, Frechet Distance and Dynamic Time Warping Distance
d. Inserting the observation object into a new cluster that has closest distance to the cluster center
e. Repeating stage b to d so there is no any observation object moving

2.2 Distances

Distance measure is used to measure the similarity of data in a cluster. The results of the clustering process will produce different results if the distance measure used is different. Used to clustering analysis with data panels (Johnson 1998).

2.1.1 Minkowski distance

The Minkowski distance between two joint variable-trajectories is given in Eq. 1 (*Genolini et al, 2015*) :

$$Dist\ (y_{1..}, y_{2..}) = \left\{ \sum_{j=1}^{T} \sum_{X=1}^{M} \left[y_{1jX} - y_{2jX} \right]^{p} \right\}^{\frac{1}{p}} \quad (1)$$

The Euclidean distance is obtained by setting p = 2, it is defined as :

$$Dist\ (y_{1..}, y_{2..}) = \sqrt{\sum_{j=1}^{T} \sum_{X=1}^{M} \left[y_{1jX} - y_{2jX} \right]^{2}} \quad (2)$$

The Manhattan distance by setting p = 1, the definition of the Manhattan distance is given by :

$$Dist\ (y_{1..}, y_{2..}) = \sum_{j=1}^{T} \sum_{X=1}^{M} \left[y_{1jX} - y_{2jX} \right] \quad (3)$$

The Maximum distance is obtained by passing to the limit $p \rightarrow +\infty$, it takes the form :

$$Dist\ (y_{1..}, y_{2..}) = \left\{ \sum_{j=1}^{T} \sum_{X=1}^{M} \left[y_{1jX} - y_{2jX} \right]^{p} \right\}^{\frac{1}{p}}, p \rightarrow \infty \quad (4)$$

In practice, the kml3d package uses Euclidean distance as the default distance. But it also allows users to define their own distance. If the units of each variable are different. then the data must be standardized in advance (Genolini 2015).

2.1.2 Frechet distance

The Frechet distance between P and Q is the maximum for all reparameterization of α and β from [0, t], thus the distance between objects P (α (t)) and Q (β (t)) where object P is a function of time α (t), and object Q is a function of time β (t), so it can be written as follows in (Eq.5) Genolini (2016):

$$DistFrechet\ (P, Q) = \underset{\alpha, \beta}{Inf} \ \underset{t \in [0,1]}{Max} \left\| P(\alpha(t)) - Q(\beta(t)) \right\| \quad (5)$$

2.1.3 Dynamic Time Warping (DTW) distance

The stages of the Dynamic Time Warping (DTW) distance are as follows (Muller 2007):

1. Calculating the local distance between elements from the two collections uses a different distance calculation technique, usually used is Euclidean Distance.
2. Determines the warping path, where the warping path is a path or path through a matrix that contains a minimum distance from the $d_{DTW}(i,j)$ element to the $d_{DTW}(N,M)$ element consisting of $d_{DTW}(i,j)$ elements themselves.
   - First Row
     $$d_{DTW}(1,j) = \sum_{k=1}^{j} d_{(x(1),y(k))}\ ,\ j \in [1,M] \quad (6)$$
   - First Column
     $$d_{DTW}(i,1) = \sum_{k=1}^{i} d_{(x(k),y(1))}\ , i \in [1,N] \quad (7)$$
   - The Other Elemen
     $$d_{DTW}(i,j) = d_{(x(i),y(j))} + min \begin{cases} DTW(i-1,j) \\ DTW(i,j-1) \\ DTW(i-1,j-1) \end{cases} \quad (8)$$

2.3 Data

The data used in this study are simulation data and real data. Simulation data obtained through data generation using statistical software, namely software R. The following is an explanation of the data used in this study:

1. Statistical simulation is done using the R programming language. The stages of the simulation data generation procedure are as follows:
   a. Setting the number of clusters is $M = 3$.
   b. Specifies the number of objects in each cluster equal to n.
   c. Sets the number of series of times ($t = 10$).
   d. Set a model for generating data :
      $$y_{1ij}(k) = \beta_{0i}(k) + \beta_{1i}(k)t + \varepsilon_{ij}\ (k), \quad (9)$$
      $i = 1, 2, 3, \ldots, n$
      $j = 1, 2, 3, \ldots, t$
      $k = 1,2,3, \ldots, M$

