# IMPLEMENTATION OF HYBRID INDEXING, CLUSTERING AND CLASSIFICATION METHODS TO ENHANCE RURAL DEVELOPMENT PROGRAMME IN SOUTH SULAWESI

**MUHAMMAD FAISAL**

**ASIA e UNIVERSITY**

**2024**

# IMPLEMENTATION OF HYBRID INDEXING, CLUSTERING AND CLASSIFICATION METHODS TO ENHANCE RURAL DEVELOPMENT PROGRAMME IN SOUTH SULAWESI

MUHAMMAD FAISAL

A Thesis Submitted to Asia e University in
Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

January 2024

## ABSTRACT

Limited involvement of communities and village institutions in planning, implementing, and supervising village development activities, as well as difficulty in monitoring development results, resulted in a lot of untapped natural potential in traditionally managed villages due to technological limitations and low levels of education. Development in rural areas faces obstacles from internal factors such as lack of initiative and knowledge among rural communities and external factors such as constraints on government support. Village development and assistance program is a strategy implemented to accelerate socio-cultural development to increase the capacity of village governance and more organized administration. The village assistance programme can motivate and engage rural communities during the participatory and transparent phase of village development. The criteria used in this study were produced through the identification, verification, and validation stages by experts consisting of academian, government, and researchers. The objectives of this research are to develop a community standard of living index based on verified criteria collected from selected communities, to cluster the village based on the community standard of living index, to classify a village based on the Developing Village Index(DVI), Human Development Index(HDI), and Community Standard of Living Index(CSLI), and to map the relevant experts with the priority needs of a village based on the input from the villages. The initial stage in this study involved the design of the questionnaire, the process of criteria weighting, and the scoring of villages by communities. CSLI was developed to represent the community welfare level for each village. Clustering techniques such as Self-Organizing Map, Fuzzy C-Means, and Xie-Beny methods are utilized to clustering villages according to the Community Standard of Living Index. The Fuzzy Tsukamoto and Smallest of Maximum methods were then used to classify villages into less development, which involved CSLI-Clusters as indicators. Using the cosine similarity algorithm for knowledge recommendation is village identified, utilizing community feedback as the foundation. Based on the clustering results using CSLI Score, Head of Family, and Number of Residence criteria, it is stated that all villages are divided into 3 clusters, which are CSLI-Good, CSLI-Average, and CSLI-Poor. The classification technique using the CSLI-Cluster, DVI, and HDI criteria showed that as many as 22 villages had the status of Less Development level, and 8 villages were declared Developed. This research identified the following recommended fields: agricultural science in 11 villages, social sciences in 11 villages, economics in 10 villages, entrepreneurship in 8 villages, marine science, forestry, and computers in 11 villages each, and regional planning in 2 villages with 82% accuracy. The result validation of decisions on the placement of accompanying experts in each village with actual data was carried out using the confusion matrix metho, which are accurate = 0.819, precision = 1, recall = 0.819, and F1-Score = 0.9. This shows that the accuracy status indicates a high percentage of correct predictions, and then the F1 Score of 0.9 indicates a well-balanced trade-off between precision and recall, demonstrating the model's overall effectiveness. The government can use the findings of this research as a decision-making tool regarding equitable village development programs.



**Keywords:** Rural, village assistant, indexing, clustering, classification, recommendation, DSS

## APPROVAL

This is to certify that this thesis conforms to acceptable standards of scholarly presentation and is fully adequate, in quality and scope, for the fulfilment of the requirements for the degree of Doctor of Philosophy

The student has been supervised by: **Professor Dr. Titik Khawa Abdul Rahman**

The thesis has been examined and endorsed by:

**Professor Dr. Azman Yassin,**
**College of Arts and Science,**
**UUM**
Examiner 1

**Professor Dr. Azlinah Mohamed**
**UiTM**
Examiner 2

This thesis was submitted to Asia e University and is accepted as fulfilment of the requirements for the degree of Doctor of Philosophy.

