Aksioma: Jurnal Matematika dan Pendidikan Matematika
Volume 17, No. 1, April 2026, pp. 171-186
ISSN: 2579-7646 (e) | 2086-2725 (p)
https://doi.org/10.26877/ncz1wh42

Identifying key determinants of students' mathematics achievement using lasso regression

Dewi Wulandari[1], Aurora Nur Aini[2], Rianti Siswi Utami[3], Agnita Siska Pramasdyahsari[4]

[1,2,4]Mathematics Education, Universitas Persatuan Guru Republik Indonesia Semarang, Central Java, Indonesia
[3]Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, Indonesia
[4]Institute for Learning Sciences and Teacher Education, Australian Catholic University, Brisbane, Australia

Correspondence: dewiwulandari@upgris.ac.id

Received: March 31, 2026 | **Revised**: Apr 10, 2026 | **Accepted**: Apr 29, 2026 | **Published Online**: Apr 30, 2026

Abstract

Cognitive, socioeconomic, and instructional factors influence students' mathematics achievement. However, modeling these determinants becomes challenging when the number of predictors is relatively large compared to the available sample size, potentially leading to unstable estimates in conventional regression models. This study aims to identify key determinants of students' mathematics achievement using the Least Absolute Shrinkage and Selection Operator (LASSO) regression approach and to construct a parsimonious predictive model. The study employed a quantitative design involving 50 second-grade students from SMA Negeri 1 Subah, Batang, Central Java, Indonesia. Predictor variables included intelligence quotient (IQ), tutoring participation, parental income, parental educational background, study duration, teacher characteristics, and gender. Categorical predictors were converted to dummy variables, and all predictors were standardized before analysis. The optimal regularization parameter was determined using $\boldsymbol{k}$-fold cross-validation, and model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination ($\boldsymbol{R^2}$). The LASSO results indicate that only three predictors retain non-zero coefficients: IQ ($\boldsymbol{\beta}$ = 3.134), tutoring participation ($\boldsymbol{\beta}$ = 6.675), and parental income ($\boldsymbol{\beta}$ = 1.299). The resulting model achieves RMSE = 9.55, MAE = 8.23, and $\boldsymbol{R^2}$ = 0.061, suggesting limited explanatory power but improved model parsimony and interpretability. Coefficient stability analysis further shows that IQ is a relatively robust predictor, while tutoring effects exhibit higher variability across cross-validation folds. These findings highlight the importance of cognitive ability and supplementary academic support in shaping mathematics achievement, while also indicating that many relevant behavioral and environmental determinants remain unobserved. Methodologically, the study demonstrates the usefulness of regularized regression in handling small-sample multivariable educational data and producing interpretable sparse models. Future research should incorporate richer behavioral variables and explore nonlinear machine learning approaches to improve predictive performance.

Keywords: mathematics achievement; LASSO regression; regularization; educational data analytics; feature selection; predictive modeling

How to Cite: Wulandari, Dewi., Aini, A.N., Utami, R.S., Pramasdyahsari, A.S. (2026). Identifying key determinants of students' mathematics achievement using lasso regression. *Aksioma: Jurnal Matematika dan Pendidikan Matematika, 17*(1), 171-186. https://doi.org/10.26877/ncz1wh42

INTRODUCTION

A combination of cognitive, socioeconomic, and instructional factors influences students' mathematics achievement. Variables such as intelligence level, study habits, parental background, and teacher characteristics have been widely examined to explain variation in academic performance. In educational research, multiple regression analysis is commonly used to investigate the relationships between predictors and academic outcomes (Kutner et al., 2005). The selection of predictor variables in the present study was guided by both empirical findings from previous educational studies and practical considerations derived from teachers' classroom observations. Prior research has shown that cognitive ability is a strong predictor of academic achievement (Deary et al., 2007), while study habits and time allocation for learning are associated with improved student performance (Nonis & Hudson, 2010). Furthermore, socioeconomic characteristics such as parental income and educational attainment significantly influence students' learning opportunities and achievement levels (Sirin, 2005). Instructional factors, including teacher-related characteristics and participation in supplementary tutoring programs, have also been reported to contribute to variations in students' academic outcomes (Hattie, 2009). Based on these theoretical and empirical considerations, this study incorporates multiple predictors, including students' IQ, study hours, tutoring participation, parental income, parental educational background, teacher characteristics, and gender, to capture the multidimensional nature of students' mathematics achievement.

