



## Classification of Family Planning Participant Status Using Support Vector Machine (SVM) Based on Age and Type of Contraceptive

Achmad Rifai<sup>1</sup>, Irmawati<sup>2\*</sup>

<sup>1</sup>Universitas Nusa Mandiri

<sup>2</sup>Universitas Bina Sarana Informatika

[achmad.acf@nusamandiri.ac.id](mailto:achmad.acf@nusamandiri.ac.id)<sup>1</sup>, [irmawati.iat@bsi.ac.id](mailto:irmawati.iat@bsi.ac.id)<sup>2\*</sup>

---

### Abstract

This research investigates the application of the Support Vector Machine (SVM) algorithm in classifying the status of Family Planning (FP) participants based on two primary input features: the wife's age and the type of contraceptive method used. The classification objective includes determining whether a participant falls under the categories of "New," "Repeater," or "Dropout." The dataset used in this study was sourced from the Population and Family Planning Control Service Unit (SATPEL PPKB) of Cilebar District, Karawang Regency. It consists of 1,402 records containing both demographic information and contraceptive usage data. Initial data preprocessing involved label encoding for categorical variables to make the dataset suitable for machine learning algorithms, followed by an 80-20 split for training and testing purposes. The SVM model was trained using the processed dataset, and its performance was evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. The model achieved a classification accuracy of 56.2%. While it demonstrated reasonable effectiveness in identifying the majority class—"New" participants—with relatively high precision and recall, the model's performance declined substantially for minority classes, particularly the "Dropout" group, which was not accurately predicted by the model. This significant discrepancy in classification performance is likely attributed to class imbalance and the limited discriminatory power of using only two input variables. The results indicate that, although SVM is capable of modeling the problem, its performance is constrained by the simplicity of the input features. The model's inability to generalize across all classes equally suggests that additional predictive features—such as education level, number of children, duration of contraceptive use, and socio-economic status—may be necessary to improve overall classification accuracy. Furthermore, addressing the imbalance among class distributions through resampling techniques or cost-sensitive learning may enhance the model's capacity to recognize underrepresented groups. In conclusion, this study serves as a foundational attempt at FP status classification using SVM and highlights the importance of comprehensive feature selection and data balancing strategies for more effective future implementations.

**Keywords:** *Family Planning, Support Vector Machine, Contraceptive Method, Classification, Machine Learning, Participant Status.*

---

### 1. Introduction

One of the Indonesian government's programs that supports population growth is the Family Planning Village Program (KB), an initiative to improve community welfare. The Family Planning (FP) Program is one of the government's strategies to control the population growth rate in Indonesia. In its implementation, FP participants use various types of contraceptives, each with different levels of success and participation. Understanding the status of FP participants—whether active, inactive, or discontinued—is crucial in evaluating the effectiveness of the program [1].

In practice, the implementation of the Family Planning Program is coordinated by the Population and Family Planning Control Agency (DPPKB) at the district or city level. At the sub-district level, the program is carried out by implementing units, one of which is the Population and Family Planning Control Service Unit (SATPEL PPKB) of Cilebar Sub-district, Karawang Regency. SATPEL PPKB's main tasks include collecting data, monitoring, and reporting the use of contraceptives by FP participants in its working area.

The data collected by SATPEL PPKB of Cilebar Sub-district includes various important information such as the identities of the FP participants (husband's name, wife's name, wife's age), participant status (new, repeat, or method switch), types of contraceptives used (IUD, implant, condom, injection, pill), and classification into long-term contraceptive methods (MKJP) and non-long-term methods (Non-MKJP). Each type of contraceptive has a different expiration or effective usage period. Contraceptives such as injections and pills have an expiration period of around 2 to 3 years and fall into the Non-MKJP category, meaning they are short-term and must be used regularly as scheduled. Condoms are also categorized as Non-MKJP with an expiration of approximately 4 years. On the other hand, contraceptives such as implants and IUDs fall into the Long-Term Contraceptive Method (MKJP) category. Implants are effective for around 2 to 3 years,

while IUDs can be used for up to 5 years. Grouping contraceptive types by expiration period and MKJP or Non-MKJP category is important to understand the duration of protection and the frequency of control or replacement required by FP participants.

With advancements in data analysis technology, machine learning approaches can be used to classify FP participant status based on certain characteristics. One effective algorithm for classification is the Support Vector Machine (SVM)[2][3]. This study aims to apply the SVM method to classify FP participant status based on the wife's age and the type of contraceptive used [4][5].

