GROUPING OF REGENCIES/CITIES IN INDONESIA BASED ON NATIONAL HEALTH INSURANCE (JKN) WITH ENSEMBLE ROCK APPROACH

Rahmania Azwarini[1*], Morina A. Fathan[2], Tri Widiantoro[3]

[1]Institut Teknologi Sepuluh Nopember, Surabaya,60111, Indonesia

[2]Tadulako University, Palu, 94119, Indonesia

[3]Badan Pusat Statistik Kabupaten Aceh Tenggara, Aceh, 24651, Indonesia

[1]rahmaniaawarini@its.ac.id, [2]morinafathan19@gmail.com, [3]tri.widiantoro@bps.go.id

*(*corresponding author)*

ABSTRACT

Health is a fundamental need for every individual, and the National Health Insurance (JKN) program in Indonesia aims to provide equitable access to healthcare services for all citizens. Despite its implementation, disparities in JKN participation persist across different regencies/cities, indicating that the implementation of JKN is not uniform. Therefore, comprehensive mapping of JKN participant profiles across regions is necessary. This study aims to group 34 regencies/cities in Indonesia based on JKN participant characteristics using a clustering approach that integrates both numerical and categorical data. The Ensemble Robust Clustering using links (ROCK) method was used, combining hierarchical clustering for numerical data and the ROCK method for categorical data. This study used eight numerical variables (such as age, household size, household expenditure, healthcare expenditure, tobacco expenditure, ability to pay for health insurance, willingness to pay for health insurance, and insurance expenditure) and six categorical variables (such as living area, sex, education, reasons for joining JKN, and categories of ability and willingness to pay). The results showed three groups based on quality: low, medium, and high, each with distinct socio-economic and demographic patterns. These findings can help policymakers in more equitably allocating healthcare resources and reducing disparities in JKN implementation across regions. The priority group for implementing the JKN program is Group 1 which still has Low Quality, after that the government can focus on districts/cities in Group 2 (Medium Quality), and finally the districts/cities in Group 2 (Medium Quality).

Keywords   : Health, Clustering, ROCK, Ensemble ROCK

I.   INTRODUCTION

Health is a basic right of every person and every citizen has the right to get health services. The government are responsible for health insurance, especially for the poor society, which is in accordance with the mandate of the 1945 Constitution. The government carries out the mandate of the 1945 Constitution by organizing social security programs, one of them is the National Health Insurance (JKN). The National Health Insurance (JKN) is a nationally administered social security program based on the principles of social insurance and the principle of equity [1].

According to Law Number 40 of 2004 concerning the National Social Security System (UU SJSN), JKN is held by 4 main actors, namely JKN participants (the society), Health Social Security Administration Agency (BPJS), Health Facilities, and the Government. The purpose of JKN is to ensure that JKN participants (the public) receive health care benefits so that people can live a healthy, productive, and prosperous life [1].

The National Health Insurance Program (JKN) aims to enhance public access to quality health services. Understanding community participation in JKN is crucial for evaluating the program's effectiveness and identifying factors influencing enrollment. Variables used in grouping JKN participation, such as education level, income, employment status, information access, and perception of health service quality, significantly explain participation patterns. For instance, education level affects individuals' understanding of JKN benefits, where individuals with higher education are more aware of the importance of health insurance [2]. Income is also a major determinant since lower-income individuals often rely on government subsidies to join JKN [3].

Additionally, employment status determines the type of participation, whether through formal, informal, or independent worker segments, impacting accessibility and affordability of [4]. Information access plays a key role in increasing public knowledge about registration procedures and program benefits, as research indicates that information campaigns boost participation by up to 15% [5]. Lastly, the perception of health service quality is a critical indicator, as negative experiences can decrease public trust in the program [6]. By analyzing these variables, the grouping approach can reveal underserved community groups and provide insights for more effective policy-making.

In fact, there are still disparities in the implementation of the JKN program in several regencies/cities in Indonesia. The disparities show that the implementation of the JKN program in each regencies/cities is not evenly distributed, meaning that the implementation of the JKN program still needs to be improved in several regions [7]. In order to overcome the gaps in the JKN program in Indonesia, it is necessary to map the profiles of regencies/cities based on the characteristics of the JKN participating community. Regencies/cities profile mapping is carried out to see the similarity of the conditions of JKN participants in one area to other regions. Mapping related to several regencies/cities in Indonesia can be done through one of the statistical methods, namely clustering analysis.

One of the common problems encountered in clustering analysis is the presence of mixed-scale data, both numerical and categorical. A frequently used method for clustering mixed-scale data involves transforming categorical data into numerical data and vice versa. [8] in [9] transformed categorical variables into numerical form, and then object clustering was performed using numerical data clustering methods. The advantage of the transformation method is that it can reduce computational complexity. However, this method has a drawback in determining the appropriate transformation to avoid losing much information from the original data.

Apart from clustering with the transformation method, an ensemble clustering method for mixed data was developed by He, Xu, and Deng [10]. Clustering for mixed data uses the ensemble ROCK (Robust Clustering using linKs) method by combining the clustering outputs for categorical and numerical data. The method applied for categorical data is ROCK, while for numerical data, the hierarchical agglomerative method is used. To determine the best clustering results, the criterion used is the ratio between within-group standard deviation (Sw) and between-group standard deviation (Sb), with the smallest ratio [11].

Data on people's ability and willingness to pay for health insurance (JKN) in Indonesia consists of mixed data, including numerical and categorical data. Therefore, the suitable clustering method for this research is the ensemble ROCK method approach. The results of this study are expected to assist the government in compiling and determining priorities for the implementation of the JKN program so that the allocation of the health assistance budget can be evenly distributed, and there are no social disparities between societies.

