BAREKENG: Journal of Mathematics and Its Applications
March 2026 Volume 20 Issue 1 Page 0197–0210
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20no1pp0197-0210

IMPLEMENTATION OF RESPONSE-BASED UNIT SEGMENTATION IN PARTIAL LEAST SQUARE (REBUS-PLS) FOR ANALYSIS AND REGIONAL GROUPING

Hairil Al-Ham [1], Neva Satyahadewi [2*], Preatin [3]

[1,2]Department of Mathematics, Faculty of Mathematics and Science, Universitas Tanjungpura
Jln. Prof. Dr. Hadari Nawawi, Pontianak, 78124, Indonesia

[3]Badan Pusat Statistik Provinsi Kalimantan Barat
Jln. Sultan Syahrir, Pontianak, 78116, Indonesia

*Corresponding author's e-mail: * **neva.satya@math.untan.ac.id***

Article Info

Article History:
Received: 25th February 2025
Revised: 26th May 2025
Accepted: 25th July 2025
Available online: 24th November 2025

Keywords:
Health;
Latent variables;
R-square;
SEM-PLS.

ABSTRACT

Housing environmental health is a key indicator of community quality of life. In West Kalimantan Province, variations in geographical and socioeconomic conditions contribute to disparities in housing conditions. This study analyzes and classifies regions based on factors influencing housing environmental health using the Response-Based Unit Segmentation in Partial Least Squares (REBUS-PLS) method. REBUS-PLS helps detect unobserved heterogeneity by identifying subgroups with different structural relationships. The exogenous latent variables include household economics, education, and housing facilities, while the endogenous variable is housing environmental health, measured through 15 indicators. The results of the SEM-PLS analysis obtained 3 paths that had a significant effect: household economics on housing facilities, household economics on education, and housing facilities on the health of the Housing environment. SEM-PLS assumes homogeneity across data, meaning all observations follow the same structural pattern. However, this assumption may not hold, especially with data representing diverse regions. To address potential heterogeneity, REBUS-PLS was applied. The analysis revealed two distinct segments, each with stronger explanatory power than the global model, as indicated by higher R^2 values (Segment 1 = 95.6%, Segment 2 = 91.4%, compared to 87.7% in the global model). Segment 1 consists of Landak, Sanggau, Sekadau, Kayong Utara, and Singkawang City. Segment 2 includes Bengkayang, Melawi, Ketapang, Kapuas Hulu, Sanggau, Sekadau, Sintang, and Pontianak City. These findings confirm the presence of structural heterogeneity and demonstrate that REBUS-PLS provides a more accurate understanding of the factors affecting housing environmental health across regions.

How to cite this article:

H. A. Ham, N. Satyahadewi and Preatin., "IMPLEMENTATION OF RESPONSE-BASED UNIT SEGMENTATION IN PARTIAL LEAST SQUARE (REBUS-PLS) FOR ANALYSIS AND REGIONAL GROUPING," *BAREKENG: J. Math. & App.,* vol. 20, iss. 1, pp. 0197-0210, Mar, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

The need for a decent place to live is one of the fundamental elements of human life [1]. A healthy home or a healthy place to live is very important to prevent health risks that originate from the environment [2]. The physical condition of housing, the quality of the surrounding environment, and the characteristics of its residents play an important role in maintaining the community's overall health. Housing quality includes aspects of building conditions, the surrounding environment, and its residents, both in terms of physical, social, and economic aspects [3]. Various factors, such as education level, income, and employment, also influence individuals' or households' ability to improve the quality of their housin [4]. Therefore, a comprehensive analysis is needed to assess the factors that influence the health of the housing environment, both physically and socio-economically.

Simple regression techniques are often inadequate in research fields involving many variables, such as in housing health analysis. This is because simple regression generally involves only one dependent variable, while in many cases, research involves several dependent variables at once. To overcome these limitations, the structural equation modeling (SEM) method was developed, which allows researchers to simultaneously model complex relationships between several latent variables and indicator or manifest variables [5]. SEM is particularly useful in research involving interactions between concepts that cannot be measured directly, such as quality of life or environmental health, which are measured through multiple variables.

SEM has two primary approaches: Partial Least Squares SEM (PLS-SEM) and Covariance-Based SEM (CB-SEM) [6]. CB-SEM relies heavily on strong theoretical assumptions, normal distribution, and large sample size and must fulfill the Goodness of Fit (GoF) test. In contrast, PLS-SEM is more flexible and less dependent on these assumptions, so it is often used in studies involving small samples or data that are not normally distributed [7]. In the measurement model, PLS-SEM also allows the use of both formative and reflective variables, which expands the flexibility in analyzing various types of variables [8].

However, a common issue in PLS-SEM applications is the problem of unobserved heterogeneity, where subpopulations within the dataset may respond differently to the modeled constructs. If ignored, this heterogeneity can lead to biased estimates, misleading conclusions, and poor model fit. Unobserved heterogeneity often arises from differences in socioeconomic status, regional characteristics, or cultural factors that are not directly measured but influence model relationships. Therefore, identifying and accounting for this heterogeneity is critical for improving the robustness and relevance of the analysis. To address this issue, the Response-Based Unit Segmentation in Partial Least Squares (REBUS-PLS) method was developed [9]. REBUS-PLS works by segmenting observation units based on the similarity of performance in the model, then estimating parameters for each segment formed. This method not only detects heterogeneity in structural and measurement equations but also provides more accurate estimation results for each segment without requiring distributional assumptions on latent and indicator/manifest variables [4].

Several studies have demonstrated the effectiveness of REBUS-PLS in uncovering hidden segments within complex data structures. For instance, [10] applied REBUS-PLS to marketing data to identify heterogeneous customer segments based on latent constructs. Similarly, [11] used it in detecting unobserved heterogeneity in the relationship between subjective well-being and satisfaction in various domains of life using the REBUS-PLS path modelling approach.

This research applies REBUS-PLS to analyze and cluster areas in West Kalimantan Province based on the quality of housing environmental health. The use of REBUS-PLS enables the identification of homogeneous area groups that exhibit similar structural relationships within the model. This segmentation provides more accurate and actionable insights for policymakers, allowing housing environmental health policies to be better tailored to the unique characteristics of each segment.

2. RESEARCH METHODS

PLS-SEM belongs to variance-based SEM and focuses on analysis that moves from testing causal models or theories to component-based predictive models [12]. PLS is utilized to address the limitations of the SEM method, particularly when dealing with challenges such as specific measurement scales, small sample sizes, missing data, and multicollinearity, by employing component-based modelling to overcome multicollinearity and using techniques such as pairwise deletion or imputation strategies to handle missing data [13].

