MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924

Index: Harvard, Boston, Sydney, Copernicus, Sinta 4, Scilit, Semantic, Google, etc

https://doi.org/10.58578/mikailalsys.v4i3.11886

# Analysis of the Determinants of the Gender Empowerment Index in West Sumatra 2024 Using the Group Lasso Method

**Devi Yopita Sipayung & Fadhilah Fitri**
Universitas Negeri Padang, Indonesia
devisipayung249@gmail.com; fadhilahfitri@fmipa.unp.ac.id

**Article Info:**

| Submitted: | Revised: | Accepted: | Published: |
|---|---|---|---|
| Jul 13, 2026 | Aug 10, 2026 | Aug 22, 2026 | Aug 27, 2026 |

**Abstract**

Gender empowerment remains a multidimensional development challenge in West Sumatra, where regional disparities reflect interconnected educational, economic, political, health, demographic, and infrastructural factors. This study aims to model the Gender Empowerment Index (GEI) across 19 regencies and municipalities in West Sumatra in 2024, identify its relevant determinants, and determine the dominant predictor using Group LASSO. A quantitative research design was employed using secondary data from Statistics Indonesia, comprising 22 predictors classified into six dimensions. All 19 regions were included through census sampling. The data were standardised, assessed for multicollinearity, and analysed using Group LASSO, with K-fold cross-validation applied to determine the optimal penalty parameter. The optimal λ value of 0.2392388 retained 14 predictors and produced a mean squared error of 1.190105 and an $R^2$ of 0.985264. The proportion of women serving in Regional People's Representative Councils emerged as the dominant predictor, with the largest coefficient of 0.892008. These findings demonstrate the utility of Group LASSO for selecting relevant predictors in a multidimensional regional dataset and provide empirical evidence for establishing gender-empowerment policy


Mikailalsys is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

priorities. The results particularly underscore the importance of women's political representation in efforts to strengthen gender empowerment across West Sumatra.

**Keywords:** Gender Empowerment Index; Group LASSO; Variable Selection; Women's Political Representation; West Sumatra

## INTRODUCTION

Gender empowerment represents an important dimension of human development because it reflects the extent to which women have equitable access to opportunities, participation, resources, and decision-making processes across social, economic, and political domains. Women's empowerment is not limited to improving material well-being but also involves strengthening women's capabilities to make choices and actively shape their own lives and communities (Jain, 2023). From this perspective, gender empowerment is a multidimensional process involving access to resources, agency, and the achievement of meaningful outcomes in social and economic life (Kaseng, 2023). The active participation of women is also essential to inclusive and sustainable development because women's capacities and involvement influence the quality of decision-making at different levels of society (Nations & Programme, 2025). Accordingly, measuring gender empowerment is necessary to assess whether development processes have provided women with adequate opportunities to participate on an equal basis.

In Indonesia, gender empowerment continues to face structural challenges related to unequal access to education, employment, economic resources, and strategic decision-making positions. These conditions indicate that formal commitments to gender equality do not necessarily guarantee equal opportunities in the implementation of development policies (KPPPA, 2020). One important indicator used to assess women's involvement in development is the Gender Empowerment Index (GEI), or Indeks Pemberdayaan Gender (IDG), which captures women's participation in decision-making, professional employment, and their contribution to labor income (BPS, 2024). Higher GEI values indicate greater participation of women in strategic economic and political domains and therefore provide an important representation of women's contribution to development (Kurnianingsih et al., 2022). The index consequently serves not merely as a statistical measure but also as an

instrument for evaluating the extent to which development processes have created opportunities for women to exercise agency and participate meaningfully in public life.

Gender empowerment becomes more complex when examined across regions because social, economic, demographic, and institutional characteristics vary considerably among geographical areas. Previous research has demonstrated differences in gender development indicators across regions, reflecting disparities in the level and characteristics of regional development (Aprilianti & Setiadi, 2020). Socioeconomic conditions are also associated with gender empowerment, particularly through education, income, and employment opportunities available to women (Yustie & Prayitno, 2024). This issue is particularly relevant in West Sumatra, where women occupy an important position within the social and family structure, while their participation in formal employment remains subject to socioeconomic constraints (Ariansyah & Satria, 2024). Examining the determinants of the GEI at the regency and municipal levels is therefore important because regional differences should not only be described but also explained through the factors that potentially shape women's participation and empowerment.

The determinants of gender empowerment are inherently multidimensional. Women's educational attainment, for instance, can increase their capacity and opportunities to participate in economic, social, and political activities (Cahyati et al., 2019). Economic participation is similarly important because women's involvement in the labor market and their contribution to household or labor income can strengthen their economic independence (Putrie & Rahman, 2021). Women's wages may further influence their ability to access economic resources and improve their socioeconomic position (Riadina & Sugianto, 2024). Employment conditions must also be considered because gender disparities in unemployment may indicate unequal access to labor-market opportunities (Novianti, 2019). These findings suggest that the GEI should be examined through a set of interconnected determinants rather than through a single explanatory indicator.