Where β_{0i} and β_{1i} are defined by the equation $\beta_{0i} = \beta_0 + u_{0i}$ and $\beta_{1i} = \beta_1 + u_{1i}$. Then, u_{0i}, u_{1i} and ε_{ij} are random factors with normal distribution $(0, \sigma^2_{uo})$, $(0, \sigma^2_{u1})$, and $(0, \sigma^2)$. With u_{0i}, u_{1i} and ε_{ij} are assumed to be mutually independent. In this study, there are 16 scenarios data generation which is consist of 4 different combinations of parameters there are β_0, β_1, σ^2_{uo} and σ^2_{u1}.

.

2. The real data is used in this research is sourced of BPS Publication. The data used is the Human Development Index (HDI) variables in Indonesia, from 2010-1019. The variables that used are Life Expectancy, School Length Expectation, Average School Length, and Percentage Expenditures. Data is used in this research is panel data with time of observation during ten years and the number of observation objects were 34 Provinces of Indonesia.

2.4 The Procedure of Data Analysis

The stages of analysis which is conducted in this study are generally divided into two parts, namely the analysis of the generation data and analysis of the real data with the following details :

1) Analysis of The Generation Data

This stage is the simulation stage, the data used are generated data with 1000 replications. The purpose of this simulation is to see the accuracy of clustering between several measures of distance with the K-Means method in classifying objects in the form of panel data. The size of the distance used in the panel data clustering are the Distance of Manhattan, Euclidean, Maximum, Frechet and DTW. The first stage is data generation, the simulation uses generation panel data with 16 scenarios of data was generated. Dataset generation consist of 3 clusters with a different parameter values. The next step was clusters analyzing with the K-Means method using raw data from generated data with k value used is k = 3. The important thing was calculating the observation object's distance to the cluster center by using several distance equation as follows: Manhattan Distance (Equation 2), Euclidean Distance (Equation 3), Maximum Distance (Equation 4), Frechet Distance (Equation 5) and Dynamic Time Warping Distance (Equation 6). After that, evaluating the five distance of measurement. Those 5 evaluation of the distance are done by calculating incorrect classification or accuracy in classifying objects based one the initial cluster that is generated, with following formula (Kohavi 1998) :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

Thus, Determining the best distance measurement between distance of Manhattan, Euclidean, Maximum, Frechet and Dynamic Time Warping (DTW) by looking at the largest cluster accuracy based on the evaluation criteria in Equation 10.

2) Analysis of Real Data

This stage is the stage of implementation the clustering method with the best distances of simulation results on real data. The first step in this section is data exploration was carried out on real data, that is human development index (HDI). The first thing to do is data imputation, the next is standardization and identified distribution of the data from each component of HDI variable. Analyzing clustering using the K-Means method based on the best distance measure obtained from the simulation results with the optimum number of clusters. Evaluated the K-Means clustering method using quality criteria of Calinski and Harabasz (Genolini, 2016) is given in :

$$C(k) = \frac{Trace(B)}{Trace(W)} \frac{n-k}{k-1} \quad (11)$$

Determine the optimum number of groups based on the largest Calinski and Harabasz criteria value. Thus, draw a conclusion from the result of evaluated clustering with the best distance measurement.

III. RESULT

3.1 Evaluation of Distance Accuracy

The accuracy of the Manhattan, Euclidean, Maximum, Frechet and Dynamic Time Warping (DTW) distances can be seen from the resulting of accuracy values. Which is the greater the accuracy value, then the better the distance size. Figure 1 presents a comparison of the distribution of accuracy for the five distances. In general based on the box plot, Manhattan Distance is superior compared to other distances, but in some data generation scenarios there is a distance that has a performance that is relatively similar to the Manhattan Distance, namely Euclidean Distance. From the simulation results, it can be seen that for Manhattan Distance and Euclidean Distance, there are relatively outliers in some scenarios, because the accuracy value only consists of 100% in every replication. Likewise for Maximum, Frechet, and DTW distances under these conditions, the results obtained are far different compared to the results of other scenario simulations.

