

PREGNANCY RISK LEVEL CLASSIFICATION USING THE CRISP-DM METHOD

Reka Dwi Syaputra^{1*}, Achmad Solichin²

Master of Computer Science,
Universitas Budi Luhur
Jakarta, Indonesia

<https://www.budiluhur.ac.id/>

^{1*}2011600075@student.budiluhur.ac.id, ²achmad.solichin@budiluhur.ac.id

(*) Corresponding Author

Abstract

Independent midwife practices have the task of reminding and maintaining the quality of standardized reproductive health services for pregnant women. Independent midwife practices have had patient visits since the covid-19 pandemic from 2020 to 2021, especially at the yetti puranama midwife, which consists of 320 pregnancy examinations, 130 delivery care, and 50 referrals. The covid-19 pandemic has impacted maternal mortality rates because there are still many restrictions on all services. Maternal health services include pregnant women who are routinely unable to go to the puskesmas or other healthcare facilities due to fear of contracting covid-19, which delays the examination of pregnancy gravida, abortion, temperature, pregnancy distance, haemoglobin, blood pressure, ideal weight, and decisions. So that the problem that occurs is an increase in the risk of pregnancy, resulting in death and increased maternal mortality. In solving this problem, the research takes a machine-learning approach. The research aims to build a classification of pregnancy risk levels that can predict early treatment in this study using the random forest method with cross-validation 2. This study obtained the results of an accuracy value of 98%, precision of 94%, and recalled 100% in the random forest method.

Keywords: pregnancy risk level, classification, decision tree, random forest.

Abstrak

Praktik bidan mandiri memiliki tugas mengingatkan dan mempertahankan kualitas pelayanan kesehatan reproduksi terstandar pada ibu hamil. Praktik bidan mandiri memiliki kunjungan pasien sejak pandemi covid-19 dari tahun 2020 sampai 2021 khususnya di bidan yetti puranama yang terdiri dari pemeriksaan kehamilan 320 orang, asuhan persalinan 130 orang dan rujukan 50 orang. Masa pandemi covid-19 memberikan dampak angka kematian ibu meningkat disebabkan masih banyak pembatasan ke semua layanan. pelayanan kesehatan ibu seperti ibu hamil yang menjadi rutinitas tidak dapat ke puskesmas atau fasilitas pelayanan kesehatan lainnya disebabkan takut tertular covid-19. Sehingga menundakan pemeriksaan kehamilan gravida, abortus, suhu, jarak kehamilan, hemoglobin, tekanan darah, berat badan ideal dan keputusan. Sehingga permasalahan yang terjadinya terdapat peningkatan tingkat risiko kehamilan berdampak kematian dan meningkatnya angka kematian ibu. Dalam memecahkan masalah ini, penelitian melakukan cara pendekatan dengan machine learning. Tujuan penelitian untuk membangun klasifikasi tingkat risiko kehamilan yang dapat memprediksi penanganan secara dini. Pada penelitian ini menggunakan metode random forest dengan cross-validation 2. Penelitian ini mendapatkan hasil nilai akurasi 98%, precision 94% dan recall 100% pada metode random forest.

Kata kunci: tingkat risiko kehamilan, klasifikasi, pohon keputusan, random forest.

INTRODUCTION

From 2007-2020, the coverage of K4 pregnant women's health services tended to increase. However, there was a decrease in 2020, from 88.55% to 84.60%. This decline is assumed to occur due to program implementation in areas

affected by the Covid-19 pandemic (Kementerian Kesehatan RI, 2021). Health services for pregnant women (K4) in 2020 in Bengkulu province by 87% (Badan Pusat Statistik Provinsi Bengkulu, 2020). Compared to 2021, the coverage of K4 pregnant women's health services tends to fluctuate. In 2021 the K4 rate was 88.8%, which is an increase



compared to the previous year. New adaptations can influence the increase in K4 coverage of the covid-19 situation in 2021 (RI, 2021). It is because in the previous year, 2020, there were still many restrictions on all routine services, including maternal health services, such as pregnant women who could not go to the health centre or other healthcare facilities for fear of contracting covid-19.

Thus delaying pregnancy checks and service uncertainty due to a lack of human resources, infrastructure, and personal protective equipment (PPE). Health services for pregnant women (K4) in 2021 in the province of Bengkulu are 89% (Kemenkes RI, 2021).

It is because in the previous year, 2020, there were still many restrictions on all routine services, including maternal health services, such as pregnant women who could not go to the health centre or other healthcare facilities for fear of contracting covid-19. So that it delays pregnancy checks and service uncertainty in terms of personnel, infrastructure, and personal protective equipment (PPE). Health services for pregnant women (K4) in 2021 in the province of Bengkulu are 89% (Badan Pusat Statistik Provinsi Bengkulu, 2020). While in 2021, the maternal mortality rate increased by 50 maternal deaths, with the cause of 22 pregnant deaths (44%), 11 maternity deaths (22%), and 17 postpartum deaths (34%) (Bengkulu, 2021).