However, methodological challenges may arise when the number of predictors becomes relatively large relative to the available sample size. In this study, the dataset comprises 50 students from second-grade classes at SMA Negeri 1 Subah, Batang, Central Java, Indonesia. The model includes eight predictor variables comprising both numerical and categorical characteristics. Several predictors contain multiple category levels, such as parental education background, which require transformation into dummy variables before regression modeling. A categorical predictor with k levels is typically represented by $k - 1$ indicator variables to avoid perfect multicollinearity and enable parameter estimation (Kutner et al., 2005). Consequently, the effective number of model parameters increases after preprocessing, reducing the degrees of freedom available for estimation and potentially affecting model stability. Methodological guidelines suggest that regression models ideally require a sufficient number of observations per predictor variable to obtain reliable estimates, with a commonly cited rule of thumb recommending at least ten observations per predictor in multivariable regression settings (VanVoorhis & Morgan, 2007). Given the relatively small sample size compared to the effective number of predictors, conventional ordinary least squares (OLS) regression may exhibit high variance, sensitivity to multicollinearity, and reduced generalization performance.

Regularized regression methods provide an alternative modeling strategy under such conditions. The Least Absolute Shrinkage and Selection Operator (LASSO) regression introduces L1 penalization to simultaneously shrink coefficients and select variables, thereby producing more parsimonious and interpretable models (Tibshirani, 1996). Penalized regression techniques have been shown to improve prediction accuracy and enhance model stability, particularly in settings involving multiple correlated predictors (Hastie et al., 2009).

Previous studies have explored both the methodological development and empirical applications of LASSO regression across various domains. Soleh & Aunudin (2013) described the procedural framework of LASSO regression, while Pardede et al. (2022), Robbani et al. (2019), and Andana et al. (2017) demonstrated its implementation in different case studies. Pardede et al. (2022) applied LASSO regression in the health sector to analyze stunting issues; Robbani et al. (2019) examined economic phenomena related to inflation during the 2014–2017 period; and Andana et al. (2017) applied LASSO in public health research on malnutrition cases. These studies indicate that LASSO has been widely utilized in fields such as health and economics. However, its application in educational research remains relatively limited, particularly in studies aiming to identify a parsimonious set of key determinants influencing students' academic performance.

Unlike previous studies that primarily focused on demonstrating the feasibility of LASSO in specific applied contexts, this study emphasizes sparse modeling for determinant identification under small-sample multivariable conditions. By addressing the potential instability of conventional regression when multiple correlated predictors and dummy-induced parameter expansion are present, this research contributes methodologically to the application of regularized regression in educational data analysis. Furthermore, the study provides empirical insights into the most influential cognitive, socioeconomic, and instructional factors affecting students' mathematics achievement. Therefore, this study aims to identify the key determinants of students' mathematics achievement using LASSO regression and to construct a parsimonious predictive model that improves model stability, predictive performance, and interpretability in educational datasets.

METHODS

1. Research Design and Data Source

This study employed a quantitative research design to identify key determinants of students' mathematics achievement using regression-based modeling. The research was conducted at SMA Negeri 1 Subah, Batang, Central Java, Indonesia. The study involved 50 second-grade students selected using purposive sampling based on data completeness and availability.

The dependent variable was the students' mathematics achievement score. Predictor variables included intelligence quotient (IQ), participation in tutoring programs, parental income, father's educational background, mother's educational background, students' study hours, teacher characteristics, and gender. Data were collected through academic records, structured documentation, and supporting questionnaires.

2. Linear Regression

As a baseline approach, multiple linear regression using the ordinary least squares (OLS) estimator was considered. The regression model is defined as:

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \varepsilon_i$$

where Y_i denotes the mathematics achievement score of the student i, X_{ij} represents the j-th predictor variable for the student i, β_j denotes the regression coefficient associated with the j-th predictor, and ε_i is the random error term.