2.1 Response-Based Segmentation Unit in Partial Least Square

REBUS PLS is an advanced SEM-PLS method that groups observation units while simultaneously estimating parameters for each local model within the identified segments [12], using an iterative partitioning algorithm that combines Partial Least Squares (PLS) estimation with a clustering technique based on residuals, such as K-means or other distance-based methods [10]. Observations are classified based on similar behavior or performance. In practice, statistical data is often heterogeneous, which can affect analysis results and lead to invalid conclusions [14]. Therefore, detecting heterogeneity is possible by clustering the observation units [4].

The grouping of observation units is determined by assessing closeness or distance through a proximity measure index (CM index) [15]. The CM index obtained from the residual communality model, which functions as a measure of goodness-of-fit (GoF) [16].

$$CM_{ik} = \sqrt{\frac{\sum_{q=1}^{Q}\sum_{p=1}^{P_q}\left[\frac{e_{ipqk}^2}{Com\left(\hat{\xi}_{qk}, x_{pq}\right)}\right]}{\frac{\sum_{i=1}^{N}\sum_{p=1}^{P_q}\left[\frac{e_{ipqk}^2}{Com\left(\hat{\xi}_{qk}, x_{pq}\right)}\right]}{(n_k - m_k - 1)}} \times \frac{\sum_{j^*=1}^{J^*}\left[\frac{f_{ijk}^2}{R^2\left(\hat{\xi}_{q*}, \hat{\xi}_q\right)}\right]}{\frac{\sum_{i=1}^{N}\sum_{j^*=1}^{J^*}\left[\frac{f_{ijk}^2}{R^2\left(\hat{\xi}_{q*}, \hat{\xi}_q\right)}\right]}{(n_k - m_k - 1)}}}, \quad (1)$$

with:

$Com(\hat{\xi}_{qk}, x_{pq})$: Communal index of the p-th variable from the q-th block in the k-th latent class
f_{ijk}^2 : Structural model residual for the i-th unit in the k-th latent class referencing the j-th endogenous block
e_{ipqk}^2 : Model residual measure for the i-th unit in the k-th latent class referencing to the p-th indicator in the q-th block
m_k : Number of components extracted from the k-th latent class
n_k : Number of units from the k-th latent class
R^2 : Coefficient of determination from the model

2.2 Determination of The Number of Segments

To overcome unobserved heterogeneity, a grouping was designed using REBUS PLS by identifying groups of observation units that showed similar behavior and performance **[17]**. Additionally, the resulting local model will demonstrate higher GoF and and R^2 values, because the CM index is based on the structure of the GoF index **[18]**. The chosen proximity measure is defined according to the GoF index structure to ensure that the global model is outperformed by the local model **[19]**.

According to (Trichera, 2007), the GoF index is defined as.

$$GoF = \sqrt{\frac{\sum_{q=1}^{Q}\sum_{p=1}^{P_q} Cor^2\left(x_{pq}, \xi_q\right)}{\sum_{q=1}^{Q} P_q} \times \frac{\sum_{j=1}^{J} R^2\left(\hat{\xi}_{q*}, \hat{\xi}_q\right)}{J}}, \quad (2)$$

With:

Q : Number of exogenous and endogenous latent variables
P_q : Number of manifest variables in the block
J : Number of endogenous latent variables

The REBUS PLS method includes the following steps:

1. The initial process involves from the global model by analyzing observation units using community and structural residuals to form preliminary groups. These initial groups are generated

through hierarchical cluster analysis of the residuals and are represented visually in a dendrogram [20].

2. For the initial class formed, based on the results of SEM PLS analysis, the structural and communality residuals in each class formed from each observation unit were recalculated, and then of each unit of each local model the CM value can be obtained [18].
3. The observation units are assigned in groups with the smaller CM value. The number of local model classes is recalculated using SEM PLS, if the distribution of observation units within a class shift [21].
4. The iteration will continue until the class composition (observations that belong to a class) does not change or until the stopping rule is reached. The stopping criterion is met when the variation in class composition falls below 0,05 and convergence is typically achieved within fewer than 15 iterations [16].

3. RESULTS AND DISCUSSION

In this research, the data obtained from publications on the website of the "Badan Pusat Statistik Kalimantan Barat", with the title of the publication, namely the "Wealth Statistics Kalimantan Barat Province 2023" and "Provinsi Kalimantan Barat in Figures 2024". The latent variables used are household economy, education, Housing facilities, and health of the Housing environment, with the number of indicators used totaling 15 indicators in percentage form with the division of regions, each of which is a regency/city in West Kalimantan Province.

Table 1. Indicators of Latent Variables

Latent Variable	Code	Indicator
Household Economics (ξ_1)	EK1	Percentage of Households with Ownership Status of Self-Owned Housing Buildings
	EK2	Percentage of Households with the Widest Floor Type Not Land
	EK3	Percentage of Households with Floor Area of Housing Buildings 150 $\geq m^2$
	EK4	The Percentage of Households with the Widest Wall Type is the Wall
	EK5	Percentage of Households That Own a Phone
	EK6	Percentage of Households That Own *Laptop/Computers*
Education (η_1)	PD1	Percentage of Population Aged above 45 Years who Are Literate
	PD2	Mean Years Schooling of Population 25 years of Age and Over (Years)
	PD3	Percentage of Population aged above 15 years According to Education Complete High School or Equivalent.
Housing Facilities (η_2)	FS1	Percentage of Households with the Main Source of Lighting is PLN Electricity
	FS2	Percentage of Households with the Use of Toilet Facilities
	FS3	Percentage of Households with Fecal Landfills are Septic Tanks
	FS4	Percentage of Households with Toilet Type in the form of Gooseneck
Housing Environmental Health (η_3)	SH1	Percentage of Households Having Access to Improved Drinking Water
	SH2	Percentage of households that have access to proper sanitation

3.1 Measurement Model Evaluation (Outer Model)

To evaluate the outer model, observations of discriminant validity, convergent validity, and composite reliability were formed.

3.1.1 Convergent Validity

The principle of convergent validity correlates each variable well with the latent variables that construct it. If the outer loading value obtained is below 0.6 [22], the value standard has not been met. So the convergent validity test results are unacceptable. Next, retest modifications are carried out by eliminating invalid indicators until have a valid outer loading value and can be continued for further analysis [23]. The structure of the model is presented in Fig. 1, and Table 2 represents outer loading values.