Beyond education and employment, political participation, health and reproductive conditions, demographic characteristics, family structure, and access to development infrastructure may also contribute to women's empowerment. Women's representation in political institutions can expand their opportunities to participate in public decision-making and influence development policies (Syukri, 2023). Reproductive empowerment is likewise associated with women's capacity to make decisions concerning fertility and reproductive

outcomes (Taqwim et al., 2025). Participation in family-planning programs can further strengthen women's decision-making capacity regarding health and family matters (Mutmainnah et al., 2021). In addition, access to higher education can improve women's human capital and expand their opportunities to participate in development activities (Valentina et al., 2022). The diversity of these determinants indicates that GEI modeling may involve substantial interrelationships among predictors, creating methodological challenges for conventional regression analysis.

Previous studies have examined gender empowerment from several empirical perspectives. Rusli and Magna (2023), for example, investigated the determinants of the GEI in Magelang City and identified the importance of economic and political dimensions in explaining gender empowerment. Research conducted in Aceh also examined factors associated with the GEI and further confirmed the relevance of socioeconomic conditions in explaining regional gender empowerment (Hafifa et al., 2026). Other research has approached gender development through multivariate classification to identify similarities and differences among Indonesian regions based on gender-related indicators (Salsabila & Hendrawan, 2021). Although these studies have contributed substantially to understanding gender empowerment, most have relied on conventional regression or classification approaches that do not explicitly preserve theoretically defined groups of correlated predictors during variable selection.

Research gap therefore arises from the limited application of a group-based regularization framework to identify determinants of the GEI while simultaneously preserving the conceptual structure of gender-related indicators. Existing GEI studies have generally examined individual relationships among predictors or applied conventional statistical approaches, whereas Group LASSO studies have predominantly addressed other substantive fields and have not specifically focused on the GEI across regencies and municipalities in West Sumatra. Strong correlations among predictors can produce unstable parameter estimates and make it difficult to distinguish the individual contributions of explanatory variables (Daoud, 2018). This problem is particularly relevant to the present study because gender empowerment indicators originate from several dimensions that are conceptually and empirically interconnected. A methodological approach capable of performing variable selection while retaining meaningful predictor groups is therefore required.

The Group Least Absolute Shrinkage and Selection Operator (Group LASSO) provides such an approach by extending penalized regression to perform variable selection according to predefined groups (Yuan & Lin, 2006). Unlike ordinary LASSO, which selects predictors individually, Group LASSO can retain or eliminate variables collectively according to their membership in a particular group, thereby preserving the theoretical structure of related indicators (Huang et al., 2012). The method is particularly useful for high-dimensional settings in which predictors may exhibit strong correlations because its penalization mechanism can reduce model complexity while supporting more stable estimation (Buch et al., 2023). Previous empirical work has also demonstrated the ability of Group LASSO to identify influential groups of predictors and produce more parsimonious models without substantially reducing predictive information (Wulandari et al., 2020). These methodological characteristics make Group LASSO appropriate for examining GEI determinants organized into education, economy and employment, politics and leadership, health and reproduction, demography and family, and infrastructure and access.

Novelty of the present study lies in applying Group LASSO to identify influential groups and dominant predictors of the GEI across all regencies and municipalities in West Sumatra in 2024. The contribution is not limited to introducing a regularization technique but also involves integrating the theoretical multidimensional structure of gender empowerment with group-based statistical variable selection. Grouped predictor methods have been increasingly applied to high-dimensional settings in which predictors share substantive or statistical relationships (Belhechmi et al., 2020). Such approaches are particularly relevant when the structure of predictors contains meaningful group information that should be incorporated into the selection process rather than ignored (Dang et al., 2021). This study therefore offers a more structured framework for determining which dimensions of gender empowerment should be retained in the model and which individual predictors contribute most strongly to explaining regional variation in the GEI.

Accordingly, this study focuses on modeling the Gender Empowerment Index across the 19 regencies and municipalities of West Sumatra in 2024 using Group LASSO. The analysis includes 22 predictor variables organized into six dimensions: education; economy and employment; politics and leadership; health and reproduction; demography and family; and infrastructure and access. Using all administrative regions as observations provides a comprehensive representation of regional variation in gender empowerment. Group LASSO is employed to address multicollinearity while simultaneously selecting relevant predictor

groups and variables, consistent with its application to structured high-dimensional data (Padhilah & Herrhyanto, 2024). The study therefore aims to apply Group LASSO to model the GEI, identify the factors associated with gender empowerment in West Sumatra, and determine the most dominant variable in explaining variations in the GEI across regencies and municipalities in 2024.

