BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0037–0042
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0037-0042

PERFORMANCE OF LASSO AND ELASTIC-NET METHODS ON NON-INVASIVE BLOOD GLUCOSE MEASUREMENT CALIBRATION MODELING

Farah Abqorunnisa [1], Erfiani[2*], Anik Djuraidah[3]

[1,2,3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University Bogor, 16680, Indonesia*

*Corresponding author's e-mail: * erfiani@apps.ipb.ac.id*

ABSTRACT

Article History:

Received: 19th July 2022
Revised: 9th November 2022
Accepted: 8th January 2023

Keywords:

Calibration;
Diabetes Mellitus;
Elastic Net c
Lasso;
Non-Invasive.

Diabetes Mellitus (DM) is a disease that can occur in humans caused by conditions of high blood glucose levels (hyperglycemia). Detection of blood glucose levels can be done using invasive methods (injuring) and non-invasive methods (with infrared rays). Analytical methods are needed to model these results to obtain estimates of blood glucose levels. An alternative approach that can be used to analyze the relationship between invasive and non-invasive blood glucose levels is the calibration model. Problems that often occur in calibration modeling are multicollinearity and outliers. These problems can be overcome by adding new data, applying principal component analysis, and using LASSO and Elastic-Net regression to overcome calibration problems. The research data used was invasive and non-invasive blood glucose data in 2019, with as many as 74 respondents. The results of the study concluded that the summarization of the trapezoidal area in calibration modeling provides a good estimate. The performance of the Elastic Net method provides better prediction results than other models, with an RMSE value of 22.39. It has the most significant positive correlation value of 0.97, which means close to 1, so the performance of the Elastic Net method can handle calibration modeling.

How to cite this article:

F. Abqorunnisa, Erfiani, and A. Djuraidah, "PERFORMANCE OF LASSO AND ELASTIC-NET METHODS ON NON-INVASIVE BLOODGLUCOSE MEASUREMENT CALIBRATION MODELING," *BAREKENG: J. Math. & App.,* vol. 17, iss. 1, pp. 0037-0042, March 2023.

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

Glucose is the primary energy source in human cells. Blood glucose is sugar from carbohydrates in food stored as glycogen in the liver and skeletal muscles [1]. Each individual has different blood glucose levels in the body. Abnormal blood glucose metabolism can cause hyperglycemia and hypoglycemia. Hyperglycemia is when blood glucose levels are at high levels > 110 mg/dl. Hypoglycemia is a blood glucose level that is too low, which is <70 mg/dl [2]. Hyperglycemia conditions that are too high and beyond normal limits can cause a person to develop Diabetes Mellitus (DM). Therefore, it is essential to detect blood glucose levels in each individual's body periodically as a prevention effort for someone who has not been exposed to Diabetes Mellitus (DM).

Detection of blood sugar levels can be detected by two types of measurement, invasive and non-invasive. Invasive methods can be performed using a glucometer (silver standard) or laboratory tests (gold standard). Meanwhile, non-invasive blood sugar detection is carried out with the help of infrared rays that produce spectral patterns. Therefore, it is necessary to summarize the data using calibration modeling to estimate blood glucose levels from the tool's output non-invasively

This study used a calibration model to formulate a function between the output of invasive and non-invasive blood glucose measurement tools. The calibration model is a function to find the relationship between units of measurement that can obtain through a relatively easy or inexpensive process [3]. The non-invasive biomarking team of IPB developed several studies of non-invasive tool calibration modeling, Artificial Neural Network (ANN) analysis [4], spline regression [5], principal component regression [6], quadratic regression least partial and support vector regression (SVR), robust regression [7], and the Gulud, LASSO, and Elastic Net calibration models [8].

The method used to estimate invasive blood sugar levels from the results of calibration data is influenced by many variables, so it is necessary to make a selection to obtain variables that have a significant effect on the many variables observed. There are various methods to handle regression analysis on high-dimensional data, one of which is selecting explanatory variables. This study uses blood sugar calibration data to overcome high-dimensional data, selecting groups of explanatory variables to obtain a simpler model.