The advantages of Support Vector Machine (SVM) apart from being popular are also very suitable for classification because they do not depend on the number of attributes and can solve problems from dimensions. In terms of calculations, Support Vector Machine (SVM) can provide education in a short time and also its learning techniques are also difficult to doubt[6]. Several previous studies have demonstrated the effectiveness of the Support Vector Machine (SVM) algorithm in classifying health and socio-demographic data, even when using a limited number of features.

First, the study [5] applied the SVM method with RBF and linear kernels to classify stunting cases in infants based on features such as age, gender, height, and nutritional status. The results showed that the RBF kernel achieved the highest accuracy of 95.26%, while the linear kernel reached 78.67%. This study proves that SVM performs optimally in health-related datasets with simple features. Second [7] implemented a linear SVM algorithm to classify stunting status among toddlers. The dataset contained thousands of records with attributes like age, birth weight, and gender. The model achieved an accuracy of 82%, with a precision of 80% and a recall of 86%. These results reinforce the reliability of SVM in handling health data with demographic characteristics. Third [3] utilized polynomial kernel SVM to classify developmental disorders in children, including autism and ADHD. Although the model yielded a relatively lower accuracy of around 73.78%, the study emphasized a key challenge in classification—data imbalance between classes—which is also relevant to the current research.

Lastly [8] compared the performance of SVM and Naive Bayes in classifying poverty levels in Indonesia. Although the domain differs, the study showed that SVM demonstrated better performance in handling multidimensional data and offered more stable prediction results. Based on these studies, it can be concluded that the SVM method has been widely applied across various domains, particularly in social and public health data classification. Therefore, the application of SVM in this research—to classify the status of Family Planning participants based on the wife's age and the type of contraceptive used—is methodologically justified and supported by a strong scientific foundation.

Moreover, SVM has been proven to work effectively even with small and imbalanced datasets, as long as proper pre-processing techniques such as feature encoding and train-test splitting are implemented. Considering that Family Planning data often includes categorical and numerical variables, SVM is suitable due to its flexibility in handling mixed data types. The insights gained from such classification models can support more data-driven decision-making in public health campaigns, contributing to more targeted and effective interventions at the community level.

## 2. Research Method

This section explains the research method used in this study. The dataset used is the Contraceptive Use Dataset. The dataset is then classified using the Support Vector Machine (SVM) method to determine its level of accuracy. The following is the flow of the research methodology which can be seen in Figure 1 below.

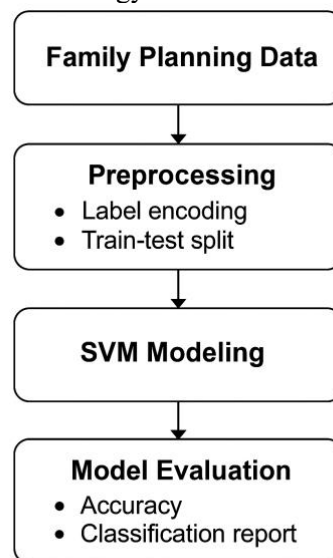


Fig. 1: Research Method

1. Data Collection  
The dataset used contains the husband's name, wife's name, wife's age), participant status (new, repeat, or change method), type of contraceptive used (IUD, implant, condom, injection, pill), and classification of long-term contraceptive methods (MKJP) and non-MKJP obtained from the PPKB SATPEL, Cilebar District.
2. Pre-Processing  
Pre-processing is the stage, the raw dataset is prepared for machine learning modeling. Two key steps were performed:

- a) Label Encoding: Categorical variables, specifically the Type of Contraceptive and FP Participation Status, were converted into numerical values using label encoding. This is essential because machine learning algorithms like SVM require input features in numeric format.
  - b) Train-Test Split: The dataset was divided into two subsets: 80% for training and 20% for testing. The training set is used to build the model, while the testing set is used to evaluate its performance. This ensures that the model can generalize well to unseen data.
3. SVM Modeling  
At this stage, the Support Vector Machine (SVM) algorithm is used to build a classification model of the status of family planning participants based on two main features: wife's age and type of contraceptive.
  4. Model Evaluation  
After the Support Vector Machine (SVM) model is run, it is evaluated using a test dataset to measure its performance. Two main metrics are used in this evaluation: accuracy and classification report.

### 3. Result and Discussion

#### 3.1. Data Collection

The dataset used contains the husband's name, wife's name, wife's age), participant status (new, repeat, or change method), type of contraceptive used (IUD, implant, condom, injection, pill), and classification of long-term contraceptive methods (MKJP) and non-MKJP obtained from the PPKB SATPEL, Cilebar District.

