Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4779

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

Predictive Modeling of Preeclampsia Risk Using Random Forest Algorithm within a Machine Learning Framework

Harizahayu[1)*], Friendly[2)] Bintarto Purwo Seputro[3)] Benar[4)] Koko Hermanto[5)]
[1)2)3)4)] Medan State Polytechnic, North Sumatra, Indonesia
[5)] Sumbawa University of Technology, East Nusa Tenggara, Indonesia
[1)]harizahayu@polmed.ac.id, [2)]friendly@polmed.ac.id, [3)]bintarto@polmed.ac.id, [4)]benar@polmed.ac.id
[5)]koko@untes.ac.id

ABSTRACT

Preeclampsia is a serious pregnancy complication characterized by high blood pressure, potentially leading to organ damage, making early risk prediction crucial to reducing maternal morbidity and mortality. This study aims to develop a preeclampsia risk prediction model using medical and clinical data from 80 patients at Rumah Bersalin Sadan. The data include demographic profiles, blood pressure, weight, maternal age, preeclampsia history, body mass index, number of previous pregnancies, as well as genetic and environmental factors. The dependent variable is the risk of preeclampsia, either as a binary outcome (yes/no) or as a continuous risk score. The predictive model was built using multivariate linear regression and the Random Forest algorithm. The results showed that the Random Forest model achieved an accuracy of 65.22%, with an F-statistic of 7.345 and a very small p-value (1.908e-06), indicating that the model effectively explains data variability. However, the low Kappa value suggests room for improvement through feature refinement, hyperparameter tuning, or exploring other algorithms. Although these findings suggest that Random Forest is a promising method, further evaluation and model optimization are needed to enhance predictive performance and determine whether this method is the most suitable for the dataset used.

Keywords: Preeclampsia; Random Forest algorithm; low Kappa; maternal morbidity and mortality; multivariate linear regression;

1. INTRODUCTION

Preeclampsia is a hypertensive disorder of pregnancy that significantly affects maternal morbidity and mortality worldwide. Preeclampsia occurs in 5-7% of all pregnancies and is the main cause of maternal death in developing countries. The most dominant factor related to the incidence of preeclampsia in pregnant women at RSUP Dr. Mohammad Hoesin Palembang in 2015 had a history of hypertension (Gustri et al., 2016). According to (Ryan Dana et al., 2024) hypertension is one of the problems that causes preeclampsia. In Indonesia, preeclampsia causes 30-40% of maternal deaths and 30-50% of perinatal deaths.Preeklampsia berhubungan dengan perubahan patologis yang signifikan dari pembuluh darah ibu dan janin serta plasenta.

Then research was carried out at the Obstetrics and Gynecology Polyclinic at Kediri District Hospital to determine health behavior and risk factors that influence the incidence of preeclampsia in pregnant women. The research design used is descriptive analytic with a cross sectional approach. This research was conducted in June-August 2017 at the Obstetrics Gynecology Polyclinic, Kediri District Hospital (Yuniarti et al., 2018).

Preeclampsia is a serious complication that can occur during pregnancy and is one of the main causes of maternal and perinatal deaths throughout the world. Preeclampsia is characterized by increased blood pressure and proteinuria after 20 weeks of pregnancy (Setyawati et al., 2018) (Kencana Dewi, 2021), and can cause a variety of serious health problems for pregnant women and their unborn babies (Budiarti et al., 2022). Despite intensive research, prediction of preeclampsia remains a challenge due to the multifactorial nature of the disease.

Sadan Medan Selayang Maternity Home, as an integral part of the community health service system in the region, has a great responsibility to provide quality health care to pregnant women. This maternity home not only functions as a birthing place, but also as a center that provides a variety of prenatal and postnatal health services to ensure the welfare of mother and baby (S, 2020). However, challenges in the world of maternal health, such as

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4779

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

preeclampsia, require improving the quality of care through a more proactive approach. Early identification of risk factors for preeclampsia and prediction of the possibility of this condition occurring in every pregnant woman is very important (Masruroh, 2021). This effort includes the use of advanced medical technology, ongoing training for health workers, as well as health education and counseling for pregnant women. With a holistic and integrated approach, the Sadan Medan Selayang Maternity Home is expected to be able to significantly reduce the risk of preeclampsia, thereby having a positive impact on the health of mothers and babies and improving the overall quality of health services in the area.

