

Machine Learning-Based Classification of Family Planning Participant Status Using Random Forest and the CRISP-DM Framework

Irmawati^{a,1}, Syaifur Rahmatullah^{b,2*}, Mohammad Syamsul Azis^{a,3},

^a Information Systems Study Program, Universitas Bina Sarana Informatika,

^b Informatics Study Program, Universitas Nusa Mandiri

¹ irmawati.iat@bsi.ac.id; ² syaifur.syl@nusamandiri.ac.id*; ³ mohammad.myz@bsi.ac.id;

* *corresponding author*

ARTICLE INFO

Article history

Received 2026-06-29

Revised 2026-06-30

Accepted 2026-06-30

Keywords

Family Planning

Classification

Random Forest

CRISP-DM

Machine Learning

Class Imbalance

ABSTRACT

The Family Planning (FP) program requires accurate information to support evidence-based decision-making and improve the quality of reproductive health services. Classification of FP participant status can assist health authorities in identifying participant patterns and monitoring program implementation. Previous research using the Support Vector Machine (SVM) algorithm on the same dataset achieved an accuracy of 56.20%, indicating that improvements in classification performance are still required. This study proposes the Random Forest algorithm within the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework to classify FP participant status. The dataset consists of 1,402 FP participant records obtained from SATPEL PPKB, Cilebar District, Karawang Regency. Data preprocessing included data transformation, One Hot Encoding for categorical predictor variables, Label Encoding for the target variable, and Hold-Out Validation with an 80:20 train-test split using stratified sampling. The predictor variables were registration month, wife's birth year, wife's age, and contraceptive method, while the target variable was FP participant status (New, Change Method, and Repeat). Model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, classification report, and feature importance analysis. The Random Forest model achieved an accuracy of 59.43%, with weighted precision, recall, and F1-score of 59.00%. However, the macro-average precision, recall, and F1-score were 45.00%, 44.00%, and 44.00%, respectively, indicating performance differences across classes caused by class imbalance. The model achieved the highest F1-score for the New class (0.63), followed by the Repeat class (0.59), whereas the Change Method class obtained the lowest F1-score (0.11). Feature importance analysis identified wife's birth year and wife's age as the most influential predictor variables. Compared with the previous SVM-based model, Random Forest provided a modest improvement in accuracy and enhanced model interpretability through feature importance analysis. Nevertheless, the low macro-average performance indicates that further research should investigate class-balancing techniques and hyperparameter optimization to improve classification performance, particularly for minority classes.

1. Introduction

The Family Planning (FP) program is one of the national strategies aimed at controlling population growth, improving reproductive health, and enhancing family welfare through the appropriate use of contraceptive methods. The effectiveness of this program depends not only on the availability of healthcare services but also on the quality of information used to support policy formulation, program monitoring, and evidence-based evaluation. Proper management of FP participant data enables health authorities to identify contraceptive utilization patterns, participant characteristics, and continuation status, thereby facilitating evidence-based decision making. Recent studies have demonstrated that data-driven approaches can improve the understanding of contraceptive utilization and reproductive health behavior among women of reproductive age [1][2]. The rapid advancement of information technology has resulted in a substantial increase in the

volume of data generated across various sectors, including healthcare. Such growth requires analytical techniques capable of transforming large datasets into meaningful information. Data mining has emerged as an effective approach for discovering hidden patterns and extracting valuable knowledge from complex datasets. One of the most widely used techniques in data mining is classification, which aims to develop predictive models by assigning data into predefined categories based on their characteristics. Classification techniques have been successfully applied in numerous healthcare-related studies, including patient length-of-stay prediction, nutritional status classification, and disease diagnosis, while similar approaches have also demonstrated their effectiveness in non-health domains such as sentiment analysis and natural disaster analysis [3][4][5][6].

The growing development of machine learning has further expanded the capability of healthcare data analysis. Machine learning algorithms are able to identify complex relationships among variables, improve prediction accuracy, and support decision-making processes through automated learning from historical data. In reproductive health research, machine learning has been increasingly adopted to analyze contraceptive utilization and reproductive behavior [7]. [1] employed several machine learning algorithms to investigate long-acting reversible contraceptive utilization in Sub-Saharan Africa, [2] developed predictive models to identify factors associated with informed contraceptive choice among women in six high-fertility African countries. Furthermore, ensemble learning techniques have demonstrated excellent performance in medical classification problems, indicating their potential for supporting healthcare decision-making systems [8].