**Professor Dr Siow Heng Loke**

Asia e University

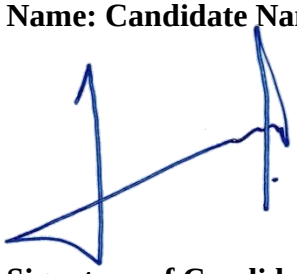
Chairman, Examination Committee

(04 January 2024)

## DECLARATION

I hereby declare that the thesis submitted in fulfilment of the PhD degree is my own work and that all contributions from any other persons or sources are properly and duly cited. I further declare that the material has not been submitted either in whole or in part, for a degree at this or any other university. In making this declaration, I understand and acknowledge any breaches in this declaration constitute academic misconduct, which may result in my expulsion from the programme and/or exclusion from the award of the degree.

**Name: Candidate Name**



**Signature of Candidate**: **Date**: **4 January 2024**

Copyright by Asia e University

## ACKNOWLEDGEMENTS

I express profound gratitude to my Doctoral Supervisor, Professor Dr. Titik Khawa Abdul Rahman, for her exceptional guidance, unwavering support, and invaluable insights throughout the entirety of this dissertation's completion process. Her profound expertise in Information Communication and Technology, commitment to academic excellence, and mentorship have played a pivotal role in shaping the trajectory of my research and academic journey.

I am indebted to STMIK Profesional Makassar and Dipa University for their collaborative support, which has enriched this research endeavor. Their significant contributions to the Information Communication and Technology field have played an instrumental role in broadening the scope and impact of this dissertation.

I wish to convey heartfelt gratitude to my family for their unwavering love, encouragement, and understanding throughout this doctoral journey. Their steadfast support and belief in my abilities have served as my greatest source of strength and motivation.

Lastly, I am thankful to all my colleagues, friends, and mentors who have provided unwavering support, encouragement, and scholarly insights throughout this journey. Your collective contributions have enriched this doctoral experience and have played a pivotal role in shaping the researcher I have become.

## TABLE OF CONTENTS


| | | |
|---|---|---|
| **ABSTRACT** | | **ii** |
| **APPROVAL** | | **iii** |
| **DECLARATION** | | **iv** |
| **ACKNOWLEDGEMENTS** | | **vi** |
| **TABLE OF CONTENTS** | | **vii** |
| **LIST OF TABLES** | | **x** |
| **LIST OF FIGURES** | | **xiv** |
| **LIST OF ABBREVIATION** | | **xvi** |
| **CHAPTER 1** | **INTRODUCTION** | **1** |
| 1.0 | Background of the Study | 1 |
| 1.1 | Problem Statement | 4 |
| 1.2 | Objectives | 5 |
| 1.3 | Research Questions | 6 |
| 1.4 | Justifications and Significance of the Study | 7 |
| 1.5 | Contribution to Methodology | 8 |
| 1.6 | Chapter Summary | 9 |
| **CHAPTER 2** | **REVIEW OF LITERATURE** | **11** |
| 2.0 | Introduction | 11 |
| 2.1 | Rural Development in Indonesia | 12 |
| 2.2 | Village Assistance | 13 |
| 2.3 | Developing Village Index (DVI) | 14 |
| 2.4 | Human Development Index (HDI) | 33 |
| 2.5 | Clustering Techniques | 38 |
| | 2.5.1 Fuzzy C-Means | 38 |
| | 2.5.2 Self-Organizing Map | 42 |
| | 2.5.3 Xie-Beni | 43 |
| 2.6 | Classification Techniques | 44 |
| | 2.6.1 Fuzzy Tsukamoto | 45 |
| | 2.6.2 Smallest of Maximum | 48 |
| 2.7 | Text Mining | 48 |
| 2.8 | Confusion Matrix | 51 |
| 2.9 | Community Standard of Living Index (CSLI) | 52 |
| 2.10 | Many Object Optimization (MOO) - Fuzzy Delphi Algorithm | 53 |
| 2.11 | Research Gap | 54 |
| 2.12 | Chapter Summary | 61 |
| **CHAPTER 3** | **METHODOLOGY** | **62** |
| 3.0 | Introduction | 62 |
| 3.1 | Research Design | 62 |
| 3.2 | Development of CLSI | 64 |
| | 3.2.1 Identification of Criteria | 64 |
| | 3.2.2 Validation of Criteria | 66 |
| | 3.2.3 Verification of Criteria | 66 |
| | 3.2.4 Evaluation of Criteria | 67 |

