CLUSTERING OF DISTRICTS AND CITIES IN INDONESIA BASED ON POVERTY INDICATORS USING THE K-MEANS METHOD

Khoirun Niswatin, Christopher Andreas, Putri Fardha Asa Oktavia Hans, M. Fariz Fadilah Mardianto

PRESENTER

Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia

4 - 5 AUGUST 2021

CONTENTS

1 INTRODUCTION

2 LITERATURE REVIEW

3 METHODOLOGY

4 RESULTS AND DISCUSSIONS

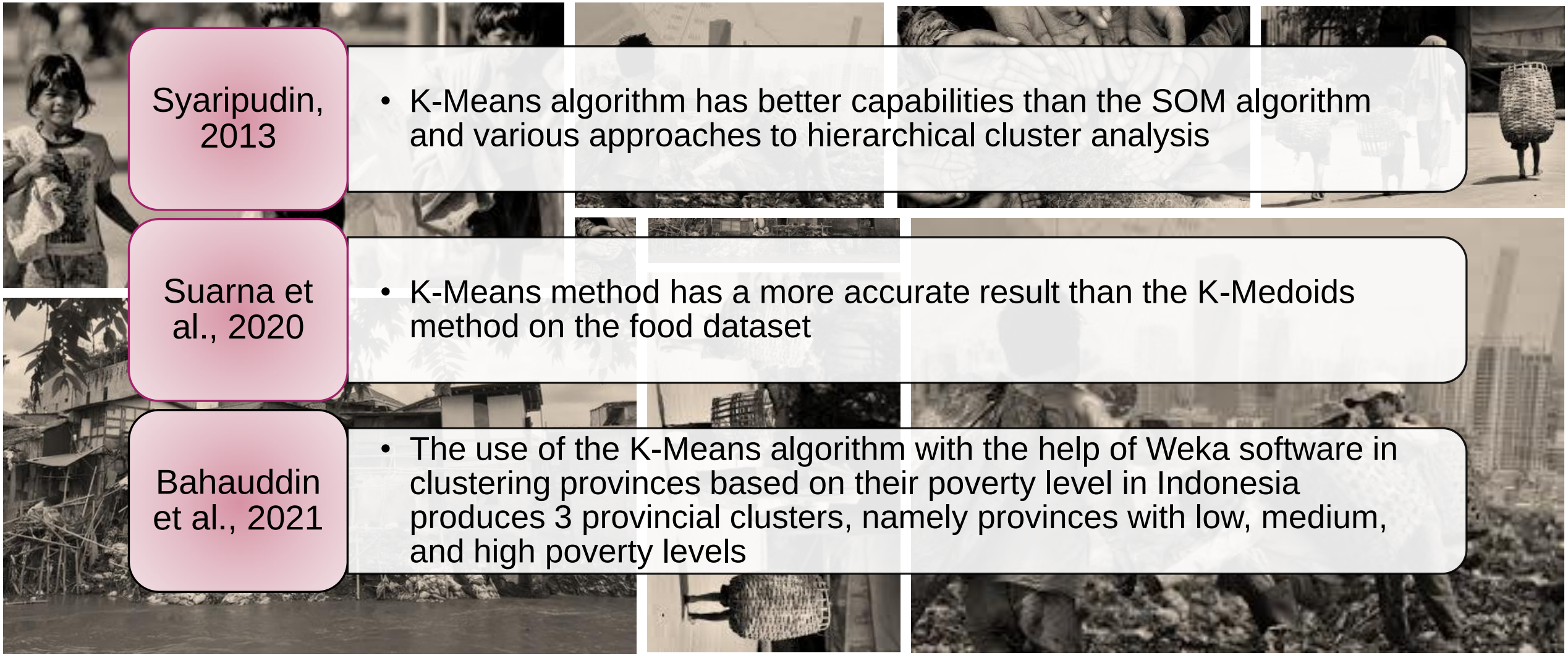
5 CONCLUSIONS

Introduction

SUSTAINABLE DEVELOPMENT GOALS

1 NO POVERTY

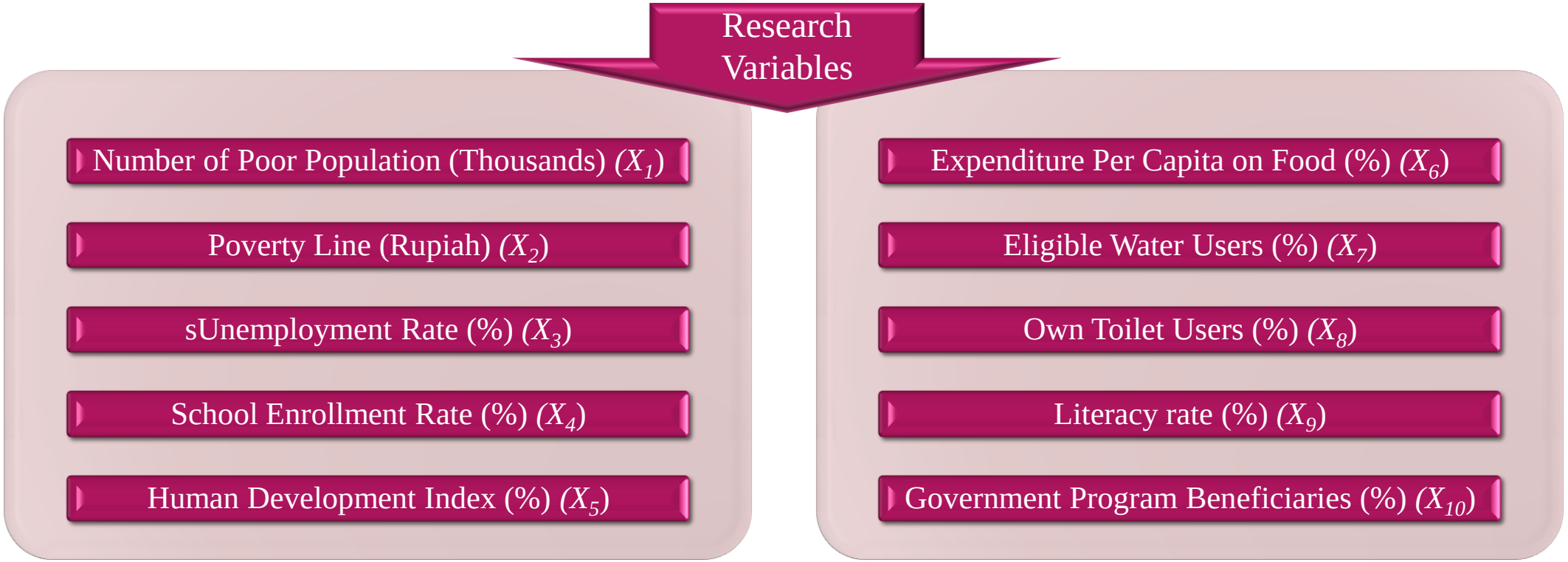
Literature Review



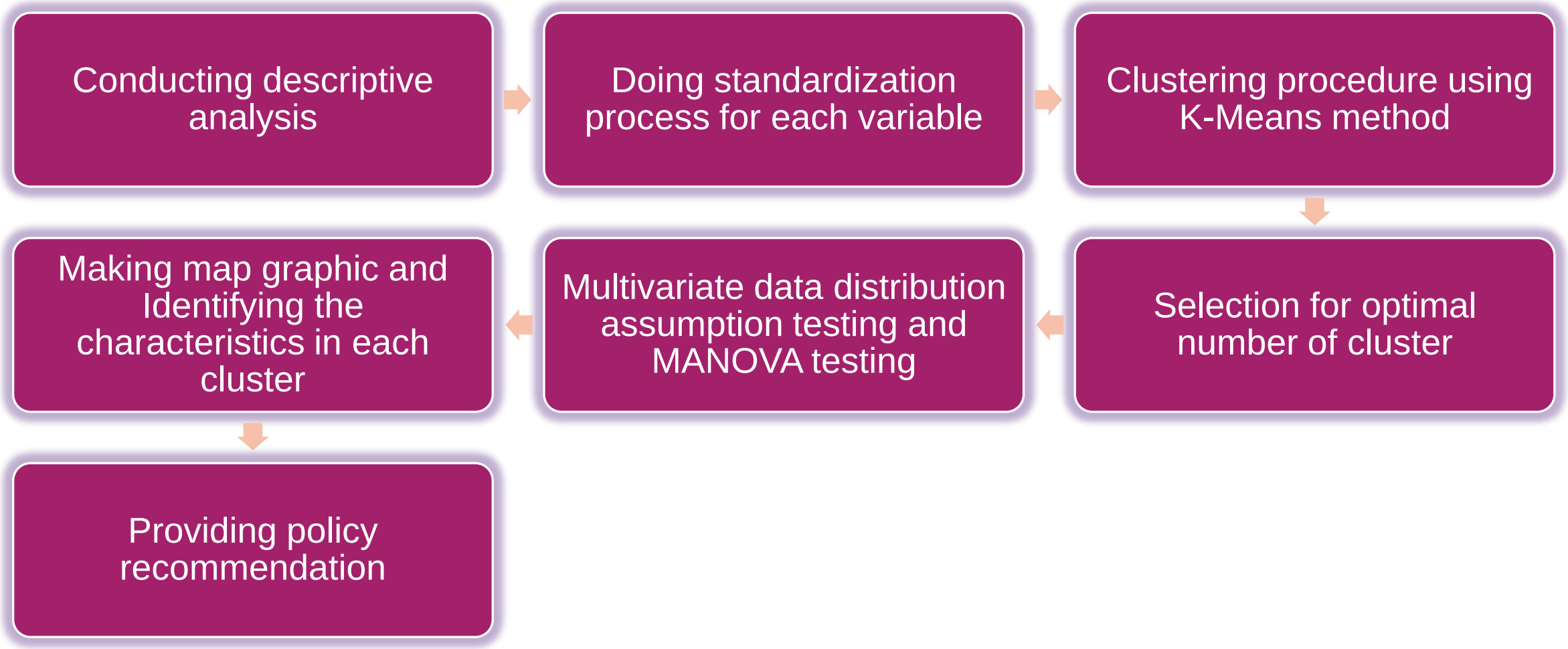
Syaripudin, 2013	• K-Means algorithm has better capabilities than the SOM algorithm and various approaches to hierarchical cluster analysis
