A Cannonical Correlation Analysis between Regional Development and the Behavior of Rural Populations in West Java, Indonesia

Reyza Farzan Rahmatsyah
School of Computer Science
Bina Nusantara University
DKI Jakarta, Indonesia
reyza.rahmatsyah@binus.ac.id

Alisha Zahra Saadiya
School of Computer Science
Bina Nusantara University
DKI Jakarta, Indonesia
alisha.saadiya@binus.ac.id

Diva Nabila Henryka
School of Computer Science
Bina Nusantara University
DKI Jakarta, Indonesia
diva.henryka@binus.ac.id

***Abstract*—**

Keywords—Canonical Correlation Analysis, Rural Development, Rural Characteristics,

I. INTRODUCTION

According to the National Census conducted by the Bureau of Statistics of Indonesia in 2022, 120 million people are currently living in rural areas. This number is approximately 44 percent of Indonesia total population. With urban areas continuing to be the intention of large scale developments, rural areas are in risk of getting left behind. The disparity between these areas are further exasperated by the amount of people currently living in poverty in said areas.

FIGURE I. PEOPLE LIVING IN POVERTY

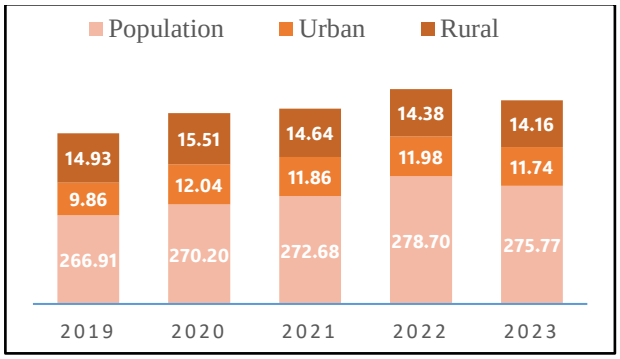

Fig. 1. A stacked bar graph depicting the difference between the total population and the total number of people living in poverty in both urban and rural areas from 2019-2023.

Fig. 1 indicates that although a majority of people in Indonesia live in urban areas, where job opportunities, education, and healthcare is more accessible, poverty is still a predominantly rural occurrence [1]. This disparity suggests that strictly focusing on urbanization is not an effective policy to counteract poverty. Urbanization further isolates the rural communities by limiting access to services aforementioned above and limits the opportunity of equal development between rural and urban areas[2, 1].

The disparity between rural and urban areas lead to an examination of each areas respective characteristics. Rural or colloquially known more as ‘left behind’ areas, are specific areas that for an undefined reason, ceases to be the object for socio-economic development. Once ‘left behind’ areas lose importance, it becomes difficult to reverse its status [2]. The disparity is then further aggravated by the different specific developmental needs of rural areas [3].

Rural areas are also characterized by their residents. Generally, the economic status of rural residents is lower than their urban counterparts [4]. Urban residents are more likely to be better educated, better paid, and taken care of [5]. This leads to rural populations not only being economically inferior, but also underprivileged in socio-economic programs [1]. Rural areas also make the bulk of domestic agricultural produce and expecting a shift from agriculture towards a more broad economic structure is unlikely to happen [6].Therefore, Urbanization and migration is the most likely solution for rural residents to obtain the necessities they have been deprived of. Contrary to this, agricultural products continue to be a valuable contributor to economics [7]. Evident in the Indonesian Oil Palm Statistics 2021 publication by Central Bureau of Statistics, the agriculture sector holds the second place of Indonesian’s GDP contributor at 13.28 percent [8].

In Kenya, education is essential for improving income and living conditions in informal settlements. Large household sizes tend to reduce long-term investment and hinder family education. To address this challenge, research by Lusweti, Job, and Annette Okoth suggested educating families about wise spending, promoting income-generating activities, and improving access to education, especially through community initiatives [9]. While in Nigeria, a significant positive correlation between poverty and literacy levels suggests that increased literacy can effectively reduce poverty. Therefore, Ebenezer and Ogunsakin, in their research, encourage policies and programs that enhance access to and use of education as a means of reducing poverty, emphasizing the important role of reading skills in improving quality of life and socio-economic participation [10]. A study conducted by Mkondiwa, Jumbe, and Wiyo about rural Malawi proved that there is a strong link between poverty and lack of water access in rural areas. It also shows a huge inequalities of both aspects between rural, semi-urban, and urban areas [11]. A study using a 2017 data to investigate the connection between economic growth and social inclusion conducted by Şanlı shows that GDP per capita is the most important for economic growth as median household income to social inclusion. The analysis also showed other factors, such as health and labor productivity can influence both aspects [12].

To narrow down the scope of the study, data used onward will be focused on specifically West Java, Indonesia. West Java is third highest province in Indonesia when it comes to the percentage of people living in poverty [13]. West Java also has the advantage of a considerable amount of natural resources in the agriculture, fisheries, livestock, and plantation sectors [14]. This study aims to use Canonical Correlation Analysis to:

i) interpret the relationship between Rural Area Community Facilities and Population Characteristics in West Java;

ii) determine what variables be it from Rural Area Community Facilities or Population Characteristics that can be used as a determining factor to help reduce poverty within rural areas.

