



UNIVERSITI
TEKNOLOGI
MARA

Cawangan Kedah
Kampus Sungai Petani



e-PROCEEDINGS

of The 5th International Conference
on Computing, Mathematics and
Statistics (iCMS2021)

4-5 August 2021

Driving Research Towards Excellence



e-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

Editor-in-Chief: Norin Rahayu Shamsuddin

Editorial team:

Dr. Afida Ahamad
Dr. Norliana Mohd Najib
Dr. Nor Athirah Mohd Zin
Dr. Siti Nur Alwani Salleh
Kartini Kasim
Dr. Ida Normaya Mohd Nasir
Kamarul Ariffin Mansor

e-ISBN: 978-967-2948-12-4

DOI

Library of Congress Control Number:

Copyright © 2021 Universiti Teknologi MARA Kedah Branch

All right reserved, except for educational purposes with no commercial interests. No part of this publication may be reproduced, copied, stored in any retrieval system or transmitted in any form or any means, electronic or mechanical including photocopying, recording or otherwise, without prior permission from the Rector, Universiti Teknologi MARA Kedah Branch, Merbok Campus. 08400 Merbok, Kedah, Malaysia.

The views and opinions and technical recommendations expressed by the contributors are entirely their own and do not necessarily reflect the views of the editors, the Faculty or the University.

Publication by
Department of Mathematical Sciences
Faculty of Computer & Mathematical Sciences
UiTM Kedah

TABLE OF CONTENT

PART 1: MATHEMATICS

	Page
STATISTICAL ANALYSIS ON THE EFFECTIVENESS OF SHORT-TERM PROGRAMS DURING COVID-19 PANDEMIC: IN THE CASE OF PROGRAM BIJAK SIFIR 2020 <i>Nazihah Safie, Syerrina Zakaria, Siti Madhihah Abdul Malik, Nur Bains Ismail, Azwani Alias Ruwaidiah Idris</i>	1
RADIATIVE CASSON FLUID OVER A SLIPPERY VERTICAL RIGA PLATE WITH VISCOUS DISSIPATION AND BUOYANCY EFFECTS <i>Siti Khuzaimah Soid, Khadijah Abdul Hamid, Ma Nuramalina Nasero, NurNajah Nabila Abdul Aziz</i>	10
GAUSSIAN INTEGER SOLUTIONS OF THE DIOPHANTINE EQUATION $x^4 + y^4 = z^3$ FOR $x \neq y$ <i>Shahrina Ismail, Kamel Ariffin Mohd Atan and Diego Sejas Viscarra</i>	19
A SEMI ANALYTICAL ITERATIVE METHOD FOR SOLVING THE EMDEN-FOWLER EQUATIONS <i>Mat Salim Selamat, Mohd Najir Tokachil, Noor Aqila Burhanddin, Ika Suzieana Murad and Nur Farhana Razali</i>	28
ROTATING FLOW OF A NANOFLUID PAST A NONLINEARLY SHRINKING SURFACE WITH FLUID SUCTION <i>Siti Nur Alwani Salleh, Norfifah Bachok and Nor Athirah Mohd Zin</i>	36
MODELING THE EFFECTIVENESS OF TEACHING BASIC NUMBERS THROUGH MINI TENNIS TRAINING USING MARKOV CHAIN <i>Rahela Abdul Rahim, Rahizam Abdul Rahim and Syahrul Ridhwan Morazuk</i>	46
PERFORMANCE OF MORTALITY RATES USING DEEP LEARNING APPROACH <i>Mohamad Hasif Azim and Saiful Izzuan Hussain</i>	53
UNSTEADY MHD CASSON FLUID FLOW IN A VERTICAL CYLINDER WITH POROSITY AND SLIP VELOCITY EFFECTS <i>Wan Faezah Wan Azmi, Ahmad Qushairi Mohamad, Lim Yeou Jiann and Sharidan Shafie</i>	60
DISJUNCTIVE PROGRAMMING - TABU SEARCH FOR JOB SHOP SCHEDULING PROBLEM <i>S. Z. Nordin, K.L. Wong, H.S. Pheng, H. F. S. Saipol and N.A.A. Husain</i>	68
FUZZY AHP AND ITS APPLICATION TO SUSTAINABLE ENERGY PLANNING DECISION PROBLEM <i>Liana Najib and Lazim Abdullah</i>	78
A CONSISTENCY TEST OF FUZZY ANALYTIC HIERARCHY PROCESS <i>Liana Najib and Lazim Abdullah</i>	89
FREE CONVECTION FLOW OF BRINKMAN TYPE FLUID THROUGH AN OSCILLATING PLATE <i>Siti Noramirah Ibrahim, Ahmad Qushairi Mohamad, Lim Yeou Jiann, Sharidan Shafie and Muhammad Najib Zakaria</i>	98

RADIATION EFFECT ON MHD FERROFLUID FLOW WITH RAMPED WALL TEMPERATURE AND ARBITRARY WALL SHEAR STRESS	106
<i>Nor Athirah Mohd Zin, Aaiza Gul, Siti Nur Alwani Salleh, Imran Ullah, Sharena Mohamad Isa, Lim Yeou Jiann and Sharidan Shafie</i>	