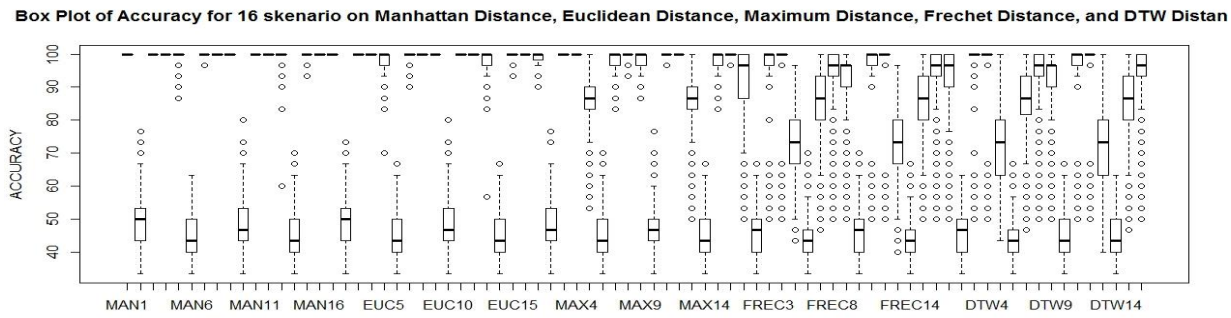

Figure 1 : The Distribution of Accuracy for Five Distances Measure

In Figure 1, it can also be seen that the position of the boxplot at Manhattan and Euclidean distances is much higher and stable compared to the boxplot position for Maximum, Frechet and DTW distances. However, the position of the boxplot for each distance looks very different when the β_0 parameter condition is close while the β_1 value is both small moreover enlarged. Under these conditions, it appears that the position of the boxplot is lower than the accuracy of the other scenarios. This shows that the accuracy value in the scenario is lower than the accuracy in the other scenario.

Table 1 : Results of Average accuracy on Manhattan Distance, Euclidean Distance, Maximum Distance, Frechet Distance, and DTW Distance for Each Scenario

Accuracy	Distances				
	Manhattan	Euclidean	Maximum	Frechet	DTW
Scenario 1	100	100	99.01	90.32	90.46
Scenario 2	49.83	48.81	47.59	45.58	45.82
Scenario 3	100	100	100	95.05	95.47
Scenario 4	100	100	100	91.96	91.98
Scenario 5	99.43	98.47	85.84	72.20	71.58
Scenario 6	45.61	45.18	45.11	44.55	44.56
Scenario 7	99.87	99.73	98.13	84.22	85.50
Scenario 8	100	100	99.85	93.53	93.90
Scenario 9	100	100	98.94	89.98	90.04
Scenario 10	48.72	48.08	47.12	45.77	45.73
Scenario 11	100	100	99.93	95.63	95.40
Scenario 12	100	100	100	91.67	92.51
Scenario 13	99.27	98.40	85.84	72.20	71.58
Scenario 14	45.45	45.14	44.98	44.59	44.89
Scenario 15	99.83	99.70	97.93	84.95	84.99
Scenario 16	100	100	99.82	92.56	93.38

The results of the clustering accuracy for the simulation data are presented in Table 1. The results of the average accuracy for each of these distances were obtained from 1000 replications. In Table 1, there are 16 types of scenarios where entirety it can be seen that the average accuracy for five distances in several scenarios is relatively the same. When viewed based on changes in parameters, when the β_0 parameter between distant clusters while the values of β_1 as same between the clusters, with varians of β_0 and β_1 are small or only variance of β_0 is enlarged (Scenarios 1 and 9) has a relatively equal average accuracy. In that scenario, Manhattan and Euclidean distances have the same average accuracy, which is 100%. This shows that in these conditions Manhattan and Euclidean distances are the best distance measurements with the most optimum performance. Meanwhile, when only variance of β_1 is enlarged while variance of β_0 is small, Manhattan Distance is the distance that has the best performance with accuracy averange is 99.43%. Likewise with the condition of variance β_0 and β_1 enlarged Manhattan Distance is also the distance that has the best performance with an average accuracy of 99.27%.