3.2.5 Weighting of Criteria 68
3.2.6 Village Input 69
3.3 Clustering of Community Standard of Living Index 77
3.3.1 Clustering Technique Using Hybrid SOM and Xie-Beny 80
3.3.1.1 Clustering Technique Using Hybrid SOM and Xie-Beny 80
3.3.1.2 Clustering Technique Using Hybrid FCM and Xie-Beny 80
3.3.1.3 Clustering Technique Using Hybrid SOM FCM, and Xie-Beny 80
3.3.1.4 Clustering Technique Using K-Means and Elbow 80
3.4 Classification of the Developing Village 86
3.4.1 Classification Fuzzy Tsukamoto and Smallest of Maximum 88
3.4.1.1 Classification Technique Using Fuzzy Tsukamoto 88
3.4.1.2 Classification Technique Using Tsukamoto and Smallest of Maximum 89
3.4.1.3 Classification Technique Using Fuzzy Mamdani 91
3.5 Determining Fields of Expertise Relevant to The Less Development Village 91
3.5.1 Text Mining Technique Using Cosine Similarity for Mapping Village Assistance 92
3.6 Chapter Summary 95

**CHAPTER 4 RESULTS AND DISCUSSION 96**

4.0 Introduction 96
4.1 Development Community Standard of Living Index based on Verified Criteria Collected from Selected Communities 96
4.1.1 Identification and Validation Criteria for CSLI 96
4.1.2 Verification and Evaluation for CSLI 98
4.1.3 Weightage for Village Input Using Rank Reciprocal 106
4.1.4 Result for Community Standard of Living Index 110
4.2 Village Clustering based on the Community Standard of Living Index Using Hybrid Fuzzy C-Means, Self-Organizing Map, and Xie-Beny 116
4.2.1 Initiation of Weight Using Self-Organizing Map 116
4.2.2 Final Weight for the Village 120
4.2.3 Clustering of Villages 123
4.2.3.1 Analysis for 4 Clusters 123
4.2.3.2 Analysis for 5 Clusters 123
4.2.3.3 Analysis for 6 Clusters 123
4.2.3.4 Comparion SOM, FCM, Xie Beny with KMeans Elbow 123
4.2.4 Optimum Clustering Using Xie-Beni and Elbow 155

4.3 Classification of a Village based on the DVI, HDI, and CSLI-Cluster Using the Fuzzy Tsukamoto + Smallest of Maximum Defuzzification 157
4.3.1 Result of Classification of Villages Using Tsukamoto and Smallest of Maximum 157
4.3.1.1 Determining of Indicators 157
4.3.1.2 Fuzzification of Indicators 159
4.3.1.3 Rule Evaluation 157
4.3.1.4 Defuzzification for Classification of Villages using Tsukamoto and Smallest of Maximum 157
4.3.1.5 Defuzzification for Classification of Villages Using Mamdani and Mean of Maximum 157
4.4 Mapping the Relevant Experts with the Priority needs of a Village based on the Input from the Villages Using Cosine Similarity Text Mining Process 193
4.5 Chapter Summary 201

**CHAPTER 5 CONCLUSION, IMPLICATION AND RECOMMENDATIONS 204**

5.0 Introduction 204
5.1 Over All Conclusion 204
5.2 Research Challenges 206
5.3 Limitation of the Study 207
5.4 Implications of the Study 208
5.5 Recommendation for Future Work 208
**REFERENCES 210**
**APPENDICES 224**
Appendix A 224
Appendix B 228
Appendix C 229
Appendix D 234
Appendix E 249
Appendix F 253