## METHODS

This study employed an applied quantitative research design aimed at identifying the determinants of the Gender Empowerment Index (GEI) across regencies and municipalities in West Sumatra Province in 2024. A descriptive approach was used to describe the characteristics of the GEI and its predictor variables, while inferential analysis was conducted to identify the factors associated with the GEI using the Group Least Absolute Shrinkage and Selection Operator (Group LASSO) method. Group LASSO was selected because it enables variable selection at the group level while addressing multicollinearity among predictor variables, thereby facilitating the development of a more parsimonious and interpretable model.

The research design was based on secondary data obtained from official publications of Statistics Indonesia (Badan Pusat Statistik/BPS) for 2024, particularly *West Sumatra in Figures 2024*. The dependent variable was the Gender Empowerment Index (GEI), while the independent variables consisted of 22 indicators organized into six dimensions reflecting the conceptual structure of gender empowerment: education; economy and employment; politics and leadership; health and reproduction; demography and family; and infrastructure and access. The education dimension comprised women's mean years of schooling, female literacy among the population aged 15–64 years, the proportion of women completing senior high school or higher education, and women's expected years of schooling. The economy and employment dimension included female labor force participation, the open unemployment rate, the proportion of women working as professionals, women's contribution to income, and the average net monthly wage of female employees. The remaining dimensions covered political participation and leadership, health and reproductive conditions, demographic and family characteristics, and infrastructure and regional access.

The study population comprised all 19 regencies and municipalities in West Sumatra Province in 2024. A census sampling technique was employed, whereby every member of

the population was included in the analysis. Consequently, the study consisted of 19 observations and 22 independent variables. The census approach was appropriate because the entire population of administrative regions was available and could be incorporated into the analysis, eliminating the need to select only a subset of regions.

Data were collected through documentation of secondary data reported in official BPS publications. The collected data were first examined for completeness and subsequently organized into a structured dataset to ensure that all variables were ready for statistical analysis. Descriptive statistics were then calculated using the minimum, maximum, mean, and standard deviation to summarize the characteristics of each variable. Multicollinearity was assessed using the Variance Inflation Factor (VIF), with a VIF value below 10 considered indicative of the absence of problematic multicollinearity. Prior to Group LASSO modeling, all predictor variables were standardized using the Z-score transformation to place variables measured on different scales on a comparable basis.

The main statistical analysis was performed using RStudio and the Group LASSO method. Predictor variables were initially classified into groups according to their shared characteristics or represented dimensions. The Group LASSO model was subsequently estimated by incorporating a penalty function to identify groups of predictors that contributed to the GEI model. The optimal penalty parameter, $\lambda$, was selected using K-Fold Cross-Validation by identifying the value that produced the lowest prediction error. Model performance was subsequently evaluated using the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination. The best-performing model was determined based on the lowest MSE and RMSE values and the highest coefficient of determination. Finally, the selected predictor groups, the direction and magnitude of regression coefficients, and the model's ability to explain variation in the GEI were interpreted.

The analysis used 2024 data; therefore, the units of analysis represented the conditions of regencies and municipalities in West Sumatra during that year. The available Chapter III does not specify the particular starting and ending months of the research or data-processing period. Accordingly, no specific research duration is stated here to avoid introducing information that is not documented in the thesis. Overall, the research procedure consisted of data collection and preparation, descriptive analysis, multicollinearity

assessment, data standardization, predictor grouping, Group LASSO modeling, optimal model selection, model evaluation, and interpretation of the resulting model.

## RESULTS

### Initial Exploration

In the LASSO group method, independent variables are grouped based on their characteristics and theoretical relationships. The variable selection process is then performed at the group level, allowing variables within the same group to be selected or eliminated simultaneously. In this study, the research variables were grouped into six main factors, as presented in Table 1.

**Table 1. Grouping of Research Variables**

| Group | Variables | Factor |
|---|---|---|
| 1 | X1, X2, X3, X4 | Education |
| 2 | X5, X6, X7, X8, X9 | Economy and Employment |
| 3 | X10, X11, X12 | Politics and Leadership |
| 4 | X13, X14, X15, X16 | Health and Reproduction |
| 5 | X17, X18 | Demographics and Family |
| 6 | X19, X20, X21, X22 | Infrastructure and Access |

*Source: Processed by the researcher*

Table 1. Grouping of Research Variables shows The research variables were grouped into six main factors based on their theoretical characteristics and relevance to gender empowerment. Group 1 consists of variables related to Education (X1–X4), Group 2 represents Economy and Employment (X5–X9), and Group 3 covers Politics and Leadership (X10–X12). Group 4 includes variables related to Health and Reproduction (X13–X16), Group 5 represents Demographics and Family (X17–X18), and Group 6 consists of variables related to Infrastructure and Access (X19–X22).