In Indonesia, several studies have investigated the application of machine learning algorithms for family planning and reproductive health classification. [9] compared Logistic Regression, Naïve Bayes, CART, C4.5, and k-Nearest Neighbor for classifying contraceptive methods among couples of reproductive age and concluded that the performance of classification algorithms is strongly influenced by dataset characteristics [10][11] Applied the Random Forest algorithm to predict community health conditions based on contraceptive utilization and reported satisfactory classification performance. In addition, [12] compared Support Vector Machine (SVM) and Backpropagation Neural Network (BPNN) for classifying couples of reproductive age in West Java Province and demonstrated that both algorithms achieved competitive results depending on the characteristics of the available data.

Among various machine learning algorithms, Random Forest has become one of the most popular classification methods because of its robustness and predictive capability [13][14]. Random Forest is an ensemble learning algorithm that constructs multiple decision trees using bootstrap sampling and combines their predictions through a majority voting mechanism. Compared with a single decision tree, Random Forest generally provides higher predictive accuracy, greater stability, and better resistance to overfitting. In addition, the algorithm provides feature importance analysis, allowing researchers to identify variables that contribute most significantly to prediction performance [15].

Previous research conducted by [16] classified Family Planning participant status using the Support Vector Machine (SVM) algorithm based on the wife's age and contraceptive method type. Using 1,402 FP participant records collected from SATPEL PPKB, Cilebar District, Karawang Regency, the proposed model achieved an accuracy of 56.20%. Although the study demonstrated the feasibility of machine learning for FP participant classification, the obtained performance indicates that further improvement is still required, particularly in achieving more balanced classification results across all participant categories.

Exploratory analysis of the dataset used in the present study revealed an imbalanced class distribution, where the number of participants in the Change Method category was substantially lower than those in the New and Repeat categories. Such class imbalance may cause classification models to become biased toward majority classes, thereby reducing their ability to correctly identify minority-class instances and affecting the overall classification performance, particularly for minority classes [17][18]. Although the dataset exhibits class imbalance, this study evaluates the performance of the Random Forest algorithm using the original data

distribution after preprocessing through One Hot Encoding without applying additional data balancing techniques[19]. This study therefore proposes the implementation of the Random Forest algorithm within the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The dataset is preprocessed using One Hot Encoding before model training to transform categorical variables into numerical representations suitable for machine learning. CRISP-DM provides a systematic methodology consisting of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment, making it one of the most widely adopted frameworks for data mining projects [20]. The proposed model utilizes the month of registration, wife's birth year, wife's age, and contraceptive method type as predictor attributes to classify Family Planning participant status. Model performance is evaluated using Accuracy, Precision, Recall, F1-score, Confusion Matrix, and Feature Importance analysis. Furthermore, the Classification Report is analyzed through Macro Average and Weighted Average values to provide a more comprehensive evaluation of model performance across all participant categories, particularly the minority class. The findings are expected to provide an alternative Random Forest-based classification model for Family Planning participant status while identifying the predictor variables that contribute most significantly to the classification process. In addition, the study is expected to provide a comprehensive evaluation of classification performance across all participant categories, particularly the minority Change Method class.

2. Method

This study uses the Cross Industry Standard Process for Data Mining (CRISP-DM) method as a framework for the development process of the classification model. CRISP-DM was selected because it provides systematic stages, starting from problem understanding to model evaluation. In this study, the implemented stages include Business Understanding, Data Understanding, Data Preparation, Modeling, and Evaluation. The Deployment stage was not performed because this research focuses only on the development and evaluation of the classification model.



Figure 1. CRISP-DM Method Stages

2.1 Business Understanding

The Business Understanding stage aims to identify the research problem and determine the objective of developing a classification model. Previous research using the Support Vector Machine (SVM) algorithm achieved an accuracy of 56.20%, indicating that further improvement is required to enhance the classification performance of Family Planning (FP) participant status. Therefore, this study applies the Random Forest algorithm as an alternative classification method to improve classification performance.