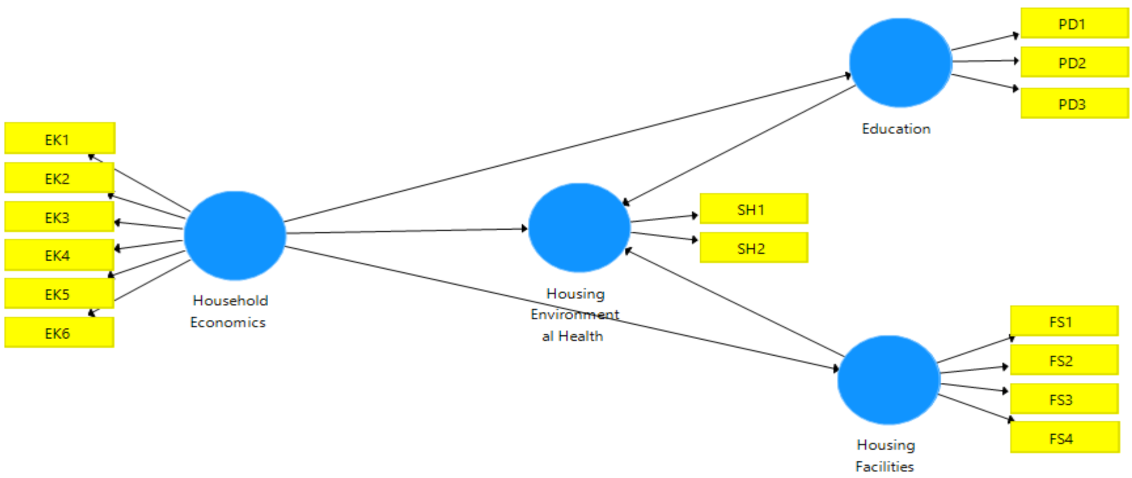

Figure 1. Model Structure on Housing Environmental Health
Source: SmartPLS version 4

Table 2. Outer Loading Value

Latent Variable	Indicator	Value of Outer Loadings	Information
Household Economics (ξ_1)	EK1	-0.936	**Invalid**
	EK2	0.876	Valid
	EK3	0.872	Valid
	EK4	0.776	Valid
	EK5	0.887	Valid
	EK6	0.814	Valid
Housing Facilities (η_1)	FS1	0.764	Valid
	FS2	0.675	Valid
	FS3	0.866	Valid
	FS4	0.593	**Invalid**
Education (η_2)	PD1	0.725	Valid
	PD2	0.934	Valid
	PD3	0.890	Valid
Housing Environmental Health (η_3)	SH1	0.864	Valid
	SH2	0.883	Valid

Table 2 shows that there are still variables with an outer loading value below 0.6, so retesting is done by removing invalid variables. Invalid indicator variables include (EK1) and (FS4). Table 3 represents the outer loading value after re-testing.

Table 3. Outer Loading Value After Re-testing

Latent Variable	Indicator	Outer Loadings	Information
Household Economics (ξ_1)	EK2	0.864	Valid
	EK3	0.876	Valid
	EK4	0.795	Valid
	EK5	0.886	Valid
	EK6	0.820	Valid
Housing Facilities (η_1)	FS1	0.774	Valid
	FS2	0.649	Valid
	FS3	0.951	Valid
Education (η_2)	PD1	0.731	Valid
	PD2	0.932	Valid
	PD3	0.889	Valid
Housing Environmental Health (η_3)	SH1	0.881	Valid
	SH2	0.866	Valid

The Outer Loadings values Table 3 exceed 0.6 for all variables, indicating that they are valid and suitable for further analysis. After testing convergent validity, a new path diagram is obtained, presented in Fig. 2.

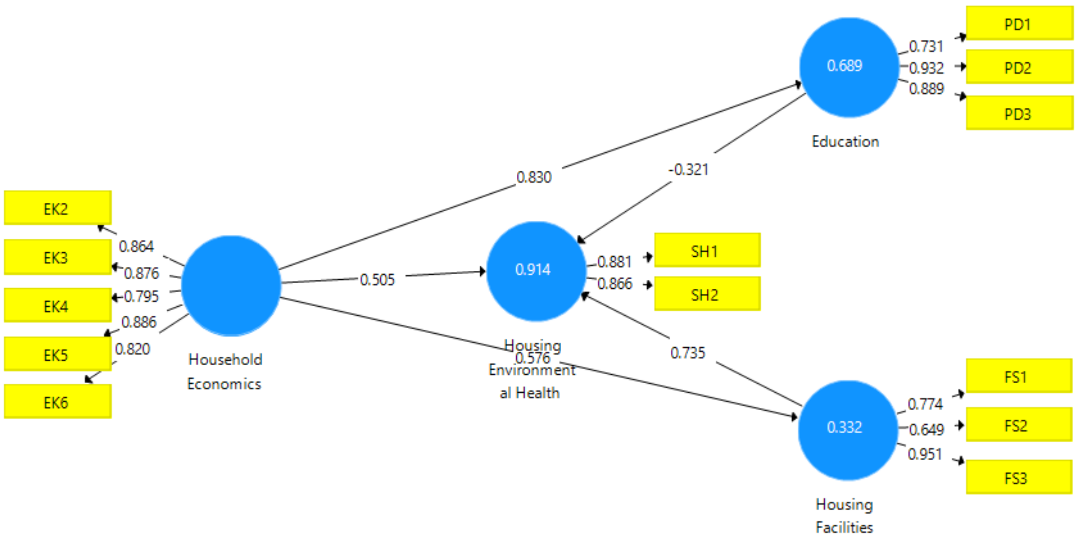

Figure 2. New Form of Conceptual Model of Housing Environmental Health
Source: SmartPLS version 4

Based on Fig. 2, the new conceptual model of the housing environmental health structure after retesting.

3.1.2 Discriminant Validity

Discriminant Validity is done to see how a latent variable differs from other latent variables, also ensure how well a construct can explain each indicator that composes it. Discriminant validity looks at the values on cross-loading, Fornell-Larcker Criterion, and AVE [24].

The reflective indicators used are valid indicator variables based on the loading factor value in the convergent validity evaluation. This is also supported by the AVE value which is more than 0.5 for each latent variable [25].