II. METHODOLOGY

A. Data Collection

To achieve the aforementioned targets of this study, a collection of variables that were used in previous studies of the resemblant topic and sourced from multiple open dataset websites. Each data is divided by 28 districts or cities in the West Java province and was last updated in 2023. Below are the identified and collected variables:

1) Rural Area Community Facilities

a) Number of School

Education is very important in regional development, as education supports one in the labor market and improves the quality of life [15]. Developing adequate educational infrastructure will boost human investment, including education and skills, which are key elements in building a prosperous society. An adequate number of schools, evenly spread across a region, is crucial because it can improve access to education for local communities. More schools will create broader learning opportunities for children and adolescents, thereby increasing literacy and the quality of education in the region. Moreover, the existence of adequate schools can also improve the competitiveness and quality of human resources, which will ultimately have a positive impact on economic growth and poverty reduction in a region. The data for this variable was collected from the Principal Education Data website.

b) Number of Health Services

The physical facilities required to support public health activities, such as hospitals, public health centers (*Puskesmas*), and integrated Healthcare Center (*Posyandu*), are known as health infrastructure. Hospitals are essential parts of the healthcare system because they offer a wide range of medical services, such as inpatient treatment, specialist care, emergency care, and surgery. Public health centers are primary care clinics that offer the community vaccines, health education, maternity and child health services, preventative care, and basic medical treatments. Concurrently, Integrated Service Post is a community-based health service in Indonesia that focuses on the health of mothers and children such as offering immunization, nutrition monitoring, and family planning services. Health infrastructure plays an important role in regional development because it can improve the quality of human resources and community welfare. This is because the development of health infrastructure, both in quantity and quality, will encourage an increase in human resource quality. Quality human resources will enhance welfare and decrease poverty. The development of a uniform health infrastructure across the region is expected to provide direct benefits to the poor and improve their quality of life [16]. The data for this variable was collected from the West Java Health Public Services website.

c) Community Settlements

A population is an individual or group that lives in a certain area. In regional development, having access to clean water, having proper sanitation, and adequate housing is very important. The availability of clean water is necessary to maintain health and daily activities such as drinking, cooking, bathing, and personal hygiene, so having access to clean water is vital for people to maintain their health. Households should also have sanitary facilities that meet health standards such as cloths and final disposal of stools using a septic tank or *Pengolahan Air Limbah* (SPAL) [17]. Aside from that, community access to adequate housing is also important in regional development because it can improve the community's quality of life and support local economic growth. Decent housing provides a safe and comfortable place for residents to live, rest, and develop, so that this can create a more productive and stable environment for local communities [18]. Thus, the availability of clean water and adequate housing for residents is an important factor in regional development because it has a direct impact on the health, welfare, and economic growth of local communities. The data for this variable was collected from the Central National of Statistics of West Java website.

d) Amount of Public Transportation

The public uses public transportation, which includes buses and travel cars. Public transportation is important for regional development because it can be a force in supporting the local economy by connecting residents with business and employment opportunities, as well as helping people access health needs, especially for the elderly [19]. Public transportation can also increase population mobility between villages and cities, which in turn will support local economic growth. As a result, public transportation plays a significant role in supporting small area development by facilitating connectivity, increasing population mobility, and supporting local economic growth. The data for this variable was collected from the National Bureau of Statistics of West Java website.

2) Rural Area Population Characteristics

a) Percentage of people living in poverty

Poverty can be defined as the inability of an individual to to maintain oneself in accordance with the standard of living of the community and to make use of one's physical and mental abilities within that group [20]. Poverty is positioned first in the Sustainable Development Goals as it has the most significant influence development globally. It is a multidimensional issue that affects not only income but also other areas of fulfilment that require fundamental services such as health, education, housing, clean water, natural resources, and the environment. Furthermore, it is linked to a sense of security and the

right to participate in social activities. All affected aspects above are affiliated with regional development indicators [21]. The data for this variable was collected from the National Bureuau of Statistics of West Java Province website.

b) Number of outmigration

Migration is commonly described as the movement of a person or group if people from one geographical unit to another with the goal of permanently or temporarily settling in their destination [22]. It could also be defined as an approach in which individuals respond to changes in said environment. Those changes could be in the form of natural circumstances, social life, economic situations, or population growth in their lives. Distinct populaton counts and socioeconomic advancements in each region means that there is distinct opportunities that could result in a better life of said individual [23]. This is backed by a specific study of regional development and internal migration in Senegal which resulted in rural migrants improving, not only their own, but also other rural communities' well-being by alleviating rising wage pressures. Despite the positive result, the gap between rich capital area residents and the rest of the population continues to widen [24]. This data for this variable in particular was collected from the West Java Open Data website.

c) Literacy Rate

Literacy refers to individuals' ability to comprehend and process information through reading and writing. Literate individuals have broader possibilities for education, career, health, and democratic involvement. As a result, literacy plays a vital role in eliminating social and economic disparities [25]. According to Mulkeen & Chen, the majority of rural parents have poor literacy levels and less educated which affects their ability to assist their children academically. They tend to be overindulged in home duties inhibited their children's progress in schoolwork, because they could not comprehend the value of education or their children's learning [26]. The data for this variable was collected from the National Bureuau of Statistics of West Java Province website.

The data comprised and used for analysis will henceforth be accompanied by their corresponding aliases, those aliases are:

FIGURE II. PEOPLE LIVING IN POVERTY

Coresponding Variable	**Alias**
Number of Schools	NoS
Number of Hospitals	NoHS1
Number of Public health center	NoHS2
Number of Integrated Health Center	NoHS3
Percentage of Households access to Sanitation	CS1
Percentage of Households access to Water Source	CS2
Percentage of Households access to Housing	CS3
Amount of Public Transportation	APB
Percentage of People Living in Poverty	PPLP
Number of Outmigrations	NoOut
Literacy Rate	LR

Fig. 2. A table depicting all the variables used in the paper with their corespoding aliases.