**Table 1:** Data Collection

No.	Wife's Age	Type of Contraceptive (Alkon)	Family Planning Participation Status (Classification Target)
1	39	Pill	Repeater
2	28	Injection	New
3	35	Implant	New
4	26	Implant	New
...	...	...	...
1401	27	IUD	New
1402	38	Condom	New

#### 3.2. Pre-processing

1. Label Encoding: Categorical variables, specifically the Type of Contraceptive and FP Participation Status, were converted into numerical values using label encoding. This is essential because machine learning algorithms like SVM require input features in numeric format.

**Table 2:** Pre-processing results Using Label Encoding

Original Value – Type of Contraceptive	Encoded Value	Original Value – FP Status	Encoded Value
Pill	3	New	1
Injection	2	Repeater	2
Implant	1	Dropout (if any)	0
IUD	0		
Condom	4		

2. Train-Test Split: The dataset was divided into two subsets: 80% for training and 20% for testing. The training set is used to build the model, while the testing set is used to evaluate its performance. This ensures that the model can generalize well to unseen data.

**Table 3:** Split Data

Data Type	Number of Records	Percentage (%)
Training Data	1,121	80%

Test Data	281	20%
<b>Total</b>	<b>1,402</b>	<b>100%</b>

### 3.3. SVM Modeling

In this stage, the Support Vector Machine (SVM) algorithm was implemented to build a classification model that predicts the status of Family Planning (FP) participants. The primary objective was to categorize each participant as either *New*, *Repeater*, or *Dropout*, based solely on two input features: the wife's age and the type of contraceptive used. These features were chosen because they are commonly recorded and accessible in FP service reports, especially at the sub-district level.

Before model training, the categorical variable representing the type of contraceptive (e.g., pill, injection, IUD, implant, condom) was converted into numerical format using label encoding. The participant status variable, which served as the classification target, was also encoded accordingly: "Dropout" as 0, "New" as 1, and "Repeater" as 2. These transformations allowed the dataset to be compatible with the requirements of machine learning algorithms, including SVM.

The SVM model was constructed using the Scikit-learn implementation of the Support Vector Classifier (SVC). For this study, the linear kernel was selected, and the model was trained using the default hyperparameter settings without additional tuning. This decision was made to establish a baseline understanding of the classification performance using minimal and straightforward input features. The dataset, which comprised 1,402 records, was split into 80% for training (1,121 records) and 20% for testing (281 records), ensuring a standard separation between model building and evaluation.

After training the model on the training set, predictions were generated on the test set. The classification results indicated that the model performed reasonably well in identifying the "New" class but struggled significantly with the "Dropout" class, which had the fewest samples in the dataset. This suggests that the model may have been biased toward the majority class during training, a common issue in imbalanced classification problems. Despite its limitations, the SVM model provided a foundational classification performance that could be further improved with additional features and data preprocessing techniques.

Overall, this modeling step served as a baseline benchmark to assess the viability of using SVM for classifying FP participant statuses and revealed critical insights regarding the limitations of using only age and contraceptive type as predictive features.

### 3.4. Model Evaluation

This study applied the Support Vector Machine (SVM) algorithm to classify the status of Family Planning (FP) participants based on the wife's age and type of contraceptive used. The model was trained and evaluated on a dataset collected from SATPEL PPKB, Cilebar District. The classification model achieved an accuracy of **56.2%**, indicating that while the method is applicable, the current input features are not strong predictors of FP status on their own.

**Table 4:** Evaluation Results

Class Label	Precision	Recall	F1-Score
0 (New)	0.58	0.72	0.64
1 (Dropout)	0.00	0.00	0.00
2 (Repeater)	0.53	0.45	0.48
<b>Accuracy</b>			<b>0.562</b>

The classification results show that the Support Vector Machine (SVM) model achieved an accuracy of 56.2% when tested on the unseen dataset. The model performed relatively well in classifying the "New" participants, with a precision of 0.58, recall of 0.72, and F1-score of 0.64. For the "Repeater" class, the model also showed moderate performance, with a precision of 0.53 and F1-score of 0.48. However, the model completely failed to classify the "Dropout" class, as indicated by precision, recall, and F1-score values of 0.00. This is likely due to the small number of dropout cases in the dataset (only 16 samples), which caused the model to underfit this minority class.

The macro average scores—Precision: 0.37, Recall: 0.39, and F1-Score: 0.38—reflect the average performance across all classes without considering class distribution. On the other hand, the weighted average scores—Precision: 0.56, Recall: 0.56, and F1-Score: 0.54—are influenced by the class sizes and emphasize the model's bias toward the majority classes.

Overall, while the model demonstrates acceptable performance for dominant classes, it struggles significantly with minority class classification. This suggests a need for class balancing techniques or additional feature inputs to improve predictive accuracy and fairness across all participant statuses.