Suarna et al., 2020	• K-Means method has a more accurate result than the K-Medoids method on the food dataset
Bahauddin et al., 2021	• The use of the K-Means algorithm with the help of Weka software in clustering provinces based on their poverty level in Indonesia produces 3 provincial clusters, namely provinces with low, medium, and high poverty levels

Methodology

The data used in this study were obtained from a publication entitled “Data and Information on District and City Poverty in 2020” published by BPS



Methodology



Result and Discussions

Score Summary of Each Research Variable

Variable	Highest		Lowest	
	Score	Location	Score	Location
Number of Poor Population (X_1)	465,670 people	Bogor District	1,360 people	Sawah Lunto Town
Poverty Line (X_2)	Rp1,021,759	Jayapura City	Rp248,184	Buton Selatan District
Unemployment Rate (X_3)	65.85 %	Tomohon Town	4.16 %	Lanny Jaya District
School Enrollment Rate (X_4)	100 %	139 cities and districts	36.91 %	Puncak District
Human Development Index (X_5)	86.61 %	Yogyakarta City	31.55 %	Nduga District
Expenditure Per Capita on Food (X_6)	82.08 %	Alor District	46.51 %	Ternate City
Eligible Water Users (X_7)	100 %	10 cities and districts	2.12 %	Mamberamo Tengah District
Own Toilet Users (X_8)	100 %	32 cities and districts	2.11 %	Puncak District
Literacy rate (X_9)	100 %	162 cities and districts	16.55 %	Lanny Jaya District
Government Program Beneficiaries (X_{10})	67.27 %	Sabu Raijua District	0 %	38 cities and districts