PART 2: STATISTICS

A REVIEW ON INDIVIDUAL RESERVING FOR NON-LIFE INSURANCE	117
<i>Kelly Chuah Khai Shin and Ang Siew Ling</i>	
STATISTICAL LEARNING OF AIR PASSENGER TRAFFIC AT THE MURTALA MUHAMMED INTERNATIONAL AIRPORT, NIGERIA	123
<i>Christopher Godwin Udomboso and Gabriel Olugbenga Ojo</i>	
ANALYSIS ON SMOKING CESSATION RATE AMONG PATIENTS IN HOSPITAL SULTAN ISMAIL, JOHOR	137
<i>Siti Mariam Norrulashikin, Ruzaini Zulhusni Puslan, Nur Arina Bazilah Kamisan and Siti Rohani Mohd Nor</i>	
EFFECT OF PARAMETERS ON THE COST OF MEMORY TYPE CHART	146
<i>Sakthiseswari Ganasan, You Huay Woon and Zainol Mustafa</i>	
EVALUATION OF PREDICTORS FOR THE DEVELOPMENT AND PROGRESSION OF DIABETIC RETINOPATHY AMONG DIABETES MELLITUS TYPE 2 PATIENTS	152
<i>Syafawati Ab Saad, Maz Jamilah Masnan, Karniza Khalid and Safwati Ibrahim</i>	
REGIONAL FREQUENCY ANALYSIS OF EXTREME PRECIPITATION IN PENINSULAR MALAYSIA	160
<i>Iszuanie Syafidza Che Ilias, Wan Zawiah Wan Zin and Abdul Aziz Jemain</i>	
EXPONENTIAL MODEL FOR SIMULATION DATA VIA MULTIPLE IMPUTATION IN THE PRESENT OF PARTLY INTERVAL-CENSORED DATA	173
<i>Salman Umer and Faiz Elfaki</i>	
THE FUTURE OF MALAYSIA'S AGRICULTURE SECTOR BY 2030	181
<i>Thanusha Palmira Thangarajah and Suzilah Ismail</i>	
MODELLING MALAYSIAN GOLD PRICES USING BOX-JENKINS APPROACH	186
<i>Isnewati Ab Malek, Dewi Nur Farhani Radin Nor Azam, Dinie Syazwani Badrul Aidi and Nur Syafiqah Sharim</i>	
WATER DEMAND PREDICTION USING MACHINE LEARNING: A REVIEW	192
<i>Norashikin Nasaruddin, Shahida Farhan Zakaria, Afida Ahmad, Ahmad Zia Ul-Saufie and Norazian Mohamaed Noor</i>	
DETECTION OF DIFFERENTIAL ITEM FUNCTIONING FOR THE NINE-QUESTIONS DEPRESSION RATING SCALE FOR THAI NORTH DIALECT	201
<i>Suttipong Kawilapat, Benchlak Maneeton, Narong Maneeton, Sukon Prasitwattanaseree, Thoranin Kongsuk, Suwanna Arunpongpaisal, Jintana Leejongpermpool, Supattra Sukhawaha and Patrinee Traisathit</i>	

ACCELERATED FAILURE TIME (AFT) MODEL FOR SIMULATION PARTLY INTERVAL-CENSORED DATA	210
<i>Ibrahim El Feky and Faiz Elfaki</i>	
MODELING OF INFLUENCE FACTORS PERCENTAGE OF GOVERNMENTS' RICE RECIPIENT FAMILIES BASED ON THE BEST FOURIER SERIES ESTIMATOR	217
<i>Chaerobby Fakhri Fauzaan Purwoko, Ayuning Dwis Cahyasari, Netha Aliffia and M. Fariz Fadillah Mardianto</i>	
CLUSTERING OF DISTRICTS AND CITIES IN INDONESIA BASED ON POVERTY INDICATORS USING THE K-MEANS METHOD	225
<i>Khoirun Niswatin, Christopher Andreas, Putri Fardha Asa OktaviaHans and M. Fariz Fadilah Mardianto</i>	
ANALYSIS OF THE EFFECT OF HOAX NEWS DEVELOPMENT IN INDONESIA USING STRUCTURAL EQUATION MODELING-PARTIAL LEAST SQUARE	233
<i>Christopher Andreas, Sakinah Priandi, Antonio Nikolas Manuel Bonar Simamora and M. Fariz Fadillah Mardianto</i>	
A COMPARATIVE STUDY OF MOVING AVERAGE AND ARIMA MODEL IN FORECASTING GOLD PRICE	241
<i>Arif Luqman Bin Khairil Annuar, Hang See Pheng, Siti Rohani Binti Mohd Nor and Thoo Ai Chin</i>	
CONFIDENCE INTERVAL ESTIMATION USING BOOTSTRAPPING METHODS AND MAXIMUM LIKELIHOOD ESTIMATE	249
<i>Siti Fairus Mokhtar, Zahayu Md Yusof and Hasimah Sapiri</i>	
DISTANCE-BASED FEATURE SELECTION FOR LOW-LEVEL DATA FUSION OF SENSOR DATA	256
<i>M. J. Masnan, N. I. Maha3, A. Y. M. Shakaf, A. Zakaria, N. A. Rahim and N. Subari</i>	
BANKRUPTCY MODEL OF UK PUBLIC SALES AND MAINTENANCE MOTOR VEHICLES FIRMS	264
<i>Asmahani Nayan, Amirah Hazwani Abd Rahim, Siti Shuhada Ishak, Mohd Rijal Ilias and Abd Razak Ahmad</i>	
INVESTIGATING THE EFFECT OF DIFFERENT SAMPLING METHODS ON IMBALANCED DATASETS USING BANKRUPTCY PREDICTION MODEL	271
<i>Amirah Hazwani Abdul Rahim, Nurazlina Abdul Rashid, Abd-Razak Ahmad and Norin Rahayu Shamsuddin</i>	
INVESTMENT IN MALAYSIA: FORECASTING STOCK MARKET USING TIME SERIES ANALYSIS	278
<i>Nuzlinda Abdul Rahman, Chen Yi Kit, Kevin Pang, Fauhatuz Zahroh Shaik Abdullah and Nur Sofiah Izani</i>	