Independent midwife practices have the task of improving and maintaining the quality of reproductive health services for pregnant women, especially in the independent midwife practice of yetti purnama, M.keb, which locate in tanah patah village, Ratu Agung sub-district, Bengkulu city. Based on reports from independent midwife practices before the outbreak of the covid-19 pandemic, health services in 1 year could have approximately 650 pregnant women patient visits. When the Covid-19 pandemic occurred, maternity health services declined from the beginning of 2020 to the end of 2021. There were 500 pregnant women patients with 320 pregnancy examinations, 130 delivery care and 50 referrals.

The problem faced during the covid-19 pandemic is the low level of pregnancy health services. Many pregnant patients do not come directly to the independent midwife practice. So there is an increase in pregnancy risk, impacting death and maternal mortality. In solving this problem, researchers took a machine-learning approach.

Machine Learning in understanding how it works aims to make computers learn data. At the same time, data mining plans to extract information

from large amounts of data (Putra, 2020). This research predicts the risk level by classifying the data set to determine whether it is risky. Then collected to build a machine learning model with a supervised learning algorithm model. This prediction occurs in a data set using the random forest method. The research entitled predicting the possibility of diabetes (Pavlov, 2018) in the early stages uses a random forest classification algorithm which produces an accuracy value of 97.88% by comparing three machine learning classification algorithms, namely support vector machine. Naive Bayes and random forest using Receiver Operating Characteristic (ROC) curve comparison (Apriliah et al., 2021).

From the citations of previous research, this research generally has similarities and differences with previous research. In the similarity of this research aims to detect disease early by using the random forest method. Also, the differences are the data and the object of research. This research focuses on pregnancy examination data, namely gravida, abortion, temperature, gestational distance, haemoglobin, blood pressure, ideal weight, and decisions so that the research aims to build a classification of pregnancy risk levels using the random forest method to detect early examination by providing an accurate level of risk and no risk during the Covid-19 pandemic.

This reason makes the design of a random forest method that can predict early detection and provide an accuracy level value, namely risk and not risk. In pregnancy, using the random forest method. With the division of datasets, namely the ratio of training 80% and testing 20% using cross-validation 2. Using the random forest method will get good accuracy results. By paying attention to the results of previous studies, it shows that the random forest method has classification results in a good category. Table 1 shows the previous related studies. Based on previous literature for solving existing problem concepts is shown in Figure 1.

RESEARCH METHODS

This research uses the CRISP-DM method (Cross-industry standard process for data mining), as seen in Figure 2. The stages refer to research entitled classification of pregnancy risk levels using the random forest method. The CRSIP-DM (Cross-Industry standard process for data mining) method contains stages carried out sequentially or in a hierarchical process, namely business understanding, data understanding, data pre-preparation, modelling, evaluation, and deployment (Mulaab, 2017).

Table 1. Literature Review Research

No.	Main Objectives	method	Results
1	Predicting the likelihood of diabetes (Apriliah et al., 2021)	Support Vector Machine, Naive Bayes and Random Forest	Using three algorithms obtained good accuracy value results, namely random forest with an accuracy value of 97%.
2	Random forest optimization for bank marketing data classification (Religia et al., 2021)	Using random forest optimization with bagging and generic algorithm	<i>Random forest with 88.30% accuracy</i>
3	Detection of patients with diabetes (Suryanegara et al., 2021)	Metode normalization (Min-Max normalization-RF), model ke 2 (Z-score normalization-RF) dan model ke 3 (Normalization-RF)	(Min-max normalization-RF) of 95.45%, model 2 (Z-score normalization-RF) of 95%, and model 3 (Normalization-RF) of 92%. These results show that model 1 (min-max normalization-rf) is better.
4	Clustering the Risk Level of Pregnant Women (Akbar et al., 2021)	K-Means Method	K-Means prediction can be modelled on clustering and is categorical.
5	Classification of preeclampsia status of pregnant women (Kustiyahningsih et al., 2020)	K-Fold Cross Validation and ID3 algorithm as data clustering	K-Fold Cross Validation and ID3 algorithm as accuracy data clustering using a confusion matrix
6	Pregnancy health diagnosis (Hikmatulloh et al., 2019)	ID3 Algorithm	This method can produce an accuracy of 80.33%
7	PSO optimization to improve the accuracy of the ID3 algorithm for pregnant women's disease prediction (Byna, 2019)	ID3 Algorithm	The ID3 classification model produces good accuracy with PSO, improving this study's accuracy.
8	Detection of pregnancy risk level (Wulandari & Susanto, 2018)	Fuzzy Mamdani method and <i>simple additive weighting</i>	Got 88% accuracy using the recognition rate model
9	Comparison of random forest and SVM classification methods (Adrian et al., 2021)	Random Forest and SVM Methods	Two data sets will be compared with the first SVM algorithm and Random Forest to get the most accurate sentiment analysis results.
10	Classification of human activities (Aziz, 2021)	Metode Ensemble Stacking	Ensemble the stacking method as classification. Compared to the SVM method for accuracy, sensitivity, and specificity.