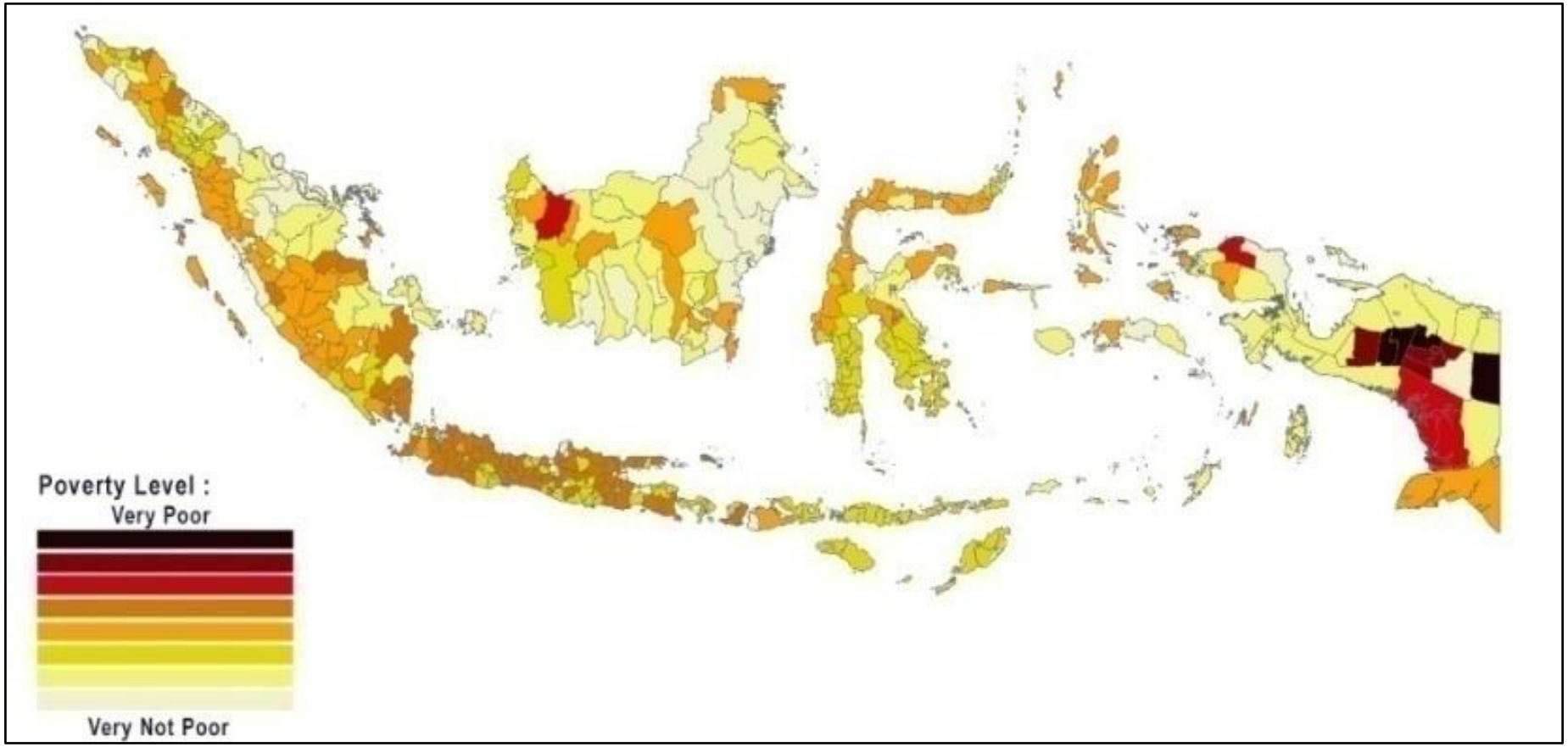
Result and Discussions

Selection for The Optimal Number of Cluster

Number of Clusters	Cluster Goodness Measures	
	Coefficient of Determination	Icd Rate
3	20.01%	79.98%
4	32.52%	67.47%
5	40.71%	59.28%
6	44.78%	55.21%
7	45.95%	54.04%
8	49.79%	50.02%

