Applying Clustering Algorithm on Poverty Analysis in a Community in the Philippines

Michelle P. Repollo
School of Information Technology
Mapua University
Manila, Philippines
michelleprepollo@gmail.com

Rex Aurelius C. Robielos
School of Industrial Engineering and Engineering Management
Mapua University
Manila, Philippines
racrobielos@yahoo.com

Abstract

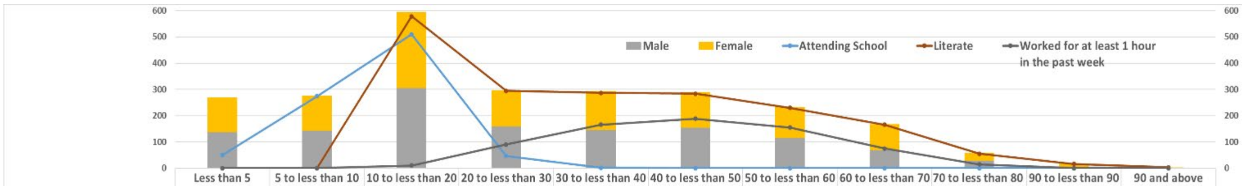
In addressing poverty, proper poverty alleviation programs must be identified. Analyzing the multidimensional aspects of poverty will lead to data-driven decision making. This study aims to analyze poverty conditions in a community in the Philippines using data mining and clustering algorithm. The model resulted in clusters described in this paper as Stable, Critical, and At-Risk clusters. Among the clusters, the Critical cluster has the highest incidents of poor conditions. In the cluster analysis, poverty alleviation programs where the household clusters should be prioritized were suggested based on the poor conditions in each cluster. The clusters analysis is intended to provide insights for use in the community's planning and program implementation.

Keywords
Data Mining, Machine Learning, Poverty Research

1. Introduction
Addressing poverty is a significant task of several countries. This makes research on poverty alleviation important and necessary (Niu et al., 2020). The research on poverty created awareness of its multidimensional nature, that is, focusing on factors other than income. (Shuhong et al., 2019). Lack of access to basic needs like education, food, housing, security, and basic health care indicates poor conditions (Talingdan, 2019). In reducing poverty, economic growth must be sustained and inclusive (NEDA, 2017). Data mining plays an essential role in explaining patterns in data and serves a decision support tool as complexity develops and the amount of information available increases (Naviamos & Niguidula, 2020). Many organizations have adopted data mining as a decision support tool (Talingdan, 2019). The results of this study may be used by government and non-government organizations in identifying and implementing poverty alleviation programs in a community in the Philippines.

The subject of this study is a community in the Philippines where the average household size is four (4) and the average monthly income per household is PhP6,238. This is below the 2015 poverty threshold at Php9,452 per month for a family of five (5) indicated in the *Philippine Statistics Authority's Special Release* dated December 5, 2019.

Six hundred thirty-one household members are employed receiving salaries and wages while some 116 received cash receipts, support, assistance and relief from domestic sources. The data with 2,502 respondents who are members of the 606 households in the community revealed that 90% of residents aged 5 to 20 years old are in school. 98% of those aged 10 and above can read and write a simple message in any language or dialect while 51% of residents who are 20 to 80 years old worked for at least an hour in the past week. Figure 1 shows the demographics, literacy rate, educational attainment, and employment in the community per age group classification.



	Less than 5	5 to less than 10	10 to less than 20	20 to less than 30	30 to less than 40	40 to less than 50	50 to less than 60	60 to less than 70	70 to less than 80	90 to less than 90	90 and above
Female	133	133	290	139	148	136	118	98	31	14	2
Male	137	143	305	159	147	154	115	69	27	3	1
Attending School	50	274	510	47	2	0	0	0	0	0	0
Literate	0	0	578	295	287	284	230	166	55	16	3
Worked for at least 1 hour in the past week	0	0	10	90	166	189	155	75	15	0	0

Figure 1. Demographics, Literacy, Education and Employment

In terms of living conditions, 445 out of the 606 households have electricity provided by an electric company, 10 of which also have a generator, solar, or battery source. About 69% of the households have owner/owner-like possession of the housing unit that they occupy. Only 34 households were beneficiaries of the various social programs covered by the survey.