In Table 1, it can be seen that the average results for Scenarios 2, 6, 10, and 14 have much lower accuracy results compared to the results of the average accuracy of other scenarios. The results of that average accuracy, obtained when the parameter β_0 between close clusters while the value of β_1 between the clusters is the same. When data is generated under these conditions, the data between the resulting clusters overlap so that the size of these distances is very difficult to distinguish preliminary data between clusters. Under these conditions, both with small β_0 and β_1 variations or with enlarged variants of β_0 and β_1, Manhattan distances as a whole produce the greatest average accuracy, this shows that overall Manhattan Distance is a measure of distance with the best performance under these conditions. Then, when the β_0 parameter between the clusters are same and the values of the β_1 parameter between clusters have a far different difference with small β_0 and β_1 variance, shows that the Manhattan, Euclidean and Maximum distances have the same average accuracy, which is 100%. Meanwhile, when the β_0 variance is small and β_1 variance is enlarged as well as the β_0 and β_1 variance conditions are equally enlarged, showing that Manhattan Distance is the best distance measurement with an average accuracy of 99.8%. Then, when variance of β_0 is enlarged while β_1 variance is small indicates that the Manhattan and Euclidean distances have the same average of accuracy, which is 100%.

When the parameter values of β_0 and β_1 between distant clusters with the variance of β_0 and β_1 are small or the β_0 variance is enlarged while the small β_1 variance shows the distance of Manhattan, Euclidean and Maximum has the same average accuracy, which is 100%. It showed that under these conditions, distances of Manhattan, Euclidean and Maximum have the best performance. Meanwhile, when the β_0 and β_1 variance are enlarged moreover the β_0 variance is small while the β_1 variance is enlarged showing the Manhattan and Euclidean Distances have the same average accuracy , which is equal to 100%. It showed that under these conditions Manhattan and Euclidean Distance have the best performance. Entirety, the Frechet Distance and DTW Distance are the distances that have the lowest average accuracy for all scenarios in this study. In general, the two distances are usually used for data generation that has a particular shape, so both distances will have optimum performance when the data generated has a certain shape.

Based on the evaluation for each scenario above, it can be seen that overall for each scenario show that the Manhattan, Euclidean and Maximum provide distances with optimum performances, when the generated data between the clusters are not overlapping. However, when there were overlapping data between clusters, the Manhattan was the most appropriate distance. Thus, based on the results of the evaluation of all these scenarios it can be concluded that the best distance for clustering panel data with the K-Means method is Manhattan Distance.

3.2 The Best Distance in The Panel Data

Panel data clustering with the K-Means method can be applied to clustering the Human Development Index (HDI) data. Where, the HDI data used in the form of panel data with a time of observation for 10 years from 2010 to 2019 and the object of observation is 34 provinces in Indonesia. Based on previous simulation results, the K-Means method with Manhattan Distance is able to provide the best results in clustering panel data. Thus, the K-Means method will be applied with this distance in clustering data with 4 components of HDI variables.

3.2.1 Exploration of Panel Data

Data exploration is carried out at an early stage to find out the pattern of data distribution and the pattern of data trends over time. Panel data plots are carried out to see the pattern of each data in the Human Development Index (HDI) variable component in Indonesia from 2010 to 2019. HDI data plots in Indonesia are presented in Figure 2. Overall HDI data patterns for Life Expectancy (AHH) variables , Expectations of School Length (HLS), Average Length of School (RLS) and Expenditures per Capita are relatively linear.