Table 4. Cross Loading Value

Indicator	Economics	Facilities	Health	Education
EK2	0.864	0.520	0.518	0.648
EK3	0.876	0.427	0.516	0.808
EK4	0.795	0.595	0.643	0.603
EK5	0.886	0.586	0.672	0.697
EK6	0.820	0.287	0.440	0.776
FS1	0.483	0.774	0.645	0.297
FS2	0.356	0.649	0.695	0.189
FS3	0.533	0.951	0.870	0.252
PD1	0.419	0.041	0.029	0.731
PD2	0.898	0.300	0.382	0.932
PD3	0.675	0.354	0.295	0.889
SH1	0.478	0.813	0.881	0.060
SH2	0.686	0.809	0.866	0.519

In Table 4, the cross-loading value for each construct of each indicator is higher than that of other constructs. Thus, for all indicators it has good discriminant validity. In addition to cross-loading value, discriminant validity can be seen through the larger criterion and AVE criteria in Tables 5 and 6.

Table 5. Fornell-Larcker Criterion Value

Latent Variable	Economics	Facilities	Health	Education
Economics	0.849			
Facilities	0.576	0.801		
Health	0.663	0.928	0.874	
Education	0.830	0.307	0.324	0.855

It can be seen that in Table 5, AVE value of the latent variable economy is 0.849, this value is the largest compared to the correlation value of Economy with Facilities, Education, and Health. This also applies for the latent variable Facility-to-Facility (0.801), a Health-to-Health (0.874) and Education-to-Education (0.855). Therefore, further observations need to be made by looking at the AVE value listed in Table 6.

Table 6. AVE Value of Each Latent Variable

Latent Variable	Average Variance Extracted (AVE)
Economics	0.721
Facilities	0.641
Health	0.764
Education	0.731

The AVE value in Table 6, of the Economic is 0.721, Facilities is 0.641, Health is 0.764, and Education is 0.731, indicating that all constructs meet the threshold for good convergent validity. An AVE value greater than 0.50 signifies that more than 50% of the variance in the indicators is explained by the latent construct, which confirms that the indicators have a strong representation of the underlying variable. Although, according to the Fornell-Larcker Criterion, the correlation between the Education construct and itself is not greater than its correlation with the health construct, this condition can still be considered acceptable. This is because convergent validity is adequately supported by the AVE values and is further reinforced by the fulfillment of the cross-loading criteria.

3.1.3 Composite Reliability

Research can be said to be reliable if it has the same data at different times. The measure states that a variable is reliable if it values larger than or equal to 0.6. Table 7 represents the Reliability construct value.

Table 7. Reliability Construct Value

Latent Variable	Cronbach's Alpha	Composite Reliability
Economics	0.903	0.928
Facilities	0.703	0.840
Health	0.691	0.866
Education	0.824	0.890

According to Table 7, for each latent variable the Cronbach's Alpha and Composite Reliability values exceed the standard threshold of 0.6, indicating an acceptable level of reliability for the study. Additionally, if the Composite Reliability values are higher than Cronbach's Alpha values, confirming that all latent variables meet the required reliability criteria.

3.1.4 Measurement Model Equation (Outer Model)

The outer model is obtained as follows based on the outer loading value.

1. Household Economic Variable Measurement Model (Reflective).
   $x_{12} = 0.864\xi_1$ $\quad$ $x_{13} = 0.876\xi_1$ $\quad$ $x_{14} = 0.795\xi_1$
   $x_{15} = 0.886\xi_1$ $\quad$ $x_{16} = 0.820\xi_1$
2. Education Variable Measurement Model (Reflective).
   $y_{11} = 0.731\eta_1$ $\quad$ $y_{12} = 0.932\eta_1$ $\quad$ $y_{13} = 0.889\eta_1$
3. Housing Facility Variable Measurement Model (Reflective).
   $y_{21} = 0.774\eta_1$ $\quad$ $y_{22} = 0.649\eta_1$ $\quad$ $y_{23} = 0.951\eta_1$
4. Housing Environmental Health Variable Measurement Model (Reflective).
   $y_{31} = 0.881\eta_1$ $\quad$ $y_{32} = 0.866\eta_1$

3.2 Structural Model Evaluation (Inner Model)

Structural evaluation in SEM-PLS measures the structural model's goodness, expressed by *R-square* (coefficient of determination), and the path coefficient value for significance testing between constructs in the structural model [26].

Table 8. Structural Model Coefficient Determination

Endogenous Latent Variables	R^2
Housing Facilities	0.332
Housing Environmental Health	0.914
Education	0.689

Table 8 shows that for housing facilities the R-square value is 0.332, which means 33.2% .the economic variable can explain the housing facilities variable (the model is categorized as moderate [27]). The resulting R^2 value on the Housing Facility latent variable is minimal because several other factors influence it, not only influenced by the Household Economy but other factors are not mentioned in this study.

Furthermore, in SEM-PLS the evaluation of the inner model is also carried out using the bootstrapping method by measuring the path coefficients between latent variables. The indirect effect path coefficient value using bootstrapping can be seen in Table 9.

Table 9. Path Coefficient Values Indirect Effect Using Bootstrap

Inner Model	Original Sample	T- Statistic	P-Value
Household Economics > Housing Facilities> Housing Environmental Health	0.424	1.706	0.089
Household Economics > Education> Housing Environmental Health	-0.266	1.031	0.303

Table 9 shows that the p-value of the indirect effect of the Household Economy on Environmental Health through Housing Facilities is significant, meaning that the Housing Facilities variable significantly mediates the variable of the Household Economy on the Environmental Health of Housing. The p-value of the latent variable on Household Economy to Housing Environmental Health through Education is also significant [28].

Table 10. Path Coefficient Values Direct Effect Using Bootstrap

Inner Model	Original Sample	T- Statistic	P-Value
Household Economics > Housing Facilities	0.576	1.966	0.050
Household Economics > Housing Environmental Health	0.505	1.458	0.145
Household Economics > Education	0.830	2.655	0.008
Housing Facilities > Housing Environmental Health	0.735	4.127	0.000
Education >Housing Environmental Health	-0.321	1.231	0.219

Table 10 shows that, 3 paths have a significant effect based on the p-value: the latent variables of Household Economy on Housing Facilities, Household Economy on Education, and Housing Facilities on Environmental Health Housing. The inner model equation is written as follows.

1. Structural Model of Education Latent Variable
   $\eta_1 = 0.830\xi_1 + \zeta_1$
2. Structural Model of Housing Facility Latent Variable
   $\eta_2 = 0.576\xi_2 + \zeta_2$
3. Structural Model of Housing Environmental Health Latent Variable
   $\eta_3 = 0.505\xi_1 - 0.321\eta_1 + 0.735\eta_2 + \zeta_3$

3.3 Group Determination with REBUS-PLS

The clustering of observation units in REBUS-PLS on communal residuals and structural residuals is based on Ward's method on the results of hierarchical cluster analysis. The results of cluster analysis on communal residuals and structural residuals can be seen in Fig. 3.