1.1 Objectives

Given the wide array of factors and data discussed, this study will focus on the different aspects of poverty and in analyzing data patterns to be able to come up with household clusters and identify priority programs for the clusters. By applying data mining and machine learning techniques to the data, this study will answer the questions (a) What are the aspects of poverty that can be used in cluster analysis? (b) What should be the focus areas for each household cluster in the community? (c) What programs can be implemented for each cluster in the community?

2. Literature Review

The diverse nature of poverty makes the handling of the problem challenging. Poverty has to be directly or indirectly addressed from the root of the real problem (Verma, 2020). The "*Philippine Development Plan (PDP) 2017-2022"* presents the foundation for more inclusive growth by lowering poverty incidence in rural areas (NEDA, 2017). Communities can promote growth and contribute to economic development at various levels of the society. By identifying and addressing hurdles in development and improving growth potential, the quality of life in communities may be improved (Senetra & Szarek-Iwaniuk, 2020). Proper poverty alleviation programs given to the community can reduce or minimize poverty (Talingdan, 2019).

Several indices have been proposed that incorporates the economic dimensions with health conditions, education, and living standards. Formulating poverty reduction-related policies requires timely and accurate research on affected areas and individuals. Poverty originally refers to the need for basic necessities when their income level is insufficient to support these needs. Poverty can be measured by various methods and data sources (Niu et al., 2020) and depends on many dimensions other than income (Bárcena-Martín et al., 2020). Although poverty has many different dimensions, not all policies follow an integrated approach in poverty reduction. Different aspects of poverty are addressed separately focusing mainly on income (Gamboa et al., 2020). It is important that multidimensional poverty is measured to be able to address poverty effectively (Bárcena-Martín et al., 2020).

Education is essential in developing human capital, enhances productivity, and contributes to economic development in a community (Afzal et al., 2011). When the education level rises, poverty decreases (Ingutia et al., 2020). Education allows young members of the community to understand and cope with the socio-economic growth in a country. Asian countries that implemented educational reforms have achieved economic growth in the past (Afzal et al., 2011). Reforms in the education system enables meeting the rapidly changing needs of industries and reduces income inequalities.

When students develop employability skills, it reduces the risk of poverty (Brown & James, 2020).
Low income results to poor housing conditions, overcrowding, and lack of access to basic needs (Doe et al., 2020). Poor living conditions affect the ability to deliver productive work. Housing is considered a social risk since it contributes to health hazards, the risk of prolonged absenteeism, and the probability of falling into extreme poverty (Frota, 2008).

The government often implements social protection programs when it aims to reduce poverty (Bhuyan et al., 2020). Poor countries are increasing implementation of social protection programs (Frota, 2008). The role of social protection is to achieve inclusiveness. Properly implemented social protection programs, therefore, promote inclusive economic growth (Bhuyan et al., 2020) and have a direct economic relevance. Interventions that provide income guarantees and social security plays a part in countering social exclusion and poverty. Policies should not only lessen the social risks but also create sustainable ways to reduce poor conditions. Members of the community should be able to live decently by having sufficient basic social protection and decent work. As such, intervention should also be able to generate jobs in the community (Frota, 2008).

Census data is considered as the main data source for poverty research. Complicated and voluminous data are handled using machine learning algorithms in recent studies (Niu et al., 2020). Decision makers such as the local government units and other government agencies need a decision support tool (Sano & Nindito, 2016). Survey data collected by the government should be processed further to support their decision making (Talingdan, 2019).

Data mining leads to discovering new patterns and trends in a database and building predictive models out of it. It involves the use of tools and techniques that can be applied to discover hidden patterns that can be used for decision making (Sano & Nindito, 2016).

A tool that is useful for explorative data analysis is cluster analysis that is used to organize data with multiple variables (Senetra & Szarek-Iwaniuk, 2020). In Gamboa et al., 2020, the household is used as the unit of analysis. Cluster analysis process was also used to define the household typologies where attributes are used to classify the households (Gamboa et al.). The goal in cluster analysis is to group objects to ensure that they belong to the same group that is most similar, but also most different from the objects in the other groups (Senetra & Szarek-Iwaniuk, 2020). It involves grouping data into groups according to their similarity (Verma, 2020).

Formulating recommendations for each multidimensional group involves dealing with several attributes that must be considered to arrive at the best alternative (Bárcena-Martín et al., 2020).