II. METHODS

The data used in this study is secondary data related to ability and willingness to pay for health insurance in Indonesia 2021. Data obtained from HARVARD Dataverse Website. The data used consists of 8 numeric variables and 6 categorical tioe data variables. The data used in this study is 34 regencies/cities in Indonesia.

There are 14 variables used. There are as many as 8 numeric variable data, while categorical variable data. The following is a table of research variables. The following is a table of research variables.

Table 1 : Research Variables

Variable	Mark	Data Type	Category
X_1	Age	Numeric	-
X_2	Household Size	Numeric	-
X_3	Household Total Expend	Numeric	-
X_4	Expend Healthcare Household	Numeric	-
X_5	Tobacco Expend Household	Numeric	-

X_6	ATP (Ability to Pay) Health Insurance in Household	Numeric	-
X_7	WTP (Willingness to Pay) Health Insurance in Household	Numeric	-
X_8	Expend Insurance Household	Numeric	-
X_9	Living Area	Categoric	1: Urban, 2: Rural
X_{10}	Sex	Categoric	1: Female, 2: Male
X_{11}	Education	Categoric	1: Graduate primary, 2: Graduate junior, 3: Graduate Secondary
X_{12}	Reason Join Jaminan Kesehatan Nasional	Categoric	1: Anticipating sick family member, 2: Mandatory government programme; sick family member;following neighbours/friends;other)
X_{13}	ATP (Group of Ability to Pay)	Categoric	1: Nearly able&able, 2: Not able
X_{14}	WTP (Group of Willingness to Pay)	Categoric	1 Not willing & nearly willing, 2:Willing

The data structure used in this study is shown in Table 2 as follows.

Table 2 : Data Structure

Regency /City	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$	$X_{4,1}$	$X_{5,1}$	$X_{6,1}$	$X_{7,1}$	$X_{8,1}$	$X_{9,1}$	$X_{10,1}$	$X_{11,1}$	$X_{12,1}$	$X_{13,1}$	$X_{14,1}$
2	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	$X_{4,2}$	$X_{5,2}$	$X_{6,2}$	$X_{7,2}$	$X_{8,2}$	$X_{9,2}$	$X_{10,2}$	$X_{11,2}$	$X_{12,2}$	$X_{13,2}$	$X_{14,2}$
3	$X_{1,3}$	$X_{2,3}$	$X_{3,3}$	$X_{4,3}$	$X_{5,3}$	$X_{6,3}$	$X_{7,3}$	$X_{8,3}$	$X_{9,3}$	$X_{10,3}$	$X_{11,3}$	$X_{12,3}$	$X_{13,3}$	$X_{14,3}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	$X_{1,34}$	$X_{2,34}$	$X_{3,34}$	$X_{4,34}$	$X_{5,34}$	$X_{6,34}$	$X_{7,34}$	$X_{8,34}$	$X_{9,34}$	$X_{10,34}$	$X_{11,34}$	$X_{12,34}$	$X_{13,34}$	$X_{14,34}$

Here are the analysis steps in grouping the regencies/cities uses the ROCK ensemble method.

1. Create descriptive statistics on variables age, household size, household total expend, expend healthcare Household, Tobacco expend Household, ATP (Ability to Pay) Household, WTP (Willingness to Pay) Household, Expend Insurance Household, Living Area, Sex, Education, Reason Join Jaminan Kesehatan Nasional, ATP (Group of Ability to Pay), and WTP (Group of Willingness to Pay).

2. Grouping regencies/cities with ensemble ROCK methods, the following are the stages:

3. Divide data into numerical and categorical data

4. Cluster numerical data with the *Agglomerative Hierarchical* method (Single Linkage, Complete Linkage, & Average Linkage) and *Non Hierarchical* method (K-Means Clustering Method) which is the distance measure used is Euclidean distance.

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(x_i - x_j)^T(x_i - x_j)} \; , i,j = 1,2,\dots,n \; and \; i \neq j \tag{1}$$

5. Determine the number of clusters in *Agglomerative Hierarchical* method (Single Linkage, Complete Linkage, & Average Linkage) based on *Dendogram* visualization. Furthermore determine the optimum number of groups using Rsquare.Calculating the best grouping performance based on the S_W/S_B ratio value of the grouping results using the single linkage, complete linkage and average linkage methods. In this case S_w (Standard Deviation within Groups) and S_B (Standard Deviation between Groups).

6. Deteremine the number of clusters in *Non Hierarchical* method (*K-Means Clustering Method*) based on *Elbow method* and *Silhouette Score*.

7. Choose the best cluster method for numerical data based on the *highest Silhouette Score*.

8. Cluster categorical data with the ROCK method with threshold value (θ). In grouping with the ROCK method, the R-project statistics program (Package R) will be used with the package name cba. This package is created by Christian Buchta and Michael Hahsler with reference to Guha, Rastogi and Shim (2000).

9. Determine the optimum number of groups based on the minimum S'W/S'B ratio value based on the results of the ROCK method grouping on each θ value.

10. Combine cluster results in stages 7 and 9.

11. Cluster the combined data using the ROCK method threshold value (θ). In grouping with the ROCK method, the R-project statistics program (Package R) will be used with the package name **cba**

12. Determine the optimum number of groups based on the minimum S_W/S_B ratio value based on the grouping results of the ROCK method on each θ value.

13. Interpreting the results of the analysis and drawing conclusions.

III. RESULTS AND DISSCUSSION

3.1. Desciptive Statistics

Descriptive statistical analysis is used to determine the characteristics of the data. In this case, descriptive statistics is for numerical data and categorical data.