## LIST OF TABLES


| Table | | Page |
|---|---|---|
| 1.1 | Relationship problem statement, research object, research question | 6 |
| 2.1 | Determination of the maximum and minimum HDI | 37 |
| 2.2 | Summary of application hybrid technique of past few review papers | 47 |
| 2.3 | Example extraction in the text mining | 51 |
| 2.4 | Research gap summary | 57 |
| 3.1 | Identification of criteria | 64 |
| 3.2 | Fuzzy scale for validation of criteria | 66 |
| 3.3 | Village input instruments | 69 |
| 3.4 | Recommendation validation status in the field of expertise | 94 |
| 4.1 | List of respondents of expert | 97 |
| 4.2 | Input agreement value for criteria 1 to criteria 19 | 99 |
| 4.3 | Input agreement value for criteria 20 to criteria 38 | 99 |
| 4.4 | Determining average m1,m2,m3 for criteria 1 to criteria 5 | 100 |
| 4.5 | Determining average m1,m2,m3 for criteria 6 to criteria 10 | 100 |
| 4.6 | Determining average m1,m2,m3 for criteria 11 to criteria 21 | 101 |
| 4.7 | Determining average m1,m2,m3 for criteria 22 to criteria 32 | 102 |
| 4.8 | Determining average m1,m2,m3 for criteria 33 to criteria 38 | 103 |
| 4.9 | Threshold value for criteria 1 to criteria 10 | 103 |
| 4.10 | Threshold value for criteria 11 to criteria 38 | 103 |
| 4.11 | Evaluation of criteria | 105 |
| 4.12 | Weight of criteria and Instrument | 107 |
| 4.13 | The demographic of the respondents | 111 |
| 4.14 | Example community scoring | 112 |

| | | |
|---|---|---|
| 4.15 | CSLI value for each village | 115 |
| 4.16 | The village input | 116 |
| 4.17 | Normalization of criteria weight | 118 |
| 4.18 | Randomly initial weight | 119 |
| 4.19 | Determining short distance | 120 |
| 4.20 | Calculation of BMU weight | 120 |
| 4.21 | Update weight on iteration 16 | 121 |
| 4.22 | The changing data on iteration 17. | 121 |
| 4.23 | Convergence data  for villages using SOM | 122 |
| 4.24 | Membership value | 123 |
| 4.25 | MiU squared | 125 |
| 4.26 | Cluster centroid | 126 |
| 4.27 | Membership degree | 127 |
| 4.28 | Objective value each iteration | 128 |
| 4.29 | Convergent data on iteration 1 | 129 |
| 4.30 | Result member value on the last iteration for 3 clusters | 130 |
| 4.31 | Cluster centroid for 4 clusters | 132 |
| 4.32 | Membership degree for 4 clusters | 133 |
| 4.33 | Result member value on the last iteration for 4 clusters | 135 |
| 4.34 | Cluster centroid for 5 clusters | 137 |
| 4.35 | Membership degree for 5 clusters | 138 |
| 4.36 | Convergent data on Iteration 1 for 5 clusters | 140 |
| 4.37 | Result member value on the last iteration for 5 clusters | 141 |
| 4.38 | Cluster centroid for 6 clusters | 143 |
| 4.39 | Membership degree for 6 clusters | 144 |

| | | |
|---|---|---|
| 4.40 | Result member value on the last iteration for 6 clusters | 146 |
| 4.41 | Initial centroid Elbow randomly | 148 |
| 4.42 | Result member value on the last iteration | 150 |
| 4.43 | Initial centroid K-means | 151 |
| 4.44 | Determining centroid distance K-means | 151 |
| 4.45 | New centroid K-means | 153 |
| 4.46 | Result clusters on the last iteration K-means | 154 |
| 4.47 | Optimum cluster Xie-Beni | 156 |
| 4.48 | List of indicators for classification | 158 |
| 4.49 | Set of CSLI - cluster indicator | 160 |
| 4.50 | Set of DVI indicators | 161 |
| 4.51 | Set of HDI indicators | 163 |
| 4.52 | Set of village development level indicators | 165 |
| 4.53 | List of rules based on Fuzzy Tsukamoto | 167 |
| 4.54 | Assessment data from each village | 170 |
| 4.55 | Result of fuzzy implication for each village using Tsukamoto | 177 |
| 4.56 | Defuzzification of SoM and weighted average | 187 |
| 4.57 | Result of fuzzy implication for village using Mamdani | 189 |
| 4.58 | Defuzzification of Mom and SoM | 192 |
| 4.59 | Dataset community comments | 194 |
| 4.60 | Building term frequency and inverse document frequency | 195 |
| 4.61 | Result of TFIDF weight for term | 196 |
| 4.62 | Counting dot product value | 196 |
| 4.63 | Result of TFIDF vector normalize value | 197 |
| 4.64 | Counting of Cosine Similarity | 198 |