B. Cannonical Correlation Analysis

Canonical correlation analysis (CCA) is a multivariate statistical technique that examines the link between numerous sets of variables by identifying linear combinations of each set such that the correlation between these combinations is maximized. It looks for common statistical variabilities among various sets of variables rather than assuming directional relationships [27]. Since correlation is a method of predicting multiple dependent variables from multiple independent variables, one of the groups typically contains independent variables while the other group contains dependent variables [28]. However, no causal conclusions can be drawn based on this analysis alone [29]. This technique is to first identify the pair of linear combinations with the highest correlation. Then identify the pair of linear combinations with the greatest correlation among all pairs that are not correlated with the originally chosen pair, and so on. These linear combinations are referred to as canonical variables or canonical variate pairs and their correlations are called canonical correlations

Let the first group with p variables and the second group with q variables be defined as random vector $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ respectively with the assumption that the first group is the smaller set so that $p \leq q$. Jointly, it has mean vector

$$\underset{((p+q)\times 1)}{\boldsymbol{\mu}} = E(\boldsymbol{X}) = \begin{bmatrix} E\left(\boldsymbol{X}^{(1)}\right) \\ E\left(\boldsymbol{X}^{(2)}\right) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{bmatrix} \quad (1)$$

and covariance matrix

$$\underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}} = \begin{bmatrix} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{bmatrix} \quad (2)$$

To attain the k-th pair of canonical variates (linear combinations), use these formulas down below

$$U_k = \boldsymbol{e}_i' \boldsymbol{\Sigma}_{11}^{-\frac{1}{2}} \boldsymbol{X}^{(1)} \quad (3)$$

and

$$V_k = \boldsymbol{f}_i' \boldsymbol{\Sigma}_{22}^{-\frac{1}{2}} \boldsymbol{X}^{(2)} \quad (3)$$

Where each $\boldsymbol{e}_i$ is proportional to the eigenvectors of the dot product

$$\boldsymbol{\Sigma}_{22}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-\frac{1}{2}} \quad (3)$$

and so is each $\boldsymbol{f}_i$ to

$$\boldsymbol{\Sigma}_{11}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-\frac{1}{2}} \quad (4)$$

The square root of the eigenvalues of (3) can be used to find the canonical correlation values. In the case of standardized variables $\boldsymbol{Z}^{(1)}$ and $\boldsymbol{Z}^{(2)}$, the same formulas can be used but replace $\boldsymbol{Z}^{(I)}$ and $\boldsymbol{\Sigma}$ with $\boldsymbol{X}^{(I)}$ and $\boldsymbol{\rho}$ accordingly [30].

III. Result And Discussion

A. Assumption Testing

Data collected within this study requires further descriptive analysis to tackle common data issues such as linearity, collinearity, and normality, Isolated test was conducted for each set. Both sets proved to be equally problematic in terms of testing results.

1) Normality

Normality of the variable is tested using the Shapiro Wilk test. The given hypothesis is, H_0 : The data is normally distributed; H_1 : The data is not normally distributed.

FIGURE III. Summary Result for CcA

Variable Set	Pr < 0.1 (Before)	Pr < 0.1 (After)
X	5.646e-16	0.9415
Y	0.0008507	0.6166

Fig. 3. A table depicting the p-value for each set for each data variable in the study before and after applying BcnTransform.

Examining the results produced in Fig.3, each variable set does not meets the assumption of normality. To counteract this issue, a transformation using Bcn-Power was used. The resulting critical value proved sufficient in stating that both variable sets are distributed normally.

2) Linearity

FIGURE IV. Linearity Plot and Correlation

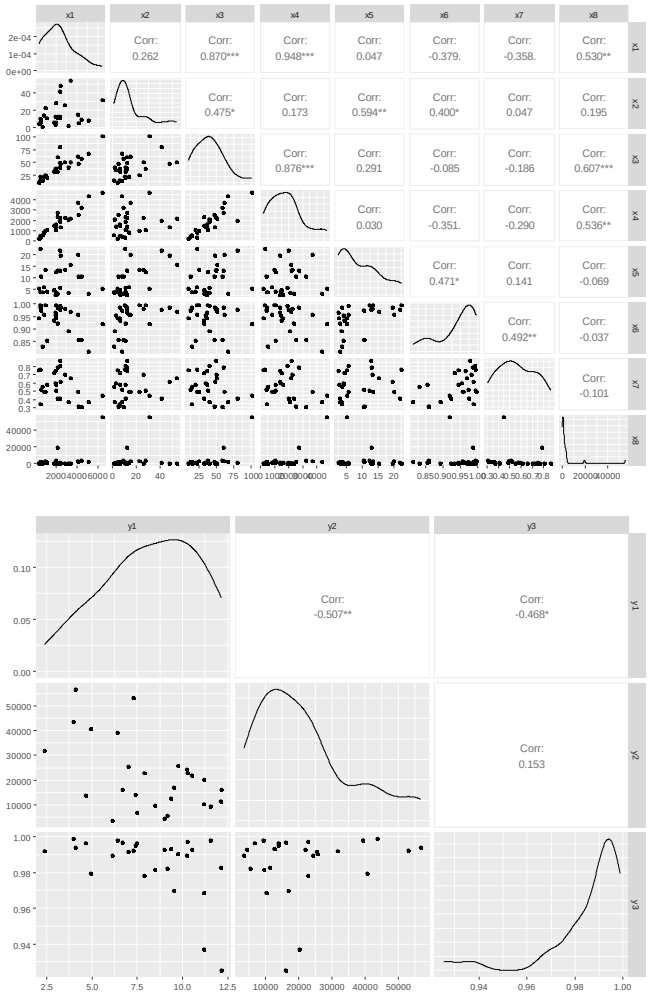

Fig. 4. Two plots depicting the distribution and the correlation of each corresponding dataset.