Based on the description above, this study aims to develop a calibration model on a non-invasive device using a comparative regression analysis of the Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net.

2. RESEARCH METHODS

The research data used is blood glucose level data from non-invasive measurements. Primary data was taken in 2019 from rays reflected by limbs. The research was carried out by the Non-Invasive Biomarking Team, located at the Biochemistry Laboratory, Department of Community Nutrition, IPB. This study involved 74 respondents from the general public in 2019.

Response variables are invasive blood glucose data obtained from Prodia laboratory testing. The independent variable results from non-invasive measurement tools in residual light intensity using the principle of spectroscopy with infrared light [9]. The lamp used uses an infrared sensor when the lamp condition is on at 1600 nm using 10 Modulation. Modulation is the level of light regulation on non-invasive devices, namely in the period 21 to 30 or modulation from 0 to 90, with each treatment time being 500 ns (nanosecond) so that one respondent takes 20 s (second) to get the results of blood sugar levels. The flow in data processing is shown as follows.

1. Exploration of invasive blood glucose level data and non-invasive data;
2. Formation of the Calibration Model;
   a. Calculating the area formed by the intensity residue on the time domain of each modulation using the trapezoidal area formula as follows:

$$L = \frac{1}{2}\sum_{i=1}^{n-1}(t_{i+1} - t_i)(y_i + y_{i+1}) \quad (1)$$

L is the area of the residual intensity curve concerning the time domain, t_i is the time domain, y_i is the residual intensity value, and n is the number of points in one modulation.

b. By adding each area obtained for each modulation, the area value will be the independent variable resulting from the intensity residue.

c. It was forming a data cluster that is applied to the 2019 data.
   - The 2019 data is 0 to 90 modulations.

3. Comparative Analysis of LASSO Regression Model and Elastic Net.

a. Data partition:
This method is done by dividing the data into training and validation data sets **[10]**. The research dataset from 74 respondents or objects will be divided into two groups: the training dataset of 70% and the testing dataset of 30%. Data analysis was repeated 100 times.

b. The selection of the explanatory variable group and the choice of model candidates are carried out by looking at the model that provides the minimum cross-validation residual value (CV).

$$CV = \sum_{k=1}^{K} \frac{MSEP_k}{K} \quad (2)$$

c. Modeling determines the best model and determines the coefficient of beta significance. **[11] [12]**. The coefficient estimator is obtained by minimizing the total squared residual **[13]**.

$$\hat{\beta}^{LASSO} = \underset{\beta}{argmin} \sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{p}|\beta_j| \quad (3)$$

$$\hat{\beta}^{Elastic\ Net} = \underset{\beta}{argmin} \sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda[(1-\alpha)\sum_{j=1}^{p}\hat{\beta}_j^{\,2} + \alpha\sum_{j=1}^{p}|\hat{\beta}_j|)] \quad (4)$$

where k is the number of groups, $\boldsymbol{X_j}$ is the j-th explanatory variable-j, and Q_j is the regression beta coefficient. While $\lambda \geq 0$ is a controller of the amount of depreciation. The model will be in standard form when $\lambda = 0$. If the value λ is greater, then the estimated value of the beta coefficient is significant, and if it is smaller, the value λ will go from zero to infinity. Where $0 < \alpha < 1$ is a composite parameter between Ridge Regression (α=0) and LASSO (α=1).

d. Make predictions based on the model obtained at the modeling stage and evaluate the model comparison by comparing the RMSE values**.**

3. RESULTS AND DISCUSSION

3.1. Exploration of Invasive and Non-Invasive Blood Glucose Levels

The descriptive analysis of the variables used in this study is presented in **Figure 1**. The plot of invasive blood glucose levels in 2019 shows that invasive blood glucose levels have a pattern that tends to fluctuate randomly with great diversity. The average blood glucose level of 74 respondents is 140.73 mg /dL and a median of 105.5 mg/dL. The lowest blood glucose level is 69 mg/dL, and the highest glucose level is 614 mg/dL and is classified as the result of measuring blood glucose levels above normal limits.