This grouping serves as the basis for the LASSO group method, in which variable selection is performed at the group level. Thus, variables within the same group are considered collectively and may be selected or eliminated simultaneously. The grouping reflects the theoretical dimensions associated with gender empowerment and facilitates the identification of the most relevant groups of variables for the subsequent analysis.

## Determination of the Optimal Penalty Parameter

The optimal penalty parameter was determined using the *cross-validation* method. The results showed that the optimal penalty parameter was **λ =** 0.2392388, which produced the smallest *cross-validation error* and was therefore selected for the Group LASSO model

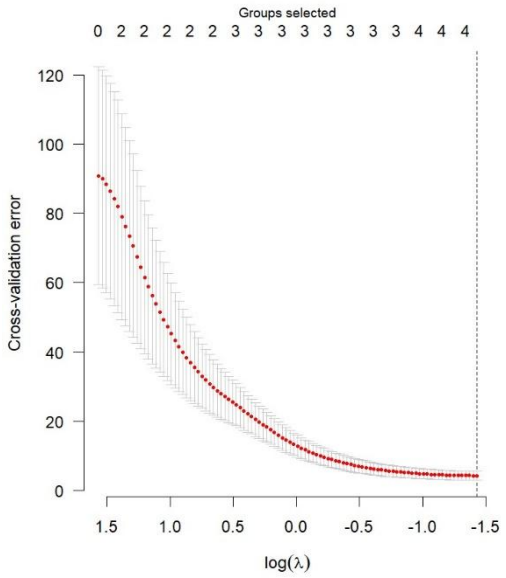



**Figure 1. *Cross-Validation* Graph of the Group LASSO Model**

Figure 1. *Cross-Validation* Graph of the Group LASSO Model shows the relationship between λ and the *cross-validation error*. The error decreases until reaching its minimum at **λ =** 0.2392388. This parameter was then used to identify the variable groups that contribute to the Gender Development Index (IDG) in West Sumatra in 2024.

## LASSO Group Model Coefficient Estimation Results

After determining the optimal penalty parameter, the coefficients of the Group LASSO model were estimated. The results are presented in the following table.

**Table 2. Results of the Group LASSO Model Coefficients**

| Variable | Coefficient | Information |
|---|---|---|
| Intercept | -6.462451 | Constant |
| X1 | 0.5411399 | Selected |
| X2 | 0.5166864 | Selected |
| X3 | -0.0598692 | Selected |
| X4 | 0.1113406 | Selected |
| X5 | -0.02356711 | Selected |
| X6 | 0.1611925 | Selected |

| Variable | Coefficient | Information |
|---|---|---|
| X7 | -0.2334996 | Selected |
| X8 | 0.5181229 | Selected |
| X9 | 0.00002456 | Selected |
| X10 | 0.8920078 | Selected |
| X11 | 0.0008164 | Selected |
| X12 | -0.00007847 | Selected |
| X13–X16 | 0 | Not Selected |
| X17 | -0.0002538317 | Selected |
| X18 | -0.01973620 | Selected |
| X19–X22 | 0 | Not Selected |

*Source: Processed by the researcher*

Table 2. Results of the Group LASSO Model Coefficients show that X13, X14, X15, X16, X19, X20, X21, and X22 have coefficients of zero and were therefore not selected by the Group LASSO model. Meanwhile, the variables with non-zero coefficients were selected as relevant variables for explaining the Gender Development Index (IDG) in West Sumatra in 2024. For the education factor, X1, X2, and X4 have positive coefficients, indicating that improvements in women's education are associated with higher IDG values. In the economy and employment factor, X5, X6, X7, X8, and X9 were selected, indicating the relevance of women's participation in economic activities. Meanwhile, in the politics and leadership factor, X10 has a positive coefficient, indicating that women's representation in political decision-making contributes to higher IDG.

**Group LASSO Regression Model**

The Group LASSO regression model was used to identify variables associated with the Gender Empowerment Index (IDG) in West Sumatra in 2024. Variables with non-zero coefficients were retained in the model, while variables with zero coefficients were excluded. The resulting Group LASSO regression model is presented as follows:

$$\begin{aligned}\hat{Y} = {} & -6{,}462451 + 0{,}5411399X_1 + 0{,}5166864X_2 - 0{,}0598692X_3 + 0{,}1113406X_4 \\ & - 0{,}02356711X_5 + 0{,}1611925X_6 - 0{,}2334996X_7 + 0{,}5181229X_8 \\ & + 0{,}00002456X_9 + 0{,}8920078X_{10} + 0{,}0008164X_{11} - 0{,}00007847X_{12} \\ & - -0.0002538317X_{17} - 0{,}01973620X_{18}\end{aligned}$$

The results show that X1, X2, X4, X6, X8, X10, and X11 have positive coefficients, indicating a positive relationship with IDG. Meanwhile, X3, X5, X7, X12, X17, and X18

have negative coefficients, indicating an inverse relationship. The variable X10 (the proportion of women in the Regional People's Representative Council) has the largest positive coefficient (0.8920078), suggesting its important contribution to the IDG. Overall, the results indicate that education, economic and employment conditions, political representation, and demographic factors are associated with gender empowerment in West Sumatra.