Result and Discussions

Mapping of Areas Based on Poverty Levels in Indonesia



Result and Discussions

Cluster Characteristics

Cluster 1

- areas with very not poor areas

Cluster 2

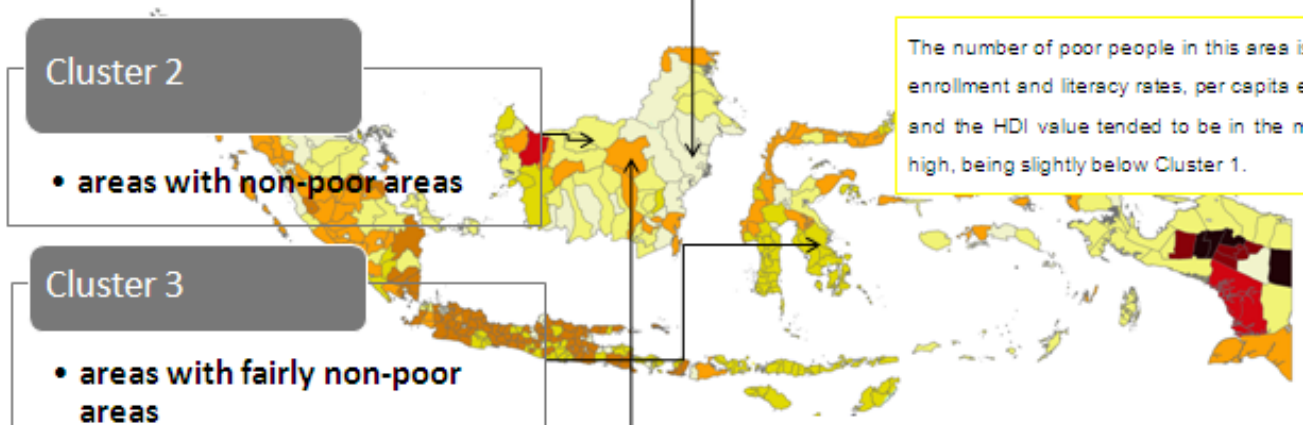
- areas with non-poor areas

Cluster 3

- areas with fairly non-poor areas

Cluster 4

- areas with areas tend to be less poor



The areas with the highest rates for school enrollment, literacy rates, HDI scores, users of safe water, and private latrines were compared to other cluster areas. The level of expenditure per capita on food is the smallest of all. The number of poor people and the percentage of people who receive assistance from the government is moderate. The poverty line is relatively high even though it is still below cluster 6 and the unemployment rate for the poor is the highest in this area.

The number of poor people in this area is the lowest. The unemployment rate for the poor, school enrollment and literacy rates, per capita expenditure on food, users of safe water and their latrines, and the HDI value tended to be in the middle. However, the poverty line in this area is relatively high, being slightly below Cluster 1.

The poverty line and the percentage of people using safe water in this area are in a good category, slightly below Cluster I. The expenditure per capita on food in this area is the largest compared to other cluster areas. The number of poor people, unemployment rates, school enrollment, and literacy rates, users of their latrines, government assistance programs, and HDI scores are in the middle category.

All indicators are in the medium value category.

Result and Discussions

Cluster Characteristics

Cluster 5

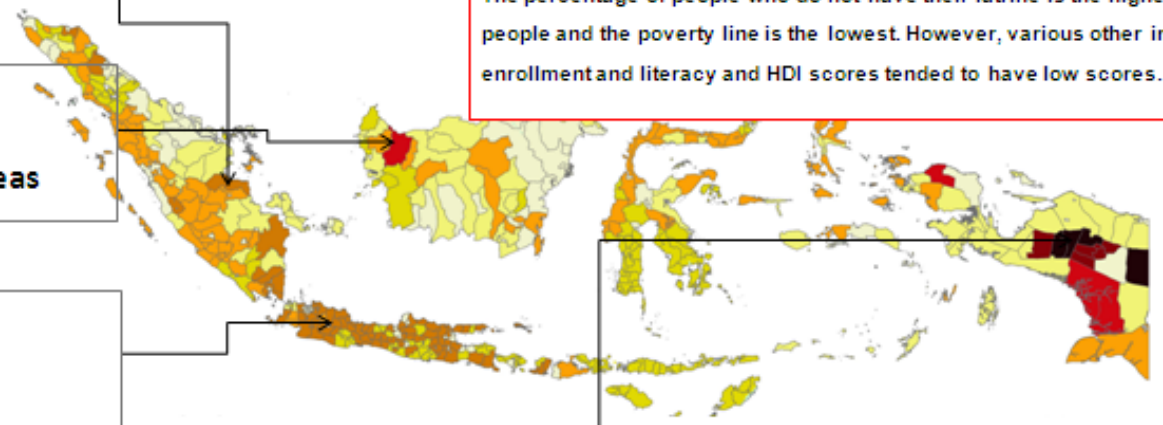
- areas with areas tend to be poor

The number of poor people and the percentage of government assistance is the highest. The unemployment rate for the poor is relatively high, with the second-highest ranking, which is one level below cluster 1. People who have their latrines are relatively high, namely one level below cluster 1.