In this context, developing a preeclampsia risk prediction model using multivariate linear regression is very important (Aminuddin et al., 2013). This model will not only allow the identification and evaluation of potential risk factors contributing to preeclampsia in a more comprehensive manner (Murniati et al., 2021), although also provides a solid basis for more informed clinical decision making. Based on data collected from the population of pregnant women who come to the Sadan Medan Selayang Maternity Home, this model will process various variables such as maternal age, health history, blood pressure, body mass index, as well as other genetic and environmental factors.

This case study aims to identify the relationship between various potential risk factors, such as maternal age, baseline blood pressure, medical history, body mass index, and other factors, with the risk of preeclampsia. By utilizing multivariate linear regression (Kusuma & Hidayat, 2024), this study will develop a prediction model that can help identify pregnant women at high risk of preeclampsia early. In addition, this study also explores how the combination of these risk factors may influence the likelihood of developing preeclampsia, providing deeper insight into the mechanisms underlying this condition. With data systematically collected from the population of pregnant women at the Sadan Medan Selayang Maternity Home, this model will be designed to provide reliable predictions and be easy to apply in daily clinical practice.Furthermore, the Random Forest algorithm, introduced by L. Breiman in 2001, has had great success as a general method for classification and regression. This method combines several randomized decision trees and combines their predictions by averaging, showing excellent performance especially when the number of variables is larger than the number of observations. Additionally, Random Forest is flexible enough to be applied to large-scale problems, is easily adapted to a variety of learning tasks, and provides a measure of variable importance (Biau & Scornet, 2016).

Random Forest is also a machine learning algorithm that combines a series of tree classifiers; each tree casts one vote for the most popular class, and the results are combined to get a final prediction (Dai et al., 2018) (Sarica et al., 2017). Random Forest has high classification accuracy, is tolerant of outliers and noise, and does not experience overfitting (Belgiu & Drăguţ, 2016). This method has become one of the most popular in data mining research and is used in the field of biology for tasks such as classification, prediction, variable importance studies, variable selection, and outlier detection (Verikas et al., 2011). Selanjutnya, penelitian yang dilakukan oleh (Ryan Dana et al., 2024) menyatkan bahwa model akurasi random forest memilki tingkat akurasi yang lebih tinggi dibandingkan dengan model decision tree.

In this study, we aim to develop a predictive model for preeclampsia risk using the Random Forest algorithm. By leveraging a machine learning framework, we intend to analyze various maternal health indicators such as systolic and diastolic blood pressure, body mass index (BMI), and genetic or environmental factors, to create a model that can assist healthcare professionals in identifying high-risk pregnancies. The study also explores the performance of the Random Forest model in comparison to other machine learning methods and seeks to provide insights into the most influential factors contributing to preeclampsia risk.

This research is structured as follows: First, we review related work on preeclampsia prediction models. Next, we describe the dataset and methodology used for modeling, followed by a detailed explanation of the Random Forest algorithm and its application to the preeclampsia dataset. Lastly, we present the results, evaluate the model's accuracy, and discuss potential improvements for future studies.

2. METHOD

This model carefully combines various variables that have significant potential to influence the risk of preeclampsia, namely: medical history, clinical parameters and lifestyle factors. Using advanced machine learning algorithms, this model is able to provide accurate and reliable predictions, enabling early detection and timely intervention.

For creating a multivariate linear regression model to predict the risk of preeclampsia in pregnant women based

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4779

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

on the provided variables, the researchers would typically set up the problem as follows:

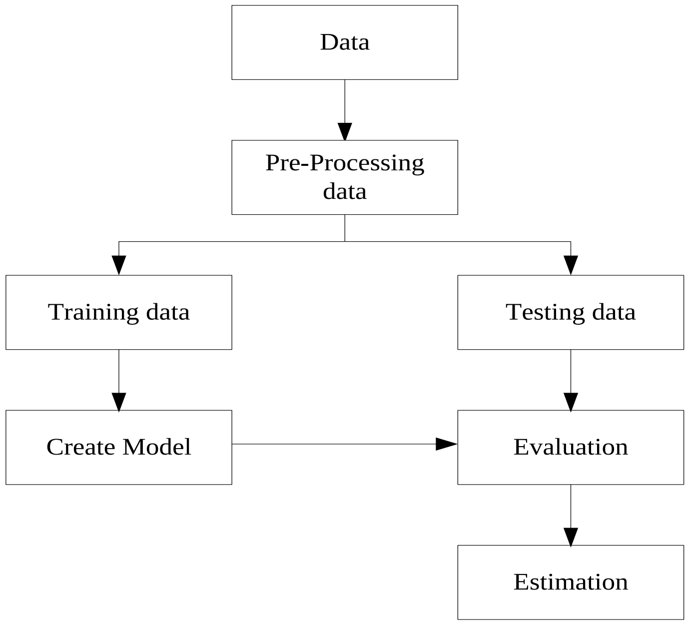
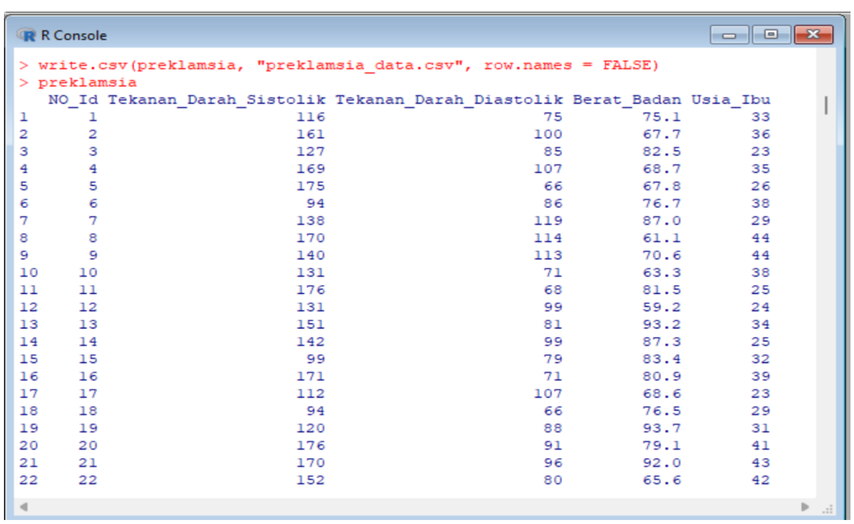


Figure 1. *The complete steps in modelling process and data evaluation*

Where variabels (feathures) consist of systolic and diastolic blood pressure (mmHg), weight (kg), mother's age (year), previous preeclampsia history (binary: yes/no), body mass index (kg), number of previous pregnancies, genetic/environmental factors (hypertension and smoking). Dependent variable (target) is risiko preeklamsia, where risk of preeclampsia can be a binary outcome: yes/no or a continuous risk score.

There are some steps for modelling:

1. Data preparation
   Firstly, it will be begun to encode categorical variables (previous preeclampsia history, genetic/environmental factors). This is data frame that is created by using R program:



R Console

```
> write.csv(preklamsia, "preklamsia_data.csv", row.names = FALSE)
> preklamsia
   NO_Id Tekanan_Darah_Sistolik Tekanan_Darah_Diastolik Berat_Badan Usia_Ibu
1      1                    116                      75        75.1       33
2      2                    161                     100        67.7       36
3      3                    127                      85        82.5       23
4      4                    169                     107        68.7       35
5      5                    175                      66        67.8       26
6      6                     94                      86        76.7       38
7      7                    138                     119        87.0       29
8      8                    170                     114        61.1       44
9      9                    140                     113        70.6       44
10    10                    131                      71        63.3       38
11    11                    176                      68        81.5       25
12    12                    131                      99        59.2       24
13    13                    151                      81        93.2       34
14    14                    142                      99        87.3       25
15    15                     99                      79        83.4       32
16    16                    171                      71        80.9       39
17    17                    112                     107        68.6       23
18    18                     94                      66        76.5       29
19    19                    120                      88        93.7       31
20    20                    176                      91        79.1       41
21    21                    170                      96        92.0       43
22    22                    152                      80        65.6       42
```

Figure 2. *Data Frame Preeclampsia*

* Corresponding author

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

And then, it will be continued by creating some scale continuous variables like systolic/diastolic blood pressure, weight, BMI, to ensure comparability. Langkah selanjutnya adalah setelah mengumpulkan data medis ibu hamil. Peneliti memastika dataset memiliki jumlah observasi yang cukup untuk melatih model. Kemudian memastikan tidak ada missing data atau handle missing data dengan metode seperti imputasi. Kemudian, melakukan normalisasi atau standarisasi pada variabel independen.