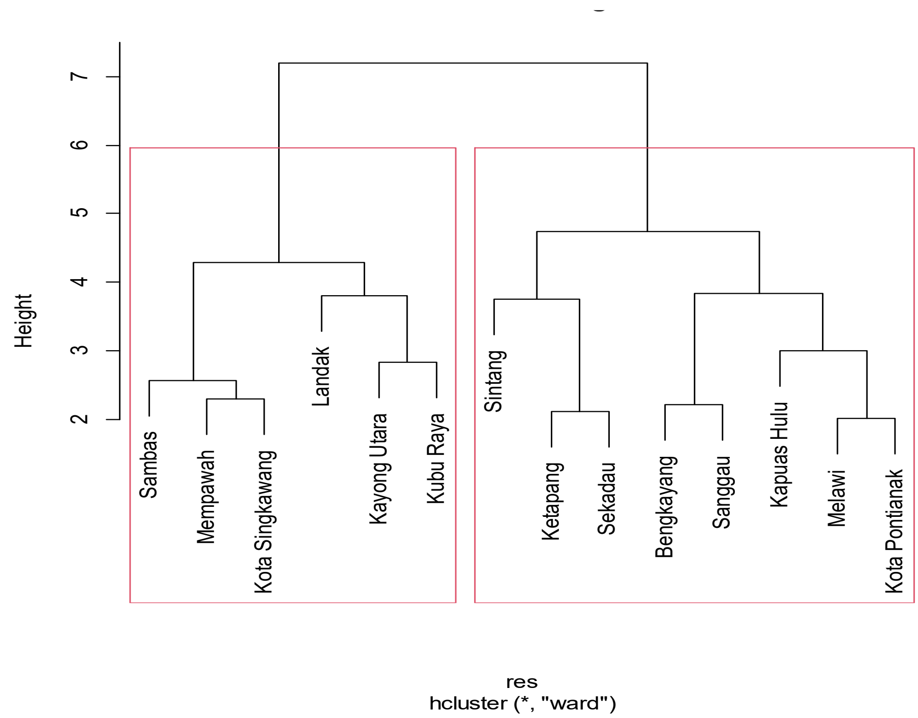

Figure 3. Hierarchical Cluster Analysis Dendrogram
Source: Rstudio

Based on Fig. 3 of the dendrogram that has been generated, it is only possible to group the observation units into 2 classes/segments. Each observation unit is put into a class that shows a smaller CM value. The composition of regencies/cities in the two segments of the cluster analysis results places 6 areas in segment 1. These include Sambas, Mempawah, Singkawang City, Landak, Kayong Utara, and Kubu Raya

Furthermore, if Regency/City has been grouped based on the segments formed, each segment estimates the model in each class/segment (local model) with ordinary PLS. The grouping of Regency/City based on REBUS-PLS can be seen in Table 11.

Table 11. Regency/City by Segment with REBUS PLS

Segment	Multiple Members	Regency/City Name
1	6	Sambas, Kayong Utara, Landak, Kubu Raya, Mempawah, Singkawang City
2	8	Bengkayang, Ketapang, Melawi, Sintang, Sanggau, Sekadau, Kapuas Hulu, Pontianak City

Based on Table 11, the regions included in Segment 1 consist of 6 regencies/cities, namely Sambas, Kayong Utara, Landak, Kubu Raya, Mempawah, Singkawang City. Regions included in segment 2 consist of 8 regencies/cities, namely Bengkayang, Ketapang, Melawi, Sintang, Sanggau, Sekadau, Kapuas Hulu, Pontianak City.

Based on Fig. 4 the Regions included in Segment 1 consist of 6 Regency/City, namely namely Sambas, Kayong Utara, Landak, Kubu Raya, Mempawah, and Singkawang City. Regions included in group 2 consist of 8 regencies/cities, namely Bengkayang, Ketapang, Melawi, Sintang, Sanggau, Sekadau, Kapuas Hulu, and Pontianak City. In developing countries, especially Indonesia, sanitation problems arise due to the lack of attention and concern of the government and health services related to the sanitation sector due to the lack of clean water, waste disposal facilities, and public services in public places such as schools, hospitals, health centers, and other places. Therefore, the availability of sanitation facilities such as drinking water, wastewater treatment, access to latrines, and waste disposal can prevent disease [29]. One of the areas with poor sanitation is the coastal area. Coastal areas often face problems such as uninhabited housing, lack of access to health services, economic problems, and environmental sanitation problems [30]. Coastal areas have unique problems, challenges, and opportunities that distinguish them from other regions [31].

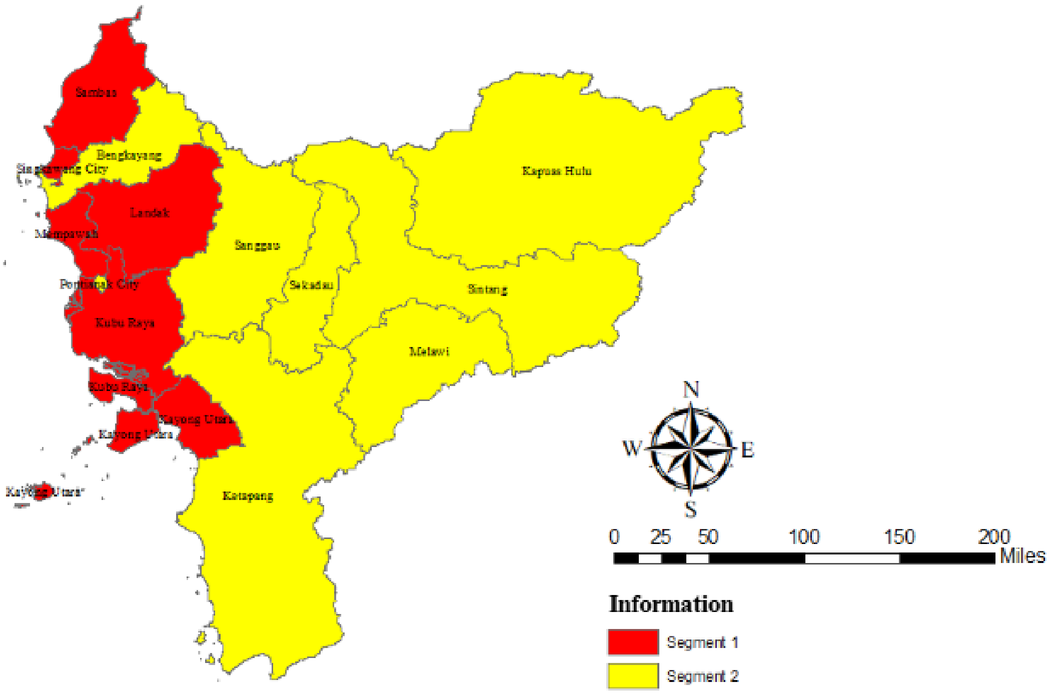

Figure 4. Map of Regency/City Segmentation Results Based on REBUS-PLS
Source: arcGIS

3.4 REBUS PLS Interpretation

3.4.1 Heterogeneity in Structural Equation

To evaluate the heterogeneity of the structural equation, the parameter coefficient values of the global model and two local model are presented in Table 12.