Fig.4 Shows all the corresponding correlations between the variables I each set. Generally, the Y-set presents itself as a heavily correlated set, thus proving linearity. Contrary to this, some characteristic variables are not correlated strongly with each other within the X-set.

3) Multicolinerity

FIGURE V. Multicolinearity

X-Set Variable	VIF	Y-Set Variable	VIF
NoS	13.790450	PLPP	1.705308
NoHS1	2.555035	NoOut	1.363588
NoHS2	10.833509	LR	1.296588

NoHS3	15.468919
CS1	2.080463
CS2	2.374641
CS3	1.430192
APB	1.890248

Fig. 5. A table depicting the the VIF values for multicolinerity

B. Cannonical Correlation Analysis

FIGURE VI. SUMMARY RESULT FOR CCA

Cannonical Variables	Cannonical Correlation	Critical Value	F
U_1V_1	0.8025233	1.790496113	1.751291078
U_2V_2	0.6454808	0.980625682	1.994857707
U_3V_3	0.5172841	1.096026704	2.661304523

Fig. 6. A table depicting the summary of the cannonical correlation that show the Correlation value, the Critical F value, and the F table value.

Based on the canonical correlation results than can be found in Fig.6, the only canonical pair of variables that has a critical value bigger than the F value is the first pair. Due to this, the first pair of variables is the maximum correlation that represents the data. The correlation between the first pair also falls between an acceptable metric (0.80). The canonical correlation value of the first pair also indicates that both sets of variables are affecting each other. To further identify the influence of certain characteristics within the variable, this study used the eigenvectors generated to determine the weight for each subset variable within their respective canonical variable.

FIGURE VII. CANNONICAL COEFFICIENT FOR Y-SET

Cannonical Variables	$PPLP$	$NoOut$	LR
U_1	0.1628413	-0.7738024	0 .7448838
U_2	0.7220701	-0.1265253	-0.3653099
U_3	-0.9722027	-0.776316	-0.7550928

Fig. 7. A Table depicting each characteristics weight within the Y-set Variable.

FIGURE VIII. CANNONICAL COEFFICIENT FOR X-SET

Cannonical Variables	NoS	$NoHS1$	$NoHS2$	$NoHS3$
V_1	2.9730	-0.4624194	2.212822	-0.1653223
V_2	1.3380	0.07511399	-1.409812	-3.127758
V_3	-1.8420	-0.060644	-1.524052	0.006301245

Cannonical Variables	$CS1$	$CS2$	$CS3$	APB
V_1	-0.264836	1.008998	-0.089454	-0.436209
V_2	-0.59915	0.2763507	-0.658705	-0.074183
V_3	-0.061659	-0.789242	-0.0422106	-0.8157169

Fig. 8. A Table depicting each characteristics weight within the X-set Variable.

Based on results found from Fig.7 and Fig.8, the optimal linear combination of the independent and dependent variables can be constructed using the respective cannonical coefficients as,

$$U_1 = (0.163PPLP) - (0.774NoOut) + (0.745LR) \quad (5)$$

$$V_1 = (2.97NoS) - (0.46NoHS1) + (2.212NoHS2) - (0.16NoHS3) - (0.26CS1) + (1.01\ CS2) - (0.08CS3) - (0.43APB) \quad (6)$$

An increase of Number of Outbound Population (NoOut) leads to an increase in the corresponding cannonical variable of the Y-set. This is an opposition with the Learning Rate (LR), where an increase in the LR leads to an increase in corresponding cannonical variable of the Y-set. Accordingly, an increase in the Number of Schools (NoS), Number of Public Health Centers (NoHS2), and Access to a water source (CS2) directly contributes in the decrease of the Number of Outmigrations (NoOut) but not the decrease of the Literacy Rate (LR) or the Percentage of People Living in Poverty (PPLP). This can be interpreted as an increase of the population within a given rural area but not accompanied by an increase in the quality of facilities provided.

FIGURE IX. CANNONICAL LOADINGS 1

Cannonical Variables	$PPLP$	$NoOut$	LR
U_1	0.152086	-0.72816	0.552812

Fig. 9. A Table depicting the cross loading of the original value of the Y-set with the corresponding cannonical variable

FIGURE X. CANNONICAL LOADINGS 2

Cannonical Variables	NoS	$NoHS1$	$NoHS2$	$NoHS3$
V_1	-0.471	0.139264	0.245649	-0.43259

Cannonical Variables	$CS1$	$CS2$	$CS3$	APB
V_1	0.337437	0.143695	0.294273	0.411377

Fig. 10. A Table depicting the cross loading of the original value of the X-set with the corresponding cannonical variable

FIGURE XI. CROSS LOADINGS 1

Cannonical Variables	$PPLP$	$NoOut$	LR
V_1	0.1220528	-0.5843665	0.4436448

Fig. 11. A Table depicting the cross loading of the original value of the X-set with the opposite cannonical variable

FIGURE XII. CROSS LOADINGS 2

Cannonical Variables	NoS	$NoHS1$	$NoHS2$	$NoHS3$
U_1	-0.379	0.1117629	0.197139	-0.3471609

Cannonical Variables	$CS1$	$CS2$	$CS3$	APB
U_1	0.2708006	0.1153186	0.2361612	0.3301397

Fig. 12. A Table depicting the cross loading of the original value of the Y-set with the opposite cannonical variable.

