



Using Bayesian Ridge Algorithm to Predict Effectiveness of Body Fat Measurement

Penggunaan Algoritma Bayesian Ridge Untuk Memprediksi Efektivitas Pengukuran Lemak Tubuh

Rachma Yuni Andari¹, Revanza Akmal Pradipta², Denny Oktavina Radianto³

¹Program Studi Informatika, Fakultas Komputer dan Teknik, Universitas Alma Ata, Indonesia

²Jurusan Teknik Otomasi, Teknik Kelistrikan Kapal, Politeknik Perkapalan Negeri Surabaya, Indonesia

³Teknik Bangunan Kapal, Teknik Bangunan Kapal, Politeknik Perkapalan Negeri Surabaya, Indonesia

E-Mail: ¹rachmayuniandari@gmail.com,

²revanzaakmalpradipta@gmail.com, ³dennyokta@gmail.com

Received Mar 26th 2023; Revised Apr 14th 2023; Accepted Apr 30th 2023

Corresponding Author: Rachma Yuni Andari

Abstract

Body fat is an important aspect in understanding and managing one's physical condition. Accurate measurement of body fat percentage is essential to help accurately plan future health plans. Currently, the method of measuring body fat is still traditional and quite difficult, so what is needed is a more effective method. The Bayesian Ridge Algorithm is a linear regression technique that uses Bayesian inference to estimate the parameters of the model. In this study, it was used to predict the effectiveness of measuring body fat, which is a method often used to evaluate a person's overall health and physical condition. This algorithm takes into account factors such as age, gender, and body mass index (BMI) to make predictions about a person's body fat percentage. The results from this study can be used to improve the accuracy of body fat measurement and help individuals better understand and manage their health. The results of this study indicate that the model has very high accuracy (more than 99%).

Keyword: Body Fat, Bayesian Ridge Algorithm, Body Mass Index, Linear Regression

Abstrak

Lemak tubuh merupakan aspek penting dalam memahami dan mengelola kondisi fisik seseorang. Pengukuran persentase lemak tubuh yang akurat sangat penting untuk membantu memantau merencanakan rencana kesehatan mendatang secara tepat. Saat ini metode pengukuran lemak tubuh masih tradisional dan tergolong susah maka yang diperlukan adalah cara / metode yang lebih efektif. Algoritma Bayesian Ridge adalah teknik regresi linier yang menggunakan inferensi Bayesian untuk mengestimasi parameter dari model. Dalam penelitian ini, digunakan untuk memprediksi efektivitas pengukuran lemak tubuh, yang merupakan metode yang sering digunakan untuk mengevaluasi kesehatan dan kondisi fisik seseorang secara keseluruhan. Algoritma ini memperhitungkan berbagai faktor seperti usia, jenis kelamin, dan indeks massa tubuh (IMT) untuk membuat prediksi tentang persentase lemak tubuh seseorang. Hasil dari penelitian ini dapat digunakan untuk meningkatkan akurasi pengukuran lemak tubuh dan membantu individu lebih memahami dan mengelola kesehatannya. Hasil dari penelitian ini menunjukkan bahwa model memiliki akurasi yang sangat tinggi (lebih dari 99%).

Kata Kunci: Algoritma Bayesian Ridge, Indeks Massa Tubuh, Lemak Tubuh, Regresi Linier

1. INTRODUCTION

In order to measuring body fat is an important aspect in understanding and managing one's health and physical condition. This can provide insight into a person's risk for certain health conditions such as obesity, heart disease, and diabetes. Traditionally, body fat percentage is measured using various methods such as skin fat measurement, bioelectrical impedance analysis, and multiple X-ray absorptiometry (DXA). This method is considered the gold standard for measuring body fat, but has several drawbacks.