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

- ANALYSIS OF THE PASSENGERS' LOYALTY AND SATISFACTION OF AIRASIA PASSENGERS USING CLASSIFICATION** 291
Ee Jian Pei, Chong Pui Lin and Nabilah Filzah Mohd Radzuan
- HARMONY SEARCH HYPER-HEURISTIC WITH DIFFERENT PITCH ADJUSTMENT OPERATOR FOR SCHEDULING PROBLEMS** 299
Khairul Anwar, Mohammed A.Awadallah and Mohammed Azmi Al-Betar
- A 1D EYE TISSUE MODEL TO MIMIC RETINAL BLOOD PERFUSION DURING RETINAL IMAGING PHOTOPLETHYSMOGRAPHY (IPPG) ASSESSMENT: A DIFFUSION APPROXIMATION – FINITE ELEMENT METHOD (FEM) APPROACH** 307
Harnani Hassan, Sukreen Hana Herman, Zulfakri Mohamad, Sijung Hu and Vincent M. Dwyer
- INFORMATION SECURITY CULTURE: A QUALITATIVE APPROACH ON MANAGEMENT SUPPORT** 325
Qamarul Nazrin Harun, Mohamad Noorman Masrek, Muhamad Ismail Pahmi and Mohamad Mustaqim Junoh
- APPLY MACHINE LEARNING TO PREDICT CARDIOVASCULAR RISK IN RURAL CLINICS FROM MEXICO** 335
Misael Zambrano-de la Torre, Maximiliano Guzmán-Fernández, Claudia Sifuentes-Gallardo, Hamurabi Gamboa-Rosales, Huizilopoztli Luna-García, Ernesto Sandoval-García, Ramiro Esquivel-Felix and Héctor Durán-Muñoz
- ASSESSING THE RELATIONSHIP BETWEEN STUDENTS' LEARNING STYLES AND MATHEMATICS CRITICAL THINKING ABILITY IN A 'CLUSTER SCHOOL'** 343
Salimah Ahmad, Asyura Abd Nassir, Nor Habibah Tarmuji, Khairul Firhan Yusob and Nor Azizah Yacob
- STUDENTS' LEISURE WEEKEND ACTIVITIES DURING MOVEMENT CONTROL ORDER: UİTM PAHANG SHARING EXPERIENCE** 351
Syafıza Saila Samsudin, Noor Izyan Mohamad Adnan, Nik Muhammad Farhan Hakim Nik Badrul Alam, Siti Rosiah Mohamed and Nazihah Ismail
- DYNAMICS SIMULATION APPROACH IN MODEL DEVELOPMENT OF UNSOLD NEW RESIDENTIAL HOUSING IN JOHOR** 363
Lok Lee Wen and Hasimah Sapiri
- WORD PROBLEM SOLVING SKILLS AS DETERMINANT OF MATHEMATICS PERFORMANCE FOR NON-MATH MAJOR STUDENTS** 371
Shahida Farhan Zakaria, Norashikin Nasaruddin, Mas Aida Abd Rahim, Fazillah Bosli and Kor Liew Kee
- ANALYSIS REVIEW ON CHALLENGES AND SOLUTIONS TO COMPUTER PROGRAMMING TEACHING AND LEARNING** 378
Noor Hasnita Abdul Talib and Jasmin Ilyani Ahmad