Large values in the cannonical loadings reinforce a strong relationship between the rural characteristics and rural community facilities. The loadings in Fig.9 demonstrate that NoOut and LR were more dominant in creating V_1 than PLPP. The loadings for the X-set in Fig.10 indicate APB and NoHS3 have both a inverse and adverse effect in creating U_1, respectively. Examining the cross loading allows the study to understand how variables from X-set relate to the canonical variates of the Y-set, and vice-versa. Fig. 11 and Fig.12 show that NoS, NoHS3, and NoOut all relate strongly to the variates of each others respective opposite set. This suggests that the The number of schools and integrated healthcare centers significantly influence the characteristics of rural populations. Adversely, the number of outgoing population from rural areas have an adverse relationship with the Rural Community Facilities.

IV. CONCLUSION

This study shows that there is a relationship between rural population characteristics and rural community facilities in West Java. As this study shows, the number of schools, integrated health services, and public transportation are critical to uncovering this relationship. These factors prove to be significant in reducing the gap between rural and urban areas. In addition, the amount of outmigration from rural areas also shows an important indication of this relationship.

The canonical correlation analysis conducted in this study shows that the number of schools and integrated health centers significantly influences rural communities' characteristics. In contrast, the number of people leaving rural areas has an opposite relationship with rural community facilities. This indicates that the more education and health facilities available, the better the rural community characteristics, while outmigration from rural areas tends to have a negative impact on rural community facilities.

According to this study, it is important to understand the factors that influence the gap between rural and urban areas. These results can later serve as a basis for policymakers to make more effective plans to improve the welfare and quality of life of rural communities. By strengthening education, healthcare, and public transportation facilities, it is expected to reduce inequality and increase the attractiveness of rural areas, thereby curbing outmigration and strengthening rural communities as a whole.

REFERENCES

[1] N. George, "Book Review: Skills Development for Inclusive and Sustainable Growth in Developing Asia-Pacific," *Journal of Learning for Development*, vol. 3, no. 1, 2016, doi: https://doi.org/10.56059/jl4d.v3i1.142.

[2] S. Fiorentino, A. Glasmeier, L. Lobao, R. Martin, and P. Tyler, "'Left behind places': What Are They and Why Do They matter?," *Cambridge Journal of Regions, Economy and Society*, vol. 17, no. 1, 2023, doi: https://doi.org/10.1093/cjres/rsad044.

[3] P. Cowie, L. Townsend, and K. Salemink, "Smart Rural futures: Will Rural Areas Be Left behind in the 4th Industrial revolution?," *Journal of Rural Studies*, vol. 79, no. 16, pp. 169–176, 2020, doi: https://doi.org/10.1016/j.jrurstud.2020.08.042.

[4] A. Sączewska-Piotrowska and J. Wąsowicz, "Income Polarization in Rural and Urban Areas," *Studia Ekonomiczne / Uniwersytet Ekonomiczny w Katowicach*, vol. 336, no. 11, pp. 146–157, 2017, Available: https://bazekon.uek.krakow.pl/rekord/171500504

[5] D. Luca, J. Terrero-Davila, J. Stein, and N. Lee, "Progressive cities: Urban–rural Polarisation of Social Values and Economic Development around the World," *Urban Studies*, vol. 60, no. 12, p. 004209802211483, 2023, doi: https://doi.org/10.1177/00420980221148388.

[6] S. Wiggins and S. Proctor, "How Special Are Rural Areas? the Economic Implications of Location for Rural Development," *Development Policy Review*, vol. 19, no. 4, pp. 427–436, 2001, doi: https://doi.org/10.1111/1467-7679.00142.

[7] P. McManus *et al.*, "Rural Community and Rural Resilience: What Is Important to Farmers in Keeping Their Country Towns alive?," *Journal of Rural Studies*, vol. 28, no. 1, pp. 20–29, 2012, doi: https://doi.org/10.1016/j.jrurstud.2011.09.003.

[8] Directorate of Food Crops, Horticulture, and Estate Crops Statistics, "Indonesian Oil Palm Statistics 2021," BPS-Statistics Indonesia, 2022. Available: https://www.bps.go.id/en/publication/2022/11/30/254ee6bd32104c00437a4a61/indonesian-oil-palm-statistics-2021.html

[9] J. Lusweti and A. Okoth, "Canonical Correlation in Modeling Short Term Expenditure versus Poverty Levels in Informal Settlements ":, *MEJ (Mathematics Education Journal)*, vol. 7, no. 2, pp. 208–215, 2023, doi: https://doi.org/10.22219/mej.v7i2.28091.

[10] O. R. Ebenezer and Iyaniwura, "Canonical Correlation Analysis of Poverty and Literacy Levels in Ekiti State, Nigeria," *Mathematical Theory and Modeling*, vol. 2, no. 6, pp. 31–38, 2012, Available: https://core.ac.uk/reader/234678990

[11] M. Mkondiwa, C. B. L. Jumbe, and K. A. Wiyo, "Poverty-Lack of Access to Adequate Safe Water Nexus: Evidence from Rural Malawi," *African Development Review*, vol. 25, no. 4, pp. 537–550, Dec. 2013, doi: https://doi.org/10.1111/1467-8268.12048.