3. Methods

In K-means clustering, similar objects are grouped into clusters that reveal patterns and predict the basic structure of the data. The algorithm groups the data or objects into k sets by minimizing the distances between data points and the nearest center of the cluster. Similarity is defined by the distance to the cluster center (Aytaç, 2020). Each object is assigned into clusters that are most similar using the Euclidean distance between the object and the center cluster or the centroid as basis (Sano & Nindito, 2016). “K-Means” is an unsupervised learning used for clustering that separates observations into “k” clusters (Talingdan, 2019).

Data normalization technique is applied to transform the numeric data into a common scale and improve the performance of the clustering algorithm (Moreno, 2020). In this study, Z-*Score method* is applied using *Python* which calculates the standard score of the data point. To compute for the standard score, the mean of the attribute is subtracted from the observation. The difference is then divided by the standard deviation as shown Figure 2. The maximum scaled values or scores per attribute as a result of the data normalization are indicated on Table 1 which were used in the K-means clustering algorithm.

$$Z = \frac{x - \mu}{\sigma}$$

Where Z = standard score, x = observation, μ= mean, and σ = standard deviation

Figure 2. Z-score

The results of the K-means clustering depend on the number of centroids defined as the k number of groups in the model. A way to determine the number of clusters to be used in clustering is the "Elbow Method". The "Elbow Method" uses the square of the distance between the sample points in each cluster and the centroid of the cluster to give a series of k values. Iteration is performed over the k-value where the Sum of Squared Errors (SSE) is computed and used as an indicator. SSE will decline rapidly when the number of groups is about to approach the number of real clusters. The inflection point in the K-SSE curve is the optimal k value or number of clusters (Yuan, 2019).

Using Python, the curve in Figure 3 was plotted by conducting iteration over the k-value and computing for the sum of squared errors. The optimal k number of groups determined is three (3). This is illustrated in the K-SSE curve.

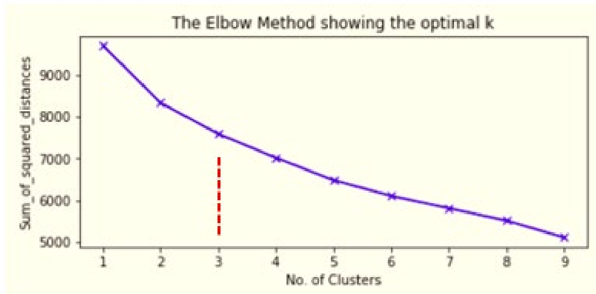


Figure 3. Using Elbow Method in determining the optimal k

The rapid decline or inflection in the curve falls at '3' in the x-axis or the number of clusters. As such, this shall be the optimal number of clusters and used as the parameter in the number of centroids in the Microsoft Azure Machine Learning Studio's K-Means++ algorithm.

Unsupervised learning techniques, such as K-Means Clustering, depend on the algorithm to detect all patterns from specific important criteria. These result to clusters and segments that will be analyzed to discover important insights (Verma, 2020).

In clustering, set of data objects are divided into subsets or cluster such that the objects within the groups are similar to each other but very different from the objects in the other clusters (Sano & Nindito, 2016). Cluster analysis is a method for numerical classification involving multivariate statistical analysis that determines similarity between samples by measuring distances between data using Euclidean distance (Dong Pengyu et. al, 2019). Figure 4 illustrates the formula for Euclidean distance and k-means algorithm. Using the data set, cluster assignments are determined for each household based on the distances of the data points from the cluster centers or centroid. Upon determining the cluster assignments for the households, the attributes will be analyzed to be able to describe and determine the characteristics of each cluster.

Euclidean Distance Formula:

$$d(p,q) = \sqrt{\sum_{i=1}^{i} (p_i - q_i)^2}$$

Where:

- *d (p, q)* = distance between p and q vectors
- p_i and q_i = correspond to the i^{th} dimension of vectors p and q

Result: A set of *k* clusters
Steps:

1. Arbitrarily choose *k* objects from the data set as the initial cluster centers.
2. Re-assign each object to the cluster to which the object is the most similar based on the mean value of the objects in the cluster using Euclidean distance.
3. Update the cluster means, that is, calculate the mean value of the objects for each cluster.
4. Repeat until no changes.

Figure 4. Euclidean Distance and K-means Algorithm
(Sources: Sano & Nindito, 2016 and Aytaç, 2020).)

4. Data Collection

Data mining involves data cleansing, data integration, and selection and transformation of data (Sano & Nindito, 2016). This study covers attributes pertaining to the socio-economic indicators in the census data. Figure 5 illustrates the steps in data mining and cluster analysis used in this study.