PART 4: OTHERS

- ANALYSIS OF CLAIM RATIO, RISK-BASED CAPITAL AND VALUE-ADDED INTELLECTUAL CAPITAL: A COMPARISON BETWEEN FAMILY AND GENERAL TAKAFUL OPERATORS IN MALAYSIA** 387
Nur Amalina Syafiqa Kamaruddin, Norizarina Ishak, Siti Raihana Hamzah, Nurfadhlina Abdul Halim and Ahmad Fadhly Nurullah Rasade
- THE IMPACT OF GEOMAGNETIC STORMS ON THE OCCURRENCES OF EARTHQUAKES FROM 1994 TO 2017 USING THE GENERALIZED LINEAR MIXED MODELS** 396
N. A. Mohamed, N. H. Ismail, N. S. Majid and N. Ahmad
- BIBLIOMETRIC ANALYSIS ON BITCOIN 2015-2020** 405
Nurazlina Abdul Rashid, Fazillah Bosli, Amirah Hazwani Abdul Rahim, Kartini Kasim and Fathiyah Ahmad@Ahmad Jali
- GENDER DIFFERENCE IN EATING AND DIETARY HABITS AMONG UNIVERSITY STUDENTS** 413
Fazillah Bosli, Siti Fairus Mokhtar, Noor Hafizah Zainal Aznam, Juaini Jamaludin and Wan Siti Esah Che Hussain
- MATHEMATICS ANXIETY: A BIBLIOMETRIX ANALYSIS** 420
Kartini Kasim, Hamidah Muhd Irpan, Noorazilah Ibrahim, Nurazlina Abdul Rashid and Anis Mardiana Ahmad
- PREDICTION OF BIOCHEMICAL OXYGEN DEMAND IN MEXICAN SURFACE WATERS USING MACHINE LEARNING** 428
Maximiliano Guzmán-Fernández, Misael Zambrano-de la Torre, Claudia Sifuentes-Gallardo, Oscar Cruz-Dominguez, Carlos Bautista-Capetillo, Juan Badillo-de Loera, Efrén González Ramírez and Héctor Durán-Muñoz

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

Khoirun Niswatin¹, Christopher Andreas², Putri Fardha Asa Oktavia Hans³, M. Fariz Fadillah Mardianto⁴

^{1,2,3,4}Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia

(¹khoirun.niswatin-2018@fst.unair.ac.id, ²christopher.andreas-2018@fst.unair.ac.id, ³putri.fardha.asa-2018@fst.unair.ac.id, ⁴m.fariz.fadillah.m@fst.unair.ac.id)

Poverty is a multidimensional problem caused by various aspects and has an impact on many aspects of life. As one of the objectives of the Sustainable Development Goals (SDG's), poverty eradication aims to improve welfare in all forms everywhere. Therefore, regional clustering is carried out based on poverty indicators as one of the alternatives to solve poverty problems in Indonesia. In this research, the mapping will be carried out based on districts and cities, so that the core problems in each region can be identified. By using the K-Means clustering method, eight clusters of districts and cities with different poverty levels were obtained. Each cluster member from the very not poor to very poor level is 130, 76, 126, 98, 66, 5, 5, and 8 districts and cities. Thus, most districts and cities in Indonesia are categorized as prosperous, only some region need special attention and treatment from the government.

Keywords: clustering, icd rate, K-Means, poverty

1. Introduction

Indonesia has a big challenge to improve the quality of basic services in preparing and implementing Sustainable Development Goals (SDG's) projects. SDG's is a sustainable development agenda launched to address world problems. SDG's has 17 target objectives and 169 measurable targets, which are expected to be achieved by 2030 (Bappenas, 2021). SDG's readiness in Indonesia has been welcomed for a long time. In 2020, the achievement of SDGs in Indonesia at the world level is in the order of 101. According to the data from the Sustainable Development Report 2020, Indonesia has progressed from the previous year with a score of 64.2 to 65.3, succeeding in raising 4 goals, one of which is to eradicate poverty, 6 goals have improved successfully, while 7 goals are stagnant and even decreased, one of which is reducing inequality (Sachs et al, 2020).

Coronavirus Disease (Covid-19) is a global pandemic that affects all aspects of life, for example world economy aspect. The Covid-19 pandemic has paralyzed the world's economic sector and suffered losses, which then resulted in layoffs, which led to the emergence of new poor groups (Rassanjani et al., 2021). As all costs of being poor, poverty is a complex problem where there is an inability to meet basic needs. The Covid-19 pandemic also caused Indonesia's economic growth to slow down and the poverty rate to increase (Susilawati et al., 2020). Based on data from the Central Bureau of Statistics or BPS, the achievement of SDGs in Indonesia, especially in alleviating poverty in 2014-2019, succeeded in reducing the poverty rate to 1.64 percent or 2.93 million people. Apart from these achievements, the presence of Covid-19 has again increased the poverty rate to 1.13 million people (BPS, 2020).

One of the characteristics of poverty in Indonesia is the huge difference between the relative poverty value and the absolute poverty value with geographic location. The measure of absolute poverty is predetermined, determined by the poverty line figures which are constant over time (Asrol and Ahmad, 2018). Relative poverty is poverty that is not connected to the poverty line, but from the perspective of each individual. If in absolute terms, more than half of the total population of Indonesia who lives in poverty are located on the island of Java, to be precise, located in the western part of Indonesia with a dense population. In the relative terms, provinces in Eastern

Indonesia show higher poverty scores (BPS, 2016). For this reason, research that conducts regional clustering based on poverty indicators is very important to do to overcome the problem of poverty in Indonesia. In this case, the mapping will be carried out on a district and city basis, so that the main problems of each region can be identified. The poverty indicators used in the clustering is based on the information given by BPS publication entitled “Data and Information on District and City Poverty in 2020” (BPS, 2020).