2. Multivariate Linear Regression Setup (Nurani & Rahmawati, 2015)
   The general form of the linear regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \epsilon$$

   Where:
   Y is the risk of preeclampsia
   $X_1, X_2, \cdots, X_n$ are the independent variables (systolic blood pressure, diastolic blood pressure, etc.)
   β_0 is the intercept
   $\beta_1, \cdots, \beta_n$ are the coefficients for each independent variable
   ϵ is the error term
3. Model Training
   For this step, it will be started by fitting the linear regression model using a dataset that includes the aforementioned variables. The following that splitting the dataset into training and testing sets to validate the model's performance.
4. Machine Learning Algorithm
   This reaserch use algorithms Linear Regression, Ridge Regression, and more advanced techniques like Random Forest which can handle more complex, nonlinear relationships and to predict the preeclampsia risk.

Finally, Machine Learning is a part of artificial intelligence that focuses on developing algorithms that allow computers to learn from data.

3. RESULT AND DISCUSSION

The results of Precalmpsia's analysis and data processing produce line diagrams so that they can help build classification models. So that it can be seen clearly, this model is used to predict categories or classes, in this case, the risk of preeclampsia which is given in binary terms, namely 0 or 1.

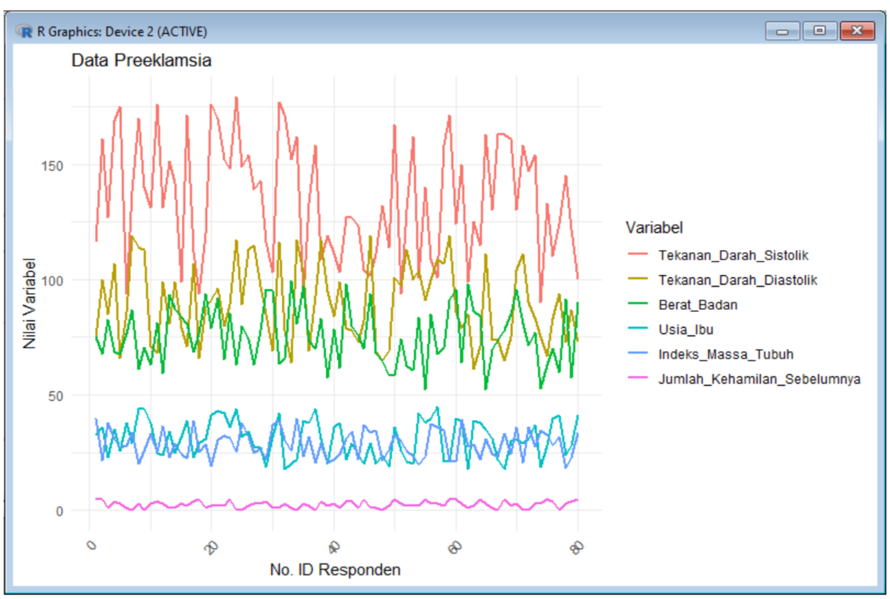


Figure 3. *Line Graph Data Preeclampsia*

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4779

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

Therefore, it is depicted in the form of a line diagram with a range of ID numbers from 1 to 80 respondents with independent variables on different scales, blood pressure, weight, age, number of pregnancies and body mass index.

```
R Console
> summary(model)

Call:
lm(formula = Preeklamsia ~ Berat_Badan_Log + Usia_Ibu + Riwayat_Preeklampsia_Se$
    Indeks_Massa_Tubuh + Jumlah_Kehamilan_Sebelumnya + Faktor_Genetik_Lingkunga$
    data = data_preeklamsia)

Residuals:
     Min       1Q   Median       3Q      Max
-12.5034  -3.5284  -0.9934   3.5236  13.7712

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -24.8513    16.5621  -1.500    0.137
Berat_Badan_Log                    5.5764     3.8575   1.446    0.152
Usia_Ibu                           0.1212     0.1099   1.103    0.273
Riwayat_Preeklampsia_Sebelumnya   -0.6147     1.1821  -0.520    0.604
Indeks_Massa_Tubuh                -0.1376     0.1423  -0.967    0.336
Jumlah_Kehamilan_Sebelumnya       -2.2578     0.3702  -6.099 2.42e-08 ***
Faktor_Genetik_Lingkungan_merokok  0.5632     1.5688   0.359    0.720
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.399 on 93 degrees of freedom
Multiple R-squared:  0.3215,    Adjusted R-squared:  0.2778
```