Table 12. Comparison of Path Coefficient Values of Global Model and 2 Local Models

Relationship Between Variables	Global Model		Local Model 1		Local Model 2	
	Coefficient Parameters	p-value	Coefficient Parameters	p-value	Coefficient Parameters	p-value
EK-PD	8.206*	0.000	8.188*	0.024	8.435*	0.017
EK-FS	6.751*	0.008	5.851	0.167	8.653*	0.011
EK-SH	3.764	1.503	3.281	9.675	4.981	1.424
PD-SH	-3.019*	-1.480	-6.418*	-2.298	2.169	5.169
FS-SH	7.940	5.025	8.060	4.078	2.838	6.313

*Note: The * sign means significant at the level α=5%*

Based on Table 12, it can be concluded that the Household Economy on education (EK-PD) has a significant positive effect in the global model, and two local model. Additionally, household Economics on housing facilities (EK-FS) has a significant positive effect in global model and local model 2. Meanwhile, in the global model and the local model 1, education on the health (PD-SH) has a significant positive effect. However, in local model 2, this effect is not significant.

3.4.2 Heterogeneity in the Measurement Equation

Observations of the loading factor values are presented in Table 13 to evaluate the heterogeneity of the measurement equation.

Table 13. Comparison of Standardized Loading Factor of Global Model and 2 Local Models

Latent Variable	Manifest Variable	Global Model		Local Model 1		Local Model 2	
		λ	Conclusion	λ	Conclusion	λ	Conclusion
Household Economics (ξ_1)	EK1	-0.936	**Invalid**	-0.935	**Invalid**	-0.970	**Invalid**
	EK2	0.875	Valid	0.829	Valid	0.923	Valid
	EK3	0.871	Valid	0.740	Valid	0.936	Valid
	EK4	0.776	Valid	0.774	Valid	0.909	Valid
	EK5	0.887	Valid	0.870	Valid	0.892	Valid
	EK6	0.813	Valid	0.745	Valid	0.924	Valid
Housing Facilities (η_1)	FS1	0.763	Valid	0.868	Valid	0.703	Valid
	FS2	0.674	Valid	0.925	Valid	0.965	Valid
	FS3	0.866	Valid	0.833	Valid	0.951	Valid
	FS4	0.593	Valid	0.674	Valid	0.929	Valid
Education (η_2)	PD1	0.724	Valid	0.648	Valid	0.788	Valid
	PD2	0.933	Valid	0.954	Valid	0.906	Valid
	PD3	0.890	Valid	0.913	Valid	0.940	Valid
Housing Environmental Health (η_3)	SH1	0.864	Valid	0.850	Valid	0.891	Valid
	SH2	0.883	Valid	0.856	Valid	0.893	Valid

Based on Table 13, in the local model or segment 1, nearly all loading factor values are ≥ 0.5; except for one manifest variable (EK1). Similarly, in the local model or segment 2, most loading factor values are also ≥ 0.5, with only one manifest variable (EK1) having a loading factor value ≤ 0.5.

3.4.2 Model Evaluation

The structural measure used in evaluating is R-squared (R^2). R-squared (R^2) testing is a way to measure the goodness of a structural model.

Table 14. Comparison of R-Squared Value

Latent Endogen Variable	R^2		
	Global Model	Local Model 1	Local Model 2
Housing Facilities (η_1)	0.455	0.342	0.748
Education (η_2)	0.673	0.670	0.711
Housing Environmental Health (η_3)	0.877	0.956	0.914

Based on Table 14, that the R^2 in each segment formed (local model) is greater for each endogenous latent. In addition, the next evaluation is GoF to validate the overall model and get a better local model than the global one.

Table 15. Comparison of the Goodness of fit Values of Global Model and 2 Local Models

Goodness of fit (GoF)		
Global Model	Local Model 1	Local Model 2
0.678	0.674	0.781

According to Table 15, the GoF value for the global model is 0.678, while for local model 1 is 0.674 and for local model 2 is 0.781. Based on the threshold values proposed by Wetzles (2009), Where GoF values of 0.1,0.25, and 0.36 are considered small, medium, and large, respectively, all three models show high GoF values, indicating that both the structural and measurement models fit the data well. Furthermore, the GoF values for each local model exceed the global model's baseline, particularly local model 2, which reaches 0.781. This indicates that the local models provide an improved fit to the data within their respective segments. It also demonstrates that the model is capable or detecting and adapting to heterogeneity among observation units, which supports the use of a segment-based approach like REBUS-PLS for capturing group-specific characteristics more effectively than a global model [32].

4. CONCLUSION

The conclusions derived from the results and discussion of this research indicate that 13 indicators are valid and reliable in modeling housing environmental health in West Kalimantan. The direct effect analysis revealed several significant relationships, including the impact of household economy on housing facilities, household economy on education, and housing facilities on environmental health. Additionally, the indirect effect analysis showed that household economy significantly influences housing environmental health through housing facilities, indicating that housing facilities serve as a significant mediator. Furthermore, the p-value confirms the significance of the household economy's impact on housing environmental health through education.

Based on the dendrogram that has been generated, the observation units are divided into two classes/segments. Based on the CM value, when 2 classes/segments are formed, the composition of class/segment 1 consists of 6 Regency/City namely Landak, Sanggau, Sekadau, Kayong Utara, Singkawang City, while for class/segment 2 consists of 8 Regency/City namely Pontianak City, Bengkayang, Sanggau, Ketapang, Sintang, Sekadau, Kapuas Hulu, and Melawi. REBUS-PLS detects heterogeneity in the SEM-PLS model with the R^2 value in each segment formed (local model) being greater for each endogenous latent, while the GoF value for both the global model and two local model is in the large GoF category which indicates a better structural and measurement model. In addition, the GoF value for each local model is better than the GoF value for the global model so that heterogeneity in this study can be detected properly.