Random Forest is selected because it is an ensemble learning algorithm that combines multiple decision trees to produce a more stable prediction model and reduce the risk of overfitting. In addition, Random Forest provides feature importance analysis, which allows the identification of the variables that contribute most significantly to the classification process.

The objective of this research is to develop a Random Forest-based classification model for predicting Family Planning participant status. The model performance is evaluated using Accuracy, Precision, Recall, F1-score, Confusion Matrix, Feature Importance analysis, and Classification Report analysis to provide a comprehensive evaluation of the Random Forest classification model.

2.2 Data Understanding

The Data Understanding stage is conducted to analyze and understand the characteristics of the dataset used in this research. The dataset was obtained from SATPEL PPKB, Cilebar District, Karawang Regency, containing Family Planning participant records. The dataset consists of 1,402 records and eight attributes related to participant characteristics and contraceptive usage. Dataset exploration was performed to examine the data structure, variable types, descriptive statistics, and missing values. The results indicated that the dataset contained no missing values, making it suitable for the preprocessing stage.

Table 1. Research Dataset

No	Month	Wife Birth Year	Wife Age	FP Participant Status	Contraceptive Method
1	January	1984	41	Repeater	Pill
2	January	1996	30	New	Injection
3	January	1998	28	Repeater	Pill
4	January	1979	47	New	Injection
...
1401	May	1980	45	New	Implant
1402	May	1980	46	New	Injection

The target variable in this study is Family Planning Participant Status, while the remaining attributes are used as predictor variables.

Table 2. Variable Description

Variable	Function
Month	Predictor Variable
Wife Birth Year	Predictor Variable
Wife Age	Predictor Variable
Contraceptive Method	Predictor Variable
FP Participant Status	Target Variable

Before modeling, the distribution of the target variable was analyzed to understand the class composition of Family Planning participant status. This analysis is important because an imbalanced class distribution can influence classification performance.

2.3 Data Preparation

The Data Preparation stage aims to transform the raw dataset into a suitable format for the Random Forest classification model. The preparation process consists of data transformation, attribute removal, feature encoding, and dataset splitting.

a) Data Transformation

The attribute Wife's Date of Birth was converted into datetime format. After conversion, the year information was extracted to create a new attribute called Year of Birth. The original date attribute was removed because machine learning algorithms require structured numerical input.

Table 4. Data Transformation

Original Attribute	Transformation	Result
Wife Date of Birth	Extract Year	Wife Birth Year
Wife Date of Birth	Remove Original Attribute	Deleted

Several attributes related to participant identity were removed before the modeling stage. The attributes Nama Suami and Nama Istri were excluded because they represent personal identification information and do not provide relevant predictive information for Family Planning participant classification.

b) One Hot Encoding

The dataset contains categorical predictor variables. Since the Random Forest algorithm requires numerical input, categorical attributes were transformed using One Hot Encoding.

The implementation uses the `pd.get_dummies()` function with the parameter `drop_first=True`. This method converts categorical values into binary numerical representations (0 and 1). The `drop_first=True` parameter removes one category from each categorical variable to reduce redundancy in the generated features.

Table 5. One Hot Encoding

Contraceptive Method	Pill	Injection	Implant
Pill	1	0	0
Injection	0	1	0
Implant	0	0	1

c) Label Encoding of Target Variable

The target variable **Status Peserta KB** was transformed into numerical values using the Label Encoding technique through the `LabelEncoder()` function.

This transformation was applied because Random Forest requires numerical labels during the training process.

The encoded target values follow the class labels generated by the Label Encoder.

Table 6. Label Encoding

Original Class	Encoded Value
New	0
Change Method	1
Repeat	2

d) Dataset Splitting

After preprocessing, the dataset was divided into training and testing data using the Hold-Out Validation method with an 80:20 ratio.

The splitting process was performed using the `train_test_split()` function with the parameters: `test_size = 0.2`, `random_state = 42`, `stratify = y`

The `stratify` parameter was applied to maintain the proportion of each class in both training and testing datasets.

Table 7. Dataset Partition

Dataset	Number of Records	Percentage
---------	-------------------	------------

Training	1121	80%
Testing	281	20%
Total	1402	100%

2.4 Modeling

The classification model was developed using the Random Forest algorithm implemented through the `RandomForestClassifier()` function from the Scikit-Learn library. The model was initialized using: `RandomForestClassifier(random_state=42)`. The training process was performed using the training dataset, while the testing dataset was used to evaluate the model performance.