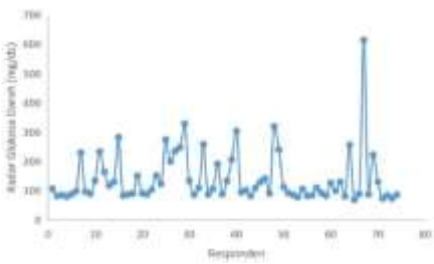


Figure 1. The plot of invasive blood glucose levels

The 2019 non-invasive tool kit is designed to capture intensity residues absorbed by the finger and captured by non-invasive tool sensors. The results of the 2019 data spectrum for one measurement in the first replication are shown in **Figure 2**.

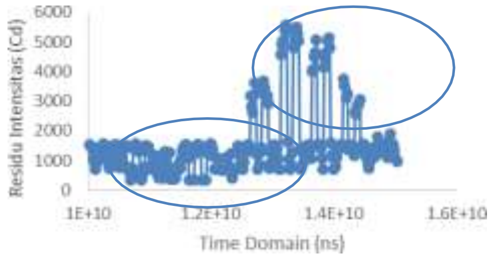


Figure 2. The spectrum of the output of a non-invasive blood glucose level measuring device

The non-invasive device is designed to perform two repetitions for each modulation. Differentiation of recurrence in modulation is done by turning off the infrared lamp and turning on the infrared light. It is what makes the formation of 2 peaks in each modulation.

3.2. Calibration Model

Establish a calibration model with a non-invasive summary of blood glucose data using the trapezoidal area principle. Each residual intensity value in 1 modulation will be drawn straight toward its time domain. It will form a trapezoidal area. Then calculate each size created using the formula for the scope of the trapezoid. The residual intensity value is used as the parallel side, while the time domain interval length is used as the height. An illustration of the calculation of the area can be seen in **Figure 3**.

After obtaining a value of the area of each area, which is added in one modulation, then the value is used as an independent variable. One independent variable means the sum of the values of the trapezoid site in one modulation. Five replications were carried out for each respondent in ten modulations, so there were 50 independent variables. Fifty independent variables were obtained from 10 modulations with five repetitions.

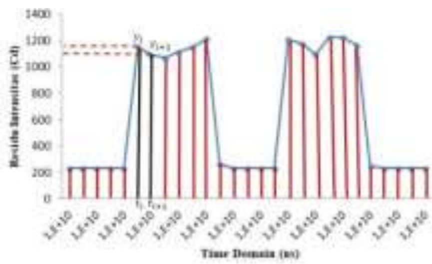


Figure 3. Illustration of the calculation of the area of one modulation

Figure 4 shows the correlation plot between variables. The blue color indicates a positive correlation between variables. The smaller the correlation, the color fades to red. From **Figure 4**, it can be seen that all the variables used in the study have a high correlation, so it can be concluded that there is multicollinearity between the independent variables used in the study.

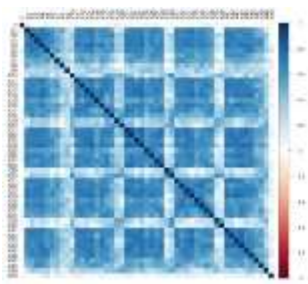
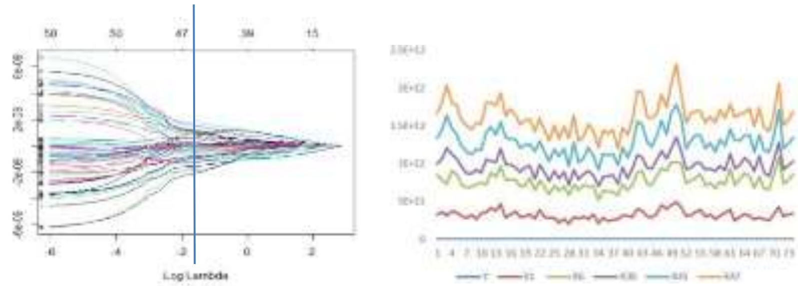