Cluster 6

- areas with non-poor areas

The percentage of people who do not have their latrine is the highest. The number of poor people and the poverty line is the lowest. However, various other indicators such as school enrollment and literacy and HDI scores tended to have low scores.



Cluster 7

- areas with poor areas

Cluster 8

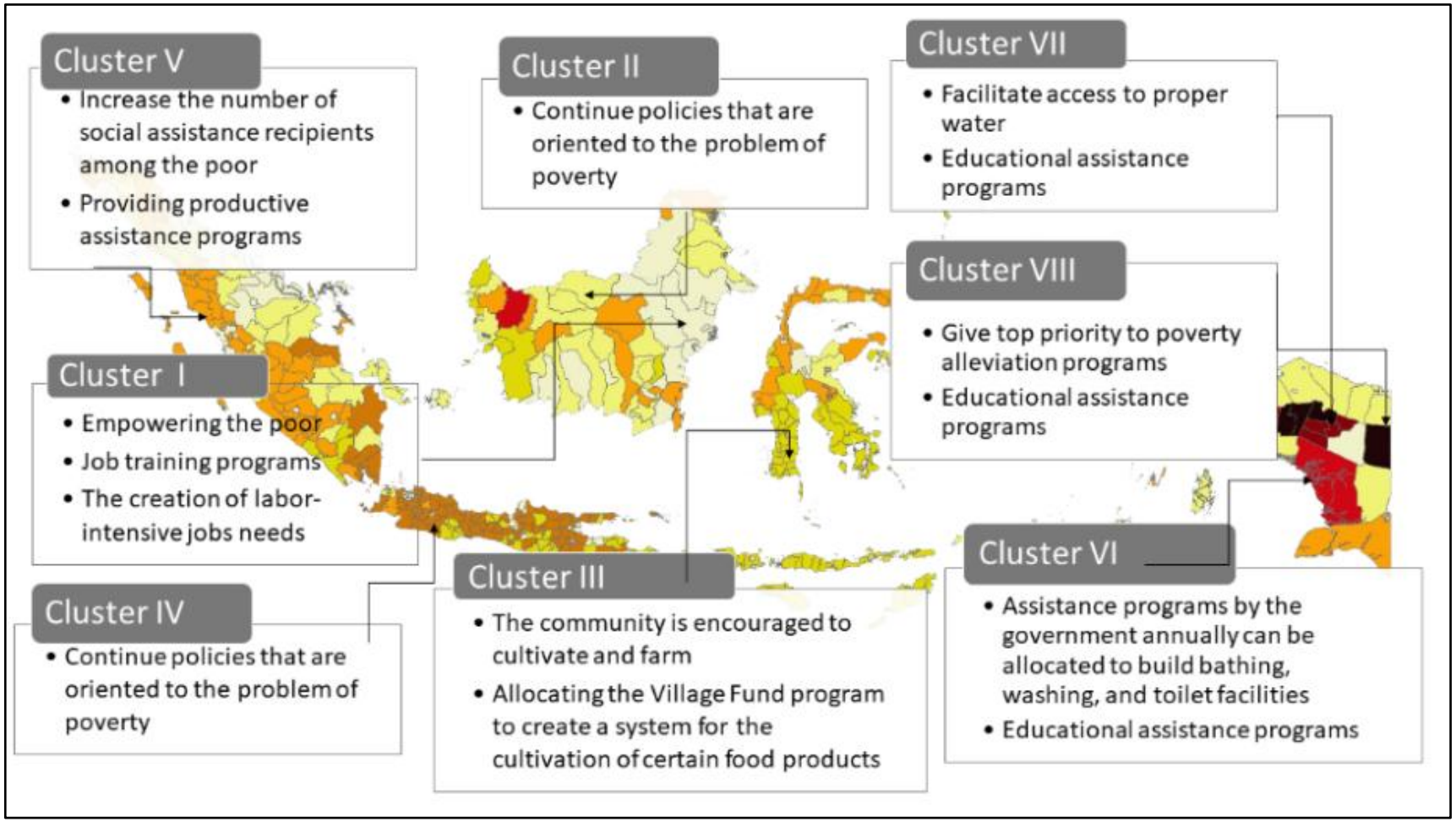
- a very poor area

The percentage of literacy rate and percentage of safe water users is the lowest. The unemployment rate and government assistance programs in the area are the lowest.

