BULLETIN OF ENGINEERING SCIENCE, TECHNOLOGY AND INDUSTRY
(BESTI Journal) | ISSN (e): 3025-5821
Website : https://bestijournal.org/index.php/go/index | Volumes 4 No. 3 (2026)

PREDICTION OF SUSTAINABILITY OF FAMILY PLANNING PARTICIPANTS BASED ON DEMOGRAPHIC CHARACTERISTICS USING RANDOM FOREST

Dara Fazila[1], Zahratul Fitri[2], Lidya Rosnita[3]
[1,2,3]Informatics Engineering Study Program, Faculty of Engineering, Universitas Malikussaleh, Lhokseumawe
Email: dara525211@gmail.com, zahratulfitri@unimal.ac.id, lidyarosnita@unimal.ac.id

Received: 24/06/2026 | **Revised:** 05/07/2026 | **Accepted:** 27/07/2026 | **Published:** 03/08/2026

Abstract

Family Planning (KB) is one of the government programs aimed at controlling population growth and improving family welfare. However, some Family Planning participants discontinued the use of contraceptives, which may affect the success of the program. Therefore, this study aims to develop a prediction system for the continuity of Family Planning participants in Cot Girek District using the Random Forest algorithm. The study used 1,000 Family Planning participant records containing demographic, socioeconomic, and contraceptive-related attributes. After the data preprocessing stage, 998 records were used to construct the prediction model. The research involved data preprocessing, Random Forest model construction, and model evaluation using Hold-Out Validation and 5-Fold Cross Validation. The results showed that the Age of the Youngest Child attribute had the highest Gini Gain value of 0.3486, indicating that it was the most influential factor in predicting the continuity of Family Planning participants. The model achieved an Accuracy of 93%, Precision of 93.33%, Recall of 96.18%, and F1-Score of 94.74%, while 5-Fold Cross Validation produced an average accuracy of 97.40% with a standard deviation of ±4.95%. In addition, Black Box Testing confirmed that all system functions are operated according to user requirements. These findings indicate that the Random Forest algorithm can effectively predict the continuity of Family Planning participants and can be used as a decision-support tool to assist Family Planning officers in monitoring and providing more targeted assistance to participants.

Keywords: Family Planning; Family Planning Continuity; Predictions; Random Forest

INTRODUCTION

The Family Planning (FP) program is one of the Indonesian government's primary strategies to reduce population growth and improve family welfare through birth control. The program's success is measured not only by the number of new participants but also by the continuation rate of contraceptive use among registered participants. Low continuation rates can increase the risk of unplanned pregnancies and reduce the overall effectiveness of the family planning program (Ali et al., 2020). At the local level, Cot Girek District, one of the working areas of the North Aceh District Health Office, also faces challenges in maintaining the continuity of family planning (KB) participants. Initial interviews with community health centers (Puskesmas) revealed a decline in the number of active participants and unplanned changes in contraceptive methods, which could impact the stability of the family planning program's outcomes in the area. These issues highlight the need for a data-driven system capable of analyzing participant continuity patterns based on the demographic characteristics of the local community, allowing for more targeted and effective outreach and mentoring efforts.

Continuing family planning (FP) is influenced by various factors, including demographic characteristics such as age, education level, number of children, occupation, and the type of contraceptive method used. Furthermore, social aspects such as partner support, experience with side effects, and quality of health care also influence whether participants will continue or discontinue contraceptive use (Bibi et al., 2023). Furthermore, the Family Planning (FP) Program is fundamentally designed not only to control population growth but also to improve the quality of life for families. Proper child-planning helps families align their economic, educational, and health needs within their means. Amid rising living costs and challenging economic conditions, continued contraceptive use is increasingly important for families to plan for a more stable future. Through data analysis, recommendations for family planning sustainability are expected to assist families, particularly in Cot Girek District, in making informed decisions

regarding child-planning, thereby reducing the risk of poverty, alleviating household economic burdens, and promoting more prosperous families. Based on preliminary data obtained from relevant agencies in Cot Girek District, the number of fertile couples (PUS) is recorded at approximately 2,500, but only around 1,700, or approximately 68% of the total, are actively participating in the Family Planning program. This indicates that approximately 32% of couples have not yet started or are discontinuing contraceptive use. Furthermore, there has been a tendency for a decline in the number of active participants over time, as well as a tendency for premature discontinuation of contraceptive use (drop-out). This situation is a serious problem because it can increase the risk of unplanned pregnancy, increase the economic burden on families, and reduce the effectiveness of population control programs at the regional level.

Previous research conducted by Rabiatul Adawiyah, Yahya, and Muhammad Saiful in November 2023 in a study entitled Predicting the Level of Community Health in Suralaga District Based on the Use of Contraceptives Using the Random Forest Algorithm. The 790 data used consisted of 10 attributes, including full name, age, last education, fertile age couples, age at first pregnancy, age at first birth, family planning readiness, type of contraceptive, participation in the contraceptive program, and birth spacing. The data processing was carried out using a classification method with K-Fold Cross Validation testing. The evaluation results showed an accuracy level of 71.99% with an Area Under Curve (AUC) value of 0.680. In addition, this model produced a precision value of 71.99% and a recall of 100%. The findings of this study conclude that the Random Forest algorithm is able to analyze patterns of family planning use and provide decision tree output that identifies the most widely used contraceptives, namely IUDs, birth control pills, and birth control injections, so that it can assist the government in determining strategies to address population growth (Adawiyah et al., 2023).