3.1.1. Numeric Data

Numeric data on district/city data in Indonesia consists of eight characteristics which include of Age, Household Size, Household Total Expend, Expend Healthcare Household, Tobacco Expend Household, ATP (Ability to Pay) Health Insurance in Household, WTP (Willigness to Pay) Helath Insurance in Household, and Expend Insurance Household. The characteristics of regencies/cities in Indonesia can be seen in Table 3.

Table 3: Numeric Data Characteristics

#	Variable Name	Average	Minimum	Maximum
1	Age	40	29	48
2	Household Size	4	3	5
3	Household Total Expend	2.412.919	1.377.654	3.711.13
4	Expend Healthcare Household	70.595	204	4
5	Tobacco Expend Household	115.679	0	520.385
6	ATP (Ability to Pay) Health Insurance in	63.219	21.314	336.240
7	Household	12.319	3.315	114.344
8	WTP (Willigness to Pay) Health	26.439	3.315	114.344
	Insurance in Household			113.868
	Expend Insurance Household			

Table 3 shows the results regarding the average value, minimum value and maximum value. In this case, it is known that the average age of respondents in the data is 40 years with a household size of 4. Among all expenditures, it is known that the highest expenditure is for household with a total of 2.412.919. Furthermore, it is also known that people tend to be less willing to pay for health insurance even though they This capability is indicated by the average value of ATP (Ability to Pay) Health Insurance in Household which is higher than WTP (Willigness to Pay) Health Insurance in Household.
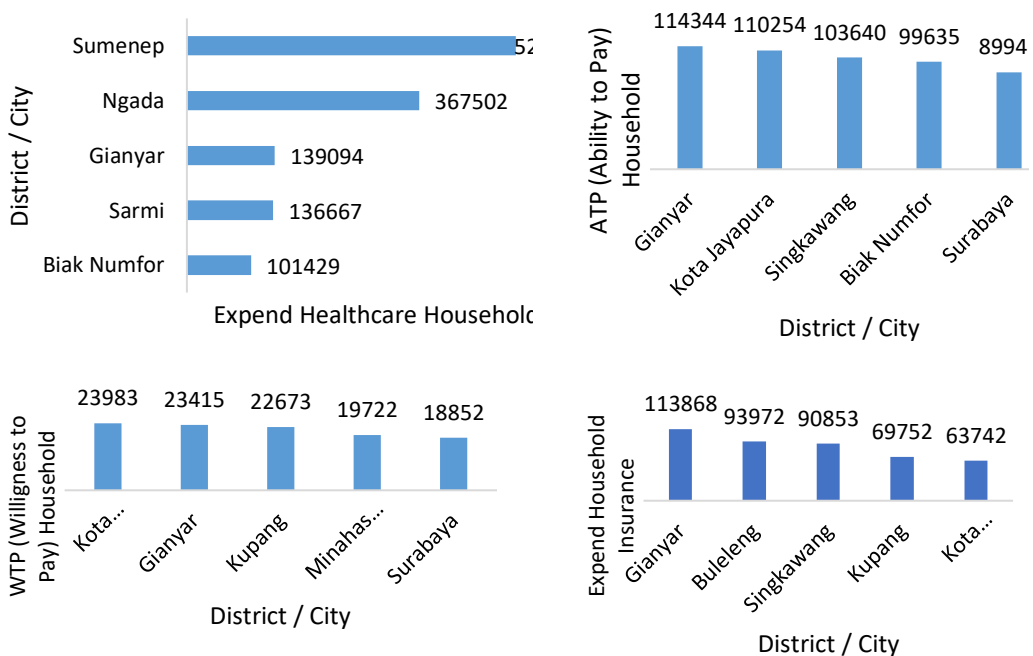

Figure 1: Bar Chart of Numerical Data (5 Highest Value)

Figure 1 shows the results of 5 districts/cities in Indonesia with the highest scores when viewed in terms of the aspects of Healthcare Expend Household, ATP (Ability to Pay) Health Insurance in Household, WTP (Williness to Pay) Health Insurance in Household , and Expend Household Insurance. In Figure 1, it can be seen that the highest scores for these aspects are dominated by Districts/Cities located in Central and Eastern Indonesia. Apart from that, it is also known that Gianyar Regency is always in the top 5 category for these aspects.

3.2.1. Categorical Data

The district/city categorical data in Indonesia consists of Living Area, Sex, Education, Reasons Join Jaminan Kesehatan Nasional, ATP (Group of Ability to Pay), WTP (Group of Willingness to Pay).