[12] S. Şanli, "A Canonical Correlation Approach in Determining Growth & Development and Social Inclusion Linkages," *Academic Review of Humanities and Social Sciences*, vol. 1, no. 3, pp. 130–146, 2018, Available: https://www.acarindex.com/pdfs/1000521

[13] D. Ardian and M. R. Destanto, "Pengaruh Faktor Sosial Ekonomi terhadap Kemiskinan di Provinsi Jawa Barat," *Seminar Nasional Official Statistics*, vol. 2020, no. 1, pp. 377–384, 2021, doi: https://doi.org/10.34123/semnasoffstat.v2020i1.356.

[14] A. W. Setiawan and M. B. N. Ariani, "Determinasi Faktor yang Mempengaruhi Indeks Pembangunan Manusia Provinsi Jawa Barat Tahun 2015 – 2019," *Jurnal Indonesia Sosial Sains*, vol. 3, no. 01, pp. 1–9, 2022, doi: https://doi.org/10.59141/jiss.v3i01.505.

[15] N. S. Isnaini and F. Z. Azizah, *Pentingnya Pendidikan dalam Menanggulangi Kemiskinan di Indonesia*. 2020. Available: https://www.researchgate.net/profile/Nurma-Isnaini/publication/343360341_Pentingnya_Pendidikan_dalam_Menanggulangi_Kemiskinan_di_Indonesia/links/5f24bb74a6fdcccc439fce2c/Pentingnya-Pendidikan-dalam-Menanggulangi-Kemiskinan-di-Indonesia.pdf

[16] S. D. Purnomo, M. Wijaya, and H. Setiawan, "Infrastruktur dan Kemiskinan di Provinsi Daerah Istimewa Yogyakarta," *Majalah Imiah Manajemen dan Bisnis*, vol. 18, no. 1, pp. 10–19, 2021, Available: https://mimb.unwiku.ac.id/index.php/mimb/article/view/116

[17] T. R. Mayasari, "Pengelompokkan Provinsi Berdasarkan Variabel Kesehatan Lingkungan Dan Pengaruhnya Terhadap Kemiskinan Di Indonesia Tahun 2018," *Jurnal Siger Matematika*, vol. 1, no. 1, Mar. 2020, doi: https://doi.org/10.23960/jsm.v1i1.2471.

[18] A. D. Prasetyo and E. Sonny, "The Analysis of Determinants of Developing Village Index in Indonesia," *The Asian Journal of Technology Management (AJTM)*, vol. 13, no. 2, pp. 158–172, 2020, doi: https://doi.org/10.12695/ajtm.2020.13.2.5.

[19] R. Nabillah, "Young Generation'S Perception of Small-Town Development in the Last Ten Years in Kalianda District South Lampung Regency," *Tunas Geografi*, vol. 10, no. 1, pp. 41–52, 2021, doi: https://doi.org/10.24114/tgeo.v10i1.25212.

[20] S. Machmud and I. Sidharta, "The Relationship between Regional development, Human Development and poverty: a Perspective from the Indonesian Province," *Nurture*, vol. 17, no. 3, pp. 157–165, May 2023, doi: https://doi.org/10.55951/nurture.v17i3.283.

[21] R. IR, J. Nawawi, N. Nara, and Muh. Tang. Abdullah, "Regional Development Planning Strategy as a Poverty Reduction Solution in Polewali Mandar Regency," *KnE Social Sciences*, vol. 9, no. 2, pp. 616–625, Jan. 2024, doi: https://doi.org/10.18502/kss.v9i2.14884.

[22] B. Mitrică, N. Damian, I. Grigorescu, I. Mocanu, M. Dumitrașcu, and M. Persu, "Out-migration and Social and Technological Marginalization in Romania. Regional Disparities," *Technological Forecasting and Social Change*, vol. 175, no. 40, p. 121370, 2022, doi: https://doi.org/10.1016/j.techfore.2021.121370.

[23] Abd. Rasyid, Mustamin, and T. B. Prasetio, "Strategi Pembangunan Masyarakat Desa Tertinggal di Kecamatan Lembang," *SOSIOLOGIA : Jurnal Agama dan Masyarakat*, vol. 3, no. 1, pp. 151–161, 2023, doi: https://doi.org/10.35905/sosiologia.

[24] A. Mukashov and J. Thurlow, "Regional Development and Internal Migration Aspects of Structural transformation: A Case Study of Senegal," Kiel Institute for the World Economy, Washington DC, USA, 2023. Available: https://www.ifw-kiel.de/fileadmin/Dateiverwaltung/IfW-Publications/fis-import/a0608d4a-2c55-4f56-8466-c62377c59e70-KWP_2258.pdf

[25] E. Kartikasari, "Improving the Quality of Life of the Community through the Regional Literacy Movement," *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*, vol. 10, no. 1, pp. 644–654, Jan. 2024, doi: https://doi.org/10.30605/onoma.v10i1.3330.

[26] E. E. Shikalepo, "Characteristics of Rural Areas and Their Effects on Teaching and Learning Dynamics," *International Journal of Social Sciences and Management Review*, vol. 2, no. 4, pp. 20–36, 2019, Available: https://ijssmr.org/uploads/ijssmr02_43.pdf

[27] X. Zhuang, Z. Yang, and D. Cordes, "A Technical Review of Canonical Correlation Analysis for Neuroscience Applications," *Human Brain Mapping*, vol. 41, no. 13, pp. 3807–3833, 2020, doi: https://doi.org/10.1002/hbm.25090.

[28] I. Akour *et al.*, "Using the Canonical Correlation Analysis Method to Study Students' Levels in Face-to-Face and Online Education in Jordan," *Information Sciences Letters*, vol. 12, no. 2, pp. 901–910, 2023, doi: https://doi.org/10.18576/isl/120229.

