JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

Machine Learning Based Prediction of Health Risks in Pregnant Women

Rahma Devi[1*], Inggih Permana[2], Rice Novita[3], Febi Nur Salisah[4]

[1,2,3,4]Faculty of Science and Technology, Information System, State Islamic University of Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1*]12250323758@students.uin-suska.ac.id, [2]inggihpermana@uin-suska.ac.id, [3]rice.novita@uin-suska.ac.id, [4]febinursalisah@uin-suska.ac.id
*Email Corresponding Author:12250323758@students.uin-suska.ac.id)
Received: January 5, 2026| Revision: January 12, 2025| Accepted: March 4, 2026

Abstract

Pregnancy is an important phase that requires optimal health monitoring to prevent complications that are risky for both mother and fetus. The high maternal mortality rate in Indonesia emphasizes the importance of early detection of pregnancy risks. The use of machine learning offers an effective predictive approach to quickly and accurately identify pregnancy risks. This study aims to compare the performance of five machine learning algorithms, namely Logistic Regression, Decision Tree C4.5, Random Forest, Support Vector Machine, and Naive Bayes, using the Maternal Health Risk Dataset. The hold-out validation method with data sharing of 80% training data and 20% test data was used in this study. Model evaluation is conducted based on accuracy, precision, recall, and F1-score metrics. The results showed that Random Forest had the best performance with an accuracy of 93%, followed by Decision Tree at 93%, SVM at 82%, Logistic Regression at 76%, and Naive Bayes at 72%. Thus, Random Forest is rated as the most optimal algorithm in predicting pregnancy risk and potentially supporting the development of decision support systems for health workers. This research is expected to be the basis for the development of a machine learning-based decision support system to increase the effectiveness of health services for pregnant women.

Keywords : Machine Learning, Pregnancy Risk, Random Forest, Classification, Maternal Health

1. INTRODUCTION

Pregnancy is an important phase in a woman's life that requires intensive health monitoring to prevent various complications that can harm both mother and fetus. Problems such as gestational hypertension, preeclampsia, and other pregnancy disorders are still the main causes of high maternal mortality rates, especially in developing countries. According to a 2023 WHO report, pregnancy and childbirth complications cause around 260,000 female deaths each year. In Indonesia, the Maternal Mortality Rate (MMR) is still at the 189 per 100,000 live births based on the 2023 Population Census Long Form. This figure shows a large gap to the target Sustainable Development Goals (SDGs), which reduces maternal mortality to 70 per 100,000 live births in 2030. [1]

Pregnancy complications that are not detected early can have serious impacts, such as premature birth, fetal developmental disorders, and perinatal death. This condition emphasizes the need for early detection to ensure that pregnant women get fast and targeted treatment. As technology develops, the *machine learning* is increasingly used in the medical field due to its ability to process health data more accurately and support clinical decision-making processes[2]. The machine learning approach allows modeling of various health parameters of pregnant women such as age, blood pressure, body temperature, heart rate, hemoglobin, and blood sugar levels to identify patterns and classify pregnancy risk levels. One of the relevant public datasets is *Maternal Health Risk Dataset* from Kaggle, which contains 1,822 data on pregnant women with low, medium, and high risk categories.[3].Various algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Naive Bayes have been used in previous studies and have shown different performance achievements .[4],Some studies report that the Decision Tree algorithm[5] and SVM [6] was able to achieve an accuracy of 90% and 90.9%, respectively, while Random Forest showed stable performance with an accuracy of between 75.2% by [7] up to 82% of the study[8], Naive Bayes obtained an accuracy of 85.62%.[9].

Nonetheless, most studies tend to focus on accuracy as a key indicator, without testing other metrics such as precision, recall, and F1-score. In fact, these metrics are very important in the context of medical data that often experiences class imbalances. Based on the review, there are several research gaps that still need to be bridged, namely model performance evaluations generally only focus on accuracy, without a comprehensive analysis of precision, recall, and F1-score, there have not been many studies that simultaneously compare the five main algorithms (Logistic Regression, Decision Tree C4.5, Random Forest, SVM, and Naive Bayes). Some studies still use small datasets so that model generalizations are limited. Seeing these gaps, this study was conducted to analyze and compare the performance of five machine learning algorithms in predicting pregnancy risk. The evaluation was conducted using four metrics of accuracy, precision, recall, and F1-score to determine the most optimal algorithm. This research is expected to contribute to the development of machine learning-based pregnancy risk prediction models and become the basis for further research in the field of maternal health.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

2. RESEARCH METHODOLOGY

The research process is the Planning Stage, Data Collection, Machine Learning Modeling, and Model Evaluation and Selection.