4.65 Results of confusion matrix testing in less development group villages 198

4.66 Recommendation expert for the village 200

## LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 2.1 | Diagram for inference fuzzy process | 45 |
| 2.2 | Flowchart of text mining | 50 |
| 3.1 | Research design | 63 |
| 3.2 | Flowchart of fuzzy Delphi methodology | 67 |
| 3.3 | Flowchart of rank reciprocal algorithm | 69 |
| 3.4 | Research design clustering community standard of living index | 77 |
| 3.5 | Flowchart for determining community standard of living index | 79 |
| 3.6 | Flowchart clustering technique using hybrid SOM and Xie-Beny | 81 |
| 3.7 | Flowchart clustering technique using hybrid FCM and Xie-Beny | 82 |
| 3.8 | Flowchart clustering technique hybrid SOM, FCM, and Xie-Beny | 84 |
| 3.9 | The research design for the classification of the developing village | 86 |
| 3.10 | The flowchart for the classification of the developing village | 87 |
| 3.11 | Flowchart classification hybrid Tsukamoto + weighted average | 88 |
| 3.12 | Flowchart classification using Tsukamoto and defuzzy smallest | 89 |
| 3.13 | The trapezium curve | 90 |
| 3.14 | Determine the fields of expertise relevant to the village | 92 |
| 3.15 | Flowchart cosine similarity algorithm in text mining | 93 |
| 3.16 | The process of seeking recommendations in the field of expertise | 94 |
| 4.1 | Evaluation on rank of criteria | 106 |
| 4.2 | Visualization 3-6 Clusters | 149 |
| 4.3 | Chart of optimal cluster using Elbow | 150 |
| 4.4 | Clustering K-Means+ Elbow and clustering SOM + XB + FCM | 155 |
| 4.5 | Fuzzy CSLI-Cluster set | 159 |

| | | |
|---|---|---|
| 4.6 | DVI fuzzy set | 161 |
| 4.7 | Fuzzy set of HDI indicators | 163 |
| 4.8 | Set of Village Development Level indicator | 165 |
| 4.9 | Comparison result classification methods for village development | 193 |
| 4.10 | Result of Village Classification | 194 |
| 4.11 | Chart number field of expertise recommendations for the village | 201 |

## LIST OF ABBREVIATION

| | |
|---|---|
| CSLI | Community Standard of Living Index |
| DVI | Developing Village Index |
| HDI | Human Development Index |
| VDL | Village Development Level |
| GR | Government Regulation |
| UNDP | United Nations Development Programme |
| VF | Village Facilitators |
| VL | Village Law |
| DDAT | Development of Disadvantaged Areas and Transmigration |
| CBS | Central Bureau of Statistics of South Sulawesi |

# CHAPTER 1

# INTRODUCTION

## 1.0 Background of the Study

Discussions on the framework have been started from 2021 - 2027, and regarding to the increasing importance given to rural areas by the European Union, the multiplicity of resources and their positive benefits can be influential factors for sustainable rural development (Abreu et al., 2019). Regarding the implementation of the framework by the European Union, all levels of government can work together to develop rural areas. The author's concern in this study is that the delivery of the latest information about the status of the village is prolonged. Besides that, there has been no source of information explaining the stages of implementing equitable development through village assistance programs. Previous research stated that data collection on village potential was carried out by the Central Bureau of Statistics of South Sulawesi Province every three times in 10 years. However, it did not explain which areas had high or low village potential (Azrahwati et al., 2022). Village potential is a resource or asset owned by a village that can be utilized to advance the village to improve community welfare, covering the social, economic, and regional facilities and infrastructure (Annisa Nur Fathia et al., 2018). Subsequent research stated that the Ministry of Rural and Transmigration had not provided the latest data regarding updating village data in Yogyakarta up to 2022, so the classification results displayed still use data in 2016. However, several studies reported that the statuses of several villages had changed (Salima & Ilham, 2022).