(a). Living Area

(b) Sex

(c). Education

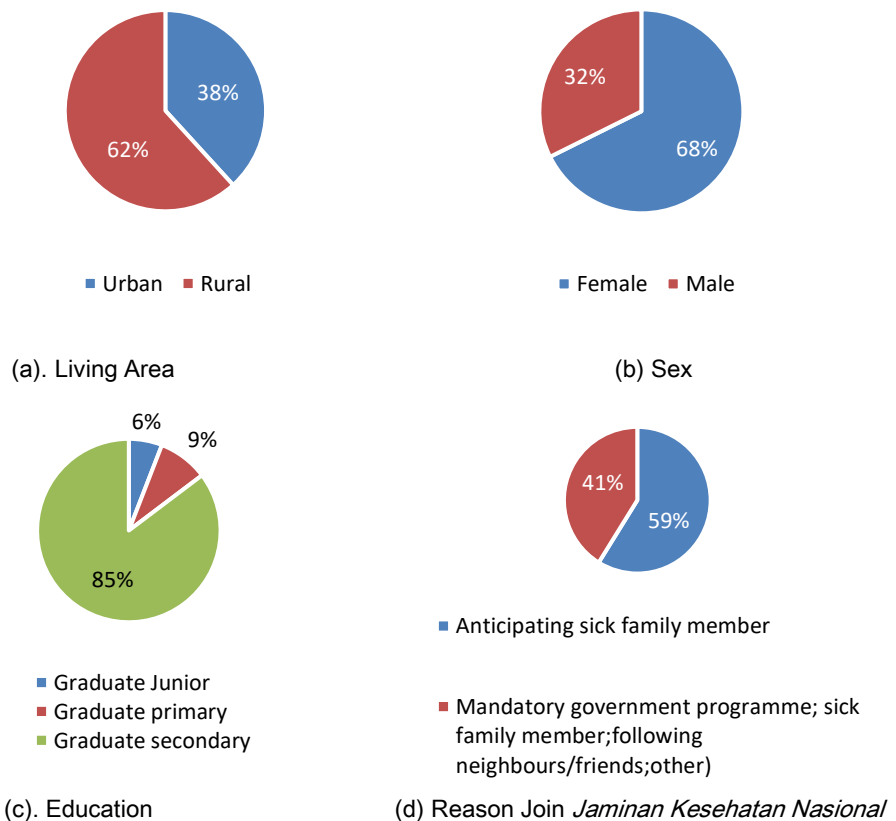(d) Reason Join *Jaminan Kesehatan Nasional*

Figure 2: Pie Chart of Categorical Data

In this study, figure 2 shows that 62 percent lived in rural areas while the remaining 38 percent lived in urban areas. In terms of gender composition, 68 percent were female, while the remaining 32 percent were male. The majority of respondent's education is secondary graduates (85 percent), the rest being primary graduates (9 percent) and junior graduates (6 percent). Then, for reasons of joining Jaminan Kesehatan Nasional, 59 percent are anticipating sick family members, while the remaining 41 percent are mandatory government programs.

3.2. Clustering with Ensemble ROCK Method

The Ensemble RObust Clustering method using linKs (ROCK) is a statistical method used to group variables with numeric data types and categorical data types. Grouping regencies/cities using the Ensemble ROCK method begins with grouping numerical variables using the Agglomerative Hierarchical Clustering (*Single Linkage, Complete Linkage, and Average Linkage*) and Non Hierarchical Clustering (*K-Means*) , then grouping categorical variables using the ROCK method and finally combined grouping using the ROCK method which is called the ensemble ROCK.

3.2.2. Clustering Analysis for Numerical Data

3.2.2.1. Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering analysis on numerical data is carried out by grouping regencies/cities using single linkage, complete linkage, and average linkage clustering methods.The distance measure that used in the Agglomerative Hierarchical Clustering is Euclidean distance. The optimum number of clusters is determined based on *Dendogram* visualization and the R-Square index value.

Every observation in *Dendogram* that has something in common will be connected by a line and unite at a point. The length of the line is different for each observation. In a dendogram, the vertical axis shows "Height" (y-axis) which describes the (dis)similarity/distance between observations (line length). Meanwhile, the horizontal axis (x-axis) shows the District /City.
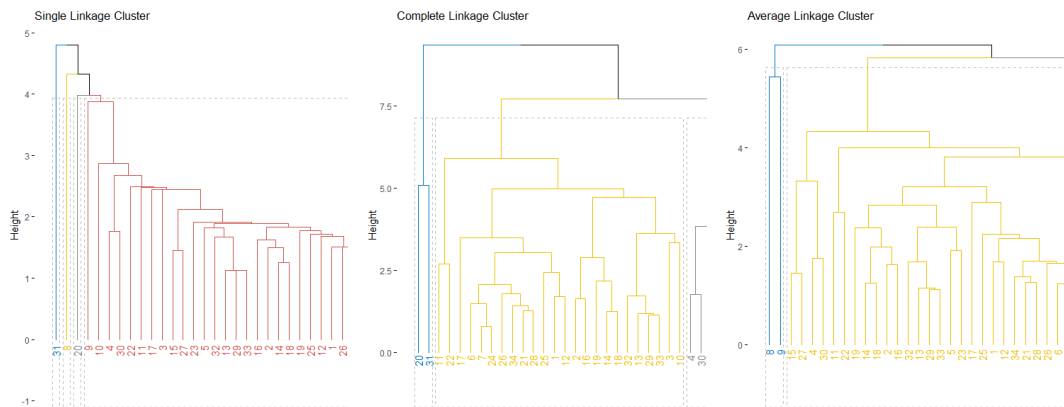

Figure 3: Dendogram Visualization for Each Agglomertive Hierarchical Clustering

Based on the Dendogram visualization in Figure 3, it can be seen that the number of clusters formed in the Single Linkage method is 4 groups, the Complete Linkage method is 3 groups, and the Average Linkage method is 3 groups. Furthermore we will determine the number of clusters using the agglomerative hierarchical clustering method with the R-Square index value. The group is said to be optimal if the number of groups with values increases very sharply and tends to be stable for the number of subsequent groups.