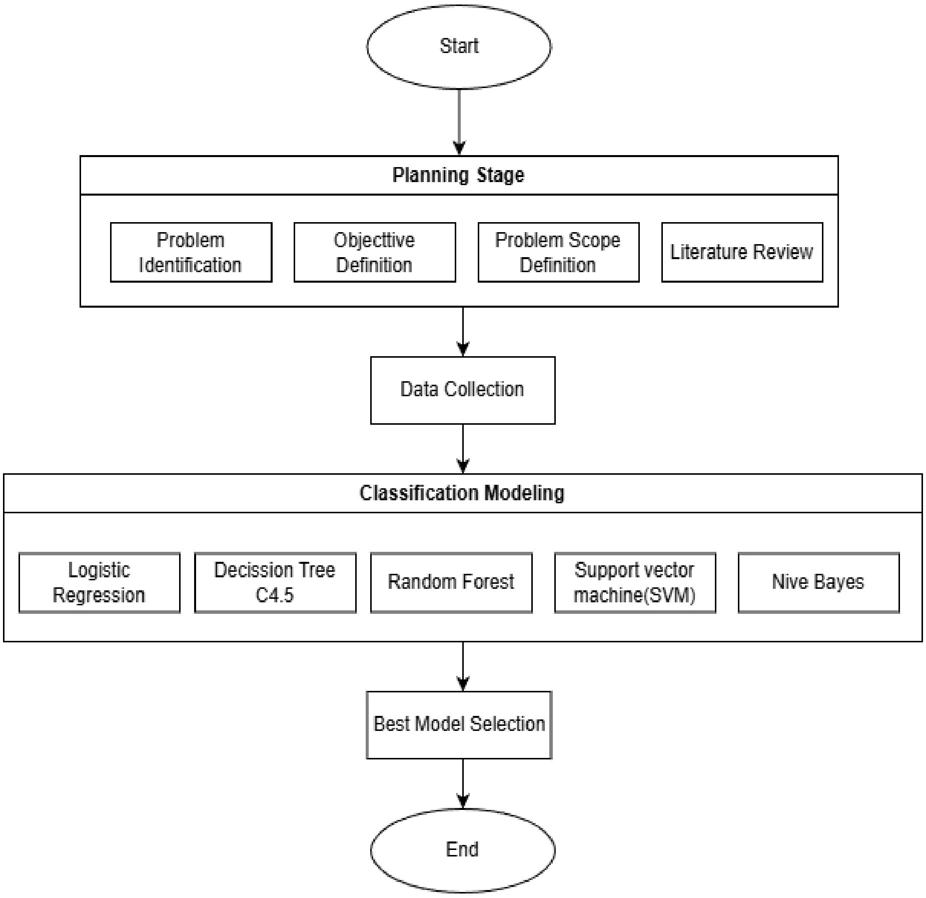

Figure 1. Research methodology

2.1 Planning Stage

The planning stage is the first step in the research process that serves as a foundation to ensure the direction, focus, and success of the research. This stage involves the determination of problems, objectives, variables, and methodological approaches so that the research can be carried out systematically and produce valid findings. In this study, the planning stage is focused on designing a comparative analysis of machine learning algorithms in predicting pregnancy risk, which aims to produce analytical and evaluative findings based on medical data. To achieve maximum results, the planning stage is divided into several steps, namely: problem identification, goal formulation, research boundary setting, literature study, and methodology design.

2.1.1 Identify the problem

The first step in planning is problem identification, which aims to understand the main issues to be researched. Based on a report by the World Health Organization (WHO, 2023), more than 295,000 women die each year from pregnancy and childbirth complications, with the majority of cases occurring in developing countries. In Indonesia, data from the Ministry of Health of the Republic of Indonesia (2023) shows that the maternal mortality rate (AKI) is still at 189 per 100,000 live births, far above the 2030 Sustainable Development Goals (SDGs) target, which is ≤70 per 100,000 live births.[10]

One of the causes of high AKI is the delay in detecting the risk of pregnancy early due to limited human resources in the health sector and lack of data-based analysis. Therefore, a machine learning-based analytical approach that can compare algorithms is needed to identify the best model in predicting the level of pregnancy risk. The problem raised in this study is the high maternal mortality rate (AKI) in Indonesia, which is still a major challenge in the health sector. One of the causative factors is the delay in recognizing the risk of pregnancy early. [11]Therefore, machine learning is needed that can help detect pregnancy risks faster by utilizing available medical data. [2]

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

2.1.2 Formulation of Research Objectives

The main purpose of this study is to analyze and compare the performance of machine learning algorithms in predicting the level of pregnancy risk based on pregnant women's health parameters such as age, systolic blood pressure, diastolic blood pressure, blood sugar levels, body temperature, hemoglobin levels, and heart rate. In addition, this study aims to identify the algorithm that provides the most optimal accuracy in classifying pregnant women into three risk categories, namely: Low Risk, Mid Risk and High Risk.

2.1.3 Determination of Problem Boundaries

This study uses secondary data in the form of *a Maternal Health Risk Dataset* available on Kaggle, consisting of 1,822 data with 7 attributes. The attributes used included six main medical parameters, namely age, systolic and diastolic blood pressure, blood sugar levels, body temperature, hemoglobin, and heart rate. This study applied and compared five *machine learning algorithms*, namely Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Naive Bayes. Evaluation of model performance was carried out using accuracy, precision, recall, and F1-score metrics.

2.14 Literature Studies

Literature studies are carried out to strengthen the theoretical foundation of research by examining relevant theories and results of previous research, as stated by Sugiyono (2019). The literature review in this study is sourced from scientific journals and official reports from related institutions, such as WHO, the Ministry of Health, and UNICEF. [12] A number of previous studies have shown that *machine learning* algorithms are effective in predicting pregnancy risk, including Decision Tree with 90% accuracy (Rahman, 2023), Random Forest with 82% accuracy and dominant variables of age and blood sugar levels (Al Mashrafi et al., 2024), and Naive Bayes with 85.62% accuracy and advantages in computing efficiency (Khoirunnisa & Lestari, 2023). However, previous research has generally focused on accuracy as the only indicator of performance, has not conducted a comprehensive comparison of the five main algorithms, and has used a lot of small-scale datasets. Therefore, this study was conducted to fill the gap by comparing the performance of Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Naive Bayes using more complete evaluation metrics in predicting pregnancy risk.

2.2 Data Collection

The data used is data obtained from the source of the Kaggle dataset. The dataset contains information on the health risks of pregnant women with three labels of low, medium, and high pregnancy risk. The variables are age, body temperature, heart rate, hemoglobin levels, blood sugar levels, and blood pressure (systolic and diastolic) The dataset used in this study amounted to 1,822 records. In general, the data collection stage in this study includes four main steps, namely identification of data sources, description of datasets, and determination of research data attributes.

2.2.1 Identification of Data Sources

The data source used in this study consists of kaggle data, the Maternal Health Risk Dataset which is publicly available on the Kaggle platform. The Maternal Health Risk Dataset (uploaded by Joakim Arvidsson in 2023) comes from research conducted in Bangladesh. Bangladesh and Indonesia have similar maternal health characteristics, as both are developing countries in South and Southeast Asia. This dataset is used because its data attributes (age, blood pressure, sugar levels, body temperature, heart rate) are universal physiological parameters used by the WHO to assess pregnancy risk worldwide. This dataset was developed to support research in the field of maternal health and contains medical data related to the condition of pregnant women, such as blood pressure, blood sugar levels, body temperature, hemoglobin, and heart rate.(World Health Organization 2013). In addition, both Indonesia and Bangladesh have similar maternal health profiles to developing countries with high maternal mortality rates and the main cause is preeclampsia. According to research in Bangladesh,[13] Preeclampsia/Eclampsia complications account for about 20-24% of maternal deaths.

2.2.2 Dataset Description

The dataset used in this study is *the Maternal Health Risk Dataset* obtained from the Kaggle platform (2021). This dataset consists of 1,822 rows of data with seven attributes (features) and one class label. The class label represents the level of pregnancy risk classified into three categories, namely *Low Risk* , *Mid Risk* , and *High Risk* .

2.3 Modeling

The modeling stage is the main part of this research which aims to build a pregnancy risk prediction model using algorithms *machine learning*. Modeling was carried out by applying five classification algorithms, namely Logistic Regression, Decision Tree, and Random Forest [14]Support Vector Machine (SVM) [15]and Naïve Bayes, to determine

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

the algorithm with the best performance. The modeling process includes determining approaches, data sharing, algorithm application, model training and testing, and initial evaluation of results. This stage is technical because it explains the data processing process until an optimal classification model is obtained. The classification modeling flow is illustrated in Figure 3.3 This stage is more technical than the general methodology, as it describes how the dataset is processed to produce the best model.