Figure 4. Correlation between variables used in research

a) LASSO Regression

The optimal shrinkage parameter (λ) for the LASSO Regression is obtained by a cross-validation method(Cross Validation-CV) which produces a minimum CV of 0.2. **Figures 5 (a)** and **Figures 5 (b)** show the level of importance of the five variables (variable importance performance/VIP) generated by the LASSO regression.The most important variable levels based on the resulting VIP values are X1(100); X41(94,41); X47(74,82); X30(73,98); X6(68,20). Variables have the same plot pattern, so they fall into the group of significant variables.

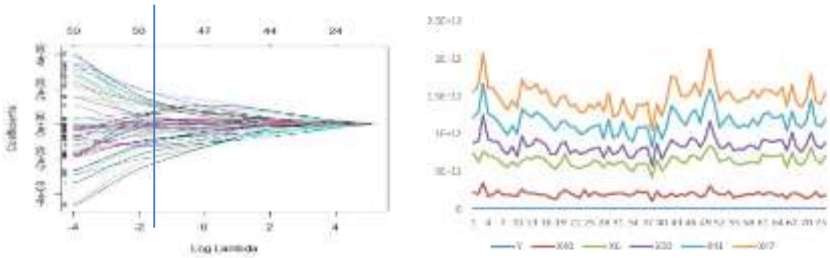


(a) (b)

Figure 5. (a) Significant Beta Coefficient and (b) the Level of Importance of the LASSO Variable

b) Elastic Net

Elastic net uses two shrinkage penalties: the Ridge Regression penalty $\left(\sum_{j=1}^{p} \beta_j^2\right)$ and the LASSO penalty$\left(\sum_{j=1}^{p} |\beta|\right)$ [14]. The optimal shrinkage parameter (λ) for the Elastic Net is obtained by a cross-validation method (Cross Validation-CV) which produces a minimum CV of 0.2. Figures 6 **(a)** and Figures 6 **(b)** show the level of importance of the five variables (variable importance performance/VIP) generated by Elastic Net.The most important variable level based on the resulting VIP value is X1(100); X41(94,38); X47(74,30); X30(73,89); X6(65,14). After forming the plot, it is seen that each variable has the same plot pattern so that it is included in the group of significant variables.



(a) (b)

Figure 6. (a) Significant Beta Coefficient, (b) Plot of Level of Interest of Elastic Net Variables

3.3. Comparison of Calibration Modeling Method

The comparative analysis of the size model's goodness using the Lasso Regression and Elastic net methods using the RMSE value. The smaller the RMSE value, the better the model obtained [15]. In addition, it can indicate that the dominant independent variable has been selected from each model. The Elastic Net method has the smallest RMSE value, which is 22.39. If the results of identifying the level of interest variables in the four models are combined, the coefficients of the variables that have implications for each other are X6, X1, and X41. The results of the comparison between the four methods can be seen in Table 1.

Table 1. Comparison of Lasso method and Elastic net

Year	Method	Lambda Optimum	RMSE	Score Correlation	Level Variables of Interest
2019	Lasso	0,2	23.57	0.968	**X1**, **X6**, X30, **X41**, X47
	Elastic Net	**0,2**	**22.39**	**0.971**	**X1**, **X6**, X30, **X41**, X47

Checking the model's goodness is not limited to looking at the RMSE value. The excellence of the model is also seen in the data pattern between the predicted value and the actual data. The graph (Figure 7) shows that the design of predicted values obtained from the lasso regression method and the elastic net has asimilar pattern following the actual data, which is shown in the orange line showing the true value and light green and dark green showing the lasso and elastic net regression.