2.5 Evaluation

The Evaluation stage aims to assess the performance of the Random Forest classification model in predicting the class of Family Planning (FP) participants. Model performance was evaluated using the testing dataset obtained from the dataset splitting stage. Several classification performance metrics were employed, namely Accuracy, Precision, Recall, F1-score, Confusion Matrix, and Classification Report analysis. These evaluation methods provide a comprehensive assessment of the Random Forest model, particularly for datasets with imbalanced class distributions. In addition, Feature Importance analysis was conducted to identify the contribution of each predictor variable to the classification results.

a) Accuracy

Accuracy represents the proportion of correctly classified instances compared with the total number of testing data.

b) Precision

Precision measures the proportion of correctly predicted positive instances among all instances predicted as positive.

c) Recall

Recall measures the ability of the model to identify actual positive instances correctly.

d) F1-Score

F1-Score is the harmonic mean between precision and recall. This metric provides a balanced evaluation when the dataset contains unequal class distributions.

e) Confusion Matrix

The Confusion Matrix is used to evaluate the classification performance by comparing the actual class labels with the predicted class labels. It summarizes the number of correct and incorrect predictions for each class, allowing the identification of classification errors such as false positives and false negatives. This evaluation provides a detailed understanding of the strengths and weaknesses of the Random Forest model in classifying Family Planning participant data.

f) Feature Importance

In addition to the classification performance metrics, this study also analyzes Feature Importance generated by the Random Forest algorithm. Feature Importance measures the contribution of each predictor variable to the classification process. Variables with higher importance values have a greater influence on the model's prediction results. This analysis provides insights into the factors that most significantly affect the classification of Family Planning participant classes.

3. Results and Discussion

3.1 Data Preprocessing Results

The data preprocessing stage was conducted to prepare the dataset before the classification process. Initially, the dataset consisted of 1,402 Family Planning participant records containing numerical and categorical attributes. During preprocessing, categorical predictor variables were transformed into numerical representations using the One Hot Encoding technique, while the target variable (**FP Participant Status**) was converted into numerical labels using Label Encoding.

The One Hot Encoding process transformed categorical predictor variables into binary numerical features, enabling the Random Forest algorithm to process categorical information without introducing artificial ordinal relationships between categories. After the transformation process was completed, the dataset was divided into training and testing datasets using the Hold-Out Validation method with an 80:20 ratio.

Table 8. Dataset after Preprocessing

Dataset	Number of Records
Training Data	1,121
Testing Data	281
Total	1,402

The preprocessing results indicate that the dataset was successfully transformed into a numerical format suitable for the Random Forest algorithm. Furthermore, stratified sampling was applied to maintain the proportion of each class in both training and testing datasets, providing a reliable basis for model training and evaluation.

3.2 Random Forest Classification

The Random Forest classification model was trained using the training dataset and evaluated using the testing dataset. Model performance was assessed using Accuracy, Precision, Recall, F1-score, Confusion Matrix, and Classification Report analysis, while Feature Importance analysis was performed to identify the contribution of each predictor variable.

The experimental results show that the Random Forest model achieved an accuracy of 59.43%. The weighted average values of Precision, Recall, and F1-Score were 59.00%, indicating that the model provided moderate performance in classifying Family Planning participant status.

Table 9. Random Forest Classification Performance

Evaluation Metric	Value
Accuracy	59.43%
Precision (Macro Average)	45.00%
Recall (Macro Average)	44.00%
F1-Score (Macro Average)	44.00%
Precision (Weighted Average)	59.00%
Recall (Weighted Average)	59.00%
F1-Score (Weighted Average)	59.00%

The evaluation results indicate that the Random Forest algorithm was able to identify classification patterns within the dataset. However, the performance was influenced by the characteristics and distribution of each class. Therefore, further analysis using the Classification Report and Confusion Matrix was performed to evaluate the model performance for each category.

3.3 Classification Report Analysis

A detailed evaluation was performed using the Classification Report, which provides Precision, Recall, F1-Score, and Support values for each classification category.