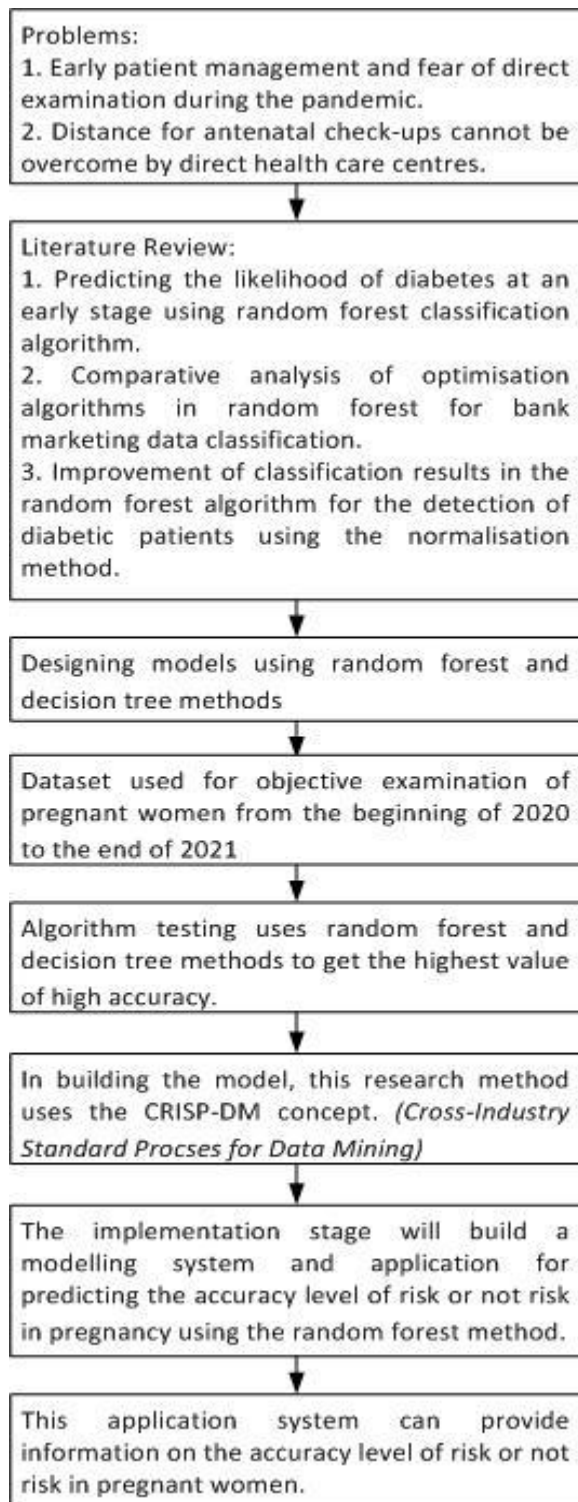


Figure 1. Conceptual framework

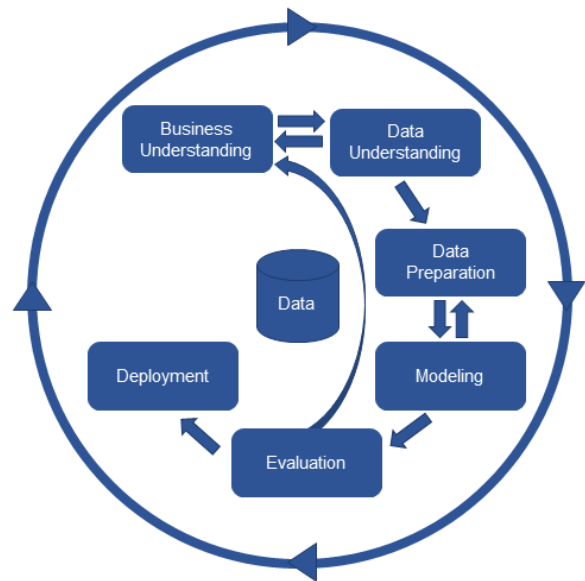


Figure 2. CRISP-DM model (Mulaab, 2017)