**Table 5:** Confusion Matrix

Actual \ Predicted	0 (New)	1 (Dropout)	2 (Repeater)
0 (New)	105	0	41
1 (Dropout)	10	0	6
2 (Repeater)	66	0	53

The confusion matrix illustrates the performance of the Support Vector Machine (SVM) model in classifying the status of Family Planning participants. From the matrix, it can be seen that among the actual "New" participants, 105 were correctly classified as "New", while 41 were incorrectly predicted as "Repeater". For the "Dropout" category, the model failed to correctly classify any instances. Instead, 10 participants were predicted as "New" and 6 as "Repeater", indicating a complete inability of the model to identify this class. In the "Repeater" category, 53 participants were correctly classified, while 66 were misclassified as "New". This pattern of misclassification reveals a strong model bias toward the majority class, which in this case is "New". The absence of correct predictions for the "Dropout" class also reflects the model's poor generalization for minority classes, likely due to the imbalance in the dataset. These results emphasize the importance of improving data representation and considering class balancing techniques to enhance the model's predictive performance.

## 4. Conclusion

This study applies the Support Vector Machine (SVM) algorithm to classify the status of Family Planning (KB) participants based on the wife's age and the type of contraceptive used. The model was built based on data obtained from the SATPEL PPKB Cilebar District and produced an accuracy of 56.2%. These results indicate that the SVM method can be used to classify the status of KB participants, but its performance is still limited if only using these two features. The findings of this study can be summarized as follows:

1. Feasibility of SVM: The Support Vector Machine (SVM) algorithm demonstrates potential in classifying KB participant status, particularly when simple and structured input features are provided.
2. Performance Limitation: The model's accuracy of 56.2% reflects that using only two features—wife's age and contraceptive type—is not sufficient to achieve high classification performance, especially in distinguishing minority classes.
3. Recommendation for Improvement: To enhance the model's effectiveness, future studies should consider incorporating additional features (e.g., duration of contraceptive use, education level, or number of children) and address class imbalance to improve minority class predictions.

## References

- [1] M. R. Ladiku, Z. Abdussamad, And F. P. Tui, "Efektivitas Program Kampung Kb Di Kecamatan Kwandang Kabupaten Gorontalo," *J. Ilmu Pemerintah. Dan Adm. Publik*, Vol. 2, Pp. 862–881, 2024.
- [2] P. Arsi And R. Waluyo, "Sentiment Analysis Of Discourse On Moving The Indonesian Capital City Using The Support Vector Machine (Svm) Algorithm," *J. Teknol. Inf. Dan Ilmu Komput.*, Vol. 8, No. 1, P. 147, 2021, Doi: 10.25126/Jtiik.202183944.
- [3] I. Monika Parapat And M. Tanzil Furqon, "Penerapan Metode Support Vector Machine (Svm) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 2, No. 10, Pp. 3163–3169, 2018, [Online]. Available: [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)
- [4] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, And R. Rahmaddeni, "Perbandingan Kinerja Naïve Bayes Dan Svm Pada Analisis Sentimen Twitter Ibukota Nusantara," *J. Ilm. Inform.*, Vol. 12, No. 01, Pp. 15–21, 2024, Doi: 10.33884/Jif.V12i01.8721.
- [5] T. R. Dewanti, R. Prathivi, And S. Susanto, "Implementasi Metode Svm Untuk Klasifikasi Penyakit Stunting Bayi," *J. Inform. Teknol. Dan Sains*, Vol. 7, No. 1, Pp. 101–106, 2025, Doi: 10.51401/Jinteks.V7i1.5070.
- [6] Irmawati, Y. Alkhalifi, A. Fazriansyah, M. Syamsul Azis, And K. Widiyanto, "Deep Learning Untuk Pendeteksian Penyakit Kanker Payudara Dengan Optimasi Adam," *J. Inf. Syst. Applied, Manag. Account. Res.*, Vol. 7, No. 1, Pp. 124–136, 2023, Doi: 10.52362/Jisamar.V7i1.1015.
- [7] A. Jalil, A. Homaidi, And Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, Vol. 8, No. 3, Pp. 2070–2079, 2024, Doi: 10.33379/Gtech.V8i3.4811.
- [8] Z. A. Mukharyahya, Y. P. Astuti, And O. N. Cahyani, "Perbandingan Naive Bayes Dan Support Vector Machine Dalam Klasifikasi Tingkat Kemiskinan Di Indonesia," *Edumatic J. Pendidik. Inform.*, Vol. 9, No. 1, Pp. 119–128, 2025, Doi: 10.29408/Edumatic.V9i1.29512.