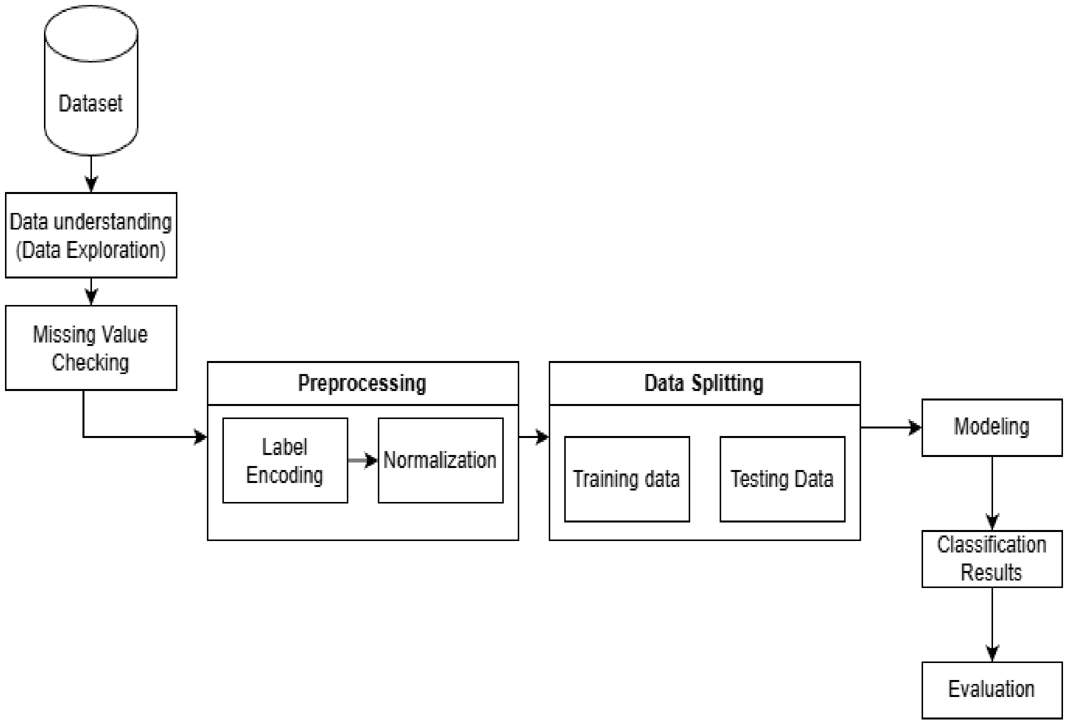

Figure 2 . Data procedure modeling flow

2.3.1 Dataset

The data used in this study is the Maternal Health Risk Dataset sourced from Kaggle. This dataset was chosen because it contains medical information on pregnant women that are relevant to the purpose that has parameters such as age, blood pressure, blood sugar levels, hemoglobin, body temperature, and heart rate. The total data of 1,822 records is considered representative enough to build and train the classification model. This dataset is then converted into CSV format so that it can be processed using the Python programming language with the help of scikit-learn, pandas, and nuppy libraries.

2.3.2 Data Understanding

The data understanding stage is carried out to obtain an initial understanding of the characteristics of the dataset. This process includes a review of data structures and types, descriptive statistics, and the identification *of RiskLevel* target variable classes.

2.3.3 Missing Value Data Check

Data checks are performed to detect the presence of missing values on each attribute in the dataset. If the results of the examination show that all data is complete, then there is no need to handle the missing value at the next stage.

2.3.4 Preprocessing Data

In the preprocessing stage, the *RiskLevel* target variable that is still in the form of categorical data is transformed into numerical form using *the Label Encoding method*. This process is carried out so that data can be processed by machine learning algorithms.

In the preprocessing stage, the *RiskLevel* target variable that is still in the form of categorical data is transformed into numerical form using *the Label Encoding method*. This process is carried out so that data can be processed by machine learning algorithms.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

a. Label Encoding
This stage is the data preprocessing stage which aims to prepare the data to be ready for use in the model training process. The step taken is to encode category variables using LabelEncoder so that all attributes can be processed by a numerical-based algorithm.

b. Normalization
Before data is used for machine learning model training, data preprocessing is an important stage in initial processing. Preprocessing is carried out with the aim of improving the quality of data and adjusting it so that the algorithm used can process it properly. Furthermore, Data Normalization, data normalization is carried out to equalize the scale of the numerical features used in the model. This is important so that no feature dominates the training process just because it has a larger range of values. The normalization method used is z-score, which converts the value of the feature into a range of 0 to 1. With normalization, machine learning models can function more effectively and produce better accuracy.

2.3.5 Data Splitting

Data Splitting, the next step is to divide the dataset into two parts, namely the training set and the test data. Train data is used to build and train models, while test data is used to test model performance. In this study, an 80:20 ratio was used, meaning that 80% of the data was used to train the model, while the remaining 20% was used to test the model. This technique is important so that the resulting model not only memorizes existing data (overfitting), but is also able to make predictions on new data. The next stage is normalization, which is the process of preprocessing data to equalize the scale between attributes. This is done so that the difference in the range of values between variables, such as age (years) and blood pressure (mmHg), does not cause bias in model formation. Normalization is done by changing the value of each variable into a certain range, for example 0–1, so that all features have a balanced contribution.

2.3.6 Algorithm Modeling

The modeling stage is at the heart of the study, where training data is used to build and test the predictive ability of five classification algorithms in classifying pregnancy risk. Each model is implemented with predefined parameters to assess the basic performance. The general modeling process includes initializing the model, training (fit) on the training data, predicting the test data, and evaluating the results.

a. Logistic Regression
Logistic Regression is a classification method that models the relationship between independent variables and categorical dependent variables using the logistic function. This algorithm does not generate continuous values such as linear regression, but rather probabilities between 0–1 that are then converted to the target class.[16]. (Scott et al., 2013)In this process, Logistic Regression is tested using several values *C* which regulates the regularization strength, which is 0.01, 0.1, 1, and 10. In addition, two penalty methods are used: *L1* (Lasso) dan *l2* (Ridge). Solver *liblinear* chosen because it is a compatible solver for both penalties. Each combination of parameters is tested by building a Logistic Regression model using the parameters being tested, then training the model on data training. The results of the prediction in the data testing were evaluated using accuracy and classification report. All results are saved to select the best parameters. Once the best parameters are obtained, the model is retrained using those parameters and evaluated using a confusion matrix.