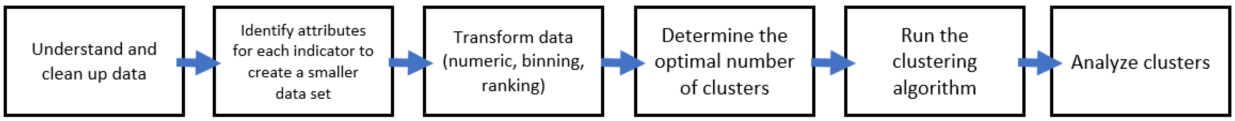


Figure 5. Data mining and cluster analysis

The detection of data issue conducted using Python and Microsoft Excel involved checking for data types, duplicate entries, and null values. Variables that are related to conditional questions were excluded from the checking of null values. Feature reduction was applied to eliminate non-significant and redundant variables (e.g., subtotals of income, free text responses, etc.). In preparing the data set to be used in clustering as illustrated in Table 1, data transformation and feature engineering methods such as binning, summation, feature creation, and averaging were used.

Table 1. Attributes pertaining to each indicator which are used in the study

Indicator	Attributes/Variables	Description	Scaled Values (Max)	Feature Engineering Applied
Income and Employment	Total Income	Total household income	10.461388	Converted (in 10,000s)
	Employment Rate	Number of employed household members who are older than 18 years old over number of members who are older than 18 years old and not studying	2.627896	Ratio
Living Conditions	House/lot tenure	Number of household members who are considered as informal settlers	4.544717	Count
	Household size	Number of household members	5.842435	None
	Number of aged 18 and below	Number of household members who are aged 18 and below	3.851493	Count
	Number of married	Number of household members who are married	4.537215	Count
	Water source	Number of household members who do not have their own tube/piped water source/system	1.898924	Count
	Access to electricity	Number of household members who do not have access to electricity	7.034049	Count
	Sanitation	Number of household members who do not have their own water-sealed, septic tank/other depository	7.01949	Count
Education	Educational Attainment	Number of household members between 5 to 22 years old who are not attending school and 22 and above who are not senior high school graduate	5.690292	Count
	Illiteracy Rate	Number of household members who are 5 years old and above but cannot read or write over the number of household members who are 5 years old and above	10.373035	Ratio
Security and Participation to Social Programs	Number of SSS/GSIS Member	Number of household members who are 22 years old and above but are not members of SSS/GSIS	4.589636	Ratio
	Recipient of CCT	Number of members whose households did not receive conditional cash transfer in the last 12 months	5.50928	Count
	Victim of crimes	No incidents of victims recorded	0	Count of victim of crimes per HH
Health and Nutrition	Number of aged 60 and above	Number of household members who are 60 years old and above	5.290316	Count
	Number of Pregnant	Number of household members who are pregnant	12.694407	Count
	Number of PWD	Number of household members who are PWD	5.558303	Count

5. Results and Discussion

5.1 Numerical Results

The clustering algorithm divided the households in the community into three (3) clusters. Using the Euclidean distance as the parameter in measuring distances between data points and the centroid or cluster centers, the average distances computed by the model are summarized in Table 2.

Table 2. Results of Clustering Algorithm

	Number of Households	Average Distance to Cluster Center	Average Distance to Other Center
Cluster 1	35	4.404216	6.39958
Cluster 2	257	3.203088	4.246216
Cluster 3	285	3.099009	4.085241
	606		

A small average distance to the cluster center indicates that the observations in the cluster has less variability and are more similar. On the other hand, the average distance to other center indicates the distance of the centroids/clusters from each other. A larger distance indicates greater variability among the different clusters.

The resulting clusters were analyzed based on the presence of poor conditions in the households belonging to such. The number of incidents of poor conditions in the households were determined per attribute. Table 3 summarizes the clusters and their corresponding levels of poor conditions based on the significant attributes. The attributes where each cluster attained high incidents of poor conditions were highlighted accordingly.

Cluster 1 is comprised of households with the lowest incidents of unemployment among the three clusters having only 9% of households without an employed member. Their main areas of concern are the living conditions such as house/lot tenure, water source, access to electricity, and sanitation. 91% of households in Cluster 1 are literate and only 28% do not have graduates of at least senior high school. Since most households in this cluster are employed or have better chances of being employed given their literacy and educational level, this cluster may be described as the "Stable" cluster.