Clustering is a part of multivariate analysis known as cluster analysis. In cluster analysis, many clustering methods can be used. Hierarchically, the most widely used methods in cluster analysis are single linkage, complete linkage, average linkage, and Ward (Afifi et al., 2020; Wierzchon and Klopotek, 2018). In contrast, the K-Means, K-Medoids, and Self Organizing Maps (SOM) methods are widely used in the non-hierarchical cluster analysis. Research related to cluster analysis with hierarchical and non-hierarchical approaches has been applied by various researchers. One of them, a study that classified minimarkets based on the level of adherence to health protocols (Andreas et al., 2021). However, several studies have shown that the K-Means algorithm has better capabilities than the SOM algorithm and various approaches to hierarchical cluster analysis (Syaripudin, 2013). Apart from that, Suarna et. al. (2020) showed that the K-Means method has a more accurate result than the K-Medoids method on the food dataset. Therefore, this study uses the K-Means method in classifying districts and cities in Indonesia based on poverty indicators.

There are several studies related to the clustering that focuses on the problem of poverty in Indonesia. According to 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. The attributes used in this study are the percentage of poor people, the poverty line, the poverty depth index, and the poverty severity index. Provinces that fall into the high poverty category are Maluku, West Papua, and Papua. In addition, K-Means method can also be used in clustering districts and cities in Central Java based on the Covid-19 case (Mahmudan, 2020).

Furthermore, in this study, the mapping results will be validated through the size of the internal cluster dispersion rate (icd rate) and the coefficient of determination to form an optimal mapping. The size of the icd rate and the coefficient of determination is a measure that measures the level of homogeneity or similarity in each cluster so that it reflects the level of goodness of the cluster formed (Rahayu et al., 2018). In this case, the variables used include the number of poor people, the poverty line, unemployment rate, school enrollment rate, human development index (HDI), expenditure per capita on food, eligible water users, users of their latrines, literacy rates, and beneficiaries of government programs. The use of various types of poverty indicators in the research variables is the novelty of this research. In addition, the clustering based on districts and cities in Indonesia based on poverty indicators is also the novelty of this research.

This study resulted in a mapping of each city and district in Indonesia based on poverty categories. The process of mapping regions that are categorized as poor is one way of implementing statistical methods to help the government determine comprehensive and targeted policies. The results of this study are very useful for formulating policies related to efforts to resolve poverty problems and improve the welfare of the people in each district and city of Indonesia according to the SDG’s goals. Thus, this research can be used as a reference and consideration for the government in determining appropriate policies to reduce poverty in Indonesia.

2. Methods

2.1 Data Sources and Research Variables

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. Research variables in this study use ten indicators from BPS that can measure poverty. The summary of information regarding the research variables can be seen in Table 1.

Table 1: Research Variables

Research Variables	Information	Research Variables	Information
X_1	Number of Poor Population (Thousands)	X_6	Expenditure Per Capita on Food (%)
X_2	Poverty Line (Rupiah).	X_7	Eligible Water Users (%)
X_3	Unemployment Rate (%)	X_8	Own Toilet Users (%)
X_4	School Enrollment Rate (%)	X_9	Literacy rate (%)
X_5	Human Development Index (%)	X_{10}	Government Program Beneficiaries (%)

2.2 Analysis Procedure

The data analysis method used in this study is a quantitative method with the following procedures:

1. Conducting descriptive data analysis.
2. Scaling measurements on research data with a standardization process is an initial step in the process of clustering cities/districts based on poverty indicators.
3. Conducting the clustering procedure using the K-Means method (Afifi et al., 2020; Wierzchon and Klopotek, 2018; Kassambara, 2017). The K-Means cluster formation algorithm is as follows:
 - a. Determining k the number of clusters to be formed.
 - b. Determining the initial k centroids randomly.
 - c. Calculating the distance of each data to each centroid.
 - d. Clustering each data to the nearest centroids.
 - e. Determining the position of the new centroids by calculating the average value of the data located on the same centroids.
 - f. Running this process iteratively until there are no more changes in the clustering.

In this case, the cluster formation procedure will be based on the euclidean distance (Afifi et al., 2020; Wierzchon and Klopotek, 2018). The euclidean distance from the i^{th} object to the j^{th} object is defined as follows:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1)$$

4. Selecting the best number of clusters by using the size of the icd rate and the coefficient of determination (Rahayu et al., 2018; Hayes and Moulton, 2017). The coefficient of determination for c groups, p variables, n_c and data in the i^{th} group is defined as follows:

$$R^2 = \frac{(SST - SSW)}{SST} \quad (2)$$

with,

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}^k)^2 \quad (3)$$

$$SSW = \sum_{i=1}^{n_c} \sum_{j=1}^c \sum_{k=1}^p (x_{ij}^k - \bar{x}_{ij}^k)^2 \quad (4)$$

In addition, the criteria for selecting the best cluster can be based on the icd rate (Rahayu et al., 2018). The better the results of clustering the smaller the icd rate. Mathematically, the icd rate can be written as follows:

$$icd\ rate = 1 - R^2 \quad (5)$$

5. Conducting multivariate data distribution assumption testing and continue with MANOVA testing if the assumptions are met. The significance level used in this study is 5%. The test statistics that apply in this test are as follows:

$$\Lambda^* = \frac{|W|}{|B + W|} \quad (6)$$

with,

W: Matrix Sum of Square Residuals

B: Matrix Sum of Square Treatment

If c is the number of groups and n_c is the number of members in the c group, then the valid test criterion H_0 is rejected if $\Lambda^* > F_{n_c-1, n-n_c}(\alpha)$

6. Displaying the best clustering results in the form of a map graphic and identifies the characteristics of each cluster formed.
7. Providing policy recommendations to improve people's welfare based on the results of the clustering that has been carried out.