For example, skin fat measurement and bioelectrical impedance analysis may be less accurate than DXA and also require trained personnel to perform the measurements. Meanwhile, DXA is considered the most accurate method, but it can be expensive, time consuming, and requires special equipment. [1] Body

composition measurements from DXA have been available since DXA technology was developed 30 years ago, but are historically underutilized

The Bayesian Ridge Algorithm is a statistical method that can be used as an alternative to these traditional methods of predicting body fat percentage. This is a linear regression technique that uses Bayesian inference to estimate the parameters of the model. This algorithm is based on factors such as age, gender, and body mass index (BMI) to make predictions about a person's body fat percentage. [2] Bayesian linear regression is one of regression modeling with parameter estimation method using Bayesian approach. In the Bayesian approach there is a prior, likelihood distribution and posterior distribution.

There are two types of supervised learning tasks: regression for continuous variables and classification for discrete variables [3]. [4] Bayesian Ridge Regression model is used in our case due to variables of OEE are correlated with each other [5]. Since this model addresses the collinearity issue, it is anticipated to have very high accuracy. Machine learning comes into play since it allows a programme to learn from example data without being programmed, which is why we picked it over exact techniques. From Arthur Samuel [6], this definition of machine learning is derived.

2. MATERIALS AND RESEARCH METHODS

The frequentist method of performing linear regression is based on giving each independent variable a weight or coefficient, representing its impact on the dependent variable. To account for noise and other outside variables, an error term is also taken into consideration. Ordinary Least Squares (OLS) is a method for fitting models that reduces residual sums of squares (see equation 1), where y is the dependent variable, w are the coefficients, and x_i is one of the M instances) [7]. In other words, by modifying the w coefficients, the aim is to reduce the model error.

$$\min \sum_{i=1}^M (y_i - \omega^T x_i) \quad (1)$$

Overfitting is an issue with this kind of regression, especially when the data shows multicollinear patterns. Regularization is frequently utilized to address this problem. A typical tactic is to use an L2 regularizer to penalize the magnitude of the coefficients, resulting in the called ridge regression [7], [8]. The w coefficient values of the OLS method have single point estimates, which means that these values are the ones most likely to be accurate given the training data. The result of the model's uncertainty are not, however, at all taken into account. By modeling the regression with probability distributions rather than single value estimations the Bayesian approach addresses this problem. Equation 2 (using the same definitions from Equation 1, X being a matrix with all M instances) [7], [8].

$$y \sim \mathcal{N}(w^T X, \sigma^2) \quad (2)$$

One technique for dealing with probability issues is the Bayesian theorem. This algorithm's working principle is that the likelihood of an event occurring in the future relies on an earlier occurrence. The Bayesian theory may forecast future events, according to statistics and probability theory. Without complicated hyperparameters, the Bayesian model is straightforward and simple to construct, making it suitable for small, medium, and big datasets. The Bayesian approach is frequently used for regression and classification even though the technique is straightforward [10].

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (3)$$

A regularized variation of the linear regression model is ridge regression. The cost aggravates a regularization term equal to $\beta \sum_{j=1}^M m_{\theta_j^2}$. During dataset training can the regularization factor be applied to the cost function [11]. The following equation represents the ridge regression cost function [11]:

$$Y(\theta) = MSE(\theta) + \beta \frac{1}{2} \sum_{j=1}^M \theta_j^2 \quad (4)$$

As a result, rather than the precise values of the w coefficients, the result will be their posterior distribution. In this case permits the use of priors, which is advantageous if pertinent information about the model is already available. Additionally, the posterior distribution provides a range of potential w coefficients based on the data and the prior [8], which accounts for the model uncertainty. Relevant aspect becomes more important when the number of cases utilized to fit the model is decreased. In reality, because there is less

uncertainty as the number of examples rises, the w coefficients approach those obtained by the OLS method. Additionally, the L2 regularization is implicitly used when the prior also has a Gaussian distribution, leading to the creation of the Bayesian ridge regression idea [6], [8].

Drawing samples from the posterior distribution to enhance and approximatively fit a Bayesian regression is one of the most used approaches, for instance using Monte Carlo methods. The Maximum A Posterior (MPA) method is another popular strategy [6]. [8]. In order to meet the goals of estimating regression coefficients and choosing variables, Bayesian Ridge Regression is presented as a solution to the problem of multicollinearity in linear regression. It is based on Bayesian knowledge. BRR adds ridge regression to PLSR in a Bayesian sense. By choosing the prior distribution, Bayes may provide the ridge regression as the mean of the posterior distribution, lowering the model's prediction error [12].