Research conducted by Budi Prasojo and Emy Haryatmi (2021) focused on the application of the random forest algorithm to predict creditworthiness in the banking sector. The dataset used was German Credit Data, which consisted of 1,000 creditors. This study involved 20 input variables (such as account status, loan duration, credit history) and 1 output variable (credit status: Good or Bad). The methodology used included data preprocessing (transformation and normalization), data division into 80% training data and 20% test data, and classification model development. Model performance was evaluated using a Confusion Matrix and a Receiver Operating Characteristic (ROC) curve. The test results showed that the Random Forest method produced an accuracy rate of 83% and an AUC (Area Under Curve) of 0.83, categorized as a "Very Good Model." These findings concluded that the random forest algorithm is effective in assisting banks in making more accurate and objective creditworthiness decisions (Prasojo & Haryatmi, 2021). Therefore, this study aims to fill this gap by implementing the Random Forest algorithm specifically to predict the sustainability status (continuity) of contraceptive use, in order to reduce the drop-out rate of family planning participants, which has not been studied in depth in previous studies.

Research Objectives

1. Implementing the Random Forest algorithm in a recommendation system to determine the sustainability of Family Planning participants based on demographic characteristics.
2. Identifying demographic factors that influence the sustainability of Family Planning participants using the Random Forest algorithm approach.
3. Evaluate the level of accuracy of the Random Forest algorithm in providing recommendations for the sustainability of Family Planning participants based on demographic characteristics.

LITERATURE REVIEW

Data Mining, Machine Learning, and Prediction

Data mining is the process of searching for and extracting important information or patterns from large, complex data sets using statistical techniques, artificial intelligence, machine learning, and database analysis to generate useful knowledge for decision-making (Srirahayu & Pribadie, 2023). Conceptually, data mining is part of the Knowledge Discovery in Databases (KDD) framework, which is the process of discovering patterns or knowledge from raw data to transform it into meaningful information that can be used for decision-making. This technology combines various disciplines such as database processing, statistics, and machine learning to generate high-quality information from previously unstructured or hidden large amounts of data (Bhaskara et al., 2022). In data mining, prediction is not a standalone technique, but rather the goal of applying specific techniques, particularly classification and regression. Prediction is achieved by utilizing patterns learned from historical data to estimate conditions or values in new data. If the prediction results are categorical, the approach used is classification, while if the results are continuous numerical values, regression is used (Han et al., 2022; Kotsiantis, 2007). In the context

of this research, predicting the continuation of Family Planning participants (continuing or discontinuing) falls under the classification technique because the target variable is categorical. Machine learning is a branch of artificial intelligence (AI) that utilizes statistical techniques and computational algorithms to automatically build models based on datasets. The primary goal of machine learning is to enable computers to learn from data without having to be explicitly programmed for each task. Through this learning process, computers are able to recognize patterns, relationships, and characteristics within data and then use this knowledge to generate predictions or make decisions based on new data. Prediction is a data mining method that aims to estimate the value or category of data based on patterns learned from previous data. Prediction methods are a scientific approach based on machine learning, utilizing learning algorithms to extract knowledge from data. This process involves a training phase using data with known results, enabling the model to make predictions on new, previously unseen data (Han et al., 2022).

In the context of data mining and machine learning, predictions can produce output in the form of either numeric or categorical values. Numeric predictions are used to estimate continuous values, while categorical predictions are used when the target variable is discrete or restricted to a specific class. Categorical predictions are generally implemented through classification techniques, which is the process of mapping new data into specific classes based on a model built from labeled training data (Ardilla, 2021). Thus, in data mining, prediction is the ultimate goal, while classification is a technique used to achieve this goal, particularly in problems with categorical target variables. Classification does not stand alone as a research objective but rather serves as a technical approach in building categorical prediction models (Han et al., 2022). In this study, prediction is used to predict whether participants will continue or discontinue contraceptive use based on demographic characteristics and other related factors. Therefore, the prediction approach in this study is categorical prediction implemented through machine learning algorithms, specifically random forests. The general stages in the prediction process include:

1. Data preparation and cleaning (data preprocessing)
2. Attribute selection (feature selection)
3. Model development (training) with a specific algorithm, namely random forest.
4. Model evaluation (testing) using test data.
5. Interpretation of results and validation to ensure the reliability of predictions.

Random Forest

Random Forest is an extension of Decision Tree for classification and regression. The model builds multiple trees from different subsets of data and features, then combines the results into a final prediction. The diversity between trees improves model stability and helps reduce the risk of overfitting (Han et al., 2022; Fhonna et al., 2023). In addition, Random Forest also applies bootstrap sampling (bagging) techniques, which randomly sample data with replacement from the training data, so that each tree has slightly different training data. At each tree node, only a subset of features is randomly selected to determine the best split, increasing diversity between trees and strengthening the model's generalization ability. With this mechanism, random forests are able to handle complex data variations, minimize the risk of overfitting, and provide more stable and reliable predictions. The stages of the random forest used refer to the standard procedures of the random forest algorithm (Fhonna et al., 2023).