b. Decision Tree C4.5
The Decision Tree C4.5 algorithm is a development of ID3 that uses *entropy* as the basis for selecting the best attributes by calculating *the information gain ratio*. In this study, the approach was applied using DecisionTreeClassifier with criterion = 'entropy', so that the node separation mechanism follows the C4.5 principle. The process starts by importing all the required libraries, including DecisionTreeClassifier, accuracy_score, classification_report, as well as visualization libraries such as matplotlib and seaborn.
Next, two main parameters are tested manually, namely max_depth (tree depth) and min_samples_split (the minimum number of samples to divide a node). These parameters were tested in several combinations, namely max_depth = [3, 5, 7, None] and min_samples_split = [2, 5, 10]. Each combination is tested through looping so that all parameters can be compared. For each combination of parameters, the Decision Tree model is constructed using criterion='entropy' and random_state=4*2 values* to ensure reproducibility. The model was then trained using training data (X_train, y_train), and the results were used to make predictions on test data (X_test). Once the prediction is made, the accuracy value is calculated and the classification report is printed so that the performance of each parameter combination can be seen. All results are stored in a list so that they can be selected later. At the end of the process, the model with the highest accuracy is selected as the best model, and the confusion matrix of the best model is visualized to see the classification pattern generated by Decision Tree C4.5.

c. Random Forest
Random Forest is a development of Decision Tree that works by building multiple random decision trees and combining the results to obtain a final prediction. Each tree is trained using a randomly selected subset of data and a subset of features (bootstrap sampling). [14]The final prediction is determined by a majority vote of all trees. Its robust nature and ability to handle complex medical data make it a strong choice. In this study,

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

the testing process begins by determining several combinations of parameters to be tested, namely the number of trees (n_estimators), the maximum depth of the tree (max_depth), the minimum number of samples for node separation (min_samples_split), and the minimum sample on leaves (min_samples_leaf). Each combination of parameters is tested through an iterative (looping) process, in which the RandomForestClassifier model is built, trained using the training data, and generates predictions on the test data to calculate its accuracy. Classification reports are printed to find out the performance of each risk class. All results are stored in a list to determine the best parameters based on the highest accuracy. Once the optimal parameters are obtained, the best Random Forest model is rebuilt and used for the final prediction. This prediction is calculated as a confusion matrix and visualized using a heatmap to determine the model error pattern in predicting high risk, low risk, and mid risk classes.

d. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that aims to find the best hyperplane to separate classes by maximum margin. SVM is effectively used for both linear and non-linear classification through the selection of appropriate kernel functions.[15]. In this study, SVM was applied to classify the level of pregnancy risk into three categories: low, moderate, and high. The model is initialized using the SVC() class with a configuration of default parameters, including kernel = rbf (radial base function), C = 1.0, and gamma = scale. Next, the model is trained on the trained data (X_train, y_train) to find the optimal separator hyperplane. Predictive results (y_pred_svm) on test data are used to evaluate model performance through accuracy metrics and classification reports. The selection of the RBF kernel helps to address the non-linearity of pregnant women's health data, so that the SVM is expected to be able to provide accurate predictions of the risk of pregnancy complications.

e. Naive Bayes

The Naive Bayes algorithm is a probabilistic-based classification method that uses Bayes' Theorem assuming that each feature is independent of each other. (Rahman, 2023)In Gaussian Naive Bayes, the data is assumed to follow a normal (Gaussian) distribution. One of the important parameters in GaussianNB is var_smoothing, which is a small value that is added to the variance to keep the model stable when finding data with very small variance. In this implementation, some values var_smoothing tested manually to find the best combination. The process begins with preparing a list of var_smoothing values (1e-9, 1e-8, 1e-7, 1e-6), then the model is trained and evaluated one by one using training data and test data. Each result obtained is saved, and then the var_smoothing value with the highest accuracy is selected as the best parameter. The model is then reconstructed using these parameters and re-evaluated using a confusion matrix to see the overall performance between classes. Here's the flowchart flow of the Gaussian Naive Bayes testing process. (Khoirunnisa & Lestari, 2023)

2.3.7 Model Classification Results

The results of the classification, after modeling was carried out with each algorithm, were obtained in the form of predictions of pregnancy risk levels for test data. Each model produces predictions of three pregnancy risk categories, namely low, medium, and high, according to the label on Kaggle's Maternal Health Risk dataset. The prediction results are compared to the actual labels to calculate the accuracy, precision, recall, and f1-score values, which are used as the basis for evaluating the model's performance.

2.3.8 Model Evaluation

The final stage is the evaluation of the model's performance. The evaluation was carried out using a Confusion Matrix which shows the number of correct and false predictions in each class (Low, Mid, High Risk). In addition, other evaluation metrics are also used, namely:

a. Accuracy

Measure the correct proportion of predictions from the overall test data. High accuracy indicates that the model is able to classify the data in general correctly.

b. Precision

Calculate the number of relevant positive predictions. To avoid mistakes in identifying pregnant women who are not at risk as at risk, it is very important to have high precision.

c. Recall (Sensitivity)

Calculate the number of cases of pregnancy risk that the model successfully detects. The high number of repetitions indicates that the model is sensitive to cases of pregnancy risk.

d. F1-Score

F1-Score is used when there is a class imbalance, such as a smaller amount of data on the risk of pregnant women than not at risk. F1-Score is a harmonious average of precision and recall.

The mathematical calculations for the confusion matrix are as follows. [17]

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3$$

$$F1 - Score = (\frac{precision\ x\ Recall}{precision + Recall} \quad (4)$$

Description:
TP (True Positive): A correct positive prediction.
TN (True Negative): A true negative prediction.
FP (False Positive): A false positive prediction.
FN (False Negative): False negative prediction. [17]

2.4 Selection of the Best Models

The final stage in the machine teaching model testing process is the selection of the best model. At this stage, the results of the previous evaluation are used to select the best-performing model by comparing the values of metrics such as accuracy, precision, recall, and F1 score. To classify the pregnancy risk status in pregnant women in the majority and minority classes, the chosen model must provide the most accurate and consistent prediction results. These decisions can also be influenced by other factors, such as model complexity, computational time, and interpretability. By choosing the best model, it is hoped that the model can be used as a powerful analytical tool to detect pregnancy risk in pregnant women and can be used in follow-up studies or decision-making processes by related parties.

3. RESULTS AND DISCUSSION

3.1 Dataset

The first step in data exploration is to analyze the structure and data types of the datasets used. Once the data was loaded, it was identified that the dataset had a total of 1,882 samples (entries) and 7 columns. RiskLevel, of type object (text), which indicates that this field is a categorical target variable and requires an encoding process before it can be used for model training. Then based on the results of the examination of the missing values, it is known that all attributes in the dataset have complete data without finding empty values in each variable. Thus, there is no need for missing value handling, and the data can be directly proceeded to the next preprocessing stage.