Author Contributions

Hairil Al-Ham: Conceptualization, Formal Analysis, Methodology, Writing-Original Draft. Neva Satyahadewi: Data Curation, Project Administration, Software, Writing- Review and Editing. Preatin: Formal Analysis, Supervision, Validation, Visualization- Original Draft. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors would like to express their gratitude and appreciation to all those who have already supported the completion of this work.

Declarations

The authors declare no competing interest.

REFERENCES

[1] Suparyanto and Rosad, *PRINSIP KESEHATAN MASYARAKAT*. Bandung: CV Media Indonesia, 2020.

[2] R. Santi, H. Hz, and H. Rawalilah, "PENYULUHAN TENTANG RUMAH SEHAT DALAM UPAYA PENCEGAHAN PENYAKIT BERBASIS LINGKUNGAN DI KELURAHAN 26 ILIR PALEMBANG TAHUN 2023," *SAFARI J. Pengabdi. Masy. Indones.*, vol. 3, no. 3, pp. 207–220, 2023.doi: https://doi.org/10.56910/safari.v3i3.854

[3] K. Risnawati, "ANALISIS KONSEP PENCEGAHAN PERMUKIMAN KUMUH DI KABUPATEN SOPPENG," *J. Al Hadarah Al Islam.*, vol. 2, no. 2, pp. 38–52, 2022.

[4] F. S. Pratiwi, S. Sudarno, and A. Rusgiyono, "PENERAPAN RESPONSE BASED UNIT SEGMENTATION IN PARTIAL LEAST SQUARE (REBUS-PLS) UNTUK ANALISIS DAN PENGELOMPOKAN WILAYAH (STUDI KASUS: KESEHATAN LINGKUNGAN PERUMAHAN DI PROVINSI JAWA TENGAH)," *J. Gaussian*, vol. 9, no. 3, pp. 364–375, 2020, doi: https://doi.org/10.14710/j.gauss.v9i3.28927.

[5] E. D. Anggita, A. Hoyyi, and A. Rusgiyono, "ANALISIS STRUCTURAL EQUATION MODELILING PENDEKATAN PARTIAL LEAST SQUARE DAN PENGELOMPOKAN DENGAN FINITE MIXTURE PLS (FIMIX-PLS) (STUDI KASUS: KEMISKINAN RUMAH TANGGA DI INDONESIA 2017)," *J. Gaussian*, vol. 8, no. 1, pp. 35–45, 2019, doi: https://doi.org/10.14710/j.gauss.v8i1.26620.

[6] D. S. Vanisa, T. R. Rahmanita, E. Ana, and A. Kurniawan, "PROVINCIAL SEGMENTATION IN INDONESIA: EXPLORING FACTORS INFLUENCING EDUCATION WITH SEM-PLS METHOD, INCORPORATING MODERATION EFFECTS AND FIMIX-PLS APPROACH," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 18, no. 3, pp.

1955–1962, 2024, doi: https://doi.org/10.30598/barekengvol18iss3pp1955-1962.

[7] R. R. Marliana, "PARTIAL LEAST SQUARE-STRUCTURAL EQUATION MODELING PADA PARTIAL LEAST SQUARES-STRUCTURAL EQUATION MODELING PADA HUBUNGAN," *J. Mat. Stat. dan Komputasi*, vol. 16, no. 2, pp. 174–186, 2020, doi: https://doi.org/10.20956/jmsk.v16i2.7851.

[8] R. S. Hamid and S. M. Anwar, *STRUCTURAL EQUATION MODELING (SEM) BERBASIS VARIAN: KONSEP DASAR DAN APLIKASI PROGRAM SMART PLS 3.2.8 DALAM RISET BISNIs*. PT Inkubator Penulis Indonesa, 2019.

[9] M. A. Zhiqiang, M. Bwabo, L. I. Mingxing, H. U. Weijun, and F. Panga, "UNOBSERVED HETEROGENEITY IN PUBLIC PROCUREMENT GOVERNANCE AND VALUE FOR MONEY," *Lex Localis*, vol. 18, no. 1, pp. 95–121, 2020, doi: 10.4335/18.1.95-121(2020).

[10] V. . Vinzi, L. Trinchera, S. Squillacciotti, and M. Tenenhaus, "REBUSPLS: A RESPONSE-BASED PROCEDURE FOR DETECTING UNIT SEGMENTS IN PLS PATH MODELLING. MODELLING. APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY, 24," 2008, doi: https://doi.org/10.1002/asmb.728.

[11] L. Zanin, "DETECTING UNOBSERVED HETEROGENEITY IN THE RELATIONSHIP BETWEEN SUBJECTIVE WELL-BEING AND SATISFACTION IN VARIOUS DOMAINS OF LIFE USING THE REBUS-PLS PATH MODELLING APPROACH: A CASE STUDY," *Soc. Indic. Res.*, vol. 110, no. 1, pp. 281–304, 2013, doi: https://doi.org/10.1007/s11205-011-9931-5.

[12] R. A. Z. Tama, M. M. Hoque, Y. Liu, M. J. Alam, and M. Yu, "AN APPLICATION OF PARTIAL LEAST SQUARES STRUCTURAL EQUATION MODELING (PLS-SEM) TO EXAMINING FARMERS' BEHAVIORAL ATTITUDE AND INTENTION TOWARDS CONSERVATION AGRICULTURE IN BANGLADESH," *Agric.*, vol. 13, no. 2, 2023, doi: https://doi.org/10.3390/agriculture13020503.

[13] M. A. Mukid, B. W. Otok, and A. Suharsono, "SEGMENTATION IN STRUCTURAL EQUATION MODELING USING A COMBINATION OF PARTIAL LEAST SQUARES AND MODIFIED FUZZY CLUSTERING," *Symmetry (Basel).*, vol. 14, no. 11, 2022, doi: https://doi.org/10.3390/sym14112431.

[14] C. Quintano and P. Mazzocchi, "EXAMINING A MULTI-DIMENSIONAL UNDECLARED WORK VIA THE REBUS-PLS," *J. Econ. Stud.*, vol. 47, no. 2, pp. 405–430, 2020, doi: https://doi.org/10.1108/JES-08-2018-0288.

[15] E. Piga, "HOME BIAS : PLS PM APPLICATION ON BANKING SECTOR," 2017.