Cluster 2 is comprised of households with the highest incidents of poor conditions in almost all attributes. 83% of households in this cluster earns below the poverty threshold. However, despite having the highest number of households falling below the poverty threshold, 283 household members in Cluster 2 did not benefit from conditional cash transfer in the last 12 months. Around a third of the households belonging to Cluster 2 do not have their own water-sealed septic tank or other depository. Around 66% of households in Cluster 2 do not have members who are covered by SSS or GSIS. This group has the highest number of elderly and PWD. Since households in this cluster ranked highest in the presence of poor conditions in almost all attributes, this cluster may be described as the "Critical" cluster.

Cluster 3 is comprised of households with the highest number of members who are married, below 18 years old, and illiterate. This group also has the highest number of households with more than 5 members. These households ranked second to highest in the number of incidents of poor conditions in almost all attributes even in the number of households without an SSS/GSIS member and households that did not receive conditional cash transfer in the last 12 months despite having 78% of the households earning below the poverty threshold. This cluster has the potential of falling into deeper poverty if current poor conditions are not addressed considering their large household size, young population, and low literacy rate. Since this cluster ranked second highest in the presence of poor conditions in most attributes and has members whose conditions must be monitored and who need more attention so as not to cause bigger poverty issues in the future, this cluster may be described as the "At-Risk" cluster.

Vulnerable residents such as persons with disability and pregnant women are almost equally distributed between Clusters 2 and 3. Security is not an issue in the community. Household members did not fall victim to any form of crime in the last 12 months. Concerns on income, water source, number of SSS/GSIS members and conditional cash transfer are consistently high among the three clusters. Table 3 shows the profile of clusters in terms of poor conditions present among the households.

Table 3. Profile of household clusters (HH) highlighting the levels of poor conditions

	% of Households	6%	47%	47%	
	Number of Households	35	286	285	
Aspect of Poverty	**Attributes/Variables**	**Stable Cluster 1**	**Critical Cluster 2**	**At-Risk Cluster 3**	**Description**
Income and Employment	Total Income	28	236	222	Below poverty threshold
	Employment Rate	3	75	15	No employed HH member
Living Conditions	House/lot tenure	31	97	53	Without rent and/or consent
	Household size	25	0	108	More than 5 HH members
	Number of aged 18 and below	137	150	810	Out of 2,502 respondents
	Number of married	76	313	531	Out of 2,502 respondents
	Water source	31	215	217	Without own faucet or tubed/piped deep well
	Access to electricity	30	82	51	No access to electricity
	Sanitation	32	86	44	Without own water-sealed septic tank or other depository
Education	Educational Attainment	10	124	68	Without a graduate of at least Senior High School
	Illiteracy Rate	3	9	29	Has Below 50% literacy rate
Security and Participation to Social Programs	Number of SSS/GSIS Member	21	188	155	No SSS/GSIS member
	Recipient of CCT	31	283	261	Did not receive CCT in the last 12 months
	Victim of crimes	0	0	0	A HH member became a victim of crime in the last 12 months
Health and Nutrition	Number of aged 60 and above	16	181	48	Out of 2,502 respondents
	Number of Pregnant	2	4	7	Out of 2,502 respondents
	Number of PWD	1	11	7	Out of 2,502 respondents

5.2 Graphical Results

Given the profiles of the household clusters in the community, the local government and various non-government organizations who plan to conduct poverty alleviation programs may prioritize and scope their projects according to the needs of the clusters. They may target household clusters who are in most need of assistance, which, based on the results of this study, is the Critical Cluster (2). Figure 6 shows the graphical representation comparing the number of poor incidents per attribute that is experienced by households in the clusters.