Table 4: R-Square Value for Each Agglomertive Hierarchical Clustering

Hierarchical Clustering	Nember of Groups (Cluster)			
	k=2	k=3	k=4	k=5
Single Linkage	0,019	0,096	0,795	0,885
Complete Linkage	0,757	0,842	0,900	0,933
Average Linkage	0,757	0,848	0,868	0,885

Table 4 shows the R-Square value for each agglomertive clustering hierarchical method. If the R-Square value is displayed in graphical form, the results obtained are as follows.
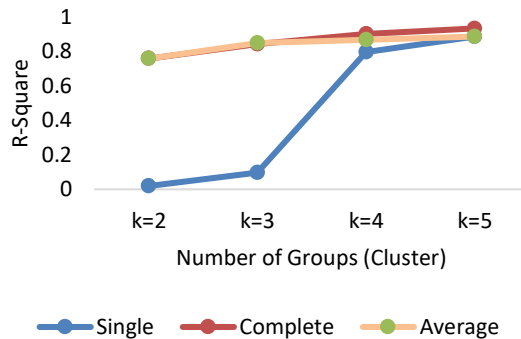

Figure 4: R-Square Value for Each Agglomertive Hierarchical Clustering

Based on Figure 4 it can be seen that the optimum number of groups in the Single Linkage method is 4 groups, then the optimum number of groups in the Complete Linkage method is 3 groups and the optimum number of groups in the Average Linkage method is 3 groups. Determination of the optimum group for each method is carried out based on the group that experienced the greatest change in the R-Square value compared to the other groups. The next step is determine the best grouping method based on the S_W and S_B ratio values for each method.

Table 5 : Ratio Value of Numerical Data Grouping Result

Hierarchical Clustering	Number of Groups (Cluster)		S_B	Ratio
Single Linkage	k = 4	21553, 73	89721, 67	0,2402288
Complete Linkage	k = 3	32067, 69	118904,7	0,2696925
Average Linkage	k = 3	29381, 49	101038,1	0,2907962

Based on Table 5, It can be seen that the optimal number of clusters formed for each agglomerative hierarchical clustering method is the same as the dendogram visualization results. Besides the result in Table 5 shows that the lowest S_W and S_B ratio is single linkAge of 0.2402288. The ratio value indicates that the standard deviation within the group is 0.2402288 times the standard deviation between groups. So it can be concluded that grouping numerical data using the single linkage method with a total

of 4 groups is the right grouping for the Agglomerative Hierarchical Clustering method. The characteristics of grouping results for numerical data with 4 groups are as follows.

3.2.2.2. Non Hierarchical Clustering (K-Means)

Clustering for numerical data is also carried out using a non-hierarchical method, namely the K-Means method. The elbow method and silhouette method are used to determine the best number of clusters that can be formed by the K-Means method.
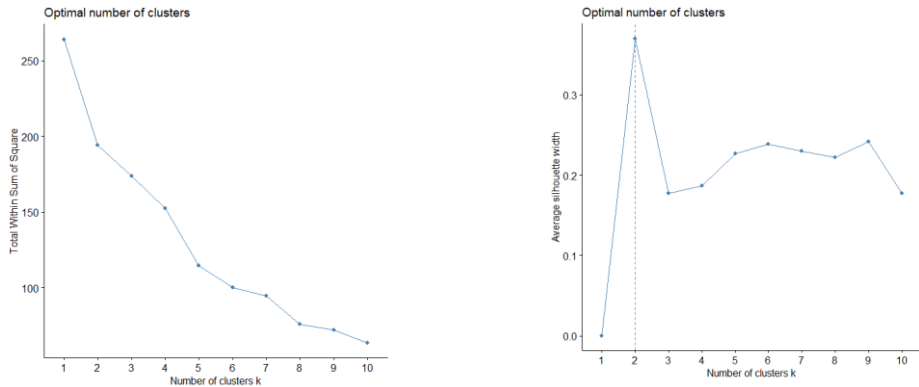

Figure 5: Elbow Method Graph (Left) & Silhouette Method Graph (Right) for K-Means Method

Determining the number of clusters using the elbow method is based on the right angles on the graph, it can be seen that the graph shows a sharp change and then slopes after k=2, then the optimal number of clusters is 2. Furthermore, if based on the silhouette method on the graph it can be seen that the optimal number of clusters is marked with the highest silhouette value , in this case occurs when the number of clusters is 2. So it can be concluded that the optimal number of clusters formed based on the K-Means method is 2.

3.2.2.3. Best Clustering Method for Numerical Data

The next stage is to determine the best clustering method for numerical data using the highest silhouette score.

Table 6: Silhouette Score for Single Linkage, Complete Linkage, Average Linkage, and K-Means

Method	Silhouette Score
Single Linkage	0,3820545
Complete Linkage	0,2101188
Average Linkage	0,3655862
K- Means	0,2298990

Table 6 shows the results that the Single Linkage method has the highest Silhouette score of 0.3820545. This means that the best method for clustering numerical data in this case is the Single Linkage Method. The next stage is to interpret the clusters formed using the Single Linkage method, namely 4 clusters as follows.

153

a.   Group 1 (Numerical Variable with Single Linkage)

Group 1 consists of 16 regencies/cities, namely Badung, Donggala, Dumai, City of Manokwari, Kuansing, Kupang, Langkat, Minahasa, North Minahasa, Palu, Pandeglang, Rote Ndao, Sambas, Sarmi, Siak, and Tojo Una-Una. Group 1 consists of regencies /cities with Jaminan Kesehatan Nasional participants with the average Age is 40 years old, the average household member is 4 people, the average total household expenditure is Rp 1.865.322, the average health expenditure is Rp 41.553, the average spending on cigarettes is Rp. 83.890, the average ATP (Ability to Pay) Household is Rp. 44.754, the average WTP (Willingness to Pay) Household is Rp. 10.907 and the average insurance expenditure is Rp. 17.710.

b.   Group 2 (Numerical Variable with Single Linkage)