3.2 Preprocessing Data

3.2.1. Encoding Variabel Trget (RiskLevel)

Analysis of the RiskLevel target variable is crucial, as it is the model's predictive output. An examination of the unique value in this column confirms that there are three risk categories to be classified: high risk, low risk, and mid risk. Because the machine learning model can only process numerical values, these target variables are then encoded into numerical codes using the LabelEncoder. After encoding, the number of samples and their proportions in each class are calculated, which is summarized in Table 3.1

Table 1. Distribution of Target Variables (RiskLevel)

Kode Encoding	Category Risk	Number of Samples	Percentage
0	high Risk	602	48.51%
1	low Risk	884	33.04%
2	Mid Risk	336	18.44%
Total		**1.822**	**100%**

3.2.2. Separation of Feature Data and Target Variables

Before normalization, the data is separated into features (X) and targets (Y). The RiskLevel target column is not included in the feature data.

Table 2 Feature data (x) before normalization

Baris	Age	Systolic BP	DiastolicBP	BS	BodyTemp	HeartRate
1	25	130	80	15.0	98	86

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

2	35	140	90	13.0	98	70
3	29	90	70	8.0	100	80
4	30	140	85	7,0	98	70
5	35	120	60	6.1	98	76
.....	.....	.....	.....	.....	.....	.....
1820	35	85	60	19.0	98.0	86
1821	43	120	90	18.0	98.0	70
1822	32	120	65	6.0	101.0	76

3.2.3. Z-Score Normalization (StandardScaler)

All numerical features on X are normalized using Z-Score Normalization via the StandardScaler of the scikit-learn library. This normalization aims to standardize the range of feature values by changing the distribution of the data so that it has an average (μ) of ≈0 and a standard deviation (σ) of ≈1. This step is important so that the model is not biased towards features with a large range of values. Normalization is carried out with the formula:

$$z = \frac{x-\mu}{\sigma} \quad (5)$$

Table 4.7 shows the normalization results for the first five rows of feature data. Note that the feature values are now around zero, representing how far the original values are from the mean in standard deviation units.

Table 3. Feature Data (X) After Z-Score Normalization

Baris	Age	Systolic BP	Diastolic BP	BS	BodyTemp	HeartRate
1	-0.379486	0.885704	0.215473	1.749531	-0.474890	1.390012
2	0.351947	1.409734	0.915404	1.169741	-0.474890	-0.510806
3	-0.086913	-1.210418	-0.484458	-0.279734	0.977115	0.677205
4	-0.013770	1.409734	0.565438	-0.569629	-0.474890	-0.510806
5	0.351947	0.361673	-1.184390	-0.830535	-0.474890	0.202001
.....	....	....	....	....	....	....
1820	351947	-1. 472434	-1.184390	2. 909111	-0.474890	1.390012
1821	0.937093	0.361673	0.915404	2.619216	-0. 474890	-0.510806
1822	0.132517	0.361673	-0.834424	-0.859525	1.703117	0.202001

3.2.4 Statistical Verification of Normalization Results

To ensure that the normalization process runs correctly, descriptive statistics are calculated from the normalized DataFrame (X_normalized_zscore).

Table 4. Statistical Verification of Normalization Results

Statistics	age	agesystolicBP	DiastolicBP	BS	Body Temp	Heart rate
mean	≈ 0	≈ 0	≈ 0	≈ 0	≈ 0	≈ **0**
hrs	1.002759	1.002759	1.002759	1.002759	1.002759	1.002759

A very small mean value (close to zero) and a standard deviation value close to 1 (1.002759 to be exact) confirm that the Z-Score Normalization process using StandardScaler has been successfully applied to the feature data. The data is now standardized and ready for use in modeling.

3.3 Data Sharing

The dataset is divided into two parts with a composition of 80% for training data (training set) and 20% for test data (test set) using the train-test split method with stratified sampling techniques. Stratification is used to keep the distribution of target classes (low, mid, high) balanced in both training data and test data. Can be seen in table 4.5

Table 5. Data sharing

Set Data	Number of Samples	Percentage

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

Training	1457	80%
Testing	365	20%

The distribution of the dataset was carried out using the train-test split method with a ratio of 80% for the training data (X_train.Y_train) and 20% for the test data (X_test.Y_test). This division resulted in 1457 samples for the training set and 365 samples for the test set. The use of stratified sampling techniques in this division ensures that the distribution of target classes (Low, Mid, and High Risk) remains proportional and balanced in both the training set and the test set, so that the model can learn from accurate and fairly evaluated class representations.

3.4 Performance Measurement Results of Each Algorithm

a. Random Forest with the best Results parameters

Tebel 6. Random Forest performance measurement results best parameters

Algoritma	Parameter name and value	Accuracy	Precision	Recall	F1- Score
Random Forest	n_estimation:50 max_depth:none min_samples_split:2 min_samples_leaf: 1	93.42%	0.93	0.93	0.93

b. Algorithm Support vector Machine with the best parameters

Table 7. Results of Best Support Vector Machine Performance Measurement Parameters

Algoritma	Prameter name and Name	Accuracy	Precision	Recall	F1-Score
Support vector machine	C=100, kernel='rbf', gamma='scale' ,degree none	81.92	0.81	0.82	0.80

c. Logistic Regression algorithm with the best parameters

Table 8. Results of Logistic Performance Measurement Regression of the best parameters

Algoritma	Parameter name and value	Accuracy	Precision	Recall	F1- Score
Logistic Regression	C:1 Penalty:l1 max_iter=1000,	75.89	0.78	0.76	0.70

d. Nive Bayes algorithm with the best parameter testing

Table 9. Results Performance measurement of the best parameters of the Nive Bayes

Algoritma	Parameter name and value	Accuracy	Precision	Reccal	F1 Score
Nive Bayes	Var_smothing, devault 1e-9 GaussianNB (default)	72.33	0.67	0.72	0.68

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

e. Decision Tree C4.5 algorithm with Best parameters

Table 10. Results Decision tree C4.5 performance measurement best parameters

Algoritma	Parameter name and value	Accuracy	Precision	Recall	F1- Score
Decision tree C4.5	Max_depth:none min_samples_split: 2	93.15%	0.93	0.93	0.93

3.5 Model Evaluation

Next, a model evaluation was carried out to classify the level of pregnancy risk using several machine learning algorithms, namely Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Naïve Bayes. The model evaluation was carried out using the Confusion Matrix and the calculation of the Accuracy, Precision, Recall, and F1-Score metrics.The results of the confusion matrix from each algorithm showed how well the model predicted *the low-risk*, *mid-risk*, and *high-risk* classes compared to the actual label. The accuracy value describes the percentage of correct predictions, while precision, recall, and F1-score are used to evaluate the model's performance in multi-class cases for more comprehensive assessment results.