The main stages of Random Forest include bootstrap sampling, random feature selection, tree formation based on impurity reduction, and aggregation of results through majority voting. Bootstrap sampling creates different training data for each tree, while feature subset selection reduces correlation between trees. In this study, the Gini Index was used to separate nodes. A low Gini value indicates that the majority of samples at a node come from a single class. After all trees have generated predictions, the class with the most votes becomes the final decision.

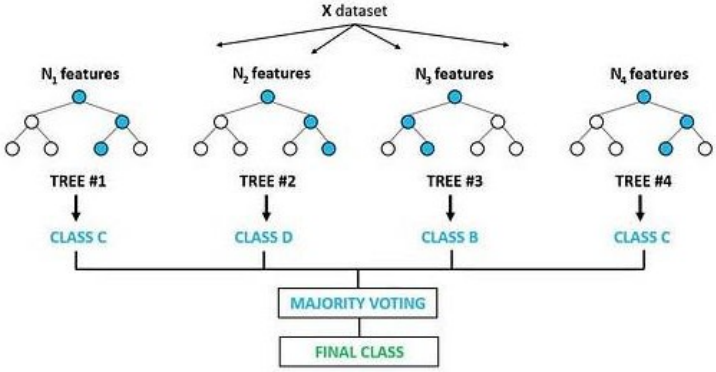

Figure 1. Illustration of the Random Forest algorithm (source: thesis manuscript)

After the final prediction (ŷ) is obtained through the majority voting mechanism in the random forest algorithm, the next stage is to evaluate the model's performance to determine the level of accuracy and reliability of the prediction results. Performance evaluation in this study was conducted using a confusion matrix, a table used to assess the accuracy and how well the algorithm produced from the classification that has been created to classify and predict the attributes of the test data. This method was developed as an assessment of machine learning algorithms used to solve classification problems. There are False Negatives (FN), False Positives (FP), True Negatives (TN), and True Positives (TP) in the confusion matrix. The following table shows information about the confusion matrix.

Family Planning Continuity and Demographic Characteristics

Cot Girek District, North Aceh Regency, Aceh Province, is an area with diverse demographic and socioeconomic characteristics. The majority of the population lives in rural areas and earns a living in the agriculture, plantation, and small-scale trade sectors. According to the publication "Cot Girek District in Figures 2024" published by the Statistics Indonesia (BPS) of North Aceh Regency, this area covers an area of approximately 189 km² with a population of over 18,000 in 2023 (BPS of North Aceh Regency, 2024). The increasing population growth each year and the uneven distribution of the population have led to differences in community access to health facilities and public services, including family planning (KB) services.

Most people live far from primary health care facilities such as community health centers (Puskesmas) and family planning clinics. This situation presents a challenge to the implementation of reproductive health programs, as services are not always easily accessible to all residents. Furthermore, the region's geographic conditions, consisting of plains and hills, also influence the distribution of contraceptives and the mobility of field health workers. These geographic and demographic factors are closely linked to the effectiveness of family planning programs, particularly in terms of equitable service distribution and the sustainability of family planning participants.

Community characteristics such as education level, family economic status, number of children, and availability of health care infrastructure are also important factors influencing participation and continued contraceptive use. Therefore, the demographic study of Cot Girek District not only provides a general overview of the region but also examines how social, economic, and geographic factors contribute to the continued participation of family planning participants in the area The Family Planning Program helps communities plan the number, spacing, and timing of births through the use of safe and effective contraceptive methods. This program is aimed at improving family welfare and strengthening family development, population management, and family planning (BKKBN, 2024). Continuity of contraceptive use, often referred to as the continuation rate, is the proportion of users who continue to use a particular contraceptive method over a specified period without prematurely stopping or switching methods. For example, a study in Indonesia by S. Sistri measured contraceptive continuation from initial use for up to 72 months (Sistri, 2020). This continuation rate is influenced by various factors, including:

Factors Related to Continuity

1. Age: Younger users tend to have higher discontinuation rates than more mature users due to instability of reproductive plans and changes in method preferences.
2. Education Level: Higher education generally correlates with more consistent method use and long-term method selection.
3. Number of Children (Parity): The number of children a family has often determines their goals regarding birth control. Participants with more children tend to prefer longer-term methods and have a higher level of sustainability.
4. Age of Last Child: The age of the last child influences the decision whether a family wants to delay or stop childbearing. The interval between previous children plays a role in the sustainability of the method used.
5. Income or Economic Status: Households with better economic conditions tend to be better able to maintain regular contraceptive use.
6. Gender of Last Child: Cultural preferences regarding the gender of a child may influence the decision to continue or discontinue family planning, especially in areas with certain norms.
7. Contraceptive Method Used: The type of method is a direct determinant of continuation. Short-term methods like the pill and injections have higher discontinuation rates, while IUDs and implants tend to have more stable continuation rates.

METHOD

This research was conducted from December 2025 until the completion of the thesis. Data collection was conducted at several agencies related to family planning services and data collection, namely the National Population and Family Planning Agency (BKKBN) in Cot Girek District, the local Community Health Center (Puskesmas), and Integrated Health Posts (Posyandu) in villages within the sub-district. In this study, the research flow was structured systematically to ensure the entire process runs smoothly and produces a valid predictive model. The stages of the research flow used are as follows:

At this stage, researchers collected various data related to data mining, machine learning, classification, random forest algorithms, and information by exploring internet sources and previous research on family planning sustainability. Literature sources came from journals, books, and scientific articles related to the research topic. This stage was conducted to identify the main problem, namely the lack of a system capable of predicting the sustainability of Family Planning (KB) participants. This problem was formulated based on field conditions and the needs of the National Population and Family Planning Agency (BKKBN), Community Health Centers (Puskesmas), and Integrated Health Posts (Posyandu).