Figure 4. *Summary dan Output Regresi Linier Multivariat*

Based on the data processing output using the R program with several independent variables to predict preclampsia in pregnant women. The following are the results and explanations of several output elements as well as steps to create the same model:

1. Residuals
   Residuals show the difference between the actual value and the value predicted by the model. In this case, the minimum residual value is -12.5034 and the maximum value is 13.7712, with the first quartile (1Q) being -3.5284 and the third quartile (3Q) being 3.5236. The median residual is -0.9934, indicating that most predictions are quite close to the actual value, with a few outliers showing large residuals.
2. Coefficients
   Each coefficient shows the influence of the independent variable on the dependent variable, namely preeclampsia. The following is an explanation of each coefficient
   1. Intercept: The intercept value of -24.8513 means that when all independent variables are 0, the predicted value for Preeclampsia is -24.8513. However, this value is not statistically significant (p-value = 0.137).
   2. Log_Body_Weight: A coefficient of 5.5764 indicates that every logarithmic increase in Body_Weight increases the Preeclampsia value by 5.5764. However, this coefficient is not significant (p-value = 0.152), which means this relationship is not strong enough to be considered statistically important.
   3. Maternal_Age: A coefficient of 0.1212 means that every additional unit of maternal age increases preeclampsia by 0.1212, but this variable is also not statistically significant (p-value = 0.273).
   4. Previous_Preeclampsia_History: A coefficient of -0.6147 indicates that a previous history of preeclampsia slightly reduces the risk of preeclampsia, but this effect is not significant (p-value = 0.604).
   5. Body_Mass_Index: A coefficient of -0.1376 indicates a very weak negative relationship between body mass index and preeclampsia, but this variable is also not significant (p-value = 0.336).
   6. Number of Previous Pregnancies: A coefficient of -2.2578 indicates that each increase in the number of previous pregnancies reduces preeclampsia by 2.2578. This variable is very statistically significant (p-value = 2.42e-08), with a significance level of $p < 0.001$, which means this variable has a strong influence on preeclampsia.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4779

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

7. Genetic_Environmental_Factors_smoking: A coefficient of 0.5632 indicates a very weak positive relationship between genetic and environmental factors (including smoking) with preeclampsia, but this variable is not statistically significant (p-value = 0.720).

3. Residual Standard Error
   The residual standard error (RSE) of 5,399 shows how far the value predicted by the model is from the actual value, with an average deviation of 5,399 units from the predicted value.
4. R-squared dan Adjusted R-squared:
   1. Multiple R-squared: The R-squared value of 0.3215 indicates that this model explains approximately 32.15% of the variability in the observed preeclampsia data.
   2. Adjusted R-squared: The adjusted R-squared value of 0.2778 indicates adjustment for the number of variables in the model, where around 27.78% of the data variability is explained by the independent variables in this model. This value indicates that the model can still be improved to explain more variance.
5. F-statistic dan P-value
   F-statistic value of 7.345 indicates that this model is overall quite good at explaining data variability, with a very small p-value (1.908e-06), indicating that this model is significant overall. This means that at least one independent variable in this model has a significant relationship with preeclampsia. Based on the following model:

lm(formula = Preeklamsia ~ Berat_Badan_Log + Usia_Ibu + Riwayat_Preeklampsia_Sebelumnya +Indeks_Massa_Tubuh+Jumlah_Kehamilan_Sebelumnya+Faktor_Genetik_Lingkungan_merokok , data = data_preeklamsia)

$$f(Preeklamsia) = \beta_0 + Berat\_Badan\_Log\ X_1 + Usia\_IbuX_2 + Riwayat\_Preeklampsia\_SebelumnyaX_3 + Indeks\_Massa\_TubuhX_4 + Jumlah\_Kehamilan\_SebelumnyaX_5 + Faktor\_Genetic\_Lingkunga\_merokokX_6$$

$$f(Preeklamsia) = -24.8513 + 5.5764X_1 + 0.1212X_2 - 0.6147X_3 - 0.1376X_4 - 2.2578X_5 + 0.5632X_6$$

Based on the model above, only the variable Number of Previous Pregnancies is statistically significant in influencing preeclampsia. Meanwhile, other variables such as Log_Body_Weight, Maternal_Age, Previous_Preeclampsia_History, Body_Mass_Index, and Smoking_Environmental _Genetic_Factors were not statistically significant. Therefore, it may be necessary to explore other approaches such as adding interactions, transforming variables, or collecting more data to increase the significance of other variables. Next, the Random Forest algorithm was used to strengthen the results of the multivariate regression model analysis.