A Bayesian approach known as Bayesian ridge regression (BRR) makes the assumption that all regression coefficients have a shared variance. However, in an additive model, all makers with the same allele frequency account for the same amount of the additive variances and have the same shrinkage effect [13]. The estimate procedure in Bayesian regression involves repeated experimental data and canonical concepts. It can avoid overfitting and data-adaptive. As more evidence is gathered, the Bayesian estimating process builds an initial estimate, which is then continuously improved [14]. The results from this study can be used to improve the accuracy of body fat measurement and help individuals better understand and manage their health. Additionally, the cost-effectiveness of this algorithm is also comparable to traditional methods, which could make it a more affordable option for people who want to measure their body fat percentage.

2.1 Dataset

Downloaded from Kaggle (kaggle.com). This data can be used to multiple regression. It is desirable to have simple methods for evaluating body fat that are neither difficult or expensive to measure body fat accurately. The abstract "Generalized body composition prediction equation for men using simple measurement techniques", by K.W. Penrose, A.G. Nelson, and A.G. Fisher, FACSM, Human Performance Research Center, Brigham Young University, Provo, Utah 84602, is listed in *Medicine and Science in Sports and Exercise*, vol. 17, no. 2, April 1985, p. 189. These data are used to create the predictive equations for lean body weight. Out of the 252 cases given below, the first 143 were used to create the prediction equation. The Bayesian Ridge regression algorithm will be applied to predict body fat according to the methodology described.

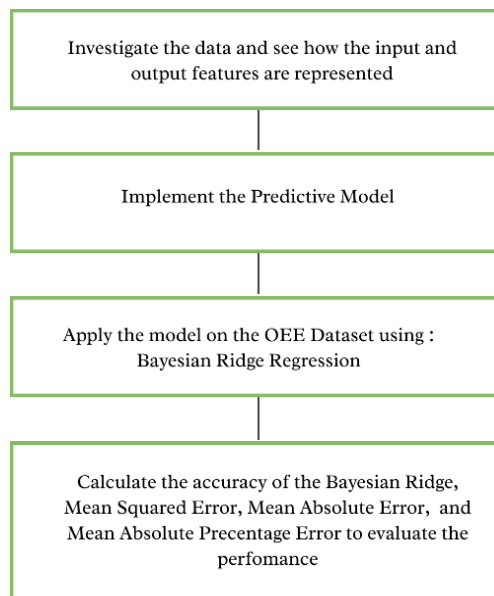


Figure 1. Research Methodology

Investigate by dividing it into sub-sections of data understanding, data loading, and explanatory data analysis, and data preparation. The steps taken are how to add and delete data that is not needed and clean and conduct initial investigations to analyze characteristics, find patterns, anomalies, and data assumptions. Next, the implementation of the model that has previously been considered for selection is carried out. The steps are in the form of evaluating the performance of each algorithm and then determining which algorithm provides the best predictions. Next is to calculate accuracy using MSE, MAE, and MAPE. The concept is that if the prediction is close to the true value, which can be concluded that the performance is good. The reason why the

Bayesian Ridge algorithm model was chosen is because Bayesian gives the closest result. After making changes to the related parameter variables.

2.2 Data Investigation (Data Understanding and Data Preparation)

This stage includes data loading, exploratory data analysis, multivariate analysis. The dataset is downloaded from the kaggle platform. Then the data is imported and explored.

2.3 Implementation of Predictive Models

Bayesian Ridge Regression is used as a development model to estimate body fat rate based on additional datasets and formulas. This stage is data preparation. Model training uses the `train_test_split` function with 70% training data sharing and testing data being the remainder. The dataset for men is attached while the data for women we consider to make.

