American Journal of Nursing and Health Sciences
2024; 5(1): 1-8
http://www.sciencepublishinggroup.com/j/ajnhs
doi: 10.11648/j.ajnhs.20240501.11
ISSN: 2994-7227 (Online)

Recursive Feature Elimination with Naive Bayes Classification of Modern Contraception in Reproductive-Aged Women in Kenya

Levi Kiplang'at Bor*, Anthony Wanjoya, Samuel Mwalili, Dennis Kirui

Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

Email address:

leviybor40@gmail.com (Levi Kiplang'at Bor), awanjoya@gmail.com (Anthony Wanjoya), samuel.mwalili@gmail.com (Samuel Mwalili), dkirui54@gmail.com (Dennis Kirui)

*Corresponding author

To cite this article:

Levi Kiplang'at Bor, Anthony Wanjoya, Samuel Mwalili, Dennis Kirui. (2024). Recursive Feature Elimination with Naive Bayes Classification of Modern Contraception in Reproductive-Aged Women in Kenya. *American Journal of Nursing and Health Sciences*, *5*(1), 1-8. https://doi.org/10.11648/j.ajnhs.20240501.11

Received: September 30, 2023; **Accepted**: October 10, 2023; **Published**: January 8, 2024

Abstract: Family planning gives the population a license to have control over its reproductive health and ultimately family size. A better understanding, therefore, of its determinants to its uptake is a necessity. The project embarked on determining these factors. It was observed that parity, marital status, age, residence, general health of an individual, education level, wealth status, and family planning awareness are significant factors that determine modern contraception. The number of children one has or is planning to have greatly impacted the use of the different modes of contraception. This research's main objective was to formulate and implement a cross-validated RFE-NB classifier on modern contraceptive data and compare its performance to that of RFE-SVM. A recursive feature elimination technique trained on the data and important features responsible for modern contraception identified. The naive Bayes classifier was then used for classification. The data was also used to train an RBF kernel SVM classifier. A comparative analysis was then done on the two models. Considering the findings, we conclude that the RFE-NB model has a relatively high accuracy of 81%, which, however, is lower when compared to that of RFE-SVM. The high Kappa value further underscores its reliability in distinguishing between different classes. The RFE-NB exhibits strong accuracy, sensitivity, and specificity, making it a valuable tool for accurate prediction and classification tasks.

Keywords: Modern Contraception, Childbearing Women, Recursive Feature Elimination, Naïve Bayes, Support Vector Machine

1. Introduction

Family planning is a critical tool for it allows individuals and couples to have control over their reproductive health and ultimately plan the family they desire. With the help of family planning, individuals and couples can decide the number of children they want and the spacing between pregnancies. By reviewing the pros and cons of each method, individuals then make an informed decision on the best approach to use. They include injectables, sterilization, implants, intrauterine devices (IUD), hormonal methods such as pills, and condoms.

It's not news that the world's population has grown at a rapid rate in recent years occasioned by high fertility rates, particularly in sub-Saharan Africa [8]. This is largely due to improved access to health care, better nutrition, and improved economic conditions, which have led to higher fertility rates and higher life expectancy. This increase in population has had a major impact on the region, with an increased demand for scarce resources such as food, water, and energy. The total fertility rate (TFR) has decreased in the majority of sub-Saharan countries. TFR, however, is still stagnant in certain nations. The nations that experienced the sharpest declines in overall fertility were Zimbabwe, Liberia, Namibia, Kenya, Senegal, Togo, Madagascar, and Ghana. Niger, Mozambique, Nigeria, and other nations exhibit minimal change or growth [1]. Kenya targeted to increase the modern CPR to 58% in 2020 and exceeded this remarkably, showing progress towards achieving universal access to FP. For such trends to continue, there is a need for a better understanding of the predictors or determinants of modern contraceptive uptake.