Interviews were conducted with BKKBN officers from Cot Girek District, Community Health Centers, and Posyandu cadres to obtain information regarding: Variables recorded in family planning services, Factors influencing the sustainability of family planning and Availability and structure of data in the field. Researchers conducted direct observations at the Cot Girek District BKKBN (National Population and Family Planning Agency) and the village community health centers (Puskesmas) and integrated health posts (Posyandu) to assess the format of recorded data, the family planning service process, and the consistency of participant data recording. Observations helped ensure that the data met modeling requirements.

The data processing stage is carried out to ensure that the dataset used to predict the sustainability of family planning participants is of good quality and ready to be processed by the random forest algorithm. The steps taken are as follows: Data is cleaned of empty values, duplicate data, and formatting errors. This process is performed to prevent interference with analysis results and maintain model accuracy. Variables in the form of categories such as type of contraception, education, occupation, or sustainability status are converted into numerical data using methods such as Label Encoding or One-Hot Encoding so that they can be processed by the random forest algorithm.

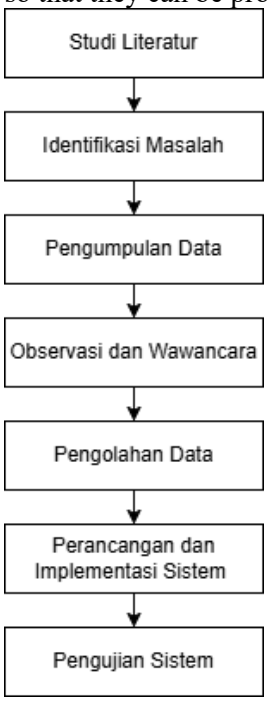

Figure 2. Research flow

Data Collection and System Analysis

The data collection methods used in developing the family planning participant sustainability recommendation system must be accurate to meet the research needs. The following are the data collection methods employed by the researchers:

Interviews were conducted directly with authorized officers, such as staff from the National Population and Family Planning Agency (BKKBN) in Cot Girek District, community health center officers, and integrated health post (Posyandu) cadres. The interviews aimed to obtain information on the types of family planning participant data recorded by the agency, variables or factors considered to influence the sustainability of contraceptive use, the data collection process in the field, and various obstacles encountered in monitoring family planning participants. Information obtained from these interviews was used to adjust the research variables to the actual data conditions available in the field.

The literature review was conducted by reviewing various scientific sources, such as national and international journals, academic books, proceedings, agency reports, and previous research related to data mining, machine learning, random forest algorithms, demographic variables in family planning analysis, and predictive systems in the health and social sectors. This literature review method assists researchers in formulating a theoretical foundation and selecting the analytical method most appropriate to the research objectives.

Direct observations were conducted at the BKKBN (National Population and Family Planning Agency) in Cot Girek District, at community health centers (Puskesmas), and at several village integrated health posts (Posyandu) to observe how family planning participant data is recorded and managed. Through these observations, researchers gained an understanding of the available data formats, their completeness and quality, the mechanism for recording family planning participants, and general patterns of family planning use and sustainability in the area. The results of these observations were used to validate the research variables and ensure the data's suitability for modeling using the random forest algorithm.

The system developed in this study is designed to support the analysis and prediction of family planning (KB) participant sustainability based on data obtained from the Cot Girek District National Population and Family Planning Agency (BKKBN), community health centers (Puskesmas), and village integrated health posts (Posyandu). This system utilizes data mining techniques and machine learning algorithms, specifically random forests, to identify patterns and tendencies in participants' continued or discontinued contraceptive use.

This prediction system functions to assist relevant parties, such as family planning officers and health workers, in understanding the factors that influence participant continuation. In developing this system, several input variables were used as the basis for modeling, including demographic characteristics and family planning service information such as maternal age, number of living children, type of contraceptive method, duration of contraceptive use, education level, occupation, and economic status. The system's output variable is the participant's continuation status, defined as a recommended continuation or discontinuation category, which is the target of the prediction model.

Table 1. Dataset Attributes

No.	Attribute	Description
1	Husband and wife age	Age of husband and wife participating in family planning
2	Education	Educational level of husband and wife
3	Occupation	Occupation of husband and wife
4	Income	Family income
5	Number of children	Number of children owned
6	Age of youngest child	Age of the youngest child
7	Sex of youngest child	Sex of the youngest child
8	Contraceptive type	Contraceptive method used

The target variable is the continuation status of family planning participants, which consists of two classes: Continuing (1) and Not Continuing (0). Input data includes demographic characteristics and contraceptive use obtained from the National Population and Family Planning Board (BKKBN), Community Health Centers (Puskesmas), and Integrated Health Posts (Posyandu).

RESULTS AND DISCUSSION

The data used in this study were Family Planning (KB) participant data obtained from a census in Cot Girek District. The data included information on the characteristics of fertile couples participating in the KB program and their continued use of contraceptives, which served as the target classification for the study.