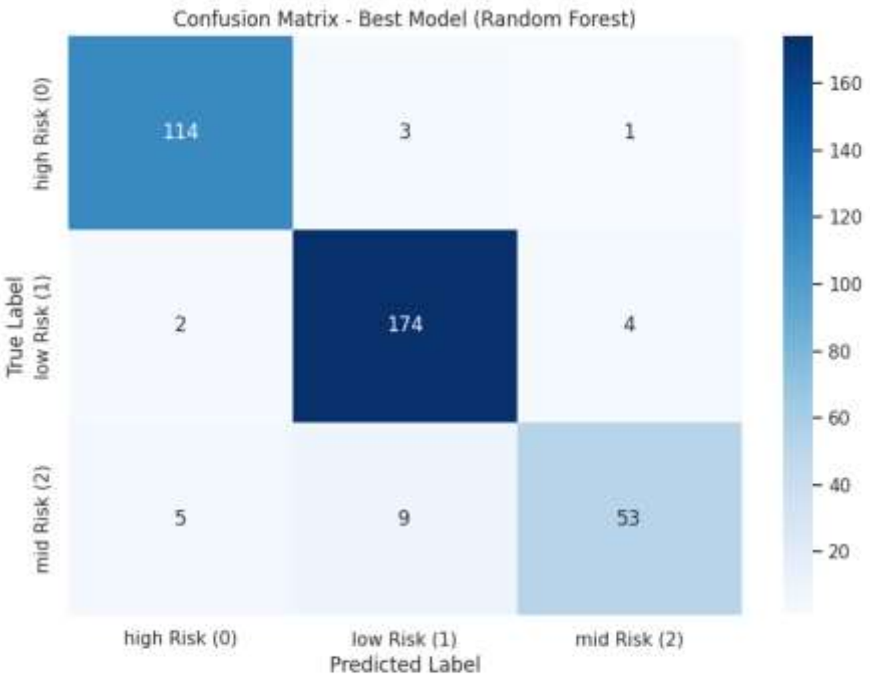

Figure 3. Confusion Matrix Rndom forest

The Random Forest model showed a very strong performance with an accuracy of 93.42%, where 342 out of 365 data were correctly predicted. In the High Risk (0) class, the model correctly predicted 114 of the 118 cases, resulting in a Recall of 0.97, Precision of 0.94, and an F1-Score of 0.95, which means that the model was able to recognize almost all High Risk cases with a very low error rate. High performance was also demonstrated in the Low Risk class (1), where 174 out of 180 data were correctly classified. The Recall values of 0.97, Precision 0.94, and F1-Score of 0.95 indicate that the model is not only accurate in detecting this class but also consistent in maintaining a balance between Precision and Recall. For the Mid Risk class (2), although the amount of data is the least, the model still shows good performance. Out of 67 data, 53 cases were correctly predicted, resulting in a Recall of 0.79, while a Precision of 0.91 indicates that most of the class 2 predictions actually came from this class data. An F1-Score value of 0.85 shows that although this class is more difficult, the model is still able to perform the classification effectively.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

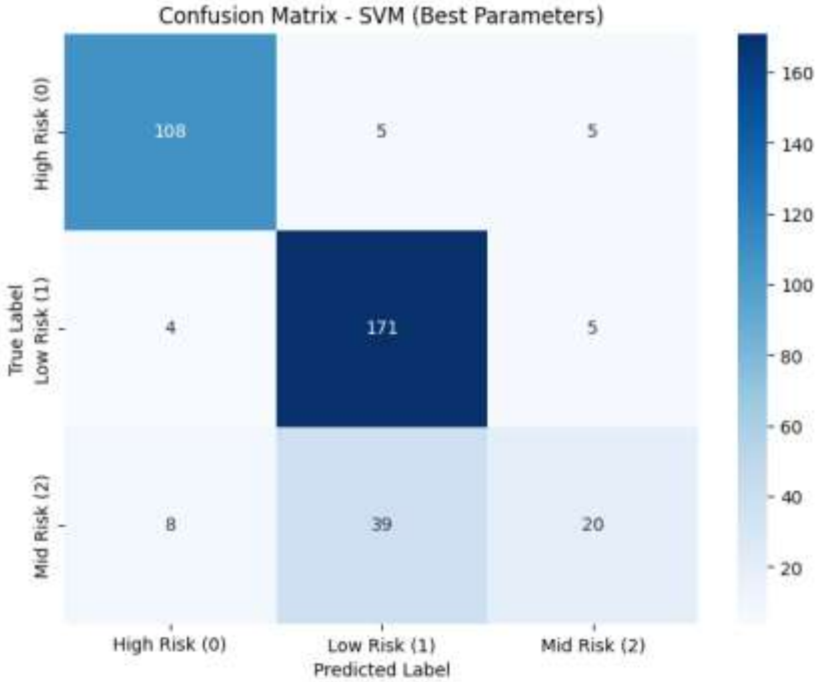

Figure 3. Confusion Matrix Super Vector Machine

The Support Vector Machine (SVM) model with the best parameters (C=100, kernel rbf, gamma scale) showed good overall performance with an accuracy of 81.92%, where out of 365 test data, 299 data were correctly predicted. For the High Risk class, the model predicted 108 of the 118 cases precisely, while 5 cases were classified as Low Risk and 5 as Mid Risk, resulting in a Recall of 0.92, Precision of 0.90, and F1-Score of 0.91, demonstrating the model's ability to recognize High Risk cases with relatively low errors. In the Low Risk class, the model correctly predicted 171 out of 180 data, while 4 cases were incorrectly categorized as High Risk and 5 as Mid Risk, resulting in a Recall of 0.95, Precision of 0.80, and an F1-Score of 0.87, indicating that the model was quite effective at detecting Low Risk, despite some incorrect predictions. Meanwhile, for the Mid Risk class, the model correctly predicted only 20 of the 67 cases, while 8 cases were incorrectly classified as High Risk and 39 as Low Risk, so that the Recall was only 0.30, Precision 0.67, and F1-Score 0.41, indicating the model's difficulty in distinguishing the Mid Risk class with fewer samples and greater overlap with other classes