2. Making a Random Forest Matrix and Making a Random Forest Model Summary

```
> model_rf <- train(Risiko_Preeclampsia ~ .,
+                   data = preklamsiaTrain,
+                   method = "rf",
+                   metric = "Accuracy",
+                   tuneGrid = tuneGrid,
+                   trControl = control)
> |
```

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4779

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

```
> print(model_rf)
Random Forest

97 samples
 9 predictor
 2 classes: '0', '1'

No pre-processing
Resampling: Cross-Validated (5 fold)
Summary of sample sizes: 78, 78, 77, 77, 78
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  1     0.7415789  0.3728679
  2     0.7926316  0.5149225
  3     0.8031579  0.5536764
  4     0.7926316  0.5235120
  5     0.8131579  0.5668178

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 5.
```

The output shows a summary of the model built:

a. 97 samples: The total amount of data used to train the model.
b. 9 predictors: Number of predictor variables.
c. 2 classes: Shows that there are 2 classes, namely 0 and 1.

3. Resampling Tunning Results

Based on the model summary results:

a. No pre-processing: No data processing is performed before model training.
b. Resampling: Using 5-fold cross-validation to evaluate the model
c. Summary of sample sizes: Sample size for each fold.

Adapun hasil akurasi berdasarkan nilai mtry:

1. mtry 1: Akurasi 0.6905, Kappa 0.2273
2. mtry 2: Akurasi 0.7837, Kappa 0.5049
3. mtry 3: Akurasi 0.7937, Kappa 0.5220
4. mtry 4: Akurasi 0.7842, Kappa 0.4963
5. mtry 5: Akurasi 0.8047, Kappa 0.5429

The model chooses the value mtry = 5 as the best model because the highest accuracy is achieved at this value (0.8131579), indicating that the model provides the best performance by considering five random variables at each division of nodes in the tree.

```
> prediksi <- predict(model_rf, preklamsiaTest)
> akurasi <- mean(prediksi == preklamsiaTest$Risiko_Preeclampsia)
> cat("Akurasi Model:", akurasi, "\n")
Akurasi Model: 0.6521739
```

The model accuracy result on the test data is 0.6522, which shows that the model is correct in 65.22% of all predictions.

4. CONCLUSION

Based on the model, only the Number of Previous Pregnancies variable is statistically significant in influencing preeclampsia.

The F-statistic value of 7.345 indicates that this model is overall quite good at explaining data variability, with a very small p-value (1.908e-06), indicating that this model is significant overall. This means that at least one independent variable in this model has a significant relationship with preeclampsia.

The Random Forest model shows quite good results with an accuracy of around 65.22%, although there is still room for improvement. A low Kappa metric indicates that the model may not be good enough at separating the classes. You may consider doing more fine-tuning to the features, hyperparameter tuning, or trying other algorithms to improve accuracy.

Based on the results obtained, Random Forest can be a good method for modeling preeclampsia data, but there is room for improvement. Further evaluation with other methods and model optimization may help in determining

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4779

Submitted : Oct 5, 2024
Accepted : Oct 16, 2024
Published : Oct 16, 2024

whether Random Forest is the best method for this dataset. If you find another model that provides better results, consider using that method.

5. Acknowledgement

The authors gratefully express their appreciation and express their gratitude for the financial support provided through the 2024 Medan State Polytechnic DIPA funds with contract number: B/433/PL5/PM.01.05/2024 dated 1 August 2024.