Machine learning (ML) is a quickly expanding field in artificial intelligence with the potential to revolutionize many areas of science and medicine [3]. For instance, a 2018 poll by Deloitte indicated that 63% of the businesses surveyed were using machine learning in their operations, out of the 1,100 US managers whose organizations were already exploring AI [5]. Machine Learning programs are designed to adapt to new information and improve their performance over time. Unlike traditional software programs that require explicit programming, ML algorithms can learn from data and make adjustments to their models accordingly.

In machine learning applications, feature selection is crucial. The principal advantage of feature selection for classification usually lies in the obliteration of overfitting problems to enhance prediction performance and improve interpretability. The Recursive Feature Elimination is a potent method for feature selection that was very recently devised. It iteratively assesses feature importance and eliminates the least significant ones based on a chosen ML algorithm and has proven relevant, especially in high dimensional data scenarios, and thus increases generalization performance. Additionally, RFE and Naive Bayes (NB), have been shown to generalize effectively for small sample classification [7]. NB classifier, a learning algorithm that utilizes Bayes' rule boasts of its computational efficiency and direct prediction of its posterior probabilities [11]. According to studies comparing classification algorithms, the NB classifier performs on par with decision trees and some neural network classifiers in terms of accuracy. Additionally, they have proven to be quick in computational time and accurate when used with enormous databases [4]. The ability to perform well even in instances when there are missing values and its' insensitivity to noise, makes NB the most commonly used ML classifier. Such advantages if harnessed can result in better exploitation of classification properties in the health fields.

In a study conducted to classify diabetes [9], a support vector machine algorithm was used to select the optimal features for prediction. Fourfold Cross-validation was applied to fairly evaluate the models. A hybrid multi-layer perceptron was then used and compared with other algorithms. The hybrid particle swarm optimization-MLP outperformed the models with an accuracy of 97%. Results also showed that when a 4-fold cross-validation framework and a 2-fold cross-validation are used an NB outshines an MLP-neural network, a Gaussian support vector machine, KNN, and a decision tree in both cases with an accuracy of 96.7% on 2-fold cross-validation and 96.81% on 4-fold cross-validation. A random forest-recursive feature elimination technique, using high dimension and integrated omics data, was used to mitigate the effects of correlated factors on the random forest significance score [2]. It was discovered that with low correlated variables, the RF-RFE performed better than a random forest when comparing importance score rankings from the two methods. However, RF-RFE was not an appropriate technique when there are numerous highly correlated features present in high-dimensional omics.

The Naive Bayes algorithm, which has been proven to be the best classifier in classifications involving binary response, is known for its simplicity and efficiency but can suffer from suboptimal performance when provided with irrelevant or redundant features. This research intends to enhance the accuracy and efficiency of a Naive Bayes classifier by employing Recursive Feature Elimination. Often applied in the context in which we have a lot of predictor variables that can be deployed to predict a particular outcome of interest while working with big data, they generate parsimonious models that have just enough significant predictors for the outcome of interest. There is a need to understand the models' performance in instances with fewer predictor variables.

2. Methods

2.1. Data Source

The study utilized secondary data obtained from the Performance Monitoring for Action (PMA) project, funded by the Bill and Melinda Gates Foundation, which conducted surveys between November 2021 and January 2022. The data collection followed a multi-stage cluster sampling design, involving both urban and rural stratification and covering eleven counties in Kenya, including Bungoma, Nairobi, West Pokot, Kericho, Kilifi, Siaya, Kiambu, Nyamira, Nandi, Kakamega, and Kitui. The enumeration areas were selected using the probability proportional to size (PPS) approach, with each county containing 25 enumeration areas, each consisting of at least 35 dwelling units. The study focused on reproductive-aged women using contemporary contraception within these enumeration areas.