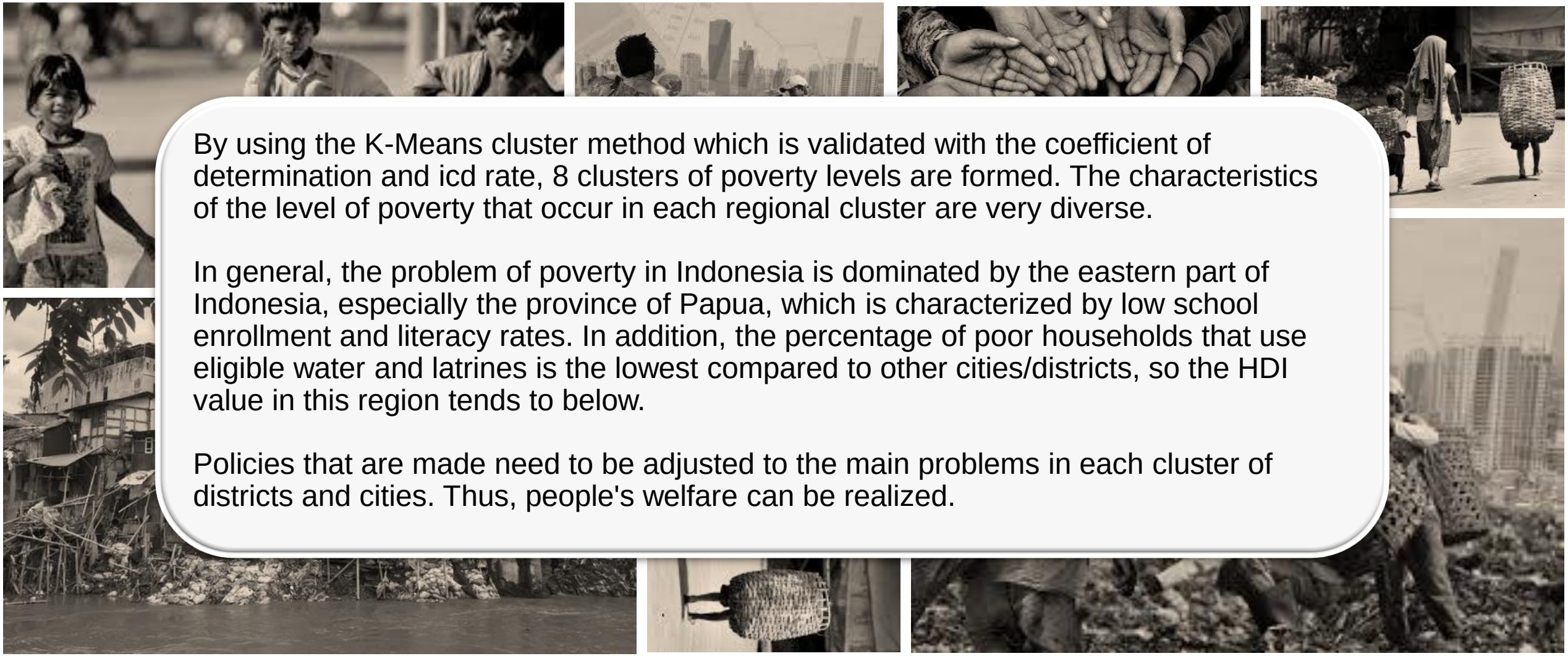
The poverty line is very high with low enrollment rates. The HDI value in this area is the lowest compared to areas in other clusters. The percentage of government aid programs in this area is very low, it is the same as in Cluster 2.

Result and Discussions

Policy Recommendations for Realizing People's Welfare.



Conclusions



By using the K-Means cluster method which is validated with the coefficient of determination and icd rate, 8 clusters of poverty levels are formed. The characteristics of the level of poverty that occur in each regional cluster are very diverse.

In general, the problem of poverty in Indonesia is dominated by the eastern part of Indonesia, especially the province of Papua, which is characterized by low school enrollment and literacy rates. In addition, the percentage of poor households that use eligible water and latrines is the lowest compared to other cities/districts, so the HDI value in this region tends to below.

Policies that are made need to be adjusted to the main problems in each cluster of districts and cities. Thus, people's welfare can be realized.

THANK YOU

2021 ICMS
INTERNATIONAL CONFERENCE ON COMPUTING,
MATHEMATICS AND STATISTICS