The dataset used consisted of 1,000 family planning participants. After data cleaning, 998 data points were deemed suitable for use in the Random Forest model development. The variables used included husband's age, wife's age, husband's education, wife's education, husband's occupation, wife's occupation, family income, number of children, age of the last child, gender of the last child, and type of contraception used. In addition, there was one target variable, namely the continuation status of family planning participants, which consisted of two classes: Continuing and Not Continuing. The attributes used in this study can be seen in Table 4.1.

After analyzing the dataset, the next step is data cleaning and data preparation. This step is carried out to produce a clean dataset ready for use in the Random Forest model development process. The data cleaning process was carried out by checking for missing values, duplicate data, and data irrelevant to the classification process. Based on the inspection results, two data points were found to not meet the criteria and were therefore removed. Thus, the total research data was reduced from 1,000 to 998. After the data cleaning process is complete, the next step is data transformation. This stage aims to convert categorical data into numerical form so it can be processed by the Random Forest algorithm. The encoding process was performed on several categorical attributes, namely education, occupation, type of contraception, gender of last child, and sustainability status. This process aims to convert categorical data into numerical form so that it can be used in the model development process.

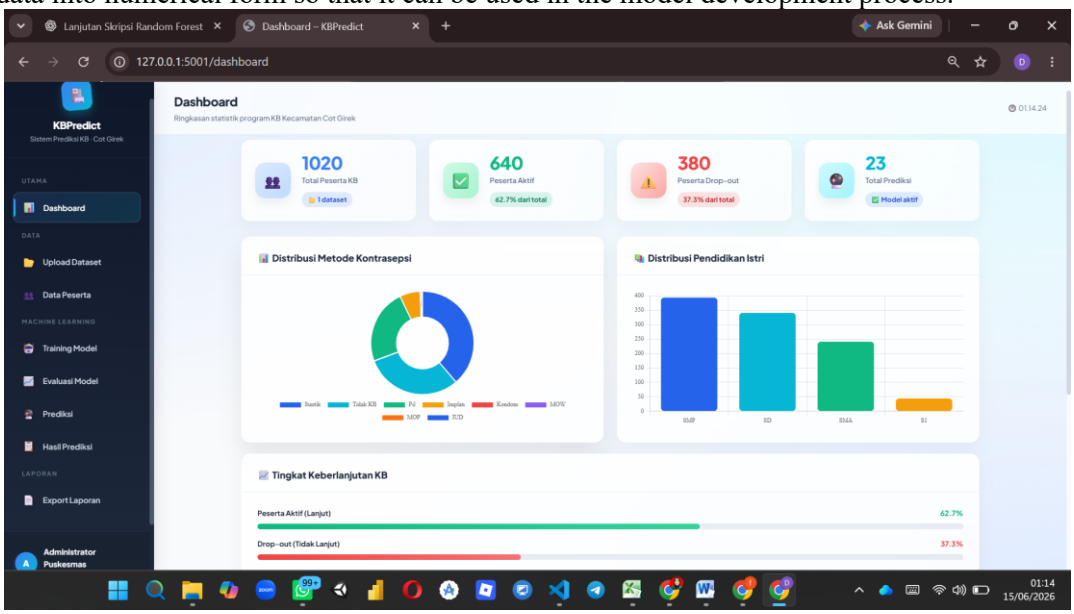

Figure 3. System dashboard for monitoring family planning participants

The system was built using Python with the Flask framework and a SQLite database. The main menu includes a dashboard, dataset upload, KB participant data, model training, model evaluation, predictions, prediction results, and report export.

Manual Decision Tree Calculation and Model Parameters

To illustrate the process of building a Random Forest model, this study performed manual calculations on a single decision tree. The manual calculations were performed using 20 sample data selected from the entire research dataset. These sample data were used as the basis for calculating the Gini Index, Gini Split, and Gini Gain values, as well as determining the best attributes for each node at each branch to form a decision tree. The data used in the manual calculations are shown in Table 4.5. The Random Forest algorithm in this study uses the Gini Index criterion to determine the best attributes. The Gini value is used to measure the purity of a node in the decision tree formation process. The Gini Index is calculated using Equation (2.3). We begin by calculating the Initial Gini of the overall distribution of Sustainability Status. The data has two classes:

Continue = 12 data

No Continue = 8 data

Total data = 20.

Gini(S) = 1 - $[(12/20)^2 + (8/20)^2]$ = 0.4800. This value indicates that the data at the initial node still contains a mixture of classes and requires separation based on the attribute that produces the highest Gini Gain.

Table 2. Gini Split and Gini Gain

Attribute	Split	Gain
Husband's age	0.3636	0.1164
Wife's age	0.4404	0.0396
Income	0.4688	0.0112
No. children	0.4791	0.0009
Youngest child age	0.4000	0.0800
Husband education	0.3422	0.1378
Wife education	0.2467	0.2333
Husband occupation	0.3882	0.0918
Wife's occupation	0.4278	0.0522
Sex of youngest child	0.4322	0.0478
Contraceptive type	0.3357	0.1443

Table 3. Random Forest Parameters

Parameter	Value
n_estimators	50
max_depth	3
criterion	like this
min_samples_split	10
min_samples_leaf	5
max_features	sqrt
max_samples	0.85

Based on the calculation results in Table 4.10, it is known that the Wife's Education attribute has the highest Gini Gain value, namely 0.2333. This value indicates that the Wife's Education attribute has the best ability to separate data compared to other attributes. Therefore, the Wife's Education attribute was selected as the root node in the formation of the decision tree in the Random Forest algorithm. The results of the root node formation show that the initial node has a Gini value of 0.4800, then a split is performed based on the Wife's Education attribute with a Gini Gain value of 0.2333. The initial structure of the resulting decision tree is shown as follows.