2.2. Recursive Feature Elimination

For models to perform better and prevent overfitting, feature selection is crucial. RFE systematically removes less significant features through iteration, employing an algorithm and an importance-ranking metric. The algorithm's layout is as displayed in Figure 1 This process usually aims to pinpoint the most pertinent set of features. For some measures, the relative importance of each feature can change significantly when assessed over a new sample of data during the step-wise elimination process (especially for highly correlated features), necessitating the recursion. A final ranking is created using the features' inverse elimination order. The feature selection process itself consists only of taking the first n features from this ranking.

Algorithm 2: Recursive feature elimination incorporating resampling

```
for Each Resampling Iteration do
    Partition data into training and test/hold–back set via resampling
    Tune/train the model on the training set using all predictors
    Predict the held–back samples
    Calculate variable importance or rankings
    for Each subset size S_i, i = 1 ... S do
        Keep the S_i most important variables
        [Optional] Pre–process the data
        Tune/train the model on the training set using S_i predictors
        Predict the held–back samples
        [Optional] Recalculate the rankings for each predictor
    end
end
Calculate the performance profile over the S_i using the held–back samples
Determine the appropriate number of predictors
Estimate the final list of predictors to keep in the final model
Fit the final model based on the optimal S_i using the original training set
```

Figure 1. *Recursive feature elimination algorithm.*

2.3. Naive Bayes

An NB is a Bayesian classifier that operates on the premise that the impact of one attribute's value on a particular class is unrelated to the effects of the other attributes' values. NB delivers competitive classification accuracy despite its independence assumption being often violated in practice [11]. It is widely applied in practice thanks to its computational efficiency as well as many other desirable features such as its accuracy, which is measured as the class-conditional information that exists between the features, and does not correlate directly with the degree of feature dependencies [10].

The class Cm for a given data point $x1, \dots, xn$ with n features is predicted using the probability

$$p(Cm|x) = p(Cm|x1, \dots, xn)\ for\ m = 1, \dots, m$$

Applying Bayes' Theorem to factor this as;

$$p(C_m|\vec{x}) = \frac{p(\vec{x}|C_m)p(C_m)}{p(\vec{x})} = \frac{p(x_1,\dots,x_n|C_m)p(C_m)}{p(x_1,\dots,x_n)}$$

The factor $p(x\_1, \dots, x\_n|C\_m)$ in the numerator can be subsequently broken down using the chain rule:

$$p(x_1,\dots,x_n|C_m) = p(x_1|x_2,\dots x_n, C_m)p(x_2|x_3,\dots,x_n,C_m)\dots p(x_{n-1}|x_n,C_m)p(x_n|C_m)$$

Given the class C_m, naïve Bayes models assume that feature x_i is independent of feature x_j for $i \neq j$. This can be stated using the previous decomposition as;

$$p(x_i|x_{i+1},\dots,x_n|C_m) = p(x_i|C_m) \Longrightarrow p(x_1,\dots,x_n|C_m) = \prod_{i=1}^{n} p\ (x_i|C_m)$$

Thus,

$$p(C_m|x_1,\dots,x_n) \propto p(C_m, x_1,\dots,x_n)$$

$$\propto p(C_m)p(x_1,\dots,x_n|C_m)$$

$$\propto p(C_m)\prod_{i=1}^{n} p\ (x_i|C_m)$$

The class C with the highest probability given the predictors is identified by maximizing.

$$\hat{C} = \max_{m\in\{1,\dots,n\}} p(C_m)\prod_{i=1}^{n} p\ (x_i|C_m)$$

2.4. Support Vector Machine

SVMs are based on statistical learning frameworks and are among the most reliable prediction techniques. SVMs are a technique for classifying linear and nonlinear data. The dataset is mapped into a higher dimension using nonlinear mapping.

Choosing the best hyperplane in the element space that maximum separates the two target classes is the core tenet of this rule. We must resolve the form's minimization problem in order to find it;

$$Minimize_{w,b}\frac{1}{2}(//w//^2)$$

subject to:

$$y_j(w^T x_j + b) \geq 1, j = 1,2,3,\cdots,n$$

The linear hyperplane for a binary classification problem having the input feature vectors X and the corresponding class labels C has the following notation:

$$w^T x + b = 0$$

where w = the normal vector to the hyperplane

b = distance of the hyperplane from the origin along the normal vector w.