## Model Prediction Results

The *Group LASSO* model was used to predict the Gender Empowerment Index (GEM) for 19 districts/cities in West Sumatra in 2024. The prediction results are presented in Table 3.

**Table 3. Group LASSO Model Prediction Results**

| Observation | Prediction Value | Observation | Prediction Value |
|---|---|---|---|
| 1 | 49.79298 | 11 | 51.58790 |
| 2 | 56.59228 | 12 | 62.08485 |
| 3 | 60.49720 | 13 | 62.87605 |
| 4 | 57.05586 | 14 | 65.33944 |
| 5 | 60.59475 | 15 | 63.56715 |
| 6 | 50.25285 | 16 | 75.24583 |
| 7 | 64.47214 | 17 | 83.10106 |
| 8 | 51.58222 | 18 | 66.67865 |
| 9 | 66.34601 | 19 | 59.55436 |
| 10 | 51.56843 | | |

*Source: Processed by the researcher*

Table 3. Group LASSO Model Prediction Results shows the predicted GEM values ranged from 49.79 to 83.10, indicating differences in gender empowerment across districts/cities. The selected variables, including education, economic and employment factors, and politics and leadership, contributed to the predicted GEM values. Overall, the *Group LASSO* model was able to capture the relationship between the predictor variables and GEM and produced reasonably close predictions to the actual values.

## Model Performance Evaluation

The performance of the *Group LASSO* model was evaluated using the Mean Squared Error (MSE) and coefficient of determination ($R^2$). The results are presented in Table 4.

**Table 4. Model Performance Evaluation**

| Evaluation Measure | Value |
|---|---|
| MSE | 1.190105 |
| $R^2$ | 0.9852636 |
| Percentage | 98.53% |

*Source: Processed by the researcher*

Table 4. Model Performance Evaluation shows the model obtained an MSE of 1.190105, indicating a relatively small prediction error. Meanwhile, the $R^2$ value of 0.9852636 (98.53%) shows that 98.53% of the variation in the Gender Empowerment Index (GEM) is explained by the variables included in the model, while the remaining 1.47% is attributed to factors outside the model. These results indicate that the *Group LASSO* model has very good predictive and explanatory performance.

## DISCUSSION

The findings demonstrate that the application of the Group Least Absolute Shrinkage and Selection Operator (Group LASSO) produced an optimal penalty parameter of $\lambda = 0.2392388$. This value was selected through *K-Fold Cross-Validation* because it yielded the lowest validation error among the evaluated penalty values. At the optimal penalty level, 14 of the 22 predictor variables were retained in the final model, whereas eight variables were assigned zero coefficients and consequently excluded. This finding indicates that the indicators theoretically associated with gender empowerment do not contribute equally to explaining variations in the Gender Empowerment Index (GEI) across regencies and municipalities in West Sumatra. Such selection is particularly important because the dataset exhibited substantial correlations among predictors, with several pairs showing correlation coefficients above 0.90. High multicollinearity can make regression estimates unstable and complicate the interpretation of individual predictors when conventional regression is applied without an appropriate selection mechanism (Sohil et al., 2022). The Group LASSO results therefore suggest that the multidimensional nature of gender empowerment should be accompanied by a statistically structured selection process.

The presence of multicollinearity further supports the methodological suitability of Group LASSO in this study. The analysis identified several predictors with VIF values exceeding 10, while some pairs of variables exhibited extremely strong linear relationships, with the highest correlation reaching 0.996. Substantively, such relationships may arise

because regional indicators of gender empowerment are not independent of one another. Educational conditions, employment opportunities, economic characteristics, and broader social conditions may develop simultaneously within the same regional context. Under such circumstances, conventional regression may have difficulty distinguishing the unique contribution of highly correlated predictors. Group LASSO offers an alternative by applying the penalization mechanism at the group level, allowing variable selection to incorporate the substantive structure established before model estimation (Yuan & Lin, 2006). The findings therefore demonstrate not only the usefulness of variable selection but also the importance of incorporating theoretically meaningful predictor structures into the statistical modeling process.

At the group level, four dimensions remained represented in the final model: women's education, economy and employment, health and reproduction, and demography and family. In contrast, the politics and leadership and infrastructure and access groups were not retained in the final model. This result should not be interpreted as evidence that political participation, leadership, infrastructure, or access are conceptually irrelevant to women's empowerment. Rather, within the specific configuration of the 19 regencies and municipalities in West Sumatra in 2024, and after accounting for correlations among predictors, these groups did not provide sufficient additional statistical information to remain in the penalized model. This illustrates an important characteristic of regularized modeling, in which relatively weak contributions can be removed to obtain a more parsimonious model (Buch et al., 2023). The findings consequently indicate that the variation in the GEI within the study context was more strongly represented by educational, economic-employment, health-reproductive, and demographic-family dimensions.