Table 2. Data of BMI (76 rows x 5 columns)

	Age	Weight	Height	Density	BMI
65	55	82.575	171.45	1.030.907	2.809.141
59	61	808.875	167.005	1.025.626	2.900.163
197	42	75.6	181.61	1.051.032	2.292.146
195	42	81	173.355	1.040.344	2.695.332
156	28	925.875	175.26	1.039.017	3.014.302
191	42	1.099.125	193.04	1.033.717	2.949.528
154	27	901.125	186.69	1.050.875	258.549
...
112	47	802.125	177.8	1.041.978	2.537.339
37	50	885.375	173.355	1.029.847	2.946.148
171	35	565.875	166.37	1.061.359	2.044.422
97	50	73.125	168.91	1.039.793	2.563.039
115	40	71.1	175.895	1.051.901	2.298.067
148	25	646.875	184.15	1.070.408	1.907.554

Table 2 is a table that contains data with total of 76 rows and 5 columns. The table shows how the BMI calculation is obtained from parameter of age and weight. The formula for the calculation:

$$BMI = \frac{weight}{(height)^2} \quad (5)$$

2.4 Applying the Model to the Dataset using: Bayesian Ridge

After the program is trained, we apply it for predictions. By providing input values, the program will provide output values. In this step, we will evaluate the Prediction chart for Bayesian Ridge Regression.

2.5 Calculation of Bayesian Ridge Accuracy, Mean Squared Error, Mean Absolute Error, and Mean Absolute Percentage Error

Up to the model evaluation stage. Metrics are used to evaluate how well a model predicts value. If the predictions are close to the true values, the performance is good. Meanwhile, if not, poor performance. By assessing the accuracy of the model, the accuracy of the three models can be determined. We also created three metrics: MAE, MSE, and MAPE which will help calculate how accurate the model is.

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \quad (6)$$

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2 \quad (7)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (8)$$

Where:

- n : Amount of data
- y : Actual value
- \hat{y} : Predicted value

3. EXPERIMENT AND ANALYSIS

3.1. Data Investigation

This dataset is related to data that is usually associated with calculating body fat for a person. This data provides estimates of the percentage of body fat determined by various parameters for 252 men. For women, a separate formula is used. Below are correlation heatmaps. Correlation heatmaps are a type of plot that visualize the strength of a numerical variable relationship. Correlation plots are used to understand which variables are related to each other in terms of the strength of the variable relationship.

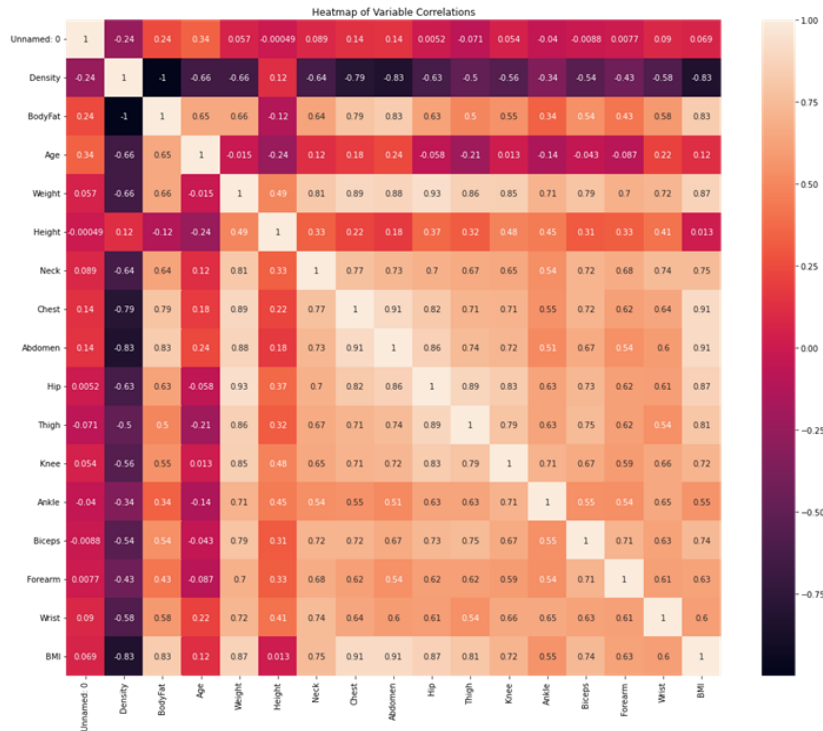


Figure 2. Correlation heatmaps body fat's dataset

In addition, there is also how the distribution of body fat so that the correlation can be seen later.