Let $\beta = \left(\beta_0, \beta_1, \dots, \beta_p\right)^T$ be the vector of regression coefficients and let $X_i = \left(1, X_{i1}, \dots, X_{ip}\right)^T$ be the predictor vector for student i. The OLS estimators are obtained by minimizing the residual sum of squares:

$$\hat{\beta}^{OLS} = \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2.$$

Although widely applied, OLS estimation may produce unstable coefficient estimates when predictors are correlated or when the sample size is relatively small relative to the number of predictors (Kutner et al., 2005).

3. LASSO Regression

To overcome these limitations, this study applied the Least Absolute Shrinkage and Selection Operator (LASSO) regression. The LASSO estimator is formulated as a penalized least squares optimization problem:

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \right\}$$

where $\lambda \geq 0$ is the regulation parameter controlling the strength of the coefficient shrinkage. The inclusion of the L1 penalty encourages sparsity by shrinking some coefficients exactly to zero, thereby performing automatic variable selection (Tibshirani, 1996).

Alternatively, the LASSO estimator can be expressed in constrained form:

$$\arg\min_{\beta} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t$$

where $t \geq 0$ is a tuning parameter that controls the amount of shrinkage imposed on the regression coefficients. Smaller values of t produce stronger regularization and encourage more coefficients to be shrunk toward zero. This formulation shows that LASSO restricts the feasible parameter space by imposing a bound on the total absolute value of the regression coefficients. The trajectory of coefficient estimates across different values of the regularization parameter is called the regularization path (Hastie et al., 2009).

4. Geometrical Interpretation and Shrinkage Mechanism

From a geometrical perspective, the LASSO solution can be interpreted as the intersection of the residual sum-of-squares contours and an L1-norm constraint region. Unlike ridge regression, whose constraint region is circular, the L1 constraint forms a diamond-shaped region in low-dimensional settings. As a result, the optimal solution often occurs at the corners of the feasible region, leading to exact zero estimates for some regression coefficients. This explains LASSO's variable selection capability (Hastie et al., 2009).

From an optimization viewpoint, the absolute value penalty modifies the first-order optimality condition by introducing a subgradient term. Consequently, small coefficient estimates are shrunk toward zero, while predictors with weak associations to the response variable may be excluded from the final model.

5. Regularization Parameter Tuning

The regularization parameter λ governs the trade-off between model fit and model complexity. Larger values of λ produce stronger shrinkage, leading to simpler and more parsimonious models, whereas smaller values yield solutions closer to ordinary least squares estimates. In this study, the optimal value of λ was determined using k-fold cross-validation, where the dataset was repeatedly partitioned into training and validation subsets to minimize prediction error (Hastie et al., 2009). Predictor variables associated with non-zero coefficients in the final model were interpreted as key determinants of students' mathematics achievement.

6. Model Evaluation

Model performance was evaluated using predictive accuracy and goodness-of-fit measures, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). These metrics are widely used in regression-based predictive modeling to assess the magnitude of prediction error and the proportion of variance explained by the model (Hastie et al., 2009; James et al., 2021). RMSE provides a quadratic measure of prediction error that penalizes large deviations more heavily, whereas MAE offers a linear, more interpretable measure of average prediction error (Willmott & Matsuura, 2005). The coefficient of determination (R^2) is used to evaluate the model's explanatory power by quantifying the proportion of variability in the response variable that can be accounted for by the predictors (Kutner et al., 2005). Models with lower prediction error values and higher explanatory power were considered more appropriate for representing the relationship between predictor variables and students' mathematics achievement.

7. Data Analysis Procedure

Prior to model estimation, categorical predictors were transformed into dummy variables to enable regression analysis and to avoid perfect multicollinearity arising from their categorical structure (Kutner et al., 2005). All predictor variables were subsequently standardized to ensure comparable coefficient penalization, as regularization methods such as LASSO are sensitive to differences in variable scales (Hastie et al., 2009).