The low participation of the communities and village institutions in planning, implementing, and controlling village development activities, as well as the

preservation of development results, has resulted a lot of natural potential in the village, which is still managed traditionally due to the inability to master technology, relatively low community education and the tendency of the villagers to accept conditions as they are. Generally, the development of rural areas has not been implemented due to internal factors originating from the village, namely the lack of initiative and knowledge from the village community and external factors originating from the government in socializing and providing assistance to rural communities (Andari & Ella, 2019). The alternative that can be performed to realize an independent village is to apply the method of coaching and mentoring or direct assistance to carry out the acceleration of development in the socio-cultural aspect, strengthen the capacity of the village government, and structure government administration (Nur Arifah & Kusumastuti, 2019). Based on the previous research results, it is necessary to have a strategic role of experts as companions who became motivators of initiative, awareness, and participation of rural communities to create independent villages that can act as development subjects starting from the planning, implementation, and evaluation stages of village development which are carried out in a participatory, transparent manner.

In order to implement village development programs aimed at improving the quality of life of people in rural areas, a study applies the Self-Organizing Map and Fuzzy C-Means methodology to classify families in the community based on socio-demographic information obtained through surveys (Morales et al., 2024). This research is directed to investigate family patterns and characteristics at the community level, with the hope that a deeper understanding of these socio-demographic structures can help plan and implement village development programs more effectively. Previous research evaluated the environmental sustainability of rural communities using the

improved TOPSIS method and entropy weighting method (Lin & Hou, 2023), where the study results provide references and guidance for policy makers for ICT-based sustainable development initiatives.

Based on the statement by the Central Bureau of Statistics (Badan Pusat Statistik) of South Sulawesi that experts are needed to assist the community in managing village assets so that all resources can be maximized for the welfare of the community, then the village assistance organizers need information sources that serve as guidelines for the placement of experts so that the implementation of village development equity is more equitable on target. The literature on village development has achieved significant progress in Infrastructure and Basic Services. However, there is a knowledge gap in terms of technology and innovation that can support the improvement of the quality of life in the village. Therefore, this research contributes to the acceleration of equitable distribution of village development through the support and role of information technology.

The strategy for equitable development through the placement of experts targeted in a village has guidelines that are generated through renewal processes, including (1) Clustering the standard of living index of the community in a village using the combination of the Fuzzy C-Means, Self-Organizing Map and Index Xie-Benny methods. (2) Classifying villages based on the developing village, human development, and community standard of living index was carried out by using the Fuzzy Tsukamoto - Smallest of Maximum Defuzzification method to determine the status of underdeveloped villages. (3) Determining the field of expertise of the village assistant that is relevant to the problems in a village through the text mining process using the cosine similarity algorithm.

## 1.1 Problem Statement

In this research, the problem statement that can be identified are :

i. A research by Azrahwati et al., (2022) stated that data collection for identifying the potential of a village is carried out by the Central Bureau of Statistics of South Sulawesi (Badan Pusat Statistik Provinsi Sulawesi Selatan) once in every 3 or 4 years. However, the results of this data collection did not cluster the villages based on the standard of living of the villages which has the ability to serve as guidelines and instructions for improving people's welfare. The implication is that the information reported to the central government cannot describe the real and current conditions which have an impact on the success of government support on villages because of lack of objective, during the decision-making (Astuti, 2017). Supandi et al., (2020) stated that developing villages need information about community needs, where clustering the villages based on characteristics can be used as a basis for development.

ii. The second problem concerns the methodology used by the Central Bureau of Statistics to determine the priority scale in providing support to disadvantaged villages has not used indicators of the actual condition approach, as stated that the indicators used by the Central Bureau of Statistics to determine the priority have not been able to represent the actual conditions of a village (Okfalisa et al., 2021). The implication is that the village development index has not been able to determine activities for the village development programme. Purwanto(2017) stated that one of the weaknesses of the village development index is that it does not involve the community in its assessment and calculation, in fact in an effort to see services that have been carried out by the government, an assessment from the community is needed.