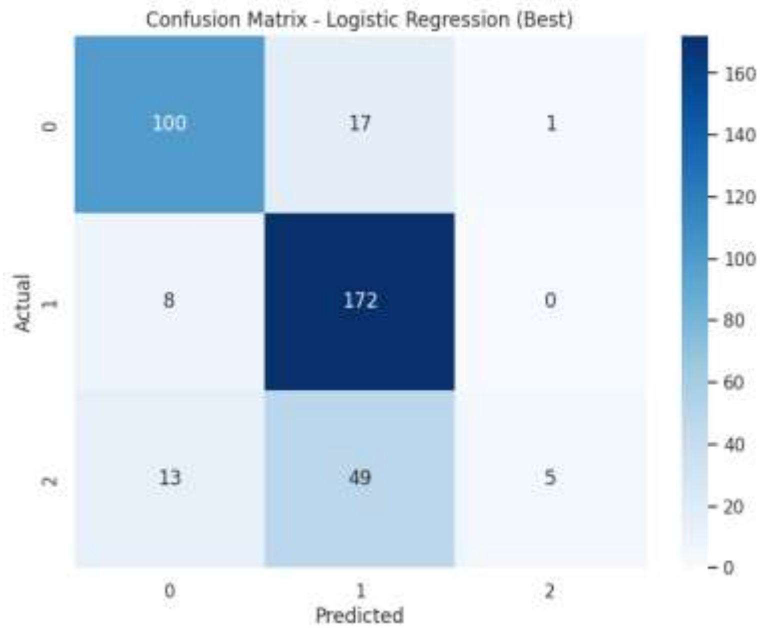

Figure 5. Confusion Matrix Logistic Regression

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

The Logistic Regression model (C = 1, penalty = L1) yielded an accuracy of 75.89%, with 277 out of 365 data correctly predicted. In the High Risk (0) class, the model recognized 100 out of 118 cases with a Recall of 0.847, Precision of 0.826, and F1-Score of 0.836, showing a fairly balanced performance. The Low Risk class (1) was the strongest, with a Recall of 0.956 because 172 of the 180 data were detected to be true, although the Precision of 0.723 still showed the presence of false positives from other classes. On the other hand, the Mid Risk class (2) is the biggest weakness. Only 5 of the 67 data were correctly predicted (Recall 0.075), although Precision reached 0.833 because some of the 2nd class predictions were correct. The low Recall drops the F1-Score to 0.137, signaling the model's difficulty learning this minority class pattern. Overall, Logistic Regression is quite effective for the majority class (High Risk and Low Risk), but less optimal for the Mid Risk class. The total accuracy remains at 75.89%.

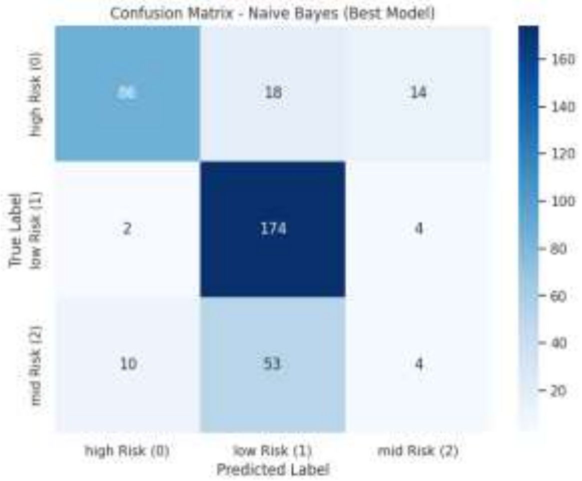

Figure 6. Confusion Matrix,Nive Bayes

The Naive Bayes Model (Best Model) obtained an accuracy of 72.33%, with 264 of the 365 data successfully classified correctly. The best performance was seen in the Low Risk class (1), where the model was able to recognize 174 out of 180 data (Recall 0.97) although the Precision was still 0.71 due to many predictions mixed with other classes. The High Risk class (0) also showed fairly good results with a Recall of 0.73, Precision of 0.88, and an F1-score of 0.80, although there were 32 cases that were misclassified to other classes. On the other hand, the Mid Risk class (2) is the biggest challenge for Naive Bayes. The model only recognized 4 of the 67 cases (Recall 0.06) with a Precision of 0.18 and a very low F1-score of 0.09, indicating that most of the Mid Risk data was confused with the Low Risk class. This condition indicates that the distribution of features between classes is quite similar so that this simple probability-based method has difficulty separating minority classes. Overall, Naive Bayes is quite competitive for the majority class, but less effective for the Mid Risk class

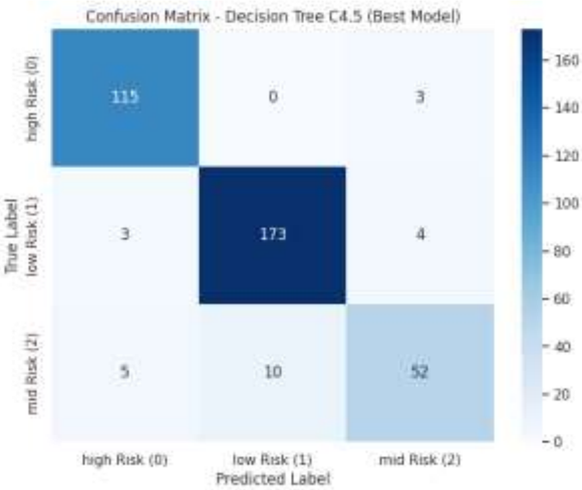

Figure 7. Confusion Matrix Decision Tree

The Decision Tree C4.5 (Best Model) model showed excellent classification performance with an accuracy of 93.15%, of which 340 out of 365 test data were correctly classified. The model works very stable in the High Risk (0) and Low

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

Risk (1) classes. For the High Risk class, the model correctly recognized 115 of the 118 data (Recall 0.974) with a Precision of 0.935 and an F1-Score of 0.954. The Low Risk class (1) also showed almost as strong performance, with 173 correct predictions from 180 actual data (Recall 0.961), Precision 0.945, and F1-Score 0.953. This shows that the model is able to recognize patterns in the two majority classes consistently and accurately. For the Mid Risk class (2), although it is the most difficult class, the model still shows good results. The model managed to correctly classify 52 of the 67 data (Recall 0.776), while the Precision 0.881 and F1-Score 0.825 showed that the predictions given were quite precise and balanced. Overall, Decision Tree C4.5 was able to deliver strong performance across the board, making it one of the most effective models in predicting pregnancy risk compared to other models.

3.6 Best Model Selection Results

The results of the model evaluation display the results of Accuracy, Precision recal, and f1 score of each algorithm.