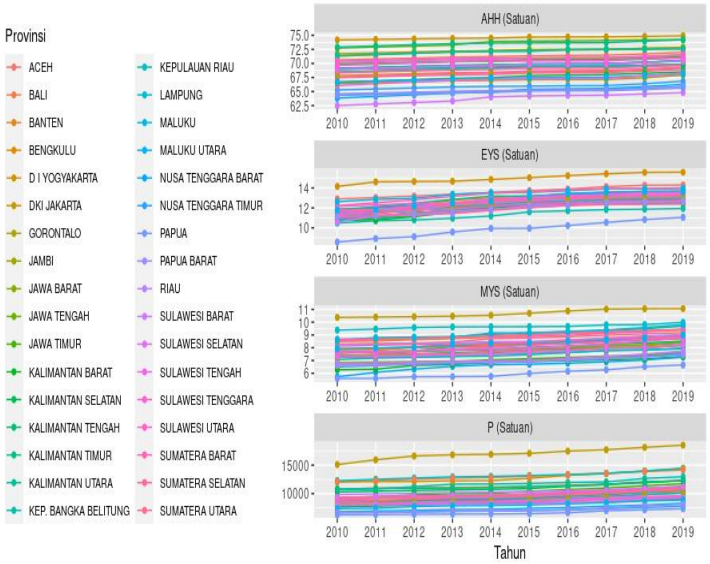

Figure 2 : Plotting of Human Development Index (IPM) Data from 2010-2019

3.2.2 The Optimal Number of Clusters

K-Means method is a type of non-hierarchical clustering, where the method requires the determination of the number of clusters (k) at the beginning. The number of clusters tried was k = 2 to k = 10. The optimum number of clusters is obtained from the Calinski & Harabatz (CH) criterion value, the higher the CH value, the better the clustering results.

Table 2 Calinski & Harabatz Criteria Value for Each Number of Cluster (k)

Cluster Number (k)	Criteria Value of Calinski & Harabatz
2	14.3267
3	14.4182
4	12.3539
5	14.5625
6	15.6631
7	11.2854
8	12.9051
9	10.0336
10	12.3361

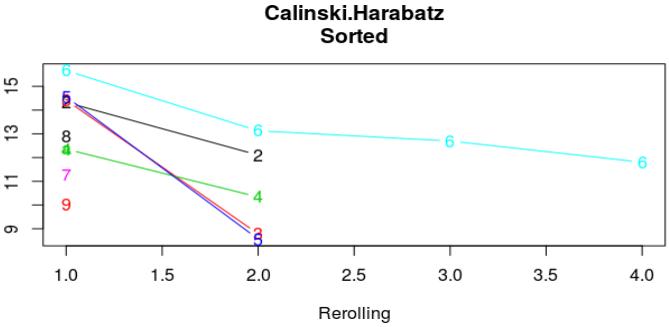

Figure 3 : Comparison of Calinski & Harabatz Criteria for Each Cluster Number (k)

Table 2 is a comparison of the value of the CH criteria for each number of clusters that have been tested in this study. Based on the table shows that the highest CH criterion value is obtained by k = 6, this can also be seen in Figure 3. In the figure it appears that the highest CH criterion value is obtained by cluster 6. Therefore the optimum cluster for HDI data clustering with Method K-Means and Manhattan Distance are k = 6.

3.2.3 The Optimum Member Of Clusters

In the previous stage, the optimum cluster produced was 6 clusters with the size of the Manhattan distance in the K-Means method, based on the Calinski & Harabatz criteria values. Table 3 presents members for each cluster. Provinces that are in one cluster have the same HDI movement pattern from year to year.

Table 3 : The Details Of Optimum Clusters

Clusters	Number of Clusters	Member of Clusters
A	15	Riau, Jambi, Sumatera Selatan, Lampung, Kep. Bangka Belitung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Bali, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan.
B	8	Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, Sulawesi Tengah, Sulawesi Tenggara, Maluku, Maluku Utara.
C	6	Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Gorontalo, Sulawesi Barat, Papua Barat.
D	2	Kepulauan Riau, Jakarta.
E	2	Di Yogyakarta, Kalimantan Timur.
F	1	Papua