Group 2 consists of 16 regencies/cities, namely Berau, Biak Numfor, Buleleng, Deli Serdang, Gianyar, Jayapura City, Manado City, Medan City, Samarinda City, Kubu Raya, North Penajam Paser, Serang, Sidoarjo, Singkawang, Surabaya, South Tangerang. Group 2 consists of regencies /cities with Jaminan Kesehatan Nasional participants with the average Age is 40 years old, the average household member is 4 people, the average total house expenditure is Rp 3.044.045, the average health expenditure is Rp 52.968, the average spending on cigarettes is Rp. 147.082, the average ATP (Ability to Pay) Household is Rp. 83.742, the average WTP (Willingness to Pay) Household is Rp. 14.511 and the average insurance expenditure is Rp. 36.004.

c.   Group 3 (Numerical Variable with Single Linkage)

Group 3 consists of 1 district, namely Ngada Regency with the average Jaminan Kesehatan Nasional participant Age is 48 years old, the average household member is 5 people, thw average total household expenditure is Rp 1.377.654, the average health expenditure is Rp 367.502, the average spending of cigarettes is Rp. 78.900, the average ATP (Ability to Pay) Household is Rp. 44.161, the average WTP (Willingness to Pay) Household is Rp. 5.750 and the average insurance expenditure is Rp. 18.734.

d.   Group 4 (Numerical Variable with Single Linkage)

Group 4 consists of 1 district, namely Sumenep Regency with the average Jaminan Kesehatan Nasional participants Age is 44 years old, the average household member is 3 people, the average total house expenditure is Rp 2.111.712, the average health expenditure is Rp 520.385, the average spending on cigarettes is Rp. 158.654, the average ATP (Ability to Pay) Household is Rp. 49.354, the average WTP (Willingness to Pay) Household is Rp. 6.423 and the average insurance expenditure is Rp. 20.769.

3.2.3. ROCK Method Analysis for Categorical Data

Grouping analysis on categorical data is using the ROCK method. The threshold value (θ) that used in the ROCK method is 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8 and 0.9. The selection of treshold is determined based on the minimum S_W/S_B ratio value. In this research, the selection of threshold values for the ROCK method was carried out selectively by considering the efficiency of the analysis. This research does not test all possible threshold values, but instead only focus on the threshold values that provide an optimal balance between within-cluster similarity (SW) and between-cluster similarity (S_B). This approach is based on the principle that the optimal combination of SW and SB values reflects the most representative cluster structure in the data. Furthermore, the value of Sw and S_B ratios for each threshold in the ROCK method can be shown as follows.

Table 7: Ratio and Treshold Value of Categorical Data Grouping Result

Tresshold (θ)	Ratio Value
0,1	0,4907930
0,2	0,3998822
0,3	0,7717977
0,4	0,5692703
0,5	0,0000000
0,6	0,0000000
0,7	0,0000000
0,8	0,0000000
0,9	0,0000000

Based on Table 7, it is known that the smallest ratio value in the ROCK method and has a value greater than 0 is 0.3998822 with a threshold value (θ) of 0.2. The ratio value indicates that the standard deviation within the group is 0.3998822 times the standard deviation between groups. This can also mean that the variance of the data within the group gives a smaller deviation value than the variance between groups.

As a note, the reason for not using a threshold value with ratio value 0 is because this value shows many objects that are considered similar even though their similarity is actually weak. As a result, the relationships between objects (links) expand greatly, creating one large cluster that includes all the data. The results of grouping the categorical data using the ROCK method with θ of 0.2 produce 3 groups with the characteristics shown below.

Table 8: ROCK Grouping Result Based on Categorical Variables

Categorical Variable	ROCK Group 1	ROCK Group 2	ROCK Group 3
Living Area (X_9)			
Urban	8	5	8
Rural	5	3	5

Sex (X_{10})			
Female	8	7	8
Male	5	1	5
Education (X_{11})			
Graduate Primary	2	1	0
Graduate Junior	0	1	1
Graduate Secondary	11	6	12
Reason Join Jaminan Kesehatan Nasional (X_{12})			
Anticipating sick family member	10	5	5
Mandatory government programme;sick family member; other	3	3	8
ATP (Group of Ability to Pay) (X_{13})			
Nearly able & able	9	6	10
Not able	4	2	3
WTP (Group of Willingness to Pay) (X_{14})			
Not willing	12	7	10
Willing	1	1	3

Based on Table 8, the groups were formed using the ROCK method for categorical data with θ of 0.2.

a.   Group 1 (Categorical Variables with ROCK Method)

Group 1 consists of 13 regencies/cities and dominated by Jaminan Kesehatan Nasional participants who live in urban areas, female, and have a Graduate Secondary education level, then the reasons for joining Jaminan Kesehatan Nasional is anticipating sick family members, the ATP category is nearly able & able, and the category WTP (Willingness to Pay) Household is not willing.

b.   Group 2 (Categorical Variables with ROCK Method)

Group 2 consists of 8 regencies/cities and dominated by Jaminan Kesehatan Nasional participants who live in urban areas, female, and have a Graduate Secondary education level, then the reasons for joining Jaminan Kesehatan Nasional are anticipating sick family members, the ATP (Ability to Pay) Household category is nearly able & able, and the category WTP (Willingness to Pay) Household is not willing. The aspect that distinguishes members from group 2 and group 1 is related to the number of membership and external aspects.

c.   Group 3 (Categorical Variables with ROCK Method)

Group 3 consists of 13 regencies/cities and dominated by Jaminan Kesehatan Nasional participants who live in urban areas, female, and have a Graduate Secondary education level, then the reason for joining Jaminan Kesehatan Nasional is the Mandatory government program; sick family members; following neighbors/friends; other, the ATP (Ability to Pay) Household

category is nearly able & able, and the WTP (Willingness to Pay) Household category is not willing.