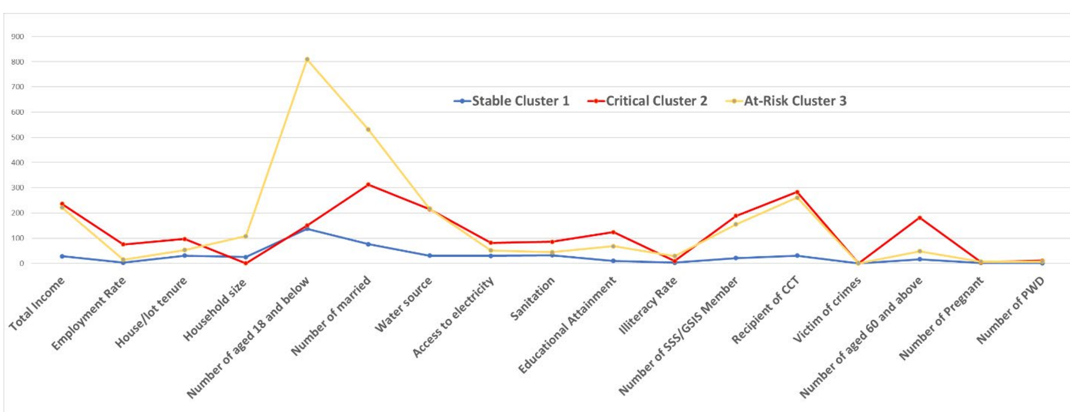


Figure 6. Graphical representation of the comparison of clusters

5.3 Proposed Improvements

In addressing poor conditions in the community, suitable programs can be identified for each household cluster using the results of the cluster analysis. Table 4 lists down the attributes where the households have high levels of poor conditions which could be considered as the main focus areas for each cluster. Cluster 1 has high incidents of poor conditions in terms of income and physical living conditions. Clusters 2 and 3 have high concerns of poor incidents mostly on income, employment, education, and literacy. Both have high incidents of married household members. All clusters have concerns on water source. Clusters 1 and 2 have high incidents of households with concerns on house/lot tenure, access to electricity and water source, and sanitation. Cluster 2 has the highest number of households without SSS/GSIS membership and who did not receive conditional cash transfer in the last 12 months. Health and nutrition are also areas of concern for Cluster 2 since they have high numbers of elderly who are aged 60 and above as well as PWDs.

Table 4. Attributes where each household has high levels of poor conditions

Aspect of Poverty	Stable Cluster (1)	Critical Cluster (2)	At-Risk Cluster (3)
Income and Employment	Total income	Total income	
		Employment Rate	
Living Conditions	House/lot tenure	House/lot tenure	
			Household size
			Number of Married
			Number of aged 18 and below
	Water source	Water source	Water source
	Access to electricity	Access to electricity	
	Sanitation	Sanitation	
Education		Highest educational attainment	
			Literacy Rate
Security and Participation to Social Programs		Number of SSS/GSIS Member	
	Recipient of CCT	Recipient of CCT	
Health and Nutrition		Number of aged 60 and above	
			Number of Pregnant
		Number of PWD	

Given the profiles of the household clusters in the community, the local government and various non-government organizations who plan to conduct poverty alleviation programs may prioritize and scope their projects according to the needs of the clusters. They may target household clusters that are in most need of assistance, that is, Critical Cluster (2).

The Stable Cluster (1) must focus on improving living conditions for the members of the households. Income augmentation and livelihood programs may be promoted to improve their levels of income. Subsidies may be provided to be able to achieve housing security and to help them gain access to basic needs such as water source, electricity, and proper sanitation.

Although this is the smallest cluster and most households in this cluster are employed or have better chance of employment compared to households in other clusters, the community still needs to address their concerns in order to achieve inclusive economic growth.

The Critical Cluster (2) is the cluster with the highest number of areas of concern. As such, they must be prioritized in the coverage and implementation of poverty alleviation programs. Aside from health and nutrition programs for the elderly, they must also focus on improving the employment rate, conducting skills training, and developing livelihood programs for the community. Low educational attainment may be addressed by scholarship programs and skills training programs that will create earning opportunities for residents. Poor living conditions may also be addressed by providing housing security through relocation of informal settlers (i.e., those who live in house/lot without paying rent and without consent from owners).

The local government may also explore alternative/more affordable sources of electricity, improve sanitation by constructing sewers for households, and improve water systems to meet the demands of the increasing population in rural areas.

Participation in social programs is very low among households in the Critical Cluster despite having the highest number of households falling below the poverty threshold. This may be improved by promoting voluntary membership to social security programs for household members who are earning from informal sources. The local government may also review if all households that are eligible to conditional cash transfer programs are given opportunity to participate in such.

The At-Risk Cluster (3) ranked second in terms of poor conditions in most attributes but current poor conditions must be monitored closely to ensure that interventions are implemented to be able to secure a better future for the young members of the households and prevent the households' poor conditions into escalating. Informal learning methods such as writing and reading workshops may also be adopted to address the cluster's low literacy rate. Considering their large household size and having the highest number of married members, the local government may focus on providing family planning programs to households in Cluster 3. They also have the highest number of children aged 18 and below who may benefit from early childhood care & development programs and comprehensive sexuality education programs for young adults.