The stages of the CRISP-DM concept are as follows:

1. Business understanding
 This stage is a research object place for initial data collection.
2. Data understanding
 The stages of the data collection process are to build the random forest method and describe the data. The data used is nominal, namely data in the form of categories.
3. Data pre-preparation
 Data preparation stage as input to analyze the random forest algorithm. Data used in the objective examination. So it uses data processing techniques to overcome data problems and get valid data. All of that uses in data mining techniques.
4. Modelling
 The stage of forming a classification method to predict the category in one object class data, namely risk or no risk. In classifying, research uses the main method in classification, namely the random forest method. At the same time, the decision tree method compares good accuracy rates. The data division of this method has a ratio of 80% as training data and 20% as testing data with a cross-validation pattern of 2. So in the modelling stage of understanding classification in general, there is a system evaluation formula in the deployment and prototype.
5. Deployment
 This stage is the final stage in forming or building an application system. This application answers the formulation of the problem of predicting the level of risk of pregnancy. The model obtained using the basic method is random forest by comparing the decision tree method to find a good level of

accuracy. There are several steps in problem-solving carried out by researchers as follows:

1. Literature Study

A literature study related to this research topic by conducting related literature objectives that discuss theories around pregnancy, midwives, RStudio, data mining, machine learning, decision tree, and random forest. In addition, also a review of previous studies. Researchers conducted a literature study stage, namely looking for various study sources in discussing theories around RStudio, data mining, machine learning, decision tree, random forest, and midwives. Also, a review of previous research studies of the literature is in table 1. research literature review.

2. Data Collection

The author collects data through secondary data at the research site, while interviews are not for primary subjects. Objective examination underlies secondary data collection, namely gravida, abortion, temperature, gestational distance, haemoglobin, blood pressure, ideal weight, and decision. At the same time, interviews were conducted with midwives for researchers to explore pregnancy knowledge. In addition, the process of collecting data by researchers, namely non-participant observation. Observation by observing the visit book of pregnancy patients in independent midwife practices.

3. Pre-Processing data

Objective examination data classify by variable names sourced from secondary data and interviews to obtain valid data. At the same time, reconstructing data to form a data structure makes it more suitable for using the random forest method.

4. Modelling

The classification algorithm modelling in this study uses predictive categories, namely Decision Tree and Random Forest. In data mining, there is an evaluation to determine the algorithm's accuracy. Comparison of these two algorithms to create patterns of train data and test data with Cross-Validation 2 to determine the level of good accuracy.

5. Model Evaluation

Comparison of two algorithms to find a good level of accuracy with predictive value or accuracy to evaluate the model requires a method to measure model performance in measuring the performance of the classification model parameters, the designer model, and represented using the confusion matrix formula.

Table 2. *Confusion Matrix*

Class	Actual: True	Actual: False
Prediction: Respondent 1	TP (True Positive)	FP (False Positive)
Prediction: Respondent n	FN (False Negative)	TN (True Negative)

There is the following explanation of the terms TP, TN, FP, and FN in this study:

a. **True Positive (TP)**

It shows that predicting pregnancy risk level is risky, and it is true that the risk level is a risk.

b. **False Positive (FP)**

It shows that predicting pregnancy risk level is not risky, but the model that has predicted it is risky.

c. **False Negative (FN)**

It shows that the pregnancy risk level is risky, but the model predicts it is not a risk.

d. **True Negative (TN)**

It shows that predicting the risk level of the pregnancy is not risky, and the model created has predicted the risk level of the pregnancy.

6. Drawing Conclusions

Data from the model evaluation results to then conclude that use as an accuracy value and model performance, including proving the theory of the research idea.

RESULTS AND DISCUSSION

In this trial, training models and processes on the RStudio platform with the main libraries, namely caret, reptime, dplyr, devtools, random forest, tree, rpart, rocr, and related to the main software in prototype development, namely the object detection framework visual code, and MySQL database.

Literature study serves as a source of study in discussing the theory surrounding previous research, as explained in table 1. At the same time, interviews with midwives and patients related to this study. The interview with the midwife is to learn the risk of pregnancy with an objective examination, namely through age, gravida, abortion, temperature, pregnancy distance, haemoglobin, blood pressure, ideal weight, and decision. Twenty questions will be given to patients through questionnaires so that the research obtains data to select and make decisions, namely risk and no risk. The objective examination results can build a classification of pregnancy risk levels using the random forest method. Below are the types of objective examination names in Table 3.