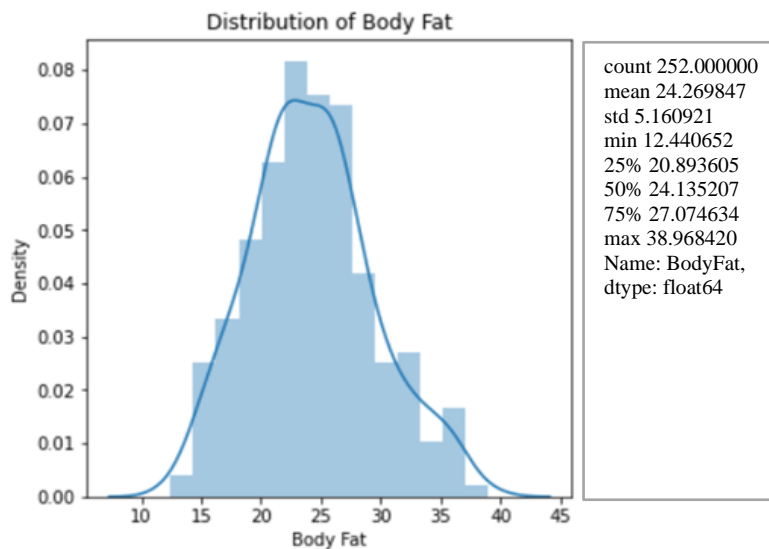


Figure 3. Distribution of Body Fit

The inputs to be defined are: Gender, Age, Weight, Height. The graph will be displayed in a table containing complete data. The dataset will be used for supervised learning. Observation is made of how the input and output are presented

3.2. Implementation of Predictive Models

The model built is Bayesian Ridge regression. For that reason we also use the NumPy, pandas, tensorflow, seaborn, and matplotlib libraries

3.3. Applying the Model to the Dataset using: Bayesian Ridge

In this step, we use graphs to visualize the prediction results. This allows us to compare the true value to the predicted value. This will tell us how strong our program is at making good predictions.

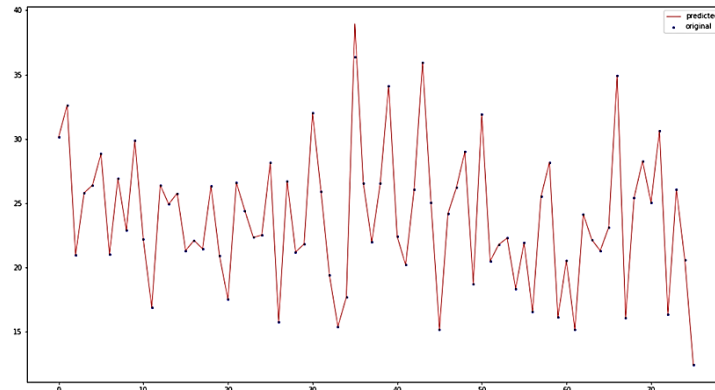


Figure 4. Body Fat Distribution based on Density and Body Fat

The actual curve and the predicted value can be compared. The level of precision according to the graph shown is very high.

3.4. Bayesian Ridge Accuracy Calculation, Mean Squared Error, Mean Absolute Error, and Mean Absolute Percentage Error

At this stage, we will calculate the accuracy, MAE, MAPE and MSE which will allow us to perform with precision.

Table 2. Score Accuracy of Bayesian Ridge Regressions (MSE, MAE, and MAPE)

MSE	MAE	MAPE	Accuracy
0,089267	0,001165	0,001166	99.6%

Mean Absolute Error (MAE) are two of the many methods to measure the level of accuracy of a forecasting model. The MAE value represents the average absolute error between the forecasting results and the actual value [15]. Using a linear regression algorithm can provide a predictive value of rice production with 2 variables of population growth and annual rice production, while the accuracy of the prediction calculation results uses the Mean Absolute Error (MAE) method which is used to measure the accuracy of a forecasting model.