Random Forest Implementation and Majority Voting

In this stage, implementation will be carried out on 998 data sets, where 80% of the data (798 data sets) are used as training data to form a rule-based model, and 20% of the data (200 data sets) are used as test data to measure model accuracy. This process aims to identify patterns of demographic attributes that contribute to the participant's KB Sustainability Status. The model will be built using a decision tree (Random Forest) based on Gini Impurity and Gini Gain, then evaluated by matching predictions to the test data. Where the attributes used: Based on the parameters in Table 4.15, the Random Forest model was built using 50 decision trees (n_estimators = 50). Each tree has a maximum depth of 3 levels (max_depth = 3) using the Gini Index as the best attribute selection criterion (criterion = gini). The node formation process is carried out if the number of data in a node is at least 10 data (min_samples_split = 10), while each leaf node must have at least 5 data (min_samples_leaf = 5).

In this study, each decision tree was constructed using the bootstrap sampling technique. This technique involves randomly sampling training data using the parameter max_samples = 0.85, so that each tree uses approximately 85% of the randomly selected training data. This method allows each tree to obtain a different combination of data, resulting in a more diverse model. In addition to using bootstrap sampling data, the Random Forest algorithm also applies random feature selection to each node formation process. In this study, the parameter max_features = sqrt was used, so the number of attributes considered in each node separation process is determined based on the square root of the number of predictor attributes.

After all decision trees are formed during the training phase, the Random Forest algorithm executes the Majority Voting method to determine the final prediction result. At this stage, the system (through an internal Scikit-Learn function) aggregates all class predictions from all 50 trees (estimators) in the "forest." Each tree casts one "vote" for the predicted class. The class with the highest cumulative vote (probability) is determined as the model's final decision.

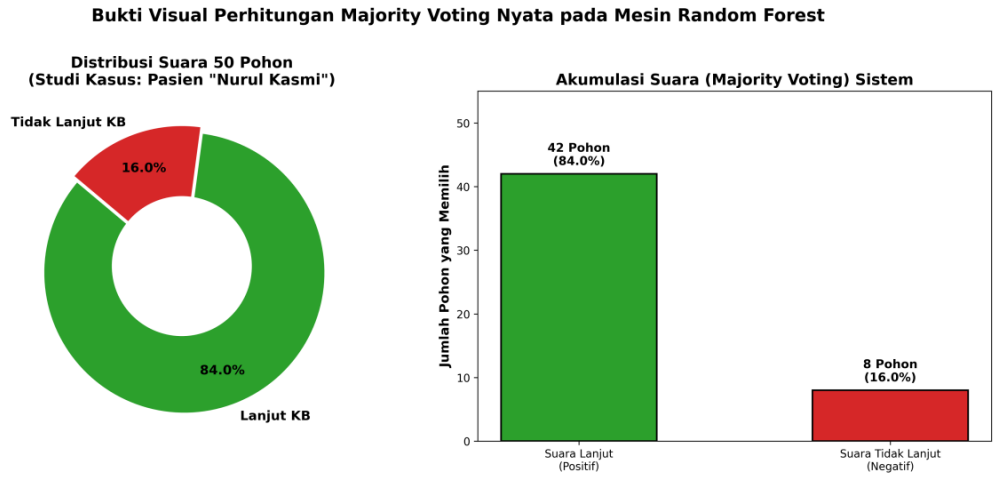

Figure 4. Majority voting of 50 decision trees for one test participant

Model Evaluation

After the Random Forest model was successfully created, the next step was to evaluate the model's performance in classifying the sustainability status of Family Planning (KB) participants. The evaluation was conducted to determine the model's ability to provide predictions that aligned with the actual data. Based on the test results, the model with a data split of 80% training data and 20% testing data was selected as the best model because it provided better classification performance. Therefore, the model evaluation in this study focused on the test results using this data split. The results of the model testing using the testing data are shown in Table 4.27.