Table 10. Classification Report of the Random Forest Model

Class	Precision	Recall	F1-Score	Support
New	0.61	0.65	0.63	146
Change Method	0.14	0.09	0.11	11
Repeat	0.6	0.57	0.59	124
Accuracy			0.59	281
Macro Average	0.45	0.44	0.44	281
Weighted Average	0.59	0.59	0.59	281

Based on Table 10, the Random Forest model achieved the highest classification performance for the New category, with a Precision of 0.61, Recall of 0.65, and F1-Score of 0.63. These results indicate that the model was able to correctly identify most participants classified as New.

For the Repeat category, the model achieved a Precision of 0.60, Recall of 0.57, and F1-Score of 0.59. This indicates that the model provided relatively balanced performance in identifying repeated participants.

The Change Method category obtained the lowest performance, with a Precision of 0.14, Recall of 0.09, and F1-Score of 0.11. This result indicates that the model experienced difficulty identifying this category, which may be influenced by the limited number of samples compared with the other categories.

Overall, the weighted average results show that the Random Forest model achieved an accuracy of 59.43% while the macro average value indicates that the classification performance varied among categories, especially for categories with fewer observations.

3.4 Confusion Matrix Analysis

The classification performance of the Random Forest model was further analyzed using the Confusion Matrix to compare the actual class labels with the predicted class labels. The Confusion Matrix provides detailed information regarding correctly classified and misclassified instances for each Family Planning participant status category.

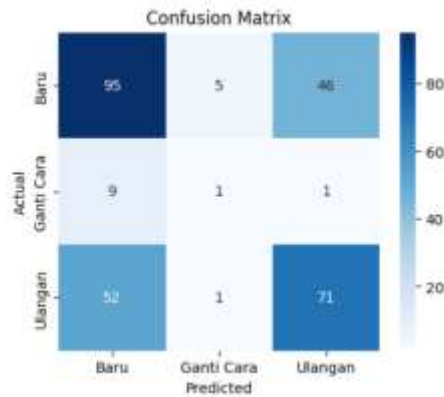


Figure 3. Confusion Matrix of the Random Forest Model

Based on Figure 3, the Random Forest model correctly classified 95 out of 146 instances in the New category. However, 5 instances were incorrectly classified as Change Method and 46 instances were predicted as Repeat. This indicates that several patterns between categories share similar characteristics, resulting in classification errors. For the Change Method category, the model correctly classified only 1 out of 11 instances. Meanwhile, 9 instances were incorrectly classified as New and 1 instance was classified as Repeat. The low classification performance of this category may be influenced by the limited number of samples available during the training process, resulting in insufficient patterns for the model to learn.

For the Repeat category, the model successfully classified 71 out of 124 instances. However, 52 instances were incorrectly classified as New and 1 instance was predicted as Change Method. This result indicates that the model still experienced difficulties in distinguishing between New and Repeat categories due to similarities in the predictor variables.

Overall, the Confusion Matrix analysis shows that the Random Forest model performed better in identifying the New and Repeat categories compared with the Change Method category. The classification errors indicate that the model performance is affected by the distribution of samples among categories, particularly for categories with fewer observations.

3.5 Feature Importance

Feature Importance analysis was conducted to identify the predictor variables that contributed most significantly to the Random Forest classification model. The importance score represents the contribution of each feature in the decision-making process of the Random Forest algorithm. Features with higher importance values indicate a greater influence on the classification results.

The Feature Importance visualization is presented in Figure 4 to illustrate the ranking of predictor variables based on their contribution to the Random Forest model.