3. Result and Discussion

3.1 Overview of Poverty Indicators in Indonesia

Indonesia has 34 provinces with total of 514 cities and districts. Each city or district has its own special characteristics. In September 2020, BPS recorded that the total number of poor people in Indonesia was 27.55 million people. The average poverty line is Rp441,580. On average, there are 41.63% of poor people over 15 years old that do not work in every city or district. Based on all of cities and districts in Indonesia, there are 27% cities and districts that has school enrollment rate of 100%. As for the literacy rate, 32% of cities and districts have reached 100%. The summary of the highest and lowest score for each research variable can be seen in Table 2. Based on Table 2, it can be seen that only 2% of cities and districts in Indonesia where all poor households have access to eligible water, and only 6% have their latrines. These percentages counted by comparing the number of cities or district showed in Table 2 (column Location) with a total 514 cities and districts that observed. The lowest score for unemployment rates, school enrollment rates, HDI, users of their latrines, eligible water users, and literacy rates are in cities or districts in Papua Province.

Table 2: Descriptive Summary of Data Score on Each Research Variable.

Variable	Highest		Lowest	
	Score	Location	Score	Location
X_1	465,670 people	Bogor District	1,360 people	Sawah Lunto Town
X_2	Rp1,021,759	Jayapura City	Rp248,184	Buton Selatan District
X_3	65.85 %	Tomohon Town	4.16 %	Lanny Jaya District
X_4	100 %	139 cities and districts	36.91 %	Puncak District
X_5	86.61 %	Yogyakarta City	31.55 %	Nduga District
X_6	82.08 %	Alor District	46.51 %	Ternate City
X_7	100 %	10 cities and districts	2.12 %	Mamberamo Tengah District
X_8	100 %	32 cities and districts	2.11 %	Puncak District
X_9	100 %	162 cities and districts	16.55 %	Lanny Jaya District
X_{10}	67.27 %	Sabu Raijua District	0 %	38 cities and districts

3.2 District and City Clustering based on Poverty Indicators

Before clustering analysis conducted, each variable of the research data must have the same measurement scale. In this case, not all research variables have the same measurement scale so that the standardization process for each variable should be carried out. Furthermore, using the K-Means procedure, the results of the clustering along with the cluster goodness measurements for each possible number of clusters were obtained as presented in Table 3.

Based on Table 3, it can be seen that the coefficient of determination increases with the increase in the number of clusters, and the value of the icd rate decreases with the increase in the number of clusters. This shows that the clustering with the number of clusters as many as 8 gives optimal results. The coefficient of determination of 49.79% indicates a value as large as the data that can be

explained by each cluster. This value almost reaches 50% which is at the intermediate level so that the clustering results can be said to have been optimal.

Table 3: Selection of The Optimal Number of Clusters.

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

The clustering results obtained with the number of clusters as many as 8 showed the distribution of districts and cities in various ways. Each poverty level cluster contains 130, 76, 126, 98, 66, 5, 5, and 8 districts and cities, respectively. In order to ensure that the results of the cluster have high homogeneity in each cluster, it is necessary to carry out a MANOVA test to see if there are differences in each cluster that has been formed. Basic assumption before conducting MANOVA test, which was normally distributed data in multivariate ways, is fulfilled. In addition, testing the correlation between the value of the Mahalanobis distance and the chi-square value shows a significant result with a Pearson correlation value of 0.823 with p-value equal to 0.000. Furthermore, the MANOVA test shows that all variables are significant with p-value 0.000. At an error rate of 5%, it can be concluded that each of these variables show a difference in the results of the clustering that has been done. In other words, the homogeneity of each cluster formed and the heterogeneity between clusters is relatively high. Thus, the results of clustering through the K-Means method with total of 8 clusters are feasible to use. Based on this mapping, each cluster is classified according to its poverty level. The results of the clustering are presented in the form of the map shown in Figure 1. Based on Figure 1, eight colors indicate the poverty level of a region. A solid red color for cluster VIII which indicates the districts or cities that are very poor, to pale white for cluster I which indicates areas of districts or cities that are very not poor.

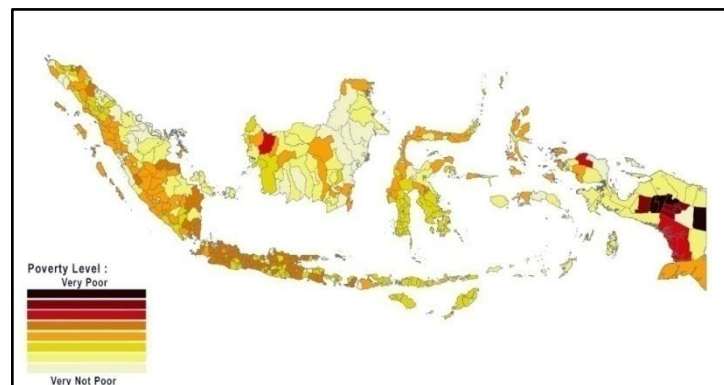


Figure 1: Mapping of Areas Based on Poverty Levels in Indonesia.