Table 3. Type Name Objective Examination

No	Name	Description	Value
1	Age	Patient Age	Number
2	Gravida	Pregnancy	Number
3	Abortus	Pregnancy Miscarriage	Number
4	Temperature	Body temperature of the patient	Celsius
5	Pregnancy Spacing	Gestational age of the expectant mother (in weeks)	Number (in weeks)
6	Haemoglobin	Haemoglobin level in pregnant women (gr/dl)	Gr/dl
7	Blood Pressure	Pregnant women's blood pressure (systole/diastole)	Systole/diastole
8	Ideal Body Weight	Normal weight gain in pregnant women (9-15 Kg during pregnancy)	Kg
9	Decision	The conclusion from the data obtained	The risk or No Risk

Validation of preparation data with an objective examination of 500 patients, from which the collection is done through secondary data, while the interviews were not with the main subject. The initial data that the type of objective examination name has grouped in table 3 forms the initial data type structure. In RStudio, the data type must first

be converted into a factor data type to perform cross-validation two so that the data processing process uses the random forest method. In table 4. the initial data type structure has been categorized as providing risk and non-risk decisions (see Table 4).

Table 4. Initial Testing Data Type Structure

No	Name	Data Type	Value
1	Age	Char	"above 35 years," "18 to 35 years."
2	Gravida	Char	"more than 4x" "below 4x"
3	Abortus	Char	"above 2x" "below 2x"
4	Temperature	Char	"Hyperthermy," "Normal"
5	Pregnancy Spacing	Char	"above two years or 24 months," "below two years or 24 months," "first child"v
6	Haemoglobin	Char	"Normal" "mild anemia"
7	Blood Pressure	Char	"Normal 100-130," "hypothermic below 100," "normal 100-130"
8	Ideal Body Weight	Char	"less than 18kg"
9	Decision	Char	The risk or No Risk

Data processing through RStudio uses a ratio of 80% as training data and 20% as testing data. Using a cross-validation two pattern that divides into train and test data with a cross-validation two ratio of 80% and 20%, which results in initial objective testing data that 52 pregnancy patients are at risk and 48 pregnancy patients are not at risk in the initial objective testing data. While in the initial objective training data, there are 192 pregnancy patients at risk and 208 pregnancy patients who are not at risk in the initial objective training data.

The data integration stage will present data analysis and interpretation after collecting data from existing data. The results of data integration analysis can produce a comparison of risk and non-risk with different variable data compositions. We can see this in Table 5.

Then, in Table 5, there is a difference in the composition of the initial dataset. Also, this data can experience a change in the variables and composition of each data before modelling and k-fold. Then the initial dataset must be changed to get the same composition and variables between risk and non-risk for modelling and k-fold.



Table 5. The initial dataset that has not been cross-validated has two risks and no risks

No	Data Name	Risk	No Risk
1	Validation Objective Initial Data	0,48	0,52
2	Objective Initial Training Data	0,52	0,48
3	Objective Initial Testing Data	0,48	0,51

Table 6. Initial Dataset with Cross-Validation 2 Risk and No Risk

No	Data Name	Risk	No Risk
1	Validation Objective Initial Data	0,50	0,50
2	Objective Initial Training Data	0,50	0,50
3	Objective Initial Testing Data	0,50	0,50

Can be seen in Table 6 shows a change in the initial objective data into balanced composition data on risk and non-risk variables. The number of data sets each data as can be seen in table 7. Data Value Results.

Table 7. Data Value Results

No	Nama Data	Jumlah
1	Validation Objective Initial Data	488
2	Objective Initial Training Data	384
3	Objective Initial Testing Data	96

Modelling employs information that has been altered by changing the number of columns and weighted to account for risk and non-risk factors. This data set uses the basic algorithm as a Random Forest model by considering the Decision Tree algorithm. These two algorithms will search and provide a good level of accuracy, namely accuracy, precision, recall, and f1-score, to get the results to build a pregnancy risk level classification system that provides accurate information on pregnant women patients experiencing a risk level or not at risk early. There will also be a ratio comparison of 80% as training data and 20% as testing data with a cross-validation pattern of 2. Can be seen in table 8. Comparison of the accuracy of random forest and decision tree algorithms.

Table 8. Comparison of Random Forest and Decision Tree Accuracy

No	Algorithm Comparison	Accuracy	Precision	recall	F1-score
1	Random Forest Training Data	85%	84%	2,95	1,31
2	Decision Tree Training Data	77%	79%	1,67	1,08
3	Random Forest Testing Data	98%	94%	1	97
4	Data Testing Decision Tree	74%	76%	0,68	0,72

Can be seen in table 8. Comparison of the accuracy of random forest and decision tree methods has different values. There are several explanations in table 8. Comparison of the accuracy of random forest and decision tree methods below:

1. Accuracy shows that the data from random forest modelling results, namely train and testing data. Compared with the data from the decision tree modelling results, namely train and testing data, it produces a conclusion on the accuracy, namely the random forest method has a good level of model accuracy and is superior in classifying correctly, which gets a data value of

85% train and 97% testing. Compared to the decision tree.