[16] G. Assaker, R. Hallak, and P. O'Connor, "EXAMINING HETEROGENEITY THROUGH RESPONSE-BASED UNIT SEGMENTATION IN PLS-SEM: A STUDY OF HUMAN CAPITAL AND FIRM PERFORMANCE IN UPSCALE RESTAURANTS. CURRENT ISSUES IN TOURISM, 23(2)," 2018, doi: https://doi.org/10.1080/13683500.2018.1490253.

[17] H. A. Reagan, "ANALISIS PENGARUH PERTUMBUHAN EKONOMI TERHADAP PEMBANGUNAN MANUSIA DENGAN RESPONSE BASED UNIT SEGMENTATION IN PARTIAL LEAST SQUARE (REBUS PLS)," Institut Teknologi Sepuluh Nopember, 2016.

[18] X. Y. Xu, S. M. U. Tayyab, Q. D. Jia, and K. Wu, "EXPLORING THE GAMIFICATION AFFORDANCES IN ONLINE SHOPPING WITH THE HETEROGENEITY EXAMINATION THROUGH REBUS-PLS," *J. Theor. Appl. Electron. Commer. Res.*, vol. 18, no. 1, pp. 289–310, 2023, doi: https://doi.org/10.3390/jtaer18010016.

[19] B. W. Otok, Purhadi, R. Sriningsih, and D. S. Dila, "SEGMENTATION OF TODDLER NUTRITIONAL STATUS USING REBUS AND FIMIX PARTIAL LEAST SQUARE IN SOUTHEAST SULAWESI," *MethodsX*, vol. 12, no. September 2023, 2024, doi: https://doi.org/10.1016/j.mex.2023.102515.

[20] X. Angerri and K. Gibert, "VARIABLE SELECTION FOR MEANINGFUL CLUSTERING OF MULTITOPIC TERRITORIAL DATA," *Mathematics*, vol. 11, no. 13, 2023, doi: https://doi.org/10.3390/math11132863.

[21] M. H. Plitt and L. M. Giocomo, "EXPERIENCE-DEPENDENT CONTEXTUAL CODES IN THE HIPPOCAMPUS," *Nat. Neurosci.*, vol. 24, no. 5, pp. 705–714, 2021, doi: https://doi.org/10.1038/s41593-021-00816-6.

[22] Ramlawati and S. Murniati, "THE INFLUENCE OF PRODUCT QUALITY AND PRICE ON PURCHASING DECISIONS," *J. Manag. Res. Stud.*, vol. 1, no. 1, pp. 86–94, 2023, doi: https://doi.org/10.61665/jmrs.v1i1.29.

[23] J. F. Hair, G. T. Hult, C. Ringle, and M. Sarstedt, *A PRIMER ON PARTIAL LEAST SQUARES STRUCTURAL EQUATION MODELING (PLS-SEM) - Joseph F. Hair, Jr., G. Tomas M. Hult, Christian Ringle, Marko Sarstedt*. 2017.doi: https://doi.org/10.15358/9783800653614

[24] M. R. Ab Hamid, W. Sami, and M. H. Mohmad Sidek, "DISCRIMINANT VALIDITY ASSESSMENT: USE OF FORNELL & LARCKER CRITERION VERSUS HTMT CRITERION," *J. Phys. Conf. Ser.*, vol. 890, no. 1, 2017, doi: https://doi.org/10.1088/1742-6596/890/1/012163.

[25] I. Ghozali and H. Latan, "PARTIAL LEAST SQUARE CONCEPTS, METHODS AND APPLICATIONS USING THE WARPPLS 5.0 PROGRAM," *Univ. Diponegoro, Semarang*, vol. 12, no. 1, pp. 1–14, 2014.

[26] M. Ardiana *et al.*, "PENGARUH SELF CONTROL TERHADAP PERILAKU KEUANGAN MAHASISWA FAKULTAS EKONOMI UNIVERSITAS HASYIM ASY'ARI," *Pros. Semin. Nas. Sains, Teknol. Ekon. Pendidik. dan Keagamaan*, vol. 7, pp. 114–121, 2023, [Online]. Available: https://ejournal.unhasy.ac.id/index.php/SAINSTEKNOPAK/article/view/4953 doi: https://doi.org/10.32682/jpekbm.v7i1.3060

[27] I. Ghozali and H. Latan, *PARTIAL LEAST SQUARE KONSEP,TEKNIK, DAN APLIKASI MENGGUNAKAN SMARTPLS 3.0 EDISI 2*. Semarang : Badan Penerbit UNDIP, 2015.

[28] S. Katmawanti, L. Nikmatasari, and S. Nurrochmah, "THE IMPACT OF EDUCATION AND INCOME ON ENVIRONMENTAL HEALTH ASPECTS IN URBAN HOUSEHOLDS," vol. 31, no. Ismophs 2019, pp. 95–100, 2020, doi: https://doi.org/10.2991/ahsr.k.201203.018

[29] C. Annisa and S. Susilawati, "GAMBARAN SANITASI LINGKUNGAN TERHADAP SARANA AIR BERSIH DAN JAMBAN KELUARGA DI KELURAHAN SUKARAJA," *PubHealth J. Kesehat. Masy.*, vol. 1, no. 1, pp. 85–90, 2022, doi: https://doi.org/10.56211/pubhealth.v1i1.46.

[30] Nurhayati, A. M. Nasution, R. Masry, A. D. Rahmadani, and D. P. Sari, "SYSTEMATIC REVIEW : SANITASI LINGKUNGAN DI WILAYAH PESISIR," *J. Kesehat. Tambusai*, vol. 5, no. 2, pp. 4439–4448, 2024.

[31] A. N. I. Berutu and S. Susilawati, "SANITASI LINGKUNGAN PESISIR DI PANTAI GUDANG GARAM DESA KOTA PARI SERDANG BEDAGAI," *Naut. J. Ilm. Multidisiplin Indones.*, vol. 1, no. 5, pp. 253–259, 2022, [Online]. Available: https://jurnal.arkainstitute.co.id/index.php/nautical/article/view/294 doi: https://doi.org/10.55904/nautical.v1i5.294

[32] M. Wetzels, G. Odekerken-Schröder, and C. Van Oppen, "USING PLS PATH MODELING FOR ASSESSING

HIERARCHICAL CONSTRUCT MODELS: GUIDELINES AND EMPIRICAL ILLUSTRATION," *MIS Q. Manag. Inf. Syst.*, vol. 33, no. 1, pp. 177–196, 2009, doi: https://doi.org/10.2307/20650284