Table 5 lists down the recommended or possible programs that may be prioritized in each household cluster.

Table 5. Recommended programs where each household may be prioritized

Household Clusters	Attributes of High Concern	Applicable progams where the Cluster may be prioritized	Notes
Stable Cluster (1)	Income	Income augmentation and livelihood programs	Although this cluster has a low priority compared to the other clusters, their poor conditions must still be addressed to achieve inclusive economic growth.
	House/lot tenure	Relocation; Housing programs and payment options	
	Water Source	Improvement of water systems to meet the demands of increasing population in rural areas	
	Sanitation	Improvement of sanitation facilities; construction of sewer or safe enclosure for households	
	Conditional Cash Transfer	Review/check if all eligible households are being covered by or are given the opportunity to participate in CCT programs	
Critical Cluster (2)	Income	Income augmentation and livelihood programs	Must be prioritized in the coverage and implementation of poverty alleviation programs.
	Employment rate	Skills training programs; job fairs	
	House/lot tenure	Relocation; housing programs and payment options	
	Water Source	Improvement of water systems to meet the demands of increasing population in rural areas	
	Access to Electricity	Explore alternative/more affordable sources of electricity	
	Sanitation	Improvement of sanitation facilities; construction of sewer or safe enclosure for households	
	Educational Attainment	Scholarship programs	
	SSS/GSIS Coverage	Promote voluntary membership for individuals who are earning from informal sources	
	Conditional Cash Transfer	Review/check if all eligible households are being covered by or are given the opportunity to participate in CCT programs	
	60 and above Number of PWD	Health and nutrition-related programs; Programs protecting the vulnerable individuals; Employment programs for PWD	
At-Risk Cluster (3)	Household Size	Family planning programs	Has the potential to fall into deeper poverty if current poor conditions are not addressed considering their large household size and lack of educational attainment.
	Number of 18 and below	Early childhood care and development programs; Comprehensive sexuality education programs for young adults	
	Number of married	Family planning and marriage counselling programs	
	Water Source	Improvement of water systems to meet the demands of increasing population in rural areas	
	Literacy Rate	Informal methods of learning such as writing or reading workshops	
	Number of pregnant	Health and nutrition-related programs; Programs protecting the vulnerable individualsD	

In further studies, data mining and clustering algorithm may be applied on census data for other areas or communities. Recommendations may differ based on the focus areas and household cluster profiles in these communities. Data on the available resources in the community or specific advocacies of organizations may also be used as considerations as to which programs should be prioritized.

6. Conclusion

Using clustering algorithm, poor living conditions of households in the community were analyzed and prioritization in implementing socio-economic programs in each household cluster was recommended in this study. Aside from income, key variables from the data were identified to address multidimensional aspects of poverty and develop comprehensive strategies for the community.

Data mining and cluster analysis were used as a decision support tool to obtain data understanding and describe the profiles of the household clusters in the community in terms of the significant socio-economic indicators present in the data.

To organize the complicated data relationships presented in census data, cluster analysis was used in this study. The households in the community were grouped into three (3) clusters described as the Stable Cluster, Critical Cluster, and At-Risk Cluster. Focus areas were identified based on the cluster profiles. These were used as basis in recommending programs that would address the most pressing poor living conditions in the clusters.

References

Afzal, M., Rehman, H. U., Farooq, M. S., & Sarwar, K. (2011). Education and economic growth in Pakistan: A cointegration and causality analysis. *International Journal of Educational Research*, *50*(5–6), 321–335. https://doi.org/10.1016/j.ijer.2011.10.004

Aytaç, E. (2020). Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey. *International Soil and Water Conservation Research*, *8*(3), 321–331. https://doi.org/10.1016/j.iswcr.2020.05.002

Bárcena-Martín, E., Pérez-Moreno, S., & Rodríguez-Díaz, B. (2020). Rethinking multidimensional poverty through a multi-criteria analysis. *Economic Modelling*, *91*(May), 313–325. https://doi.org/10.1016/j.econmod.2020.06.007

Bhuyan, B., Sahoo, B. K., & Suar, D. (2020). Nutritional status, poverty, and relative deprivation among socio-economic and gender groups in India: Is the growth inclusive? *World Development Perspectives*, *18*(January), 100180. https://doi.org/10.1016/j.wdp.2020.100180

Brown, P., & James, D. (2020). Educational expansion, poverty reduction and social mobility: Reframing the debate. *International Journal of Educational Research*, *100*(March), 1–9. https://doi.org/10.1016/j.ijer.2020.101537

Doe, B., Peprah, C., & Chidziwisano, J. R. (2020). Sustainability of slum upgrading interventions: Perception of low-income households in Malawi and Ghana. *Cities*, *107*(September). https://doi.org/10.1016/j.cities.2020.102946

Dong Pengyu, Wang Hongwei, C. Y. (2019). *Improved System Cluster Analysis Based on Rough Set and General Euclidean Distance*.