3.3 Characteristics of Poverty Level Clusters in Indonesia

Based on the cluster analysis in the previous section, it was found that the poverty rate in terms of various poverty indicators in each district and city formed 8 clusters. This indicates that there is a gap which is indicated by the level of poverty between districts and cities, which varies greatly by forming 8 levels. Efforts to overcome poverty problems in Indonesia must be carried out following the problems faced by each cluster. By identifying the characteristics of each poverty level cluster, efforts to overcome poverty can be carried out on target. The characteristics of each cluster that has been formed are presented in Table 4.

Table 4: Cluster Characteristics.

<p>Cluster I (very not poor)</p> <ol style="list-style-type: none"> 1. The regions with the highest school enrollment and literacy rates, and had the highest HDI scores compared to other cluster areas. 2. In the housing facilities sector, people in this area have had access to safe water and have their latrines with the highest percentage compared to areas in other clusters. 3. The level of expenditure per capita for food in this cluster is the smallest so that the community provides a relatively larger portion of expenditure for other needs besides food. 4. The number of poor people and the percentage of people who receive assistance from the government is moderate. 5. The poverty line is relatively high even though it is still below Cluster VI and the unemployment rate for the poor is the highest in this region.
<p>Cluster II (not poor)</p> <ol style="list-style-type: none"> 1. The number of poor people in this area is the lowest. 2. The unemployment rate for the poor, school enrollment and literacy rates, per capita expenditure on food, users of eligible water and latrines themselves, and the HDI score tended to be in the middle by not being included in the highest or lowest categories. 3. However, the poverty line in this region is relatively high, being slightly below Cluster 1.
<p>Cluster III (fairly not poor)</p> <ol style="list-style-type: none"> 1. The poverty line and the percentage of people using eligible water in this area are in a good category, slightly below Cluster I. 2. The per capita expenditure on food in this region is the largest compared to other cluster areas. 3. 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.
<p>Cluster IV (tend not to be poor)</p> <ol style="list-style-type: none"> 1. There is no single poverty indicator that stands out, either the highest or the lowest, in this cluster area. 2. In other words, all indicators are in the medium value category.
<p>Cluster V (tend to be poor)</p> <ol style="list-style-type: none"> 1. The number of poor people in this cluster area is very high and is accompanied by a very high percentage of government assistance to the poor. 2. The unemployment rate for the poor is relatively high with the second-highest ranking, which is one level below Cluster I. 3. People who already have their latrines are relatively high, which is one level below Cluster I.
<p>Cluster VI (fairly poor)</p> <ol style="list-style-type: none"> 1. The percentage of people who do not have their latrine is very high. 2. The number of poor people and the poverty line is the lowest. 3. However, various other indicators such as school enrollment and literacy rates and IPM scores tended to have low scores.
<p>Cluster VII (poor)</p> <ol style="list-style-type: none"> 1. The percentage of literacy rate is very low and the percentage of eligible water users is very low. 2. The unemployment rate for the poor and government assistance programs in the region is the lowest.
<p>Cluster VIII (very poor)</p> <ol style="list-style-type: none"> 1. The poverty line is very high and is accompanied by a low enrollment rate. 2. The HDI value in this region is the lowest compared to areas in other clusters. 3. The percentage of government aid programs in this region is very low, which is the same as in Cluster II.

3.4 Public Welfare Policy Recommendations

People's welfare can be achieved if the poverty problem can be resolved. However, the problem of poverty is complex so that a partial policy will find it difficult to solve the main problem. By considering various poverty indicators as a whole, the main problems of poverty in each cluster can be seen in Table 5.

Policy recommendations in actualizing people's welfare will be based on each characteristic and main problem of the poverty level mapping that has been carried out. Thus, each policy recommendation will be appropriate and answer every major problem of poverty in different clusters. In summary, policy recommendations that need to be carried out to overcome the poverty

problem are presented in Figure 2. By implementing a policy strategy that is following the main problems in each cluster, efforts to resolve the problem of poverty will run more precisely and more comprehensively so that people's welfare can be realized.

Table 5: Main Problem in Each Cluster

Cluster	Main Problem
Cluster I	Very high unemployment rate (number of unemployed).
Cluster II	In general, there are no poor indicators of poverty.
Cluster III	The per capita expenditure on food is very high.
Cluster IV	Various aspects of poverty indicators in this region show a score that is not bad, although it is still lower than Cluster II.
Cluster V	Dominated by a high number of poor people.
Cluster VI	The number of households that use their latrines is low.
Cluster VII	Decent eligible water users and low literacy rates.
Cluster VIII	High poverty line, very low enrollment rates, and low HDI scores.