[29] N. D. Wickramasinghe, "Canonical Correlation analysis: an Introduction to a Multivariate Statistical Analysis," *Journal of the College of Community Physicians of Sri Lanka*, vol. 25, no. 1, p. 37, 2019, doi: https://doi.org/10.4038/jccpsl.v25i1.8204.

[30] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Upper Saddle River, New Jersey: Pearson, 2019.

V. APPENDIX

A. *Dataset*

1) People Living in Poverty (Introduction)

Tahun	Population	Urban	Rural
2019	266.91	9.86	14.93
2020	270.20	12.04	15.51
2021	272.68	11.86	14.64
2022	278.70	11.98	14.38
2023	275.77	11.74	14.16

2) Rural Community Facilities

Wilayah	NoS	NoHS1	NoHS2	NoHS3	CS1	CS2	CS3	APB
Kab. Bogor	6432	31	101	4685	0.0533	0.919	0.4456	55485
Kab. Sukabumi	4460	9	58	3233	0.1043	0.8572	0.3051	3554
Kab. Cianjur	4159	5	47	2528	0.1064	0.8565	0.315	2032
Kab. Bandung	4060	15	62	3696	0.2001	0.9573	0.4948	2847
Kab. Sumedang	1959	3	35	1431	0.0401	0.9542	0.5963	607
Kab. Garut	5149	8	67	4334	0.0293	0.8113	0.3667	2114
Kab. Tasikmalaya	3211	2	40	2112	0.0343	0.8903	0.3695	461
Kab. Ciamis	2000	6	37	1597	0.0278	0.9439	0.5776	957
Kab. Kuningan	1905	12	37	1393	0.0239	0.9657	0.7444	824
Kab. Majalengka	1779	6	32	1500	0.0442	0.9338	0.7219	470
Kab. Cirebon	2157	12	60	2706	0.1294	0.9852	0.7878	19489
Kab. Indramayu	2383	12	49	2220	0.1287	0.9777	0.8635	593
Kab. Subang	2370	11	40	1907	0.0561	0.9922	0.8035	270
Kab. Purwakarta	1093	11	20	1047	0.0407	0.854	0.5757	1398
Kab. Karawang	2812	26	50	2225	0.1332	0.9811	0.6996	200
Kab. Bekasi	3387	54	51	2153	0.1538	0.9695	0.6549	405
Kab. Bandung Barat	2137	11	32	2340	0.0293	0.9227	0.4114	459
Kab. Pangandaran	665	1	15	532	0.0353	0.8295	0.5498	50
Kota Bandung	2374	41	80	1995	0.2183	0.9767	0.4117	1131
Kota Bogor	1001	22	25	982	0.1336	0.9961	0.493	3065
Kota Sukabumi	464	6	15	461	0.0368	0.9874	0.3402	2092
Kota Cirebon	513	10	22	328	0.2242	0.9957	0.7649	261
Kota Bekasi	2453	48	48	1359	0.1958	0.9877	0.6161	247
Kota Depok	1976	28	38	1059	0.124	0.9956	0.5079	3213
Kota Cimahi	540	8	13	410	0.105	0.9736	0.5142	1708
Kota Tasikmalaya	830	14	22	819	0.0566	0.9584	0.4882	1002
Kota Banjar	360	4	10	200	0.0533	0.9434	0.7521	203

3) Rural Community Characteristics

Wilayah	PPLP	NoOut	LR
Kab. Bogor	6.4	39254	0.9981
Kab. Sukabumi	10.52	21860	0.9928
Kab. Cianjur	4.93	40625	0.9794
Kab. Bandung	7.27	53081	0.9924
Kab. Sumedang	7.42	13963	0.9952
Kab. Garut	10.22	24334	0.9895
Kab. Tasikmalaya	11.2	20215	0.937
Kab. Ciamis	9.77	25719	0.9902
Kab. Kuningan	12.13	16090	0.9252
Kab. Majalengka	7.87	22909	0.9782

Kab. Cirebon	3.96	43701	0.9987
Kab. Indramayu	6.14	3574	0.9895
Kab. Subang	4.1	56670	0.994
Kab. Purwakarta	6.67	16046	0.9969
Kab. Karawang	4.66	13856	0.9964
Kab. Bekasi	9.16	5543	0.9823
Kab. Bandung Barat	2.38	31773	0.9919
Kab. Pangandaran	7.5	6752	0.9964
Kota Bandung	11.53	9295	0.9981
Kota Bogor	12.12	11340	0.9829
Kota Sukabumi	11.21	10196	0.9687
Kota Cirebon	8.98	4435	0.9929
Kota Bekasi	8.46	9618	0.9813
Kota Depok	9.52	16920	0.9697
Kota Cimahi	7.01	25464	0.9917
Kota Tasikmalaya	9.36	12566	0.9934
Kota Banjar	10.28	22741	0.9971

B. Manual Calculations

k	df1	df2	w	t	F
1	24	47.00608	20	2.900304	1.790496113
2	14	33.58973274	20	1.979486637	0.980625682
3	6	18	20	1	1.096026704

C. R Script

1) Libraries

```
library(expm)
library(MASS)
library(readxl)
library(dplyr)
library(car)
library(tidyr)
library(purrr)
library(GGally)
```