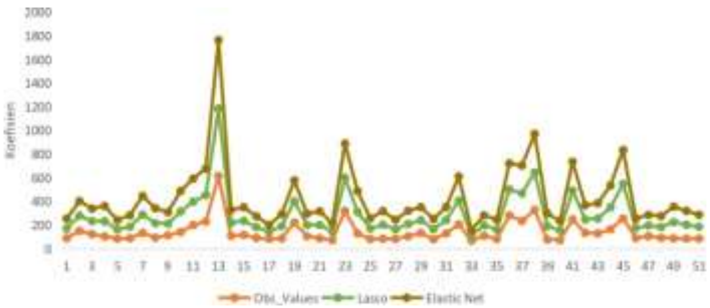


Figure 7. The plot of predicted Y against actual Y between models

4. CONCLUSIONS

Summarizing the area of the trapezoid in calibration modeling provides a reasonable estimate because itcan utilize the general information from the data correctly and is appropriately used on the residual intensitygraph, which tends to fluctuate, as in 2019. Based on several approach methods, the analysis of the Elastic Net Model produces the smallest RMSE value and the lowest value. The most significant correlation for predicting calibration data compared to other methods. Researchers can improve the performance of non-invasive blood glucose level detectors in terms of time and intensity residual spectrum results by paying attention to selected variables and groups of variables that can explain the total blood glucose levels in 2019.

REFERENCES

[1] J. Jiang, Linier and Generalized Liniear Mixed Models and Their Applications, Springer, New York, 2007.

[2] S. A. P, P. Anugrah, A. Dharma, C. Wijaya and W. Lorraine McCarty, Patofisiologi : konsep klinis proses-proses penyakit, Jakarta, 1995.

[3] T. Naes, T. Isaksson, T. Fearn and T. Davies, A User-Friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, 2002.

[4] Avan, "Pendugaan Konsentrasi Glukosa dalam Darah Menggunakan Jaringan Syaraf Tiruan pada Alat Non-Invasive," Repository IPB, 2017.

[5] R. H. Nur, "Pemodelan Kalibrasi Alat Non-Invasive Glukosa dalam Darah dengan Pendekatan Regresi Spline," Repository IPB, 2019.

[6] Rosni, "Pendekatan regresi komponen utama dan regresi kuadrat terkecil parsial untuk menduga kadar glukosa darah non invasif," Repository IPB, 2019.

[7] L. Rassiyanti, "Analisis Regresi Ridge Robust pada Pengukuran Kadar Glukosa Darah Non-Invasif," Repository IPB, 2020.

[8] Herianti, "Pemodelan Kalibras Regresi Kekar (Pengukuran Kadar Glukosa Darah Non-Invasif)," Repository IPB, 2020.

[9] Y. Sinaga, "Penerapan Metode Loopy Multiplicative Scatter Correction Dan Orthogonal Signal Correction Pada Spektrum Keluaran Alat Non-Invasif," Repository IPB, 2019.

[10] T. T. R. &. F. J. Hastie, The Elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed.), Springer, New York, 2009.

[11] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological),* pp. 267-288, 1996.

[12] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B (Statistical Methodology),* pp. 301-320, 2005.

[13] M. Y. a. Y. Lin, "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society. Series B (Statistical Methodology),* pp. 49-67, 2006.

[14] F. M. Sari, K. A. Notofiputro and B. Sartono, "Analisis Tingkat Kemiskinan di Provinsi Sumatera Barat Melalui Pendekatan Regresi Terkendala (Ridge Regression, LASSO, dan Elastic Net)," *STATISTIKA: Forum Teori dan Aplikasi Statistika,* pp. 29-36, 2021.

[15] Sudjana, Metoda Statistika, Bandung: Tarsito, 2005.