To determine how far a data point x_i is from the decision boundary;

$$d_i = \frac{w^T x_i + b}{//w//}$$

where //w// represents the Euclidean norm of the weight vector w.

The linear SVM classifier is then;

$$\hat{y} = \begin{cases} 1 & : w^T x + b \geq 0 \\ 0 & : w^T x + b < 0 \end{cases}$$

2.5. Performance Metrices

Cohen's kappa, which is usually symbolized by k is a robust statistic useful for either interrater or intra-rater reliability testing- degree of agreement among repeated administrations of a diagnostic test performed by a single rater [6]. Sensitivity and specificity are also used to evaluate the tests and results' reliability. Sensitivity (recall, true positive rate, or probability of detection) is described as the test's ability to recognize positive results as such. The ability of a test to recognize negative results as such, on the other hand, is referred to as specificity.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Specificity = \frac{TN}{TN+FP}$$

$$Sensitivity = \frac{TP}{TP+FN}$$

$$Kappa = \frac{p_o - p_e}{1 - p_e}$$

Where;
TP = True Positives
TN = True Negatives
P_o= (TP + TN) / Total
FP = False Positives
FN = False Negatives

P_e= [(TP + FP) * (TP + FN) + (FP + TN) * (FN + TN)] / (Total^2)

3. Results

3.1. Feature Selection

The study that consisted of 50,938 respondents. The females were 23,441 with 10,960 falling into our desired category. To understand the use of modern contraceptives, we targeted females aged 15 to 49 years (childbearing age). Pregnant women and those who were in the process of conceiving were left out of our study. After considering those who used modern contraception and relevant data-cleaning practices, a sample of 8291 was obtained. The dataset consisted of 10 categorical variables.

We used Recursive Feature Selection (RFE) to identify optimal features crucial to building a parsimonious model for modern contraceptive classification and prediction. The control options were configured for $rfeControl$ and utilized the $rfFuncs$ function, known for its robust feature importance computation. A repetitive (5 times) 10-fold cross-validation approach was used so as to improve the performance of the feature selection model. The output was as shown on figure 2.

```
Recursive feature selection

Outer resampling method: Cross-Validated (10 fold, repeated 5 times)

Resampling performance over subset size:

 Variables Accuracy  Kappa AccuracySD KappaSD Selected
         1   0.7021 0.4000    0.01643 0.03333
         2   0.7081 0.4127    0.01570 0.03171
         3   0.7082 0.4135    0.01603 0.03239
         4   0.7216 0.4412    0.01652 0.03323
         5   0.7180 0.4342    0.01694 0.03402
         6   0.7194 0.4369    0.01656 0.03332
         7   0.7197 0.4374    0.01652 0.03318
         8   0.7219 0.4416    0.01632 0.03284        *
         9   0.7157 0.4296    0.01516 0.03043

The top 5 variables (out of 8):
   Parity, Marital_Status, Age, EducationLevel, Wealthstatus
```

Figure 2. *Feature selection by RFE.*

Recursive Feature Elimination (RFE) determined that retaining 8 features namely, parity, marital status, age, education level, wealth status, general health, residence, and family planning knowledge in the model is optimal. This decision was supported by the fact that both Kappa and accuracy metrics reached their maximum values when using this subset of features. The visualization of accuracy and kappa is as shown in figures 3 and 4 respectively.