Frota, L. (2008). Securing decent work and living conditions in low-income urban settlements by linking social protection and local development: A review of case studies. *Habitat International*, *32*(2), 203–222. https://doi.org/10.1016/j.habitatint.2007.08.016

Gamboa, G., Mingorría, S., & Scheidel, A. (2020). The meaning of poverty matters: Trade-offs in poverty reduction programmes. *Ecological Economics*, *169*(May 2019). https://doi.org/10.1016/j.ecolecon.2019.106450

Ingutia, R., Rezitis, A. N., & Sumelius, J. (2020). Child poverty, status of rural women and education in sub Saharan Africa. *Children and Youth Services Review*, *111*(February). https://doi.org/10.1016/j.childyouth.2020.104869

Moreno, Amanda Iglesias (2020). *Data Normalization with Pandas and Scikit-Learn*

National Economic and Development Authority. (2017). *Philippine Development Plan 2017-2022*.

Naviamos, M. P., & Niguidula, J. D. (2020). A study on determining household poverty status: SVM based classification model. *ACM International Conference Proceeding Series*, 79–84. https://doi.org/10.1145/3378936.3378969

Niu, T., Chen, Y., & Yuan, Y. (2020). Measuring urban poverty using multi-source data and a random forest algorithm: A case study in Guangzhou. *Sustainable Cities and Society*, *54*(June 2019), 1–12. https://doi.org/10.1016/j.scs.2020.102014

Philippine Statistics Authority. (2017). *Philippine Statistics Authority, 2015 Census of Population, Report No. 2 – Demographic and Socioeconomic Characteristics* (p. 157). https://psa.gov.ph/sites/default/files/03_Aurora.pdf

Sano, A. V. D., & Nindito, H. (2016). Application of K-Means Algorithm for Cluster Analysis on Poverty of Provinces in Indonesia. *ComTech: Computer, Mathematics and Engineering Applications*, *7*(2), 141. https://doi.org/10.21512/comtech.v7i2.2254

Senetra, A., & Szarek-Iwaniuk, P. (2020). Socio-economic development of small towns in the Polish Cittaslow Network — A case study. *Cities*, *103*(May 2019), 102758. https://doi.org/10.1016/j.cities.2020.102758

Shuhong, L., Li, J., & Tao, H. (2019). Logistic model-based measurement and analysis of factors affecting poverty in underdeveloped areas. *ACM International Conference Proceeding Series*, 39–46. https://doi.org/10.1145/3324033.3324050

Talingdan, J. A. (2019). Data mining using clustering algorithm as tool for poverty analysis. *ACM International Conference Proceeding Series*, *Part F1479*, 56–59. https://doi.org/10.1145/3316615.3316672

Verma, P. K. (2020). Application of K-Means Algorithm to Mapping Poverty Outline by Province in India. *International Journal of Recent Technology and Engineering*, *8*(6), 1045–1049. https://doi.org/10.35940/ijrte.f7357.038620

Yuan, C. (2019). *Research on K-Value Selection Method of K-Means Clustering Algorithm*.

Biographies

Michelle Repollo is currently taking up Master in Business Analytics in the School of Information Technology in Mapua University. She is a graduate of Bachelor of Science in Business Administration from the University of the Philippines-Diliman. She is a compensation, benefits, and information systems specialist in a leading fast-moving consumer goods company in the Philippines.

Prof. Rex Aurelius C. Robielos is the Dean of the School of Industrial Engineering and Engineering Management. He is currently taking his PhD in Industrial Management from the National Taiwan University of Science and Technology in Taiwan. He graduated with a diploma and Master of Science in Industrial Engineering from the University of the Philippines-Diliman and his Bachelor of Science in Applied Mathematics (Operations Research) from the University of the Philippines-Los Baños. Prof. Robielos is a director of the Philippine Institute of Industrial Engineers, Treasurer of the Operations Research Society of the Philippines and Secretary of the Human Factors and Ergonomics Society-Philippines.