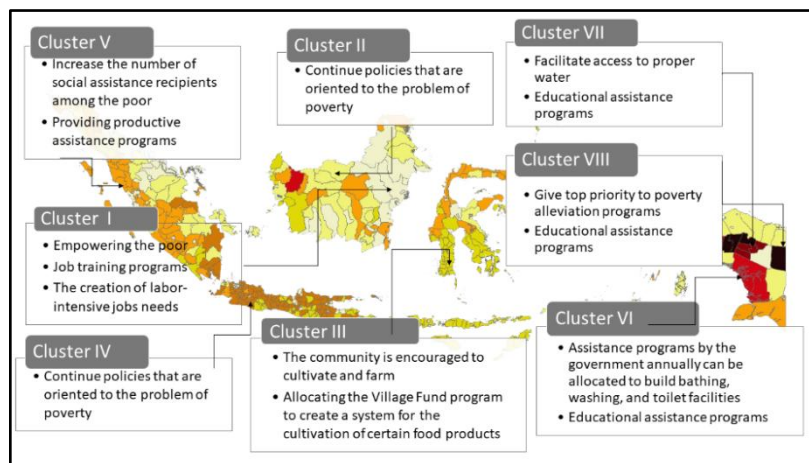


Figure 2. Policy Recommendations for Realizing People's Welfare.

4. Conclusions

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 be low. 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 numbers of members for each cluster from least poor to very poor are 130, 76, 126, 98, 66, 5, 5, and 8 districts or cities. In addition, cluster analysis testing through the MANOVA test also showed significant results so that the results of the area mapping carried out were feasible to use. The characteristics of the level of poverty that occur in each regional cluster are very diverse. In cluster VIII which is included in the very poor category, the poverty line is very high. Meanwhile, the school enrollment rate and the HDI score are very low. The opposite occurred in the cluster I which was categorized as not very poor. In this region, various poverty indicators show good scores, even though the unemployment rate for the poor is very high. 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.

Acknowledgment

The author would like to thank all parties who played a role during the research process. In addition, the authors also thank the Faculty of Science and Technology for providing support in this research activity.

References

- Afifi, A., May, S., Donatello, R. A., and Clark, V. A. (2020). *Practical Multivariate Analysis, Sixth Edition*. Boca Raton: Taylor & Francis Group.
- Andreas, C., Susanti, R., Mardianto, M. F. F., and Huda, A. M. (2021). Classification of Minimarkets Based On Compliance Level with Health Protocols Using A Comparison of Hierarchic and Non-Hierarchic Approaches. *International Academic Conference*, 5: 25-30.
- Asrol and Ahmad, H. (2018). Analysis of Factors that Affect Poverty in Indonesia. *Revista ESPACIOS*, 39(45): 14-24.
- Bahauddin, A., Fatmawati, A., Sari, F. (2021). Clustering Analysis of Province in Indonesia Based on Poverty Levels Using The K-Means Algorithm. *Jurnal Manajemen Informatika dan Sistem Informasi*, 4(1): 1-8.
- Bappenas. (2020). Sekilas SDG's. Retrieved May 7 2021, from <http://sdgs.bappenas.go.id/sekilas-sdgs/>.
- Central Bureau of Statistics. (2016). *Statistik Kemiskinan dan Ketidaksetaraan di Indonesia*. Jakarta: BPS Indonesia.
- Central Bureau of Statistics. (2020). *Data dan Informasi Kemiskinan Kabupaten/Kota*. Jakarta: BPS Indonesia.
- Hayes, R and Moulton, L. (2017). *Cluster Randomized Trials, Second Edition*. Boca Raton: Chapman & Hall / CRC Biostatistics Series.
- Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning, Edition 1*. sthda.com.
- Mahmudan, A. (2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering. *Jurnal Matematika, Statistika, & Komputasi*, 17(1): 1-13.
- Rahayu, S. P., Febryani, S., Kusuma, H. B., Suhartono, and Prastyo, D. D. (2018). The Approach of Fuzzy C-Means Cluster and Factor Analysis on Economic Potential Mapping of Regency/ City in East Java Province. *AIP Conference Proceedings*, 2014(1): 020082.
- Rassanjani, S., Risky, N., Maz, D., Alqarni, W., and Tharis, M. A. (2021). Activities and Poverty Threats in the Asia-Pacific Region. *Policy & Governance Review*, 5(1): 82-96.
- Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., Fuller, G., and Woelm, F. (2020) *The Sustainable Development Goals and COVID-19. Sustainable Development Report 2020*. Cambridge: Cambridge University Press.
- Suarna, N., Wijaya, Y. A., Mulyawan, Hartati, T., and Suprpti, T. (2020). Comparison K-Medoids Algorithm and K-Means Algorithm for Clustering Fish Cooking Menu from Fish Dataset. *IOP Conference Series: Materials Science and Engineering*, 1088(1): 012034.
- Susilawati, Falevi, R. and Purwoko, A. (2020). Impact of COVID-19's Pandemic on the Economy of Indonesia. *Budapest International Research and Critics Institute (BIRCI-Journal)*, 3(2): 1147-1156.
- Syaripudin, U. (2013). Studi Komparatif Penerapan Metode Hierarchical, K-Means dan Self Organizing Maps (SOM) Clustering pada Basis Data. *Jurnal ISTEK*, 7(1): 132:149.
- Wierzchon, S. And Klopotek, M. (2018). *Modern Algorithms of Cluster Analysis, Studies in Big Data, Vol. 34*. Gewerbestrasse: Springer.



**20
21** **ICMS**
INTERNATIONAL CONFERENCE ON COMPUTING,
MATHEMATICS AND STATISTICS

e ISBN 978-967-2948-12-4



9 7 8 9 6 7 2 9 4 8 1 2 4