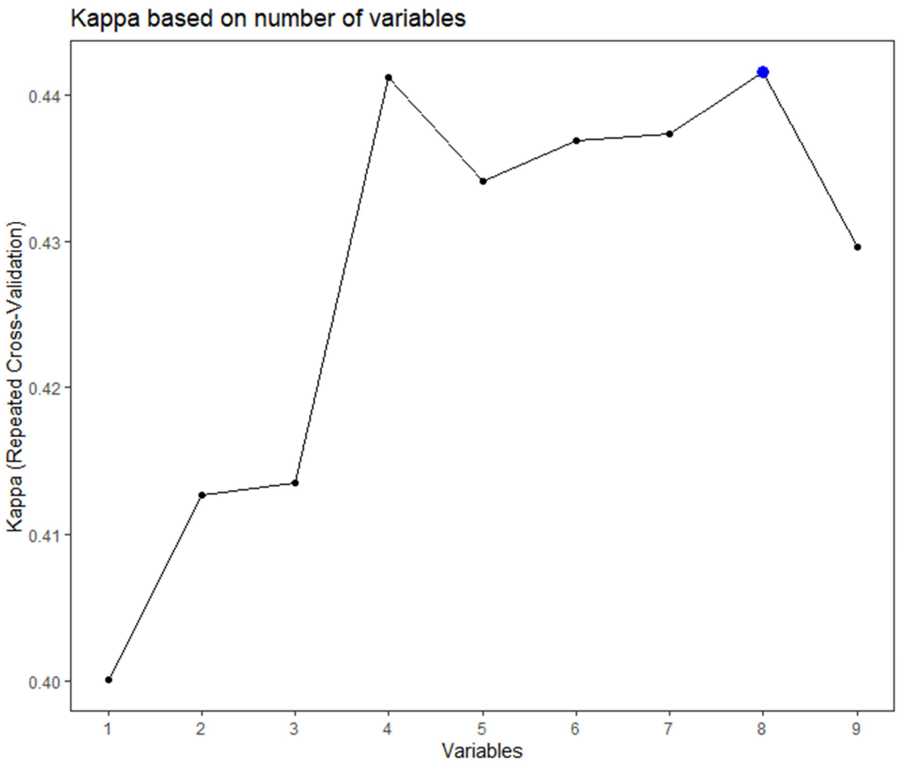

Figure 3. *Kappa of the model based on the number of features.*

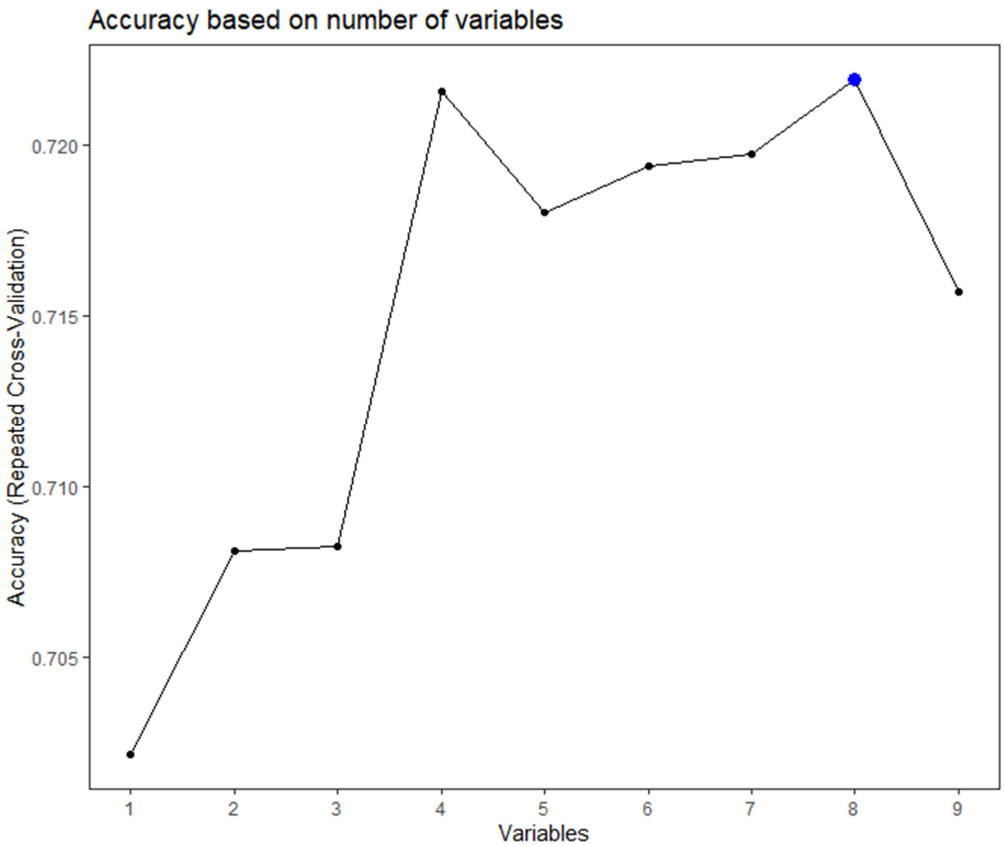

Figure 4. *Accuracy of the model based on the number of features.*

The optimal levels reached is at the blue point. This is where the accuracy and kappa are maximized.

3.2. Variable Importance

Evaluating the variable importance of the selected features allows us to identify the features that have the most significant impact on modern contraception. High importance scores indicated a strong influence on the model's predictions, signifying their role as the most informative factors in the classification and prediction process.