2. Precision shows that the random forest method has a good accuracy between the data requested and the prediction results given by the model, getting a value of 84% train data and 94% testing data. Compared to the decision tree, the value obtained by the training data is 79%, and the testing data is 76%.

3. Recall shows that random forest modelling gets the model's success in finding good information with a data train value of 2.95 and testing data 1.

At the same time, compared to decision tree modelling getting a data train value of 1.67.

4. F-1 Score shows that random forest modelling can provide a good algorithm performance value, namely data train 1.31 and data testing 97. While compared to decision tree modelling, the algorithm performance value gets a data train value of 1.08 and data testing 0.72.

Testing tools in the form of visual code programs, like those used for constructing web-based models, and a database used for the random forest method, are among the many model implementations available. So that it answers the purpose of this research is to detect early on the risk of pregnancy, both risk and not risk to pregnant women patients. This implementation will provide information on the accuracy level of pregnant women in the form of risk or no risk.

In the concept of model implementation, researchers use the SDLC (System Development Life Cycle) waterfall. The waterfall model provides a sequential software flow or cycle approach. We can see this in Figure 3. Waterfall illustration below:

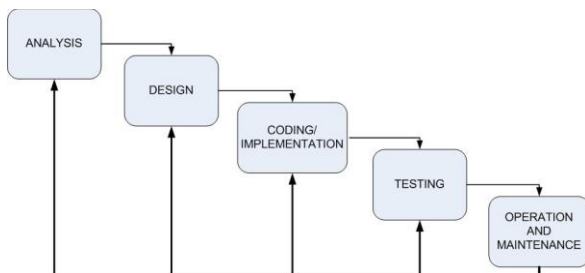


Figure 3. Modified Waterfall Model

The second analysis stage occurs during the translation of software requirements into a design representation for later incorporation into the program.

So this research is built using a use case diagram to classify pregnancy risk levels using the random forest method (Figure 4).

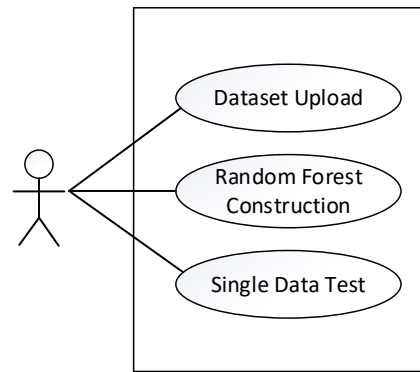


Figure 4. Use Case Diagram

Some actors or users have several functions, as shown in Figure 4. Use-case-diagram there is a process of uploading data sets, building random forests, and testing data, namely single testing data with a cross-validation rule of 2 and displaying accuracy values. The initial view on implementing the random forest model is still empty to run the single data system provided. Also has a menu function in the form of data sets to split data and modelling used in random forest models. We can see the initial appearance of the application in Figure 5.

Tingkat Resiko Ibu Hamil

Menggunakan Algoritma Random Forest

[Dataset](#) | [Split Data](#) | [Modelling](#) | [Prediksi](#) | [History Prediksi](#)

Dataset

Upload Dataset

No file chosen

Show 10 entries

Search:

Usia : Gravida : Abortus : Suhu Celcius : Jarak Kehamilan Bulan : Hemoglobin : Tekanan Darah : Berat Badan Ideal : Keputusan :

No data available in table

Showing 0 to 0 of 0 entries

[Previous](#) [Next](#)

Figure 5. Initial View of the Menu

Can be seen in Figure 5. It is a display of the results of uploading datasets into the system. The data set shown is raw or has no training and test data partitions yet. So, to do random forest modelling, first split the data or divide the training and testing data with a ratio of 80% training data and 20% testing data with cross-validation 2 in the

application of pregnancy risk level classification using the random forest method. The above dataset divides training and testing data using the random forest model. Use training data to train the algorithm while using test data to determine the performance of a previously trained algorithm when it encounters new, never-before-seen data.