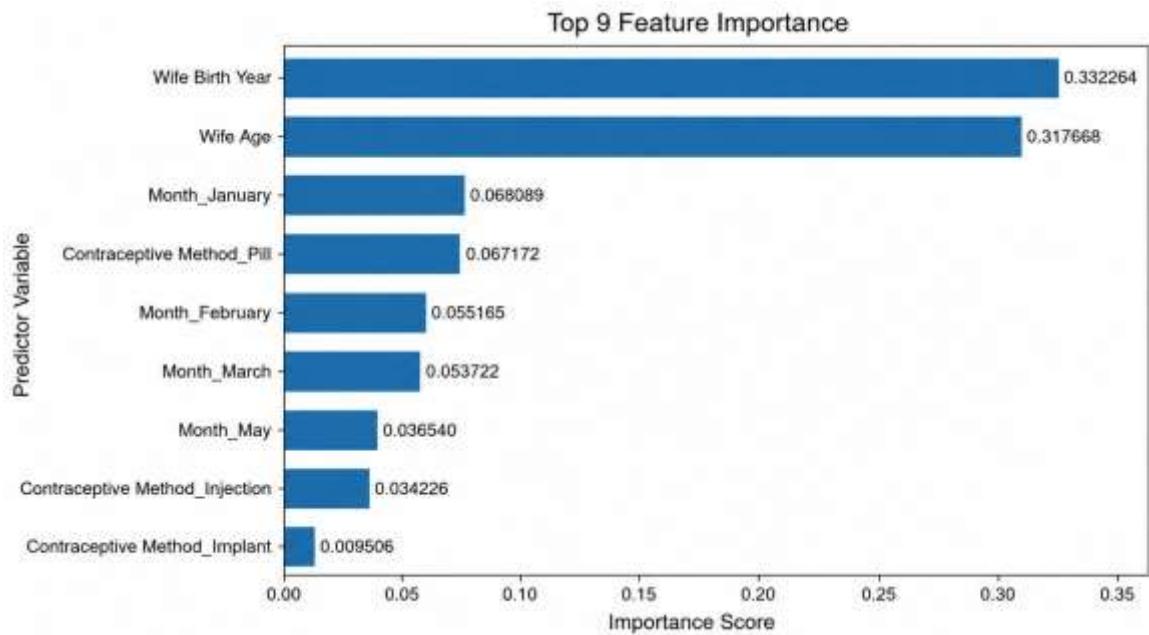


Figure 4. Feature Importance Ranking

Based on Figure 4, the highest contributing features were **Wife Birth Year** and **Wife Age**, indicating that demographic characteristics were the most influential variables in predicting Family Planning participant status.

The detailed importance values of each predictor variable are presented in Table 11.

Table 11. Feature Importance Values of the Random Forest Model

Rank	Predictor Variable	Importance Score
1	Wife Birth Year	0.332264
2	Wife Age	0.317668
3	Month_January	0.068089
4	Contraceptive Method_Pill	0.067172
5	Month_February	0.055165
6	Month_March	0.053722
7	Month_May	0.03654
8	Contraceptive Method_Injection	0.034226
9	Contraceptive Method_Implant	0.009506

Based on Table 11, Wife Birth Year obtained the highest importance score of 0.332264, followed by Wife Age with an importance score of 0.317668. These results indicate that demographic characteristics, particularly birth year and age, have the greatest influence on the classification process.

The importance of Wife Birth Year and Wife Age suggests that demographic factors may influence Family Planning participation patterns, including contraceptive preferences and participant decisions regarding Family Planning services.

In addition to demographic variables, contraceptive method features such as Contraceptive Method_Pill, Contraceptive Method_Injection, and Contraceptive Method_Implant also contributed to the classification model. This indicates that contraceptive usage patterns provide relevant information for distinguishing between Family Planning participant categories.

The month-related features, including Month_January, Month_February, Month_March, and Month_May, also contributed to the model, although their importance values were lower compared with demographic variables.

Overall, the Feature Importance analysis demonstrates that demographic characteristics and contraceptive method variables were the main factors influencing the Random Forest model in predicting Family Planning participant status.

3.6. Discussion

The results of this study demonstrate that the Random Forest algorithm was able to classify Family Planning participant status with an overall accuracy of **59.43%**. Although the model achieved moderate classification performance, it was able to identify meaningful patterns related to participant characteristics and contraceptive usage. The evaluation results indicate that the model performance varied across different participant categories, suggesting that the characteristics and distribution of each class influenced the classification results.

To provide a clearer comparison, the performance of the proposed Random Forest model is compared with the previous study conducted by [16], which applied the Support Vector Machine (SVM) algorithm to the same Family Planning dataset.

Table 12. Comparison of the Previous SVM-Based Study and the Proposed Random Forest Model

Item	Previous Study	This Study
Algorithm	SVM	Random Forest
Dataset	1,402 records	1,402 records
Preprocessing	One Hot Encoding	One Hot Encoding
Accuracy	56.20%	59.43%
Feature Importance	No	Yes

As presented in Table 12, the Random Forest model achieved a slightly higher accuracy (59.43%) than the previous SVM model (56.20%). Although the improvement in accuracy was relatively small, the Random Forest model provided additional interpretability through Feature Importance analysis, which was not available in the previous SVM-based model.