3.2.4. Ensemble ROCK Method Analysis for Mixed Data

Grouping the ROCK ensemble on mixed data is done by grouping each data type. The grouping results for each data type are expressed as categorical variables which are then regrouped using the ROCK ensemble method. In grouping mixed data using the ROCK ensemble method the value of θ used is the same as in the categorical grouping analysis using the ROCK method, namely between 0 and 1. The reasons for using the SW and SB ratio values for stresshold selection in this method are the same as those explained in subsection 3.2.2 . Table 9 shows the value of the S_W and S_B ratio for θ of 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8 and 0.9.

Table 9: Ratio and Treshold Value of Ensemble Rock Grouping Result

Tresshold (θ)	Ratio Value
0,1	0,974805
0,2	0,385119
0,3	0,575577
0,4	0,594092
0,5	0,000000
0,6	0,000000
0,7	0,000000
0,8	0,000000
0,9	0,000000

Table 9 shows that the smallest ratio value and has value greater than 0 in the ROCK ensemble grouping is 0.385119 with an θ value of 0.2. This value indicates that the standard deviation within the group is 0.385119 times the standard deviation between groups or the variance of the data within the group gives a smaller deviation value than the variance between groups. The reason for not using a threshold value with ratio value 0 is because this value shows many objects that are considered similar even though their similarity is actually weak. As a result, the relationships between objects (links) expand greatly, creating one large cluster that includes all the data. Grouping results on mixed data using the ROCK ensemble method with θ of 0.2 produces 3 groups with the characteristics of each group as follows.

Table 10: Ensemble Rock Result

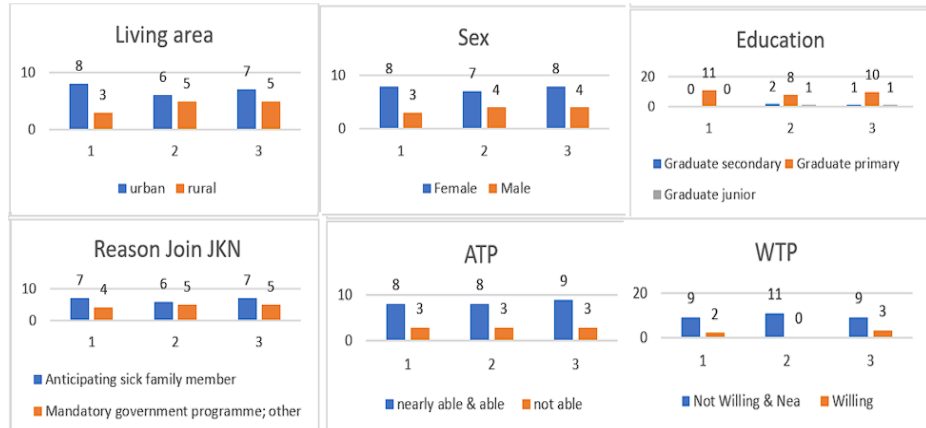| Group 1	Group 2	Group 3
Berau, Dumai, Kota Samarinda, KubuRaya Kupang, Minahasa Utara, Palu, Serang, Siak, Tangsel, Tojo Una- Una	Biak Numfor, Buleleng, Gianyar, Kuansing, Langkat, Ngada, Pandeglang, Penajam Paser Utara, Rote Ndao, Sambas, Singkawang	Badung, Deli Serdang, Donggala, Kota Jayapura, Kota Manado, Kota Manokwari, Kota Medan, Minahasa, Sarmi, Sidoarjo, Sumenep, Surabaya

Figure 6: ROCK Ensemble Grouping Results Based Variables Data

a. Low quality (Group 1)

Group 1 is a group consisting of 11 regencies/cities. This group has an average Age quality value of 39 years, Household Size of 4, household total expend of 2.311.038, expend health care of 38.571, tobacco expend of 129.085, atp of 60.651, wtp of 12.930 and expend insurance of 18.027. Based on Table 10 and Figure 6, the most dominant living areas in this group are urban, namely 8 regencies. the most dominant sex is female, namely 8 regencies, the most dominant education is graduate secondary in 11 regencies, the most dominant reason for joining JKN is Anticipating sick family members in 7 regencies, the most dominant ATP is nearly able & able in 8 regencies, the most dominant WTP are not willing & nearly willing as many as 9 regencies.

b. Medium Quality (Group 2)

Group 2 is a group consisting of 11 regencies/cities. This group has an average Age quality value of 41 years, Household Size of 4, household total expend of 2.419.160, expend health care of 98.353, tobacco expend of 133.980, ATP of 66.189, WTP of 11.683 and expend insurance of 41,865.Based on Table 10 and Figure 6, the most dominant living area in this group is urban, namely 6 regencies. the most dominant sex is female, namely 7 regencies, the most dominant education is graduate secondary in 8 regencies, the most dominant reason for joining JKN is Anticipating sick family members in 6 regencies, the most dominant ATP is nearly able & able in 8 regencies, the most dominant WTP are not willing & Nearly willing as many as 11 regencies.

c. High quality (Group 3)

Group 3 is a group consisting of 12 regencies/cities. This group has an average Age quality value of 40 years, Household Size of 4, household total expend of 2,500,588, expend healthcare of 74.504, tobacco expend of 86,615, ATP of 62.851, WTP of 12.343 and expend insurance of 20.009. Based on Table 10 and Figure 6, the most dominant living area in this group is urban, namely 7 regencies. the most dominant sex is female, namely 8 regencies, the

most dominant education is graduate secondary in 10 regencies, the most dominant reason for joining JKN is Anticipating sick family members in 7 regencies, the most dominant ATP is nearly able & able in 9 regencies, the most dominant WTP are not willing & nearly willing as many as 9 regencies.