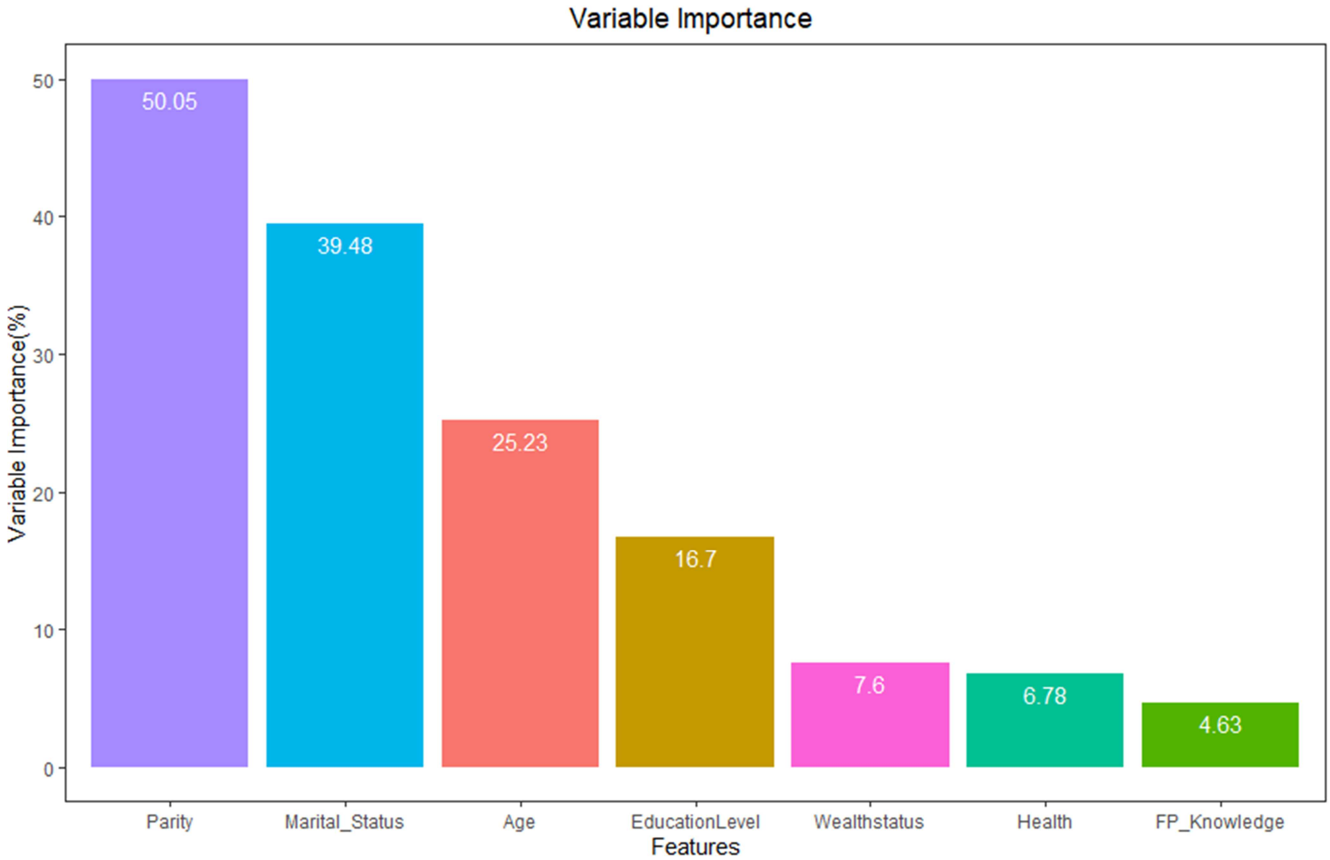

Figure 5. *Variable importance scores.*

Parity stood out as the most critical variable, boasting an importance score of 50.05% making it twice as influential as age, 25.23%. Marital status closely followed with an accrued importance of 39.48%. The relative significance of marital status was 39.47%, whereas the importance of wealth status and level of education was 16.7% and 7.6%, respectively. Despite the fact that health status is a key factor in determining modern contraception, its influence was just 6.78%. Awareness of family planning (4.63%) was the least significant feature.

It is observed that the metrics derived from both the training and testing datasets exhibited a remarkable similarity as shown in table 1, indicating the robustness and reliability of the RFE model.

Table 1. *Table of comparison for RFE performance metrices during training and testing.*

Table on accuracy and Kappa		
	Accuracy	**Kappa**
Training set	0.7219	0.4416
Testing set	0.71997	0.4374

3.3. Classification

Based on the features selected as determinants of modern contraception we developed and trained the RFE-NB classifier. To perform classification, the trained model was fitted to the testing dataset. For each data point, the model computed the likelihood of each class and assigned the class with the highest probability as the predicted class. This is summarized as a confusion matrix in figure 6.

There were 702 true positive predicted values and 643 true negative predicted values and with the off-diagonal values being greater than the leading diagonal (FP, FN), the model appeared to be more accurate in predictions. The RFE-NB model demonstrates robust performance with an accuracy of approximately 81.17%, reflecting a strong overall agreement between predictions and actual outcomes. The model performed highly in predicting modern contraceptive use as sensitivity was at 78.99%. Specificity, 83.27%, signifies that the model performs better predicting the modern contraceptive non-use. The model therefore can be used to predict modern contraceptive discontinuation.