Particular attention should be given to the political dimension because the proportion of women in regency/municipal legislative councils (DPRD) recorded the largest coefficient in the model, at 0.892008. Among the predictors retained by the model, this coefficient indicates the strongest contribution to explaining variation in the GEI after penalization and standardization. The finding highlights the strategic relevance of women's representation in decision-making institutions within the context of gender empowerment. This result is consistent with Rusli and Magna (2023), who found that women's representation in parliament was among the significant factors associated with the GEI in Magelang City. The present study extends that evidence by showing that women's political representation

remains prominent when a broader set of socioeconomic, educational, health, and demographic indicators is considered simultaneously through a Group LASSO framework.

Nevertheless, an important distinction emerges between the group-selection results and the interpretation of individual coefficients. Although the study's conclusion states that the politics and leadership group was eliminated, the coefficient table identifies the proportion of women in the DPRD as a selected variable with the largest coefficient. This discrepancy should be interpreted carefully because the two statements cannot be treated as equivalent without clarifying the underlying group assignment and model output. Based on the available coefficient table, the DPRD representation variable has a coefficient of 0.892008 and is classified as selected. Therefore, in a journal article, it would be methodologically safer to report the individual coefficient and the group-selection result precisely rather than simultaneously stating that the entire political group was eliminated and that its DPRD indicator was the dominant predictor without qualification. This distinction is important for maintaining consistency between the statistical output and the substantive interpretation.

The comparison with previous literature indicates that the present findings broadly reinforce the relevance of socioeconomic and political dimensions to gender empowerment. Hafifah et al. (2025), in their study of Aceh Province, found that economic and political factors were significantly associated with the GEI, while Rusli and Magna (2023) identified women's professional participation, women's income contribution, and parliamentary representation as important determinants of the GEI. The present study supports this general empirical pattern while extending it methodologically. Rather than examining predictors separately through conventional linear regression, the study evaluates a structured set of correlated indicators and applies group-based penalization to determine which predictors should remain in the model. Its contribution therefore lies not merely in confirming the relevance of socioeconomic and political conditions, but in demonstrating how these determinants can be statistically screened within a correlated, multidimensional predictor structure.

From a methodological perspective, the findings highlight the value of Group LASSO for gender empowerment research. Usman (2024) demonstrated that Group LASSO can effectively perform variable selection when predictors can be organized into meaningful groups. Similarly, Sunandi et al. (2021) highlighted the usefulness of Group LASSO for

grouped variable selection in regression models. In the present study, the approach reduced the original predictor set from 22 variables to 14 retained variables, thereby simplifying the model while preserving predictors from the dimensions that contributed to explaining GEI variation. The resulting RMSE of 1.090919 and $R^2$ of 0.985264 indicate that the model explains approximately 98.53% of the variation in the GEI within the analyzed data, with an average prediction error of approximately 1.09 index points. However, the very high $R^2$ should not be interpreted as evidence of causality, because the coefficient of determination primarily reflects the proportion of observed response variation explained by the model (Akoglu, 2018).

Theoretically, this study contributes to gender empowerment research by demonstrating that GEI determinants can be conceptualized simultaneously as a multidimensional substantive construct and as a statistically structured predictor system. Group LASSO enables education, economy and employment, health and reproduction, demography and family, politics and leadership, and infrastructure and access to be incorporated as conceptually meaningful groups rather than treated exclusively as unrelated individual variables. Practically, the findings may provide supporting evidence for regional policymakers when identifying areas requiring greater attention in gender empowerment programs. The prominent coefficient associated with women's representation in the DPRD suggests that women's participation in decision-making institutions deserves continued policy attention. At the same time, the retention of education, economic-employment, health-reproductive, and demographic-family dimensions indicates that gender empowerment should not be approached through a single policy sector.

Several limitations should nevertheless be considered when interpreting these findings. First, the analysis covers only 19 regencies and municipalities in West Sumatra and relies on data from a single year, 2024, resulting in a relatively small number of observations compared with the number of predictors examined. Second, the cross-sectional design does not allow the analysis to capture changes in the determinants of gender empowerment over time. Third, the substantial correlations among predictors require caution when interpreting individual coefficients because the retained coefficient reflects the model's penalized structure rather than an isolated causal effect. These limitations constrain the extent to which the findings can be generalized beyond the study context. Consistent with the recommendations in the thesis, future research should consider broader geographical coverage, longer observation periods, panel data, and direct comparisons between Group

LASSO and alternative variable-selection methods. The principal contribution of the present study therefore lies in applying group-based regularization to identify the structure of GEI determinants in West Sumatra while providing a methodological foundation for more extensive longitudinal and comparative research.