3.3.   Analysis of Grouping Result

If National Health Insurance assistance is specifically and prioritized to help underprivileged people, then groups with low variable scores on economic indicators, such as ATP (Ability to Pay), Household Total Expend, and Household Insurance Expend, will be given more priority. This is because this group is considered to have economic limitations and therefore needs more assistance. So if we link the results of the Ensemble ROCK cluster method and the results of descriptive statistics in subchapter 3.1, it can be explained that the priority group for implementing the JKN program is Group 1 which still has Low Quality, after that the government can focus on districts/cities in Group 2 (Medium Quality), and finally the districts/cities in Group 2 (Medium Quality). As a result, it is hoped that the allocation of the health aid budget will be evenly distributed and there will be no social disparities between communities.

VI.   CONCLUSIONS

The conclusions obtained based on the research results above are as follows.

1.   Based on descriptive statistics analysis, it is known that the average Age of Jaminan Kesehatan Nasional participants is 40 years, the average number of household members is 4 members, and the average household income is Rp. Rp. 115,679, then the average ATP (Ability to Pay) Household is Rp. 63,219, the average WTP (Willingness to Pay) Household is Rp. 12,319, and the average Expend Insurance Household is Rp. 26,439.

2.   The grouping of regencies/cities in Indonesia regarding participation in Jaminan Kesehatan Nasional is based on the type of data. The results of grouping the numerical data using the best hierachical clustering method, namely single linkage, form 4 groups. Furthermore, the grouping results on categorical data using the ROCK method with an θ value of 0.2 produce 3 groups.

3.   Grouping the ROCK ensemble on mixed data is done by grouping each data type. The grouping results for each data type are expressed as categorical variables which are then regrouped using the ROCK ensemble method, with an θ value of 0.2 to produce 3 groups. Group 1 is a low quality group consisting of 11 regencies/cities, then group 2 is a group with middle quality consisting of 11 regencies/cities, and group 3 is a high quality group consisting of 12 regencies/cities.

4.   The priority group for implementing the JKN program is Group 1 which still has Low Quality, after that the government can focus on districts/cities in Group 2 (Medium Quality), and finally the districts/cities in Group 2 (Medium Quality).

REFERENCES

[1].   Dewan Jaminan Sosial Nasional. (2014). *Paham JKN Jaminan Kesehatan Nasional.* Jakarta: Friedrich-Ebert-Stiftung Kantor Perwakilan Indonesia.

[2].   Adhikari, R., Aryal, N., & Regmi, P. R. (2019). Factors influencing enrollment in health insurance schemes: A systematic review. *Journal of Public Health Research, 8*(3), 153-161. https://doi.org/10.4081/jphr.2019.153

[3].   Susilo, T., & Prasetyo, E. (2021). Analisis faktor ekonomi terhadap partisipasi dalam program Jaminan Kesehatan Nasional (JKN) di Indonesia. *Jurnal Ekonomi dan Kebijakan Kesehatan, 5*(2), 101-112. https://doi.org/10.xxxx/jekk.2021.52

[4].   Wulandari, R. D., Nugroho, A., & Santoso, D. (2020). Determinan status keikutsertaan program Jaminan Kesehatan Nasional pada pekerja informal. *Jurnal Kesehatan Masyarakat Nasional, 14*(4), 241-250. https://doi.org/10.21109/kesmas.v14i4.2020

[5].   Nuraini, A., Suryandari, D. A., & Wardhana, S. A. (2018). The role of information dissemination in increasing National Health Insurance participation. *Journal of Health Communication, 7*(1), 25-30. https://doi.org/10.xxxx/jhc.2018.71

[6].   Putri, A. F., & Handayani, P. W. (2022). Pengaruh kualitas layanan kesehatan terhadap tingkat kepercayaan peserta Jaminan Kesehatan Nasional. *Jurnal Administrasi dan Kebijakan Kesehatan, 12*(1), 45-56. https://doi.org/10.xxxx/jakk.2022.121

[7].   BPJS Kesehatan. (2022). BPJS Kesehatan. Available at: https://www.bpjs-kesehatan.go.id/bpjs/post/read/2022/2360/BPJS-Kesehatan-Mendengar-2022-Jaring-Masukan-tentang-Pengelolaan-JKN-ke-Depan#:~:text=%E2%80%9CTahun%202021%2C%20cakupan%20kepesertaan%20Program,RPJMN%20tahun%202022%20yaitu%2087%25 (accessed 24 July 2022).

[8].   Dewangan, R. R., Sharma, L. K., & Akasapu, A. K. (2010). Fuzzy clustering technique for numerical and categorical dataset. *International Journal on Computer Science and Engineering, 2*(1), 75-80.

[9].   Alvionita. (2017). *Metode Ensemble ROCK dan SWFM untuk Pengelompokan Data Campuran Numerik dan Kategorik pada Kasus Aksesi Jeruk* (Thesis). Institut Teknologi Sepuluh November.

[10].   He, Z., Xu, X., & Deng, S. (2005a). A cluster ensemble method for clustering categorical data. *Information Fusion, 6*, 143-151.

[11].   Prasetya, R. A. (2020). *Analisis Klaster dengan Metode Ensemble ROCK untuk Data Campuran: Studi Kasus Stunting di Provinsi Jawa Barat* (Tugas Akhir). Universitas Pendidikan Indonesia.