To assess model generalization performance, the dataset was split into training and test sets following common predictive modeling practices (James et al., 2021). Data analysis was conducted using the Python programming language in the Google Colaboratory environment with support from the scikit-learn library, which provides efficient implementations of regression modeling, cross-validation, and performance evaluation procedures (Pedregosa et al., 2011). The analytical steps included: 1) Data preprocessing, including dummy encoding and standardization; 2) Splitting the dataset into training and testing subsets; 3) Estimating the baseline multiple linear regression model; 4) Fitting the LASSO regression model with k-fold cross-validation to determine the optimal regularization parameter; 5) Identifying selected predictors based on non-zero coefficient estimates; and 6) Evaluating model performance using RMSE, MAE, and R^2.

RESULTS AND DISCUSSION

1. Model Specification and Data Preprocessing

This study aims to identify the determinants of students' mathematics achievement using the Least Absolute Shrinkage and Selection Operator (LASSO) regression approach. The dependent variable is students' math score, while the independent variables include gender, teacher factor, IQ, tutoring participation, mother's educational background, father's educational background, study duration, and parental income.

Prior to model estimation, all categorical variables were transformed into dummy variables to avoid imposing ordinal linear assumptions across categories. Furthermore, all predictor variables were standardized to ensure scale comparability, as LASSO regression is sensitive to differences in variable magnitudes.

The dataset was divided into training and testing subsets using an 80:20 ratio. The optimal regularization parameter (λ) was determined through k-fold cross-validation, allowing the model to minimize prediction error and improve generalization performance.

2. Multicollinearity Assessment

Multicollinearity was evaluated using the Variance Inflation Factor (VIF) statistic. The results are presented in Table 1.

Table 1. Variance Inflation Factor (VIF) Results

Variable	VIF
Gender	1.248
Teacher	1.244
IQ	1.932
Tutoring	3.827
Mother's Education	5.761
Father's Education	5.335
Study Duration	1.490
Parental Income	2.238

The results indicate that all predictor variables have VIF values below the commonly accepted threshold (VIF < 10), suggesting no severe multicollinearity issues in the initial regression model. Nevertheless, the application of LASSO remains appropriate as it not only mitigates potential multicollinearity effects but also performs automatic variable selection through L1 regularization.

3. Variable Selection Using LASSO Regression

The LASSO model retained three predictors with non-zero coefficients: IQ, tutoring participation, and parental income. This finding suggests that, among the variables included in the present dataset, these three predictors were the most relevant for explaining variation in students' mathematics achievement.

The resulting regression model can be expressed as:

Math Score = 81.771 + 3.134 (IQ) + 6.675 (Tutoring) + 1.299 (Parental Income).

These findings demonstrate that LASSO produces a more parsimonious and interpretable model than traditional linear regression approaches that include all predictors.

The coefficient for IQ (3.134) indicates that a one-unit increase in IQ is associated with approximately 3.13 points higher mathematics performance, holding other variables constant. This finding is consistent with educational research highlighting the importance of cognitive ability in academic achievement.

Tutoring participation exhibits the largest coefficient (6.675), suggesting that students who engage in supplementary learning programs tend to achieve substantially higher mathematics scores than those who do not. This result supports the effectiveness of academic intervention programs in enhancing student performance.

Parental income also shows a positive association with mathematics achievement, with a coefficient of 1.299. This relationship may reflect differences in access to educational resources, learning environments, and academic support across socioeconomic backgrounds.

The model's predictive performance was assessed using the test dataset. The results show a Root Mean Squared Error (RMSE) of 9.55, Mean Absolute Error (MAE) of 8.23, and coefficient of determination (R^2) of 0.061.

The relatively low R^2 value indicates that the model explains only about 6.1% of the variation in students' mathematics scores. This suggests that although IQ, tutoring participation, and parental income contribute to academic performance, a large proportion of the variability is driven by unobserved factors. These may include learning motivation, teaching quality, school environment, peer influence, psychological conditions, and study habits.

Coefficient stability was examined using repeated estimation under k-fold cross-validation. The standard deviations of coefficients across folds are presented in Table 2.

Table 2. Coefficient Stability (Standard Deviation)

Variable	Std. Dev
Tutoring	1.635
IQ	0.896
Parental Income	0.886
Gender	0.171
Study Duration	0.027

The results indicate that the tutoring variable exhibits relatively higher coefficient variability than the other predictors, suggesting sensitivity to sample partitioning. In contrast, IQ demonstrates greater coefficient stability, suggesting that it is a more robust predictor of mathematics achievement within the model.