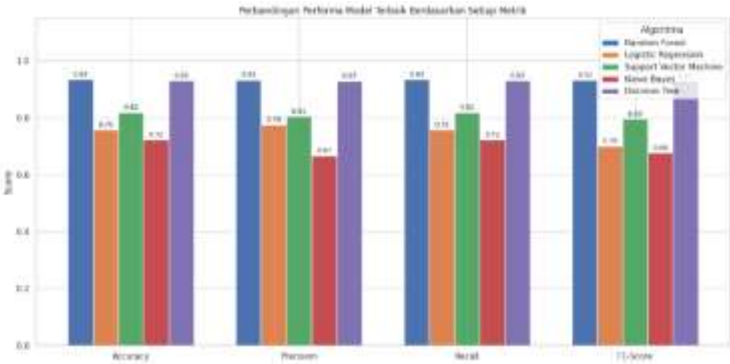

Figure 8. results Accuracy, Precision recal, and f1 score of each algorithm

Table 11. Results of Machine Learning Model Evaluation

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	93%	93%	93%	93%
Decision Tree C4.5	93%	93%	93%	93%
SupportVector Machine	82%	81%	82%	80%
LogisticRegression	76%	78%	76%	70%
Naive Bayes	72%	67%	72%	68%

Based on the results of the evaluation of five *machine learning* algorithms using the Accuracy, Precision, Recall, and F1-Score (*weighted average*) metrics, it was found that the tree-based algorithm showed the best performance. Random Forest became the best model with Accuracy, Precision, Recall, and F1-Score values of 93% each, demonstrating the stability and consistency of predictions across all pregnancy risk classes. Decision Tree came in second with a relatively equivalent metric value after rounding, albeit slightly below Random Forest. The Support Vector Machine (SVM) showed medium performance with an accuracy of 82%, while Logistic Regression and Naïve Bayes had lower performance, with an accuracy of 76% and 72%, respectively. The low performance of both models shows limitations in handling complex data patterns. Overall, Random Forest is recommended as the most optimal model for pregnancy risk prediction because it has the highest performance and the most stable across all evaluation metrics.

4. CONCLUSION

The results of this study are about the prediction of health risks of pregnant women using five machine learning algorithms: Random Forest, Decision Tree C4.5, Support Vector Machine (SVM), Logistic Regression, and Naive Bayes. The model evaluation process is with the 80:20 hold-out method. The study successfully identified cases of pregnancy risk in pregnant women by utilizing health datasets containing variables such as age, body temperature, heart rate, hemoglobin levels, blood sugar levels, and blood pressure (systolic and diastolic). The results of the evaluation show that patterns in the data can be well studied by machine learning models, so that the three risk categories (low, medium, high) can be accurately classified. Analysis and Comparison of the performance of five Machine Learning algorithms. This study successfully analyzed and compared five machine learning algorithms, namely Logistic Regression, Decision Tree

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 5 No 1 June 2026
Page: 114-127
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v5i1.766

C4.5, Random Forest, Support Vector Machine (SVM), and Naïve Bayes, using the evaluation metrics of Accuracy, Precision, Recall, and F1-Score (weighted average). The comparison shows that there is a significant difference in performance between algorithms. The Random Forest and Decision Tree C4.5 tree-based algorithms appear the most superior and stable compared to other algorithms. Random Forest achieved the best performance with an Accuracy of 93%, Precision of 93%, Recall of 93%, and F1-Score of 93%. This model shows stable performance across all risk classes and is the most superior to other algorithms. Decision Tree C4.5 is ranked second, with a high performance and close to Random Forest. The advantage of Decision Tree lies in its simpler and easier to understand model structure, so it can be a more practical alternative to implementation. Support Vector Machine showed medium performance with an accuracy of 82%, while Logistic Regression with an accuracy of 76% and Naïve Bayes with an accuracy of 72% had the lowest performance. This suggests that the two algorithms are less able to handle more complex data patterns in predicting pregnancy risk.

REFERENCES

[1] B. G. Sadikin, "Laporan Kinerja Kementrian Kesehatan RI," *Lap. Kinerja Kementrian Kesehat. RI*, pp. 1–23, 2023.
[2] N. Modi and Y. Kumar, "Automated Machine Learning-Based System for the Prediction of Maternal Health Indicators and High-Risk Pregnancy," in *2025 7th International Conference on Energy, Power and Environment (ICEPE)*, IEEE, 2025, pp. 1–6.
[3] K. Tomar, C. M. Sharma, and T. Prasad, "A Machine Learning-Based Risk Prediction Model During Pregnancy in Low-Resource Settings †," pp. 1–9, 2024.
[4] A. Hennessy, T. H. Tran, S. N. Sasikumar, and Z. Al-Falahi, "Machine learning, advanced data analysis, and a role in pregnancy care? How can we help improve preeclampsia outcomes?," *Pregnancy Hypertens.*, vol. 37, no. March, p. 101137, 2024, doi: 10.1016/j.preghy.2024.101137.
[5] R. A. F. W. A. H. W. K. Rahman, "Analisis Perbandingan Algoritma Machine Learning untuk Klasifikasi Tingkat Risiko Ibu Hamil," *Student Res. J.*, vol. 1, no. 6, pp. 246–261, 2023.
[6] R. Raja, I. Mukherjee, and B. K. Sarkar, "A Machine Learning-Based Prediction Model for Preterm Birth in Rural India," *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6665573.
[7] S. S. Assaduzzaman, Al Mashrafi, L. Tafakori, and M. Abdollahian, "Predicting maternal risk level using machine learning models," *BMC Pregnancy Childbirth*, vol. 24, no. 1, 2024, doi: 10.1186/s12884-024-07030-9.
[8] K. E. Setiawan, A. Kurniawan, and S. Y. Prasetyo, "Comparative analysis of machine learning decision tree-based Comparative analysis of machine learning decision models for predicting maternal health risks models for predicting maternal," *Procedia Comput. Sci.*, vol. 245, pp. 57–64, 2024, doi: 10.1016/j.procs.2024.10.229.
[9] V. Khoirunnisa and S. Lestari, "Implementasi Klasifikasi Kehamilan Beresiko Dengan Metode Naive Bayes Pada Puskesmas Kelurahan Malaka Jaya," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 1680–1693, 2023, doi: 10.35870/jimik.v4i3.396.
[10] WHO, *Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division.* 2023. [Online]. Available: https://www.who.int/publications/i/item/9789240068759
[11] D. Hawale, A. Chavan, D. Timalsina, and A. B. Thatere, "The Role of Artificial Intelligence in Healthcare: A Review," *AIP Conf. Proc.*, vol. 3188, no. 1, pp. 1–22, 2024, doi: 10.1063/5.0240194.
[12] World Health Organization 2013, "Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy," pp. 1–62.
[13] S. Khan *et al.*, "Preeclampsia and eclampsia-specific maternal mortality in Bangladesh : Levels , trends , timing , and care-seeking practices," vol. 13, pp. 1–12, 2023, doi: 10.7189/jogh.13.07003.
[14] M. Sandri and P. Zuccolotto, "A bias correction algorithm for the gini variable importance measure in classification trees," *J. Comput. Graph. Stat.*, vol. 17, no. 3, pp. 611–628, 2008, doi: 10.1198/106186008X344522.
[15] L. Saitta, "Support-Vector Networks," vol. 297, pp. 273–297, 1995.
[16] A. J. Scott, D. W. Hosmer, and S. Lemeshow, *Applied Logistic Regression.*, vol. 47, no. 4. 1991. doi: 10.2307/2532419.
[17] M. Ghanem *et al.*, "brain sciences Limitations in Evaluating Machine Learning Models for Imbalanced Binary Outcome Classification in Spine Surgery : A Systematic Review," 2023.