Tingkat Resiko Ibu Hamil

Menggunakan Algoritma Random Forest

Dataset | Split Data | Modelling | Prediksi | History Prediksi

Dataset

Upload Dataset

Choose File No file chosen

Upload

Show 10 entries

Search:

Usia	Gravida	Abortus	Suhu Celcius	Jarak Kehamilan Bulan	Hemoglobin	Tekanan Darah	Berat Badan Ideal	Keputusan
18 sampai 35 tahun	Dibawah 4x	Dibawah 2x	Normal	anak pertama	Normal	Normal_100-130	Kurang dari 18 Kg	Berisiko
18 sampai 35 tahun	Dibawah 4x	Dibawah 2x	Normal	Di atas 2 tahun atau 24 bulan	Normal	Hipotermi_dibawah 100	Kurang dari 18 Kg	Tidak Berisiko
18 sampai 35 tahun	Dibawah 4x	Dibawah 2x	Normal	Di atas 2 tahun atau 24 bulan	Normal	Hipotermi_dibawah 100	Kurang dari 18 Kg	Tidak Berisiko
18 sampai 35 tahun	Dibawah 4x	Diatas 2x	Normal	Di atas 2 tahun atau 24 bulan	Anemia Sedang	Normal_100-130	Kurang dari 18 Kg	Tidak Berisiko
18 sampai 35 tahun	Dibawah 4x	Dibawah 2x	Normal	anak pertama	Anemia Ringan	Hipotermi_dibawah 100	18 Kg sampai 25 Kg	Tidak Berisiko
18 sampai 35 tahun	Dibawah 4x	Dibawah 2x	Hipotermi	Di bawah 2 tahun atau 24 bulan	Anemia Sedang	Normal_100-130	Kurang dari 18 Kg	Berisiko
18 sampai 35 tahun	Lebih dari 4x	Dibawah 2x	Normal	Di atas 2 tahun atau 24 bulan	Normal	Diatas 140_Hipotermi	18 Kg sampai 25 Kg	Berisiko
18 sampai 35 tahun	Lebih dari 4x	Dibawah 2x	Hipertermi	Di atas 2 tahun atau 24 bulan	Normal	Normal_100-130	18 Kg sampai 25 Kg	Tidak Berisiko
18 sampai 35 tahun	Dibawah 4x	Dibawah 2x	Normal	Di bawah 2 tahun atau 24 bulan	Normal	Diatas 140_Hipertermi	Kurang dari 18 Kg	Berisiko
18 sampai 35 tahun	Dibawah 4x	Diatas 2x	Normal	anak pertama	Normal	Normal_100-130	18 Kg sampai 25 Kg	Tidak Berisiko

Showing 1 to 10 of 500 entries

Previous 1 2 3 4 5 ... 50 Next

Reset Data

Figure 6. Dataset view

Tingkat Resiko Ibu Hamil

Menggunakan Algoritma Random Forest

Dataset | Split Data | Modelling | Prediksi | History Prediksi

Training Data

K-Fold

- 2 Kali
- 3 Kali
- 4 Kali
- 5 Kali
- 6 Kali
- 7 Kali
- 8 Kali
- 9 Kali
- 10 Kali

Table 9. K-Fold values 2 to 10 below.

No.	K-Fold Sequence	Testing Data Accuracy	Training Data Accuracy
1	K-Fold 2	96%	85%
2	K-Fold 3	96%	79%
3	K-Fold 4	95%	78%
4	K-Fold 5	95%	77%
5	K-Fold 6	94%	77%
6	K-Fold 7	94%	75%
7	K-Fold 8	93%	75%
8	K-Fold 9	93%	72%
9	K-Fold 10	93%	71%

Figure 7. K-Fold value input display



Figure 7 displays the pregnancy risk level classification system using the random forest method. There are several k-folds of 2 to 10, which have different accuracy values for testing and training. The k-fold 2 to 10 is k-fold 2, which has a high accuracy value in both testing and training data in Figure 8.



Figure 8. K-Fold 2 Result Display

Aplikasi Prediksi Tingkat Resiko Ibu Hamil

Silahkan Masukan Data Kebutuhan Berikut :

Nomor RM Pasien

Nama Pasien

Usia

Gravida

Abortus

Suhu_Celcius

Jarak_Kehamilan_bulan

Hemoglobin

TekananDarah

Berat_Badan_Ideal

Figure 9. Display of Pregnancy Risk Level Prediction Application



Figure 10. Single Data Result View

In the next stage, researchers will predict the classification of pregnancy risk levels using the random forest method from the testing data that has been k-fold 2. Display of pregnancy risk level prediction application.

In testing the prediction of pregnancy risk level classification applications using the random forest method using single data on testing data with risk status. By having several factors, namely 18 to 35 years, under 4x, above 2x, hyperthermia, under two years or 24 months, mild anemia, normal 100-130, and less than 18 kg. To get the results at 87% and not at risk at 13%. Can be seen in Figure 10. Display of single data prediction results.

CONCLUSIONS AND SUGGESTIONS

The results of the application of pregnancy risk level classification using the random forest method for early pregnancy prediction. In the random forest, the method can get a good accuracy value. In the future, this research should use subjective and objective examination data. Also, the software can be built based on android and perform data analysis in hybrids.