Based on the Classification Report, the New category achieved the highest classification performance, with an F1-score of 0.63, indicating that the model was able to recognize the characteristics of new Family Planning participants relatively well. The Repeat category also showed moderate performance, with an F1-score of 0.59, indicating that the available predictor variables were sufficient to distinguish repeated participants in many cases.

However, the Change Method category achieved the lowest classification performance, with an F1-score of 0.11. This result is likely influenced by the substantially smaller number of samples in this category compared with the New and Repeat categories. The limited number of observations reduced the model's ability to learn

representative patterns for the minority class, resulting in lower Recall and F1-score. These findings indicate that class imbalance remained one of the primary challenges affecting the classification performance.

The Confusion Matrix analysis further supports these findings, showing that most classification errors occurred between the New and Repeat categories. This may be attributed to similarities in demographic characteristics and contraceptive usage patterns shared by participants in both categories. Meanwhile, the Change Method category was more difficult to identify because of its limited representation in the dataset.

Consistent with the Feature Importance results presented in Table 11, Wife Birth Year and Wife Age were identified as the two most influential predictor variables. These findings indicate that demographic characteristics play an important role in determining Family Planning participant status. Age-related characteristics may influence reproductive conditions, contraceptive preferences, and participant decisions regarding Family Planning services. Furthermore, contraceptive method variables, including Contraceptive Method_Pill, Contraceptive Method_Injection, and Contraceptive Method_Implant, also contributed to the classification process, indicating that contraceptive utilization patterns provide relevant information for distinguishing participant status categories.

Overall, the Random Forest model successfully identified meaningful patterns within the Family Planning dataset and demonstrated slightly better overall performance than the previous SVM-based approach. Nevertheless, the model still experienced difficulty in correctly classifying the minority Change Method category. Future studies may improve the classification performance by applying data balancing techniques such as the Synthetic Minority Oversampling Technique (SMOTE), optimizing Random Forest hyperparameters, or investigating other ensemble learning approaches to improve predictive performance, particularly for minority classes.