## CONCLUSION

This study aimed to model the Gender Empowerment Index (GEI) across regencies and municipalities in West Sumatra Province in 2024 using the Group Least Absolute Shrinkage and Selection Operator (Group LASSO), identify the factors contributing to the GEI, and determine the most dominant predictor retained in the model. The findings indicate that the optimal Group LASSO model was obtained at a penalty parameter (*lambda*) of 0.239 through a cross-validation procedure. Of the 22 predictor variables initially examined, 14 were retained in the final model, while the remaining variables were eliminated because their estimated coefficients were equal to zero. These findings demonstrate that Group LASSO can simplify the predictor structure while performing group-based variable selection and addressing multicollinearity within the model.

The selected predictors indicate that women's education, economy and employment, health and reproduction, and demography and family constitute the factors represented in the final GEI model. In contrast, the politics and leadership and infrastructure and access dimensions were not retained by the Group LASSO procedure. Among the variables selected, the proportion of women serving in regency/municipal legislative councils (DPRD) had the largest coefficient, at 0.892008, making it the most dominant predictor in explaining variations in the GEI across the study area. This finding highlights the importance of women's representation in decision-making institutions within the context examined.

The study contributes methodologically by applying Group LASSO to GEI modeling while organizing predictors according to the multidimensional structure of gender empowerment. In practical terms, the findings may provide supporting evidence for regional stakeholders in identifying priorities for strengthening women's empowerment, particularly by considering the dimensions and predictors retained in the model. Future research should compare Group LASSO with alternative variable-selection methods to assess differences in model performance and accuracy. Further studies may also extend the geographical coverage or observation period and employ panel data to increase the number of observations,

improve model stability, and capture changes in the determinants of gender empowerment over time.

## REFERENCES


Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine, 18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Aprilianti, S., & Setiadi, Y. (2022). Faktor-Faktor yang Memengaruhi Indeks Pembangunan Gender di Indonesia Tahun 2020. *Seminar Nasional Official Statistics, 2022*(1), 245–254. https://doi.org/10.34123/semnasoffstat.v2022i1.1351

Ariansyah, F., & Satria, D. (2024). Faktor-Faktor yang Mempengaruhi Partisipasi Tenaga Kerja Wanita di Indonesia. *Media Riset Ekonomi Pembangunan (MedREP), 1*(4), 738–747. https://medrep.ppj.unp.ac.id/index.php/MedREP/article/view/123

Belhechmi, S., De Bin, R., Rotolo, F., & Michiels, S. (2020). Accounting for grouped predictor variables or pathways in high-dimensional penalized Cox regression models. *BMC Bioinformatics, 21*, Article 277. https://doi.org/10.1186/s12859-020-03618-y

Buch, G., Schulz, A., Schmidtmann, I., Strauch, K., & Wild, P. S. (2023). A systematic review and evaluation of statistical methods for group variable selection. *Statistics in Medicine, 42*(3), 331–352. https://doi.org/10.1002/sim.9620

Cahyati, C., Herrhyanto, N., & Puspita, E. (2019). Pemodelan Indeks Pembangunan Gender (IPG) dengan Menggunakan Regresi Probit Ordinal (Studi Kasus IPG Kabupaten/Kota di Pulau Sumatera Tahun 2015). *Jurnal EurekaMatika, 7*(2), 83–99. https://doi.org/10.17509/jem.v7i2.22137

Dang, X., Huang, S., & Qian, X. (2021). Penalized Cox's proportional hazards model for high-dimensional survival data with grouped predictors. *Statistics and Computing, 31*(6), Article 77. https://doi.org/10.1007/s11222-021-10052-4

Daoud, J. I. (2017). Multicollinearity and regression analysis. *Journal of Physics: Conference Series, 949*, Article 012009. https://doi.org/10.1088/1742-6596/949/1/012009

Hafifa, R., Nabila, D. R., Hanum, N., Asnidar, A., Andiny, P., & Safuridar, S. (2026). Faktor-Faktor yang Mempengaruhi Indeks Pemberdayaan Gender di Provinsi Aceh. *Jurnal Ilmiah Ekonomi dan Manajemen, 4*(1), 132–143. https://doi.org/10.61722/jiem.v4i1.7151

Huang, J., Breheny, P., & Ma, S. (2012). A selective review of group selection in high-dimensional models. *Statistical Science, 27*(4), 481–499. https://doi.org/10.1214/12-STS392

Jain, M. (2023). Women empowerment: A multidimensional approach. *Journal of Women Empowerment and Studies, 3*(4), 36–42. https://doi.org/10.55529/jwes.34.36.42

Kaseng, E. S. (2023). Keberdayaan Perempuan: Konsep Pemberdayaan dalam Pembangunan Masyarakat Berbasis Gender. *Jurnal Kajian Sosial dan Budaya: Tebar Science, 7*(3), 117–127. https://doi.org/10.36653/jksb.v7i3.178