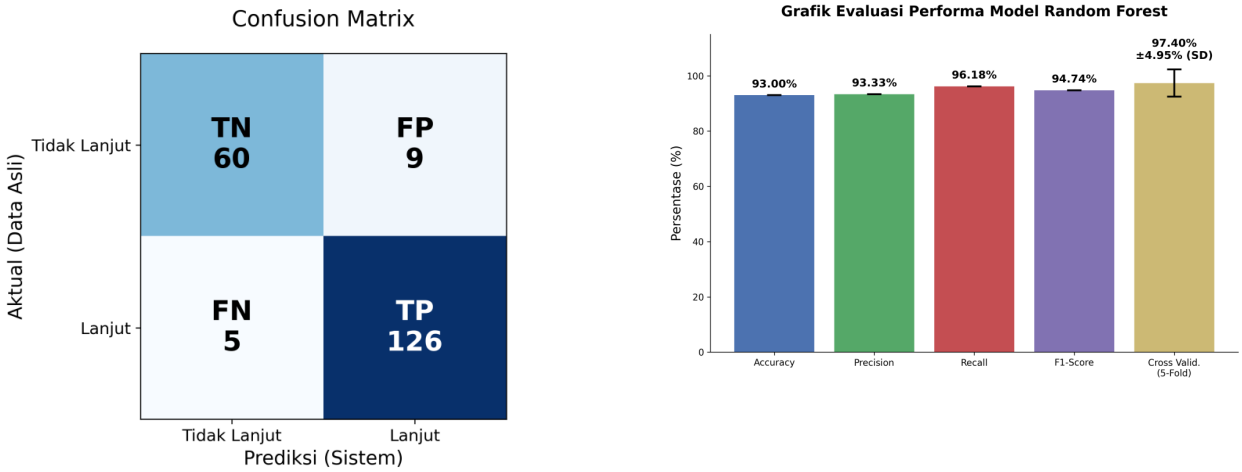

Figure 5. Confusion matrix and performance evaluation of the Random Forest model

Table 4. Model Evaluation Results

Metric	Value
Accuracy	93.00%
Precision	93.33%
Recall	96.18%
F1-Score	94.74%
5-Fold Cross Validation	97.40%
Standard Deviation	±4.95%

The 5-Fold Cross Validation results yielded an average accuracy of 97.40% with a standard deviation of ±4.95%. This value indicates a good level of model stability, as its performance was relatively consistent across different data sets. A comparison of the model evaluation results can be seen in Table 4.28.

The predictions generated by the system can be used as a basis for recommendations for family planning officers in providing support to participants. Participants predicted to be in the Advanced category or at high risk can be prioritized for further education, counseling, and monitoring. Thus, the system functions not only as a predictive tool but also as a decision-making tool in the Family Planning program.

Feature Importance, System Testing, and Conclusion

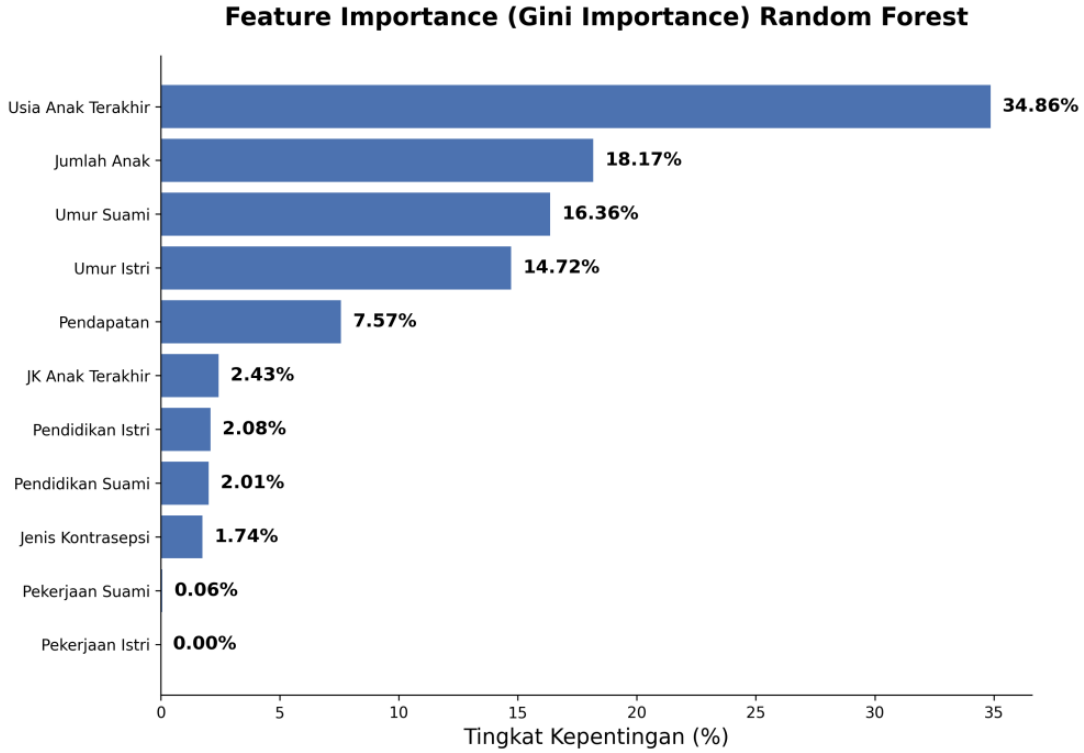

Figure 6. Feature importance of demographic attributes

After the Random Forest model was successfully created, a Feature Importance analysis was conducted to determine the level of contribution of each attribute to the prediction of the sustainability of family planning participants. The Feature Importance value was obtained from the contribution of each attribute in the decision tree formation process in the Random Forest model. The combination of Feature Importance and expert weights provides a more comprehensive picture of the factors influencing the sustainability of family planning participants. Based on Figure 4.17, the Age of Last Child attribute has the highest Feature Importance value, at 34.86%. These results indicate that this attribute makes the greatest contribution to the classification process of the sustainability of family planning participants compared to other attributes. The high Feature Importance value indicates that the age of the last child is one of the factors most frequently used by the Random Forest model in forming decision rules.