2) Rural Community Facilities Data Checking

```
##### Rural Community Facilities ######
data.RawCom <- read_excel("C:/Users/reyza/OneDrive - Bina Nusantara/Sem.6/MS-
Paper/RuralCommunityFacilities.xlsx")
data.Community <- data.RawCom[,2:ncol(data.RawCom)]
names(data.Community)  <- c('x1','x2','x3','x4','x5','x6','x7','x8')
data.Community
#Testing
library(mvShapiroTest)
#H0 = variable pada Community terdistribusi multivariate normal
#H1 = variable pada Community Tidak terdistribusi normal multivariate
mvShapiro.Test(as.matrix(data.Community))

hasil <- summary(powerTransform(data.Community))
hasil
x1 <- data.Community[,1]^hasil$result[1]
x2 <- data.Community[,2]^hasil$result[2]
x3 <- data.Community[,3]^hasil$result[3]
x4 <- data.Community[,4]^hasil$result[4]
x5 <- data.Community[,5]^hasil$result[5]
x6 <- data.Community[,6]^hasil$result[6]
x7 <- data.Community[,7]^hasil$result[7]
```

```
x8 <- data.Community[,8]^hasil$result[8]

transformed.Com <- data.frame(x1,x2,x3,x4,x5,x6,x7,x8)
mvShapiro.Test(as.matrix(transformed.Com))

# Multicolinearitas
data.Community$DummyY <- seq(1)
model <- lm(DummyY ~., data = data.Community)
vif(model)
```

3) Rural Population Characteristics Data Checking

```
##### Rural Characteristics#####
data.RawChar <- read_excel("C:/Users/reyza/OneDrive - Bina Nusantara/Sem.6/MS-
Paper/ruralCommunityCharacteristics.xlsx", sheet = "All")
data.RawChar
data.Characteristic <- data.RawChar[,2:ncol(data.RawChar)]

names(data.Characteristic) <- c("y1","y2","y3")
#Testing
library(mvShapiroTest)
#H0 = variable pada Community terdistribusi multivariate normal
#H1 = variable pada Community Tidak terdistribusi normal multivariate
mvShapiro.Test(as.matrix(data.Characteristic))

hasil <- summary(powerTransform(data.Characteristic))
hasil
y1 <- data.Characteristic[,1]^hasil$result[1]
y2 <- data.Characteristic[,2]^hasil$result[2]
y3 <- data.Characteristic[,3]^hasil$result[3]

transformed.Char <- data.frame(y1,y2,y3)
mvShapiro.Test(as.matrix(transformed.Char))

## Multicolinearitas
data.Characteristic$DummyY <- seq(1)
library(car)
model <- lm(DummyY ~., data = data.Characteristic)
vif(model)

#Lineritas
ggpairs(data.Characteristic[1:3])
```

4) Cannonical Correlation Analysis

```
##### Canonical Analysis #####

library(dplyr)
combined_df <- cbind(transformed.Char %>% dplyr::select(y1,y2,y3), transformed.Com  %>%
dplyr::select(x1,x2,x3,x4,x5,x6,x7,x8))
rho.all<- cor(combined_df)
rho.all
rho.11 <- rho.all[1:3,1:3]
rho.22 <- rho.all[4:11,4:11]
rho.22
rho.12 <- rho.all[1:3,4:11]
rho.12
rho.21 <- rho.all[4:11,1:3]
rho.21


##### U(Y --> anggota paling kecil) bisa juga yang dependent#####
inv2.rho.11 <- solve(sqrtm(rho.11))
inv.rho.22 <- solve(rho.22)

A = inv2.rho.11 %*% rho.12 %*% inv.rho.22 %*% rho.21 %*% inv2.rho.11

#ini  untuk cari yang kanonik kuadrat $values
cuadratic.Can <- eigen(A)$values
```

```
cuadratic.Can

#untuk nilai korelasi kanonik
sqrt(cuadratic.Can)

#Bobot Kanonik for U
Weight.CanonicalA <- eigen(A)$vectors
Weight.CanonicalA

#Uk
U.1 <- t(Weight.CanonicalA[,1]) %*%  inv2.rho.11
U.2 <- t(Weight.CanonicalA[,2]) %*%  inv2.rho.11
U.3 <- t(Weight.CanonicalA[,3]) %*%  inv2.rho.11
U.1
U.2
U.3
##### V(X->anggotanya paling banyak) bisa juga yang independent ####
inv2.rho.22 <- solve(sqrtm(rho.22))
inv.rho.11 <- solve(rho.11)

B <- inv2.rho.22 %*% rho.21 %*% inv.rho.11 %*% rho.12 %*% inv2.rho.22

#Bobot Kanonik for V
Weight.CanonicalB <- eigen(B)$vectors
Weight.CanonicalB

#Vk
V.1 <- t(Weight.CanonicalB[,1]) %*%  inv2.rho.22
V.2 <- t(Weight.CanonicalB[,2]) %*%  inv2.rho.22
V.3 <- t(Weight.CanonicalB[,3]) %*%  inv2.rho.22

V.1
V.2
V.3
##### Simultaneous Testing ####
lambda.stat <- det(rho.all)/(det(rho.11) * (det(rho.22)))
lambda.stat
```

5) Cannonical Loadings

```
#Loadings
library(CCA)
library(CCP)
cc1 <- cc(transformed.Com, transformed.Char)
cc2 <- comput(transformed.Com, transformed.Char, cc1)

x_scores <- cc2$xscores
y_scores <- cc2$yscores

# Calculate canonical loadings
yx_loadings <- cor(transformed.Char, x_scores)
xy_loadings <- cor(transformed.Com, y_scores)
xx_loadings <- cor(transformed.Com, x_scores)
yy_loadings <- cor(transformed.Char, y_scores)
```

D. Link

MS-Paper