The inclusion of the LASSO penalty reduces model complexity by eliminating less relevant predictors, resulting in a more parsimonious specification. This regularization process introduces additional bias while reducing model variance, thereby enhancing the model's generalization capability. However, the trade-off also implies that the model may not fully capture complex nonlinear or interaction effects present in educational performance data.

The findings highlight the significant role of supplementary academic support programs in improving students' mathematics achievement. Policies aimed at expanding access to tutoring services may therefore contribute to improved academic outcomes. Additionally, the positive relationship between parental income and academic performance underscores the importance of addressing socioeconomic disparities in access to educational resources.

The present study applies LASSO regression to identify key determinants of students' mathematics achievement. The findings indicate that cognitive ability (IQ), participation in tutoring programs, and parental income are the only predictors retained in the final parsimonious model. These results provide several important insights into the drivers of academic performance.

First, the positive association between IQ and mathematics achievement is consistent with a substantial body of educational research emphasizing the role of cognitive ability in learning outcomes. Cognitive skills such as reasoning, problem-solving, and information processing speed have long been identified as strong predictors of academic success (Deary et al., 2007; Rohde & Thompson, 2007). The relatively stable coefficient estimates for IQ across cross-validation folds further suggest that this variable is a robust determinant of mathematics performance in the present dataset.

Second, tutoring participation emerges as the most influential predictor in the model. This finding aligns with previous studies showing that supplementary instruction can significantly improve academic achievement by providing individualized attention, structured practice, and reinforcement of classroom material (Cheung & Slavin, 2013). The relatively high variability in tutoring coefficients observed across different training splits may indicate heterogeneity in the effectiveness of tutoring programs, which could depend on factors such as tutoring quality, duration, and student engagement.

Third, parental income demonstrates a positive relationship with mathematics performance, reflecting broader socioeconomic disparities in educational outcomes. Students from higher-income households often have better access to learning resources, private instruction, and supportive learning environments (Sirin, 2005). This result supports existing evidence that socioeconomic status remains an important contextual factor influencing academic achievement.

Despite the successful identification of relevant predictors, the overall predictive performance of the model is relatively low, with an R^2 value of approximately 0.06. This suggests that unobserved factors drive a large proportion of variation in mathematics achievement. Prior research indicates that non-cognitive variables such as motivation, self-regulated learning, classroom climate, and teacher effectiveness play critical roles in shaping student outcomes (Duckworth & Seligman, 2005; Hattie, 2009). The absence of these variables in the dataset may partly explain the model's limited explanatory power.

From a methodological perspective, applying LASSO improves model interpretability by reducing dimensionality and mitigating overfitting. Regularization techniques have been increasingly adopted in educational data mining and learning analytics to handle high-dimensional predictors and enhance generalization performance (Romero & Ventura, 2010). However, the bias–variance trade-off inherent in LASSO can lead the model to underfit complex relationships, particularly when nonlinear interactions or hierarchical structures are present in the data.

Another important consideration relates to sample size and measurement limitations. The relatively large variation in cross-validation error indicates that model performance may be sensitive to data partitioning, a phenomenon commonly reported in predictive modeling studies involving small samples or noisy predictors (James et al., 2021; Kuhn & Johnson, 2013). Limited sample diversity and measurement error can reduce model stability and generalization capability. Therefore, future studies should employ larger, more comprehensive datasets that incorporate behavioral indicators, longitudinal learning trajectories, and school-level contextual variables, which have been shown to improve explanatory power in educational research (Hattie, 2009; Romero & Ventura, 2010).

From a practical perspective, the findings emphasize the importance of targeted academic support programs, particularly tutoring interventions, in improving mathematics achievement. Prior meta-analytic evidence suggests that structured supplementary instruction can significantly enhance student learning outcomes through individualized feedback and additional practice opportunities (Cheung & Slavin, 2013). Policymakers and educators may therefore consider expanding access to tutoring services, especially for students from lower socioeconomic backgrounds, as socioeconomic disparities remain strongly associated with academic achievement (Sirin, 2005). Furthermore, strengthening cognitive skill development through curriculum innovation and evidence-based instructional strategies may yield long-term academic benefits (Deary et al., 2007; Hattie, 2009).