Table 5. Summary of Black Box Testing

Features	Expected Output	Result
Login	Dashboard or error message is displayed	Valid
Upload dataset	Dataset is stored and displayed	Valid
Training model	Random Forest model is generated	Valid
Model evaluation	Metrics and feature importance are displayed	Valid
Prediction and export	Predictions and Excel reports are produced	Valid

Based on the results of black box testing, all functions of the Random Forest algorithm-based family planning participant sustainability prediction system performed according to user requirements. The system was able to produce output consistent with expected results, thus confirming its proper functioning and readiness for use.

CONCLUSION

The Random Forest algorithm was successfully applied to 998 family planning participant data. Age of Last Child was the most influential attribute with a feature importance of 34.86%. Hold-Out Validation produced 93% accuracy, 93.33% precision, 96.18% recall, and 94.74% F1-Score, while 5-Fold Cross Validation produced an average accuracy of 97.40% with a standard deviation of ±4.95%. Black Box Testing showed that all system functions ran according to user requirements. The model is suitable for use to assist family planning officers in monitoring and providing more targeted assistance to participants.

REFERENCES

Adawiyah, R., Yahya, Y., & Saiful, M. (2023). Prediksi Tingkat Kesehatan Masyarakat Berdasarkan Penggunaan Alat Kontrasepsi Menggunakan Algoritma Random Forest. Jurnal PRINTER: Jurnal Pengembangan Rekayasa Informatika dan Komputer, 1(2). https://doi.org/10.29408/jprinter.v1i2.22007

Ali, M. (2020). Ensuring Contraceptive Security Through Effective Supply Chains (Family Planning Evidence Brief). World Health Organization, Geneva.

Ardilla, Y. (2021). Konsep Machine Learning. Dalam N. Rismawati (Ed.), Data Mining dan Aplikasinya. Widina Bhakti Persada, Bandung.

Badan Kependudukan dan Keluarga Berencana Nasional, https://www.bkkbn.go.id/, diakses pada 15 Desember 2025 pukul 14.30 WIB.

Badan Pusat Statistik Kabupaten Aceh Utara, https://acehutarakab.bps.go.id/, diakses pada 15 Desember 2025 pukul 10.00 WIB.

Bhaskara, I. M. W., Suputra, I. P. G. H., Widiartha, I. M., Kadyanan, I. G. A. G. A., Putra, I. G. N. A. C., & Dwidasmara, I. B. G. (2022). Klasifikasi Serangan Distributed Denial of Service (DDoS) Menggunakan Random Forest Dengan CFS. Jurnal Elektronik Ilmu Komputer Udayana, 11(2), 215–222. https://doi.org/10.24843/JLK.2022.v11.i02.p01

Bibi, F., Saleem, S., Rozi, S., & Tikmani, S. S. (2023). Factors Associated With Continuation of Hormonal Contraceptives Among Married Women of Reproductive Age in Gilgit, Pakistan: A Community-Based Case–Control Study. BMJ Open, 13(11), e075490. https://doi.org/10.1136/bmjopen-2023-075490

Fhonna, R. P., Afrillia, Y., Zulfan, Z., Aqmal, J., & Abadi, S. (2023). Klasifikasi Penentuan Jenis Tanah yang Sesuai terhadap Tanaman Pangan sebagai Solusi Ketahanan Pangan di Kabupaten Pidie Jaya Menggunakan Metode Random Forest. Jurnal Informasi dan Teknologi, 5(4). https://doi.org/10.60083/jidt.v5i4.402

Han, J., Kamber, M., & Pei, J. (2022). Data Mining: Concepts and Techniques (4th ed.). Morgan Kaufmann, Waltham.

Nurhalizah, R. S., Ardianto, R., & Purwono. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. Jurnal Ilmu Komputer dan Informatika (JIKI), 4(1), 61–72. https://doi.org/10.54082/jiki.168

Nurullah, F. A. (2021). Perkembangan Metode Kontrasepsi di Indonesia. Cermin Dunia Kedokteran, 48(3), 166–172. https://doi.org/10.55175/cdk.v48i3.53

Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. Jurnal Nasional Teknologi dan Sistem Informasi, 7(2), 79-89. https://doi.org/10.25077/TEKNOSI.v7i2.2021.79-89

Rizkianti, A., Amaliah, N., & Rachmalina, R. (2020). Penggunaan Kontrasepsi pada Remaja Perempuan Kawin di Indonesia (Analisis Riskesdas 2018). Buletin Penelitian Kesehatan, 45(4), 257–266. https://doi.org/10.22435/bpk.v45i4.7369.257-266

Sistri, S. (2020). Kelangsungan Penggunaan Kontrasepsi di Indonesia. Kesmas: National Public Health Journal, 3(5), 202–209. https://doi.org/10.21109/kesmas.v3i5.211

Srirahayu, A., & Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. Jurnal Ilmiah Informatika Global, 14(1). https://doi.org/10.36982/jiig.v14i1.2981

Suryanti, Y. (2021). Hubungan Tingkat Pengetahuan dengan Penggunaan Metode Kontrasepsi Jangka Panjang (MKJP) pada Pasangan Usia Subur. Jambura Health and Sport Journal, 3(1), 20–28. https://doi.org/10.35971/jjhsr.v1i1.1795