References

- [1] A. K. Mengistu *Et Al.*, “Insights Into Long-Acting Reversible Contraceptive Practices In Sub-Saharan Africa: A Machine Learning Perspective,” *Plos One*, Vol. 21, No. 1, P. E0330960, 2026, Doi: 10.1371/Journal.Pone.0330960.
- [2] M. S. Melaku, L. Yohannes, B. Sharew, M. H. Derseh, And E. A. Taye, “Application Of Machine Learning Algorithms To Model Predictors Of Informed Contraceptive Choice Among Reproductive Age Women In Six High Fertility Rate Sub Sahara Africa Countries,” *Bmc Public Health*, Vol. 25, No. 1, 2025, Doi: 10.1186/S12889-025-23242-W.
- [3] I. Irmawati, H. Hermanto, E. H. Juningsih, S. Rahmatullah, And F. Aziz, “Prediksi Lama Tinggal Pasien Rawat Inap Di Rumah Sakit Pada Masa Pandemi Covid-19 Menggunakan Metode Ensemble Learning Dan Decision Tree,” *J. Inform. Kaputama*, Vol. 5, No. 2, Pp. 391–397, 2021, Doi: 10.59697/Jik.V5i2.276.
- [4] A. Prasetyo, M. M. Effendi, And M. N. Dwi M, “Analisis Gempa Bumi Di Indonesia Dengan Metode Clustering,” *Bull. Inf. Technol.*, Vol. 4, No. 3, Pp. 338–343, 2023, Doi: 10.47065/Bit.V4i3.820.
- [5] I. Putri And A. Razi, “Implementasi Metode Svm Rbf (Radial Basis Function) Kernel Untuk Klasifikasi Status Gizi Ada Balita,” *J. Teknol. Terap. Sains 4.0*, Vol. 5, No. 2, 2024, Doi: 10.29103/Tts.V5i2.19156.
- [6] T. W. W. Mutiah, “Machine Learning Untuk Identifikasi Gizi Balita Menggunakan Algoritma Naïve Bayes,” 2026.
- [7] R. Sholehurrohman And I. Sabda, “Machine Learning Regression Model: Exploring Regression Algorithms For Mercedes-Benz Price Prediction,” Vol. 18, No. 1, Pp. 152–162, 2025.
- [8] A. S. Assiri, S. Nazir, And S. A. Velastin, “Breast Tumor Classification Using An Ensemble Machine Learning Method,” *J. Imaging*, Vol. 6, No. 6, May 2020, Doi: 10.3390/Jimaging6060039.
- [9] L. Khikmah, “Frameworks Comparative Study Of Classification Models Based On Variable Extraction Model For Status Classify Of Contraception Method In Fertile Age Couples In Indonesia,” *Indones. J. Artif. Intell. Data Min.*, Vol. 2, No. 1, Pp. 52–60, 2019, Doi: 10.24014/Ijaidm.V2i1.7568.
- [10] R. Adawiyah, Y. Yahya, And M. Saiful, “Prediksi Tingkat Kesehatan Masyarakat Kecamatan Suralaga Berdasarkan Penggunaan Alat Kontrasepsi Menggunakan Algoritma Random Forest,” *J. Print. J. Pengemb. Rekayasa Inform. Dan Komput.*, Vol. 1, No. 2, Pp. 89–102, 2023, Doi: 10.29408/Jprinter.V1i2.22007.
- [11] A. S. Yahya, Nurhidayati, “Infotek : Jurnal Informatika Dan Teknologi Prediksi Tingkat Kesehatan Masyarakat Berdasarkan Penggunaan Alat Kontrasepsi Menggunakan Algoritma Random Forest Infotek : Jurnal Informatika Dan Teknologi Kecamatan Suralaga Merupakan Kecamatan Yang Berada Di K,” Vol. 7, No. 1, 2024.
- [12] P. R. Sihombing And I. F. Yuliati, “Penerapan Metode Machine Learning Dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah Di Indonesia,” *Matrik J. Manajemen, Tek. Inform. Dan Rekayasa Komput.*, Vol. 20, No. 2, Pp. 417–426, 2021, Doi: 10.30812/Matrik.V20i2.1174.
- [13] P. Handayani And A. Charis Fauzan, “Klik: Kajian Ilmiah Informatika Dan Komputer Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest,” *Media Online*, Vol. 4, No. 6, Pp. 3064–3072, 2024, Doi: 10.30865/Klik.V4i6.1909.
- [14] C. Haryanto, N. Rahaningsih, And F. Muhammad Basysyar, “Komparasi Algoritma Machine Learning Dalam Memprediksi Harga Rumah,” *Jati (Jurnal Mhs. Tek. Inform.)*, Vol.

- 7, No. 1, Pp. 533–539, 2023, Doi: 10.36040/Jati.V7i1.6343.
- [15] L. Breiman, “Rfrsf: Employee Turnover Prediction Based On Random Forests And Survival Analysis,” *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 12343 Lncs, Pp. 503–515, 2020, Doi: 10.1007/978-3-030-62008-0_35.
- [16] A. Rifai, “Classification Of Family Planning Participant Status Using Support Vector Machine (Svm) Based On Age And Type Of Contraceptive,” Vol. 5, No. 1, 2025.
- [17] N. Wuryani And Sarifah Agustiani, “Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis Smote,” *J. Tek. Komput. Amik Bsi*, Vol. 8, No. 2, Pp. 174–180, 2022, Doi: 10.31294/Jtk.V4i2.
- [18] C. Mulia And A. Kurniasih, “Teknik Smote Untuk Mengatasi Imbalance Class Dalam Klasifikasi Bank Customer Churn Menggunakan Algoritma Naïve Bayes Dan Logistic Regression,” *Pros. Semin. Ilm. Nas. Online Mhs. Ilmu Komput. Dan Apl.*, Vol. 0, Pp. 552–559, 2023.
- [19] N. V. Chawla, Kevin W. Bowyer, And Lawrence, “Deep Synthetic Minority Over-Sampling Technique,” Vol. 16, Pp. 321–357, 2020.
- [20] S. Amri, “Information Science And Library Perbandingan Kerangka Model Klasifikasi Untuk Pemilihan Metode Kontrasepsi Dengan Pendekatan Crips-Dm Info Artikel,” *J. Ilm. Univ. Semarang*, Vol. 1, No. 1, Pp. 14–23, 2020.