6. REFERENCES

Aminuddin, Sudarno, & Sugito. (2013). Pemilihan Model Regresi Linier MUultivariat Terbaik dengan Kriteria Mean Square Error. In *JURNAL GAUSSIAN* (Vol. 2, Issue 1). http://ejournal-s1.undip.ac.id/index.php/gaussian

Biau, G., & Scornet, E. (2016). *A Random Forest Guided Tour*. http://www.kaggle.com/c/dsg-hackathon

Budiarti, I., Rohaya, R., & Silaban, T. D. S. (2022). Faktor-Faktor yang Berhubungan dengan Kejadian Bayi Berat Lahir Rendah (BBLR) di Rumah Sakit Muhammadiyah Palembang Tahun 2020. *Jurnal Ilmiah Universitas Batanghari Jambi*, *22*(1), 195. https://doi.org/10.33087/jiubj.v22i1.1927

Dai, B., Chen, R.-C., Zhu, S.-Z., & Zhang, W.-W. (2018). Using Random Forest Algorithm for Breast Cancer Diagnosis. *2018 International Symposium on Computer, Consumer and Control (IS3C)*, 449–452. https://doi.org/10.1109/IS3C.2018.00119

Gustri, Y., Januar Sitorus, R., & Utama, F. (2016). Determinants Preeclamsia in Pregnancy at RSUP DR. Mommadd Hoesin Palembang. *Jurnal Ilmu Kesehatan Masyarakat*, *7*(3), 209–217. https://doi.org/10.26553/jikm.2016.7.3.209-217

Kencana Dewi, N. M. R. (2021). Pola Pengobatan Antihipertensi pada Pasien Preeklampsia di Rumah Sakit Harapan Bunda. *Jurnal Sosial Sains*, *1*(7), 637–644. https://doi.org/10.59188/jurnalsosains.v1i7.143

Kusuma, M. D. H., & Hidayat, S. (2024). Penerapan Model Regresi Linier dalam Prediksi Harga Mobil Bekas di India dan Visualisasi dengan Menggunakan Power BI. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *5*(2), 1097–1110. https://doi.org/10.35870/jimik.v5i2.629

Masruroh, N. (2021). Determinan Maternal Kejadia Preeklampsia Pada Ibu Hamil Trimester III di RS Prima Husada Sidoarjo. *Jurnal Bidang Ilmu Kesehatan*, *11*(1), 94–104. https://doi.org/10.52643/jbik.v11i1.1072

Murniati, L., Taherong, F., & Syatirah, S. (2021). Manajemen Asuhan Kebidanan pada Bayi Baru Lahir dengan Asfiksia (Literatur Review). *Jurnal Midwifery*, *3*(1), 32–41. https://doi.org/10.24252/jmw.v3i1.21028

Nurani, V., & Rahmawati, R. (2015). Pada Studi Kasus Nilai Ujian Nasional 2014 SMP Negeri 1 Sayung. *Jurnal Gaussian*, *4*(3), 697–704. http://ejournal-s1.undip.ac.id/index.php/gaussian

Ryan Dana, A., Valentino Kristananda, R., Bagas Satrio Wibowo, M., & Arman Prasetya, D. (2024). Perbandingan Algoritma Decision Tree dan Random Forest dengan Hyperparameter Tuning dalam Mendeteksi Penyakit Stroke. In *Seminar Nasional Informatika Bela Negara (SANTIKA)* (Vol. 4).

S, H. H. (2020). Estimasi Maksimum Likelihood Melalui Algoritma Ekspektasi Maksimasi Untuk Model Regresi Linear dengan Data Hilang. *Jurnal Matematika*, *10*(2), 105. https://doi.org/10.24843/jmat.2020.v10.i02.p127

Sarica, A., Cerasa, A., & Quattrone, A. (2017). Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. In *Frontiers in Aging Neuroscience* (Vol. 9, Issue OCT). Frontiers Media S.A. https://doi.org/10.3389/fnagi.2017.00329

Setyawati, A., Widiasih, R., & Ermiati, E. (2018). Faktor-Faktor yang Berhubungan dengan Kejadian Preeklampsia di Indonesia. *Jurnal Perawat Indonesia*, *2*(1), 32. https://doi.org/10.32584/jpi.v2i1.38

Yuniarti, F., Wijayati, W., Ivantarina, D., Studi, P. D., & Karya Husada Kediri, S. (2018). Analisis Perilaku Kesehatan Dan Faktor Resiko Kejadian Preeklampsia Pada Ibu Hamil Di Poliklinik Obstetri Gynekologi RSUD Kabupaten Kediri. *Journal of Issues in Midwifery*, *1*(3), 1–17.

* Corresponding author