Table 4 : Average of HDI Data Based on Optimal Clusters

Gerombol	AHH	HLS	RLS	PP	IPM
A	70.414	12.146	7.883	10270	68.990
B	68.048	13.013	8.512	8798	68.011
C	66.128	12.201	6.923	8225	63.662
D	70.744	12.398	10.176	15116	76.060
E	74.064	13.977	9.013	12090	75.719
F	64.933	9.872	6.000	6631	57.330

Table 4 presents the average HDI data based on the optimal cluster for 10 years (2010-2019). Table 4 shows that the largest average Life Expectancy (AHH) and School Old Expectancy (HLS) variables were obtained by Cluster E, consisting of DIY and East Kalimantan Provinces. Whereas, cluster D which consists of Riau Islands Province and Jakarta has the largest average in the variable Length of School (RLS) and Expenditures per Capita (PP) for the past 10 years. Based on the average HDI in Table 4 obtained from BPS publications for 10 years (2010-2019), it shows that the highest average HDI obtained by D cluster with a value of 76,060, this is influenced by the high average values of the RLS and PP variables in the D cluster every year. Meanwhile, cluster F which consists of Papua Province has the lowest average of 57,330. So it can be concluded that the Riau Islands Province and Jakarta are the Provinces that have the best Human Development Index for the past 10 years. Meanwhile, Papua Province has the lowest Human Development Index in Indonesia for the past 10 years.

IV. CONCLUSION

Manhattan, Euclidean and Maximum distances are both distances that have optimum performance, when the data between the clusters generated are not overlapping. However, when the resulting data overlaps between clusters, Manhattan Distance is the distance that has the most optimum performance compared to other distances. Meanwhile, Frechet Distance and DTW Distance are the distances that have the lowest average accuracy for all scenarios in this study. In general, the two distances are usually used for data generation that has a particular shape, so both distances will have optimum performance when the data generated has a certain shape. So, based on Scenarios 1 to 16, overall the distance that has the best performance with the highest accuracy value in each panel data generation is obtained by Manhattan Distance. So, the best distance for panel data clustering with the K-Means method is Manhattan Distance.

References

[BPS] Badan Pusat Statistik RI. 2017. *Indeks Pembangunan Manusia (IPM)*. Jakarta (ID): [Downloaded 2019 February 25] from https://www.bps.go.id/subject/26/indeks-pembangunan manusia.html

Genolini C, Alacoque X, Sentenac M, Arnaud C. 2015. *KML and KML3D:R Packages to Cluster Longitudinal Data*. Journal of Statistical Software. Volume 65, Issue 4.

Genolini C, Ecohard R, Benghezal M, Arnaud C, *et al.* 2016. *kmlShape:An Efficient Method to Cluster Longitudinal Data(Time-Series) According to Their Shapes*. DOI:10.1371/journal.pone.0150738.

Johnson RA, Wichern DW. 1998. *Applied Multivariate Statistical Analysis* Fourth Edition. New Jersey: Prentice-Hall International.

Kohavi R, Provost F, 1998. *Special Issue on Applications of Machine Learning and the Knowledge Discovery Process*. Boston: Machine Learning, 30, 271-274 (1998).

Matjik AA, Sumertajaya IM. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS.* Bogor: IPB Press

Muller M. 2017. *Dyanmic Time Warping Information Retrieval For Music and Motion*. Berlin: Springer.

Zheng B, Li S. 2014. *Multivariable Panel Data Cluster Analysis and Its Application.* China: Anhui University.

ACKNOWLEDGEMENTS

This work is fully supported by Kemenristek DIKTI (Kementerian Riset Teknologi dan Pendidikan Tinggi) of Indonesia.

Authors

First Author – Adella Sari Cahyani Sugiono S.Pd, college student, Department of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, Email: adellasugiono23@gmail.com
Second Author – Dr. Ir. Indahwati, M.Si, Lecturer, Departement of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, Email: indah.stk@gmail.com
Third Author – Dr. Ir. I Made Sumertajaya, M.Si, Lecturer, Departement of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, Email: imsjaya.stk@gmail.com

Correspondence Author - Dr. Ir. Indahwati, M.Si, Lecturer, Department of Statistics, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor,16680, Indonesia, Email:indah.stk@gmail.com