iii. The third problem is on matching of experts to suit the needs of the villages. Currently the skills in the fields of science possessed by village assistants are not relevant to the problems that are currently happens in a village (Dianto, 2019). The findings from this research that the village assistants who were suppose to be professional experts were still lacking in knowledge and mastery of community empowerment and mentoring (Suheryadi, B., Shomad, A., & Djatmiati, 2020). In addition, that the expertise or skills possessed by the village assistant are not in line with the expectations.

## 1.2 Objectives

After outlining the issues related to the strategy for implementing equitable village development, four objectives that have been identified for this research are stated as follows:

i To develop a community standard of living index based on verified criteria collected from selected communities.

ii To cluster the village based on the community standard of living index using hybrid Fuzzy C-Means, Self-Organizing Map, and Xie-Beny.

iii To classify a village based on the Developing Village Index(DVI), Human Development Index(HDI), and Community Standard of Living Index(CSLI) using the Fuzzy Tsukamoto – Smallest of Maximum Defuzzification, relevant expertise in term of his/her area of specialization that suit to the needs of prioritised villages.

iv To map the relevant experts with the priority needs of a village based the input from the villages using cosine similarity text mining process.

## 1.3 Research Questions

The research questions of this study are stated as follows:

i Who is the source of information for the data needed?

ii What are the criteria used to determine the community standard of living index?

iii What is the best clustering technique to cluster the villages based on its community standard of living index?

iv What are the input criteria to determine the best clustering technique?

v What are the input indicator to determine the classification the villages?

vi What are the rules to classification the villages?

vii What are the input data to determine the field of expertise needed by a village?

viii What are the indicators to validate the right placement of experts in a village?

**Table 1.1: Relationship problem statement, research object, research question**

| Problem Statement | Research Objectives | Research Questions |
|---|---|---|
| i. Village potential data collection is carried out by the Central Bureau of Statistics of South Sulawesi (Balai Pusat Statistik Provinsi Sulawesi Selatan) every three times in 10 years. However, the data collection results did not explain village clustering, which can serve as guidelines and instructions for improving people's welfare. | i. To develop a community standard of living index based on verified criteria collected from selected communities.<br>ii. To cluster the village based on the community standard of living index using hybrid Fuzzy C-Means, Self-Organizing Map, and Xie-Beny | i. Who is the source of information for the data needed?<br>ii. What are the criteria used to determine the community standard of living index?<br>iii. What is the best clustering technique to cluster the villages based on its ccommunity standard of living index?<br>iv. What are the criteria input to determine the best clustering technique? |

| | | |
|---|---|---|
| ii. The second problem concerns the methodology used by the Central Bureau of Statistics to determine the priority scale in providing support to disadvantaged villages has not used indicators of the actual condition approach, however the indicators used by Central Statistical Agency to determine the priority have not been able to represent the actual conditions of a village. | iii. To classify a village based on the Developing Village Index(DVI), Human Development Index(HDI), and Community Standard of Living Index(CSLI) using the Fuzzy Tsukamoto – Smallest of Maximum Defuzzification, relevant expertise in term of his/her area of spesialization that suit to the needs of a prioritised villages. | v. What are the input indicator to determine the classification the villages?<br>vi. What are the rules to classification the villages? |
| iii. Matching of experts to suit the needs of the villages, currently the skills in the fields of science possessed by village assistants are not relevant to the problems that currently happen in a village. | iv. To map the relevant experts with the priority needs of a village based on the input from the villages using the cosine similarity text mining process | vii. What are the input data to determine the field of expertise needed by a village?<br>viii. What indicators validate the right placement of experts in a village? |

### 1.4 Justifications and Significance of the Study

After interviews with authorized sources, it can be said that this research is fundamental because it is imperative to support the implementation of equitable village development. Based on village ranking status in 2020 submitted by the Ministry of