The machine learning Regression Prediction Method is one of the most effective methods for predicting various features based on the input dataset. Because it provides a model with input and output features, the model will be able to predict the output data for each input. We chose to utilize the Bayesian Ridge model, with the following inputs: gender, age, height, and weight, and body fat rate as output. The use of this algorithm will help stakeholders improve the effectiveness of body fat measurement predictions. We import the data set and serve the input feature. After that we create a model using the listed parameters. After running the model, it displays two graphs: the predicted value provided by the model and the actual value. Figure 3 shows the small difference between the two graphs, and to make the difference visible, we calculated the accuracy, MAE and we found 99.6% accuracy which is very high. To justify the performance of our model, we have compared the precision with other methods. The table below shows how much the results are for the various jobs performed.

4. CONCLUSION

An indicator of the effectiveness of the parameters as a whole is how to use the bayesian ridge algorithm to predict the size of body fat with good decisions. This is the reason, we chose a machine learning algorithm that is known for its ability to learn how to predict a good system. We chose Bayesian Ridge because it deals with the collinearity problem. This model shows very high accuracy (over 99%). We also calculate the three indicators MAE, MSE and MAPE. The perspective of this research is to determine how well this model can

predict body fat. In this case, we have a dataset divided into training and test sets and also implementation of formulas.

REFERENCES

- [1]. J. A. Shepherd, B. K. Ng, M. Sommer, and S. B. Heymsfield, "Body composition by DXA," *Bone*, vol. 104, pp. 101–105, Nov. 2017, doi: 10.1016/j.bone.2017.06.010.
- [2]. S. D. Permai and H. Tanty, "Linear regression model using Bayesian approach for energy performance of Residential Building," *Procedia Computer Science*, vol. 135, pp. 671–677, 2018. doi:10.1016/j.procs.2018.08.219
- [3]. Scikit Learn, "Linear Models," scikit, https://scikit-learn.org/stable/modules/linear_model.html (accessed Jun. 11, 2023).
- [4]. S. Patnaik, X.-S. Yang, and I. K. Sethi, *Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019*. Singapore: Springer, 2021.
- [5]. K. P. Burnham, D. R. Anderson, and K. P. Burnham, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer, 2010.
- [6]. A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210.
- [7]. K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT press, 2012.
- [8]. M. H. Kutner, C. J. Nachtsheim, J. Neter, W. Li et al., *Applied linear statistical models*. McGraw-Hill Irwin New York, 2005, vol. 5.
- [9]. M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. Jun, pp. 211–244, 2001.
- [10]. W.N. van Wieringen Lecture Notes on Ridge Regression arXiv preprint arXiv:150909169 (2015)
- [11]. A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastapol, CA: O'Reilly, 2023.
- [12]. G. Assaf, M. Tsionas, and A. Tasiopoulos, "Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression," *Tourism Management*, vol. 71, pp. 1–8, Apr. 2019, doi: <https://doi.org/10.1016/j.tourman.2018.09.008>.
- [13]. D. Gianola, "Priors in Whole-Genome Regression: The Bayesian Alphabet Returns," *Genetics*, vol. 194, no. 3, pp. 573–596, May 2013, doi: <https://doi.org/10.1534/genetics.113.151753>.
- [14]. S. Le-qu, "Anomaly Detection Based on Sparse Bayesian Regression," *Computer and Modernization*, Jan.2015.
- [15]. A. Suryanto, "PENERAPAN METODE MEAN ABSOLUTE ERROR (MEA) DALAM ALGORITMA REGRESI LINEAR UNTUK PREDIKSI PRODUKSI PADI," *SAINTEKBU*, vol. 11, no. 1, pp. 78–83, Feb. 2019, doi: <https://doi.org/10.32764/saintekbu.v11i1.298>.