REFERENCES

- Adrian, M. R., Putra, M. P., & Rakhmawati, N. A. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM pada Analisis Sentimen PSBB. *Informatika UPGRIS*, 7(1), 36–40. <https://doi.org/10.26877/jiu.v7i1.7099>
- Akbar, F. M., Ali, M., & Zahro, H. Z. (2021). K-Means Clustering Untuk Pengelompokan Tingkat Resiko Ibu Hamil Di Praktik Mandiri Bidan Upt Puskesmas Pandanwangi Malang. *Jurnal*

- Mahasiswa Teknik Informatika*, 5(2), 581–588. <https://doi.org/10.36040/jati.v5i2.3772>
- Aprilia, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *Jurnal Sistem Informasi*, 10(1), 163–171. <https://doi.org/10.32520/stmsi.v10i1.1129>
- Aziz, F. (2021). Klasifikasi Aktivitas Manusia menggunakan metode Ensemble Stacking berbasis Smartphone. *Journal of System and Computer Engineering*, 1(2), 106–111. <https://doi.org/10.47650/jsce.v1i2.171>
- Badan Pusat Statistik Provinsi Bengkulu. (2021). *Profil Kesehatan Ibu dan Anak Provinsi Bengkulu 2020*. Bengkulu: Badan Pusat Statistik Provinsi Bengkulu. Retrieved from <https://bengkulu.bps.go.id/publication/2021/07/02/e23d0b377ed0964bb9fc06c4/profil-kesehatan-ibu-dan-anak-provinsi-bengkulu-2020.html>.
- Bengkulu, D. K. P. (2021). *Profil Kesehatan Provinsi Bengkulu Tahun 2021* (provbengkulu dinkes, Ed.). dinas kesehatan provinsi bengkulu. dinkes.bengkuluprov.go.id
- Byna, A. (2019). Penerapan Optimasi PSO untuk meningkatkan Akurasi Algoritma ID3 pada Prediksi Penyakit Ibu Hamil. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat*, 4(2), 65–70. <https://doi.org/10.20527/jtiulm.v4i2.40>
- Hikmatulloh, Rahmawati, A., Wintana, D., & Ambarsari, D. A. (2019). Penerapan Algoritma Iterative Dichotomiser Three (ID3) dalam mendiagnosa Kesehatan Kehamilan. *Kumpulan Jurnal Ilmu Komputer*, 6(2), 116–127. <https://doi.org/10.20527/klik.v6i2.189>
- Putra, J. W. G. (2020). *Pengenalan Konsep Pembelajaran Mesin dan Deep Learning* (1.4). <https://wiragotama.github.io/resources/ebook/intro-to-ml-secured.pdf>
- Kementerian Kesehatan RI. (2021). Profil Kesehatan Indonesia Tahun 2020. In *Kementerian Kesehatan RI*. <https://doi.org/10.1524/itit.2006.48.1.6>
- Kustiyahningsih, Y., Mula'ab, & Hasanah, N. (2020). Metode Fuzzy ID3 Untuk Klasifikasi Status Preeklamsi Ibu Hamil. *Teknika*, 9(1), 74–80. <https://doi.org/10.34148/teknika.v9i1.270>
- Mulaab, M. (2020). *Data Mining Konsep dan Aplikasi*. Malang: MNC Publishing.
- Religia, Y., Nugroho, A., & Hadikristanto, W. (2021). Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI*

- (*Rekayasa Sistem Dan Teknologi Informasi*), 5(1), 187–192. <https://doi.org/10.29207/resti.v5i1.2813>
- RI, K. (2021). *Profil Kesehatan Indonesia 2021*. <https://www.kemkes.go.id/downloads/resources/download/pusdatin/profil-kesehatan-indonesia/Profil-Kesehatan-2021.pdf>
- Suryanegara, G. A. B., Adiwijaya, & Purbolaksono, M. D. (2021). Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi. *Jurnal Rekayasa Sistem Dan Teknologi Informasi*, 5(1), 114–122. <https://doi.org/10.29207/resti.v5i1.2880>
- Wulandari, T., & Susanto, A. (2018). Deteksi Tingkat Risiko Kehamilan dengan Metode Fuzzy Mamdani dan Simple Additive Weighting. *Jurnal Teknologi Dan Sistem Komputer*, 6(3), 110–114. <https://doi.org/10.14710/jtsiskom.6.3.2018.110-114>
- Pavlov, Y. L. (2018). *Random Forests*. Berlin: De Gruyter. Retrieved from <https://www.degruyter.com/document/doi/10.1515/9783110941975/html?lang=en>