Kementerian Pemberdayaan Perempuan dan Perlindungan Anak. (2020). *Pembangunan Manusia Berbasis Gender Tahun 2020*.

https://www.kemenpppa.go.id/lib/uploads/list/50a46-pembangunan-manusia-berbasis-gender-2020.pdf

Kurnianingsih, F., Mahadiansar, Putri, R. A., & Azizi, O. R. (2022). Perspektif Analisis Indeks Pemberdayaan Gender Kota Tanjungpinang dalam Masa Pandemi COVID-19. *Jurnal Ilmu Sosial dan Humaniora, 11*(1), 45–55. https://doi.org/10.23887/jish.v11i1.37594

Mutmainnah, M., Mawarti, I., & Yusnilawati, Y. (2021). Pemberdayaan Perempuan Akseptor KB dalam Mempertahankan Kesehatan, Kualitas Hidup dan Peran yang Optimal dalam Keluarga. *Medical Dedication (Medic): Jurnal Pengabdian kepada Masyarakat FKIK UNJA, 4*(1), 241–247. https://doi.org/10.22437/medicaldedication.v4i1.13493

Novianti, E. (2019). Kesenjangan Gender Tingkat Pengangguran Terbuka di Indonesia. *Jurnal Pendidikan dan Ekonomi, 8*(2), 166–174. https://journal.student.uny.ac.id/ekonomi/article/view/13053

Padhilah, R., Herrhyanto, N., & Lukman. (2024). Analisis Regresi Logistik Biner dengan Metode Group LASSO dalam Data Berdimensi Tinggi (Studi Kasus: Indeks Pembangunan Manusia Kota/Kabupaten di Jawa Barat). *BIAStatistics Journal of Statistics Theory and Aplication, 18*(1), 1–14. https://biastatistics.statistics.unpad.ac.id/?journal=biastatistics&page=article&op=view&path%5B%5D=266

Putrie, D. A., & Rahman, A. (2021). Analisis dan Pemodelan Pendapatan Pekerja Perempuan di Indonesia Menggunakan Data Panel. *Seminar Nasional Official Statistics, 2020*(1), 1269–1276. https://doi.org/10.34123/semnasoffstat.v2020i1.688

Riadina, B. P., & Sugianto. (2024). Analisis Pengaruh Indikator Kebijakan Moneter terhadap Tingkat Inflasi di Indonesia. *Journal of Development Economics and Digitalization, 3*(1), 1–16. https://doi.org/10.59664/jded.v3i1.7666

Rusli, A. S., & Magna, M. S. (2023). Faktor-Faktor yang Mempengaruhi Indeks Pemberdayaan Gender (IDG) Kota Magelang Tahun 2011–2021. *EVOKASI: Jurnal Kajian Administrasi dan Sosial Terapan, 2*(1), 1–10. https://doi.org/10.20961/evokasi.v2i1.778

Salsabila, D., & Hendrawan, M. Y. (2021). Analisis Kondisi Pemberdayaan Gender di Indonesia Tahun 2020 dengan Agglomerative Hierarchical Clustering dan Biplot. *Seminar Nasional Official Statistics, 2021*(1), 204–213. https://doi.org/10.34123/semnasoffstat.v2021i1.839

Syukri, M. (2023). Gender policies of the new developmental state: The case of Indonesian new participatory village governance. *Journal of Current Southeast Asian Affairs, 42*(1), 110–133. https://doi.org/10.1177/18681034221149750

Taqwim, S. F., Vaezghasemi, M., Castel-Feced, S., Dewi, F. S. T., & Schröders, J. (2025). The role of women's empowerment in fertility preferences and outcomes: Analysis of the 2017 Indonesia Demographic and Health Survey. *BMC Women's Health, 25*(1), Article 211. https://doi.org/10.1186/s12905-025-03748-6

Valentina, R. F. (2022). Studi tentang Akses Menempuh Pendidikan Tinggi bagi Perempuan. *Jurnal Dialektika Pendidikan IPS, 2*(2), 35–47. https://doi.org/10.26740/penips.v2i2.47413

Wulandari, I., Notodiputro, K. A., & Sartono, B. (2020). Variable selection in analyzing life infant birth in Indonesia using group lasso and group SCAD. In *Proceedings of the 1st*

*International Conference on Statistics and Analytics (ICSA 2019)*. EAI. https://doi.org/10.4108/eai.2-8-2019.2290485

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68*(1), 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x

Yustie, R., & Prayitno, B. (2024). Faktor Ekonomi dan Sosial yang Mempengaruhi Pemberdayaan Gender. *Equilibrium: Jurnal Ekonomi-Manajemen-Akuntansi, 20*(1), 80–87. https://doi.org/10.30742/equilibrium.v20i1.3633