Methodologically, future research should explore alternative regularization techniques such as Elastic Net regression, which combines L1 and L2 penalties and has been shown to perform better than LASSO in the presence of correlated predictors (Zou & Hastie, 2005). In addition, advanced machine learning approaches, including tree-based ensemble methods and nonlinear regression techniques, may capture complex interaction patterns that are not adequately modeled by linear regularization frameworks (Hastie et al., 2009; Kuhn & Johnson, 2013).

Overall, while LASSO regression is effective in identifying key predictors and producing interpretable sparse models, academic performance remains a multidimensional phenomenon influenced by cognitive ability, behavioral engagement, instructional quality, and environmental context (Duckworth & Seligman, 2005; Romero & Ventura, 2010). This highlights the need for integrative modeling approaches that combine statistical regularization with richer educational data sources.

CONCLUSION

This study investigates the determinants of students' mathematics achievement using the Least Absolute Shrinkage and Selection Operator (LASSO) regression approach. The empirical findings indicate that among the considered predictors, cognitive ability (IQ), participation in tutoring programs, and parental income are the only variables retained in the final parsimonious model. The results suggest that students with higher cognitive ability and access to supplementary academic support tend to achieve better mathematics performance. In addition, socioeconomic conditions, as reflected in parental income, also contribute positively to academic outcomes. These findings reinforce the importance of both individual cognitive factors and external educational support mechanisms in shaping student achievement.

Although the LASSO model successfully performs automatic variable selection and improves model interpretability, its predictive performance remains relatively limited. This indicates that mathematics achievement is influenced by a broader set of latent cognitive, behavioral, and environmental factors that were not captured in the current dataset. Overall, the study demonstrates the usefulness of regularization-based regression methods in educational data analysis, particularly for identifying key predictors and developing interpretable predictive models.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by the Institute for Research and Community Service (LPPM), Universitas PGRI Semarang (UPGRIS), which made this research possible. The authors also thank colleagues and reviewers who offered helpful comments and suggestions that improved the quality of this manuscript.

DECLARATIONS

Author Contribution	: DW: Conceptualization, Data Analysis, Visualization, and Drafting of the manuscript. ANA: Methodology Development, Editing. RSU: Reviewing and Editing, ASP: Reviewing and Editing.
Funding Statement	: This work was supported by the Institute for Research and Community Service (LPPM), Universitas PGRI Semarang (UPGRIS), Indonesia.
Conflict of Interest	: The authors declare no conflict of interest.
Additional Information	: Additional information is available for this paper.

REFERENCES

Andana, A. P., Safitri, D., & Rusgiyono, A. (2017). Model regresi menggunakan Least Absolute Shrinkage and Selection Operator (LASSO) pada data banyaknya gizi buruk kabupaten/kota di Jawa Tengah. *Jurnal Gaussian*, 6(1), 21–30.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K–12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88–113.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35(1), 13–21.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12), 939–944.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R* (2nd ed.). Springer.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill.

Nonis, S. A., & Hudson, G. I. (2010). Performance of college students: Impact of study time and study habits. *Journal of Education for Business*, 85(4), 229–238.

Pardede, T. T., Sumargo, B., & Rahayu, W. (2022). Penerapan regresi Least Absolute Shrinkage and Selection Operator (LASSO) untuk mengidentifikasi variabel yang berpengaruh terhadap kejadian stunting di Indonesia. *Jurnal Statistika dan Aplikasinya*, 6(1), 37–48.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Robbani, M., Agustiani, F., & Herrhyanto, N. (2019). regresi least absolute shrinkage and selection operator (LASSO) pada kasus inflasi di Indonesia tahun 2014–2017. *Jurnal Eureka Matika*, 7(2), 1–16.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35(1), 83–92.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*, *Part C*, 40(6), 601–618.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453.

Soleh, A. M., & Aunuddin, A. (2013). LASSO: Solusi alternatif seleksi peubah dan penyusutan koefisien model regresi linier. *Forum Statistika dan Komputasi: Indonesian Journal of Statistics*, 18(1), 21–27.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288.

VanVoorhis, C. R. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43–50.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B*, 67(2), 301–320.