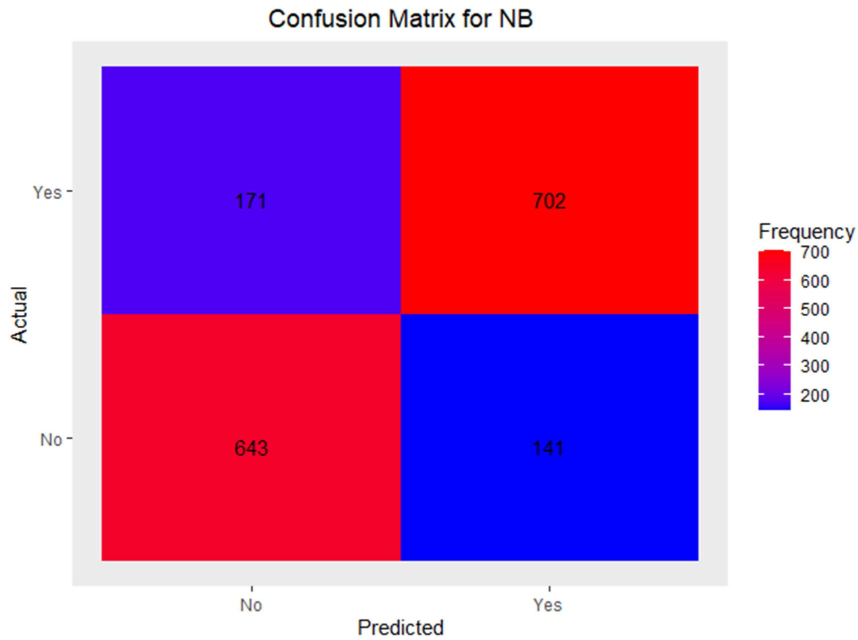

Figure 6. *Confusion matrix for RFE-NB.*

The confusion matrix generated by RFE-SVM (figure 7) classified 769 as TP and 654 TN. Similarly, the leading diagonal was way less than the off-diagonal implying the model's accuracy in predictions. An accuracy score of 0.838 indicates that the model correctly predicted the class of 81.17% of the samples in the dataset, which is a high overall correctness rate. The model was highly dependable when it came to recognizing instances of contraception use due to its high sensitivity, which was roughly 80.34% indicating the model's capacity to recognize positive cases. On average, 86.99% of contraception non-use were correctly classified by the model. Negative case misclassification was minimized with such great specificity.

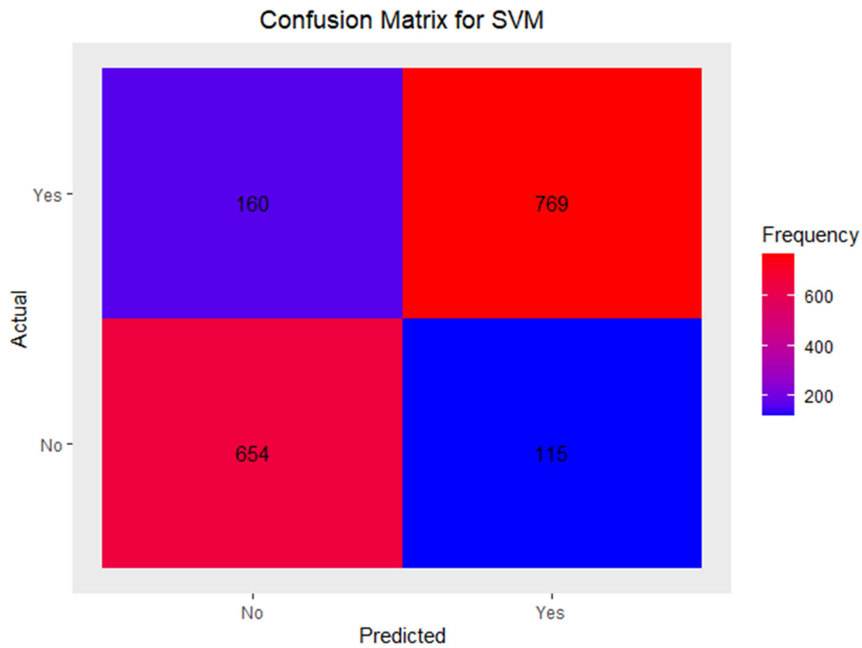

Figure 7. *Confusion matrix for RFE-SVM.*

3.4. Comparing Performances of the Two Models

Both models were reliable in classifying modern contraceptives. The RFE-SVM was however substantial in its inter-reliability measure score. With a higher accuracy score, the RFE-SVM proved to be more fit in classification. In spite of the RFE-SVM having better performance metric scores, the departure between these two models was minimal. The comparison is as in table 2.

Table 2. *Table of comparison between RFE-SVM and RFE-NB.*

Performance Metrices						
	Accuracy	**Sensitivity**	**Specificity**	**F_1 score**	**Precision**	**Kappa**
RFE-NB	0.8117	0.7899	0.8327	0.8048	0.8202	0.6231
RFE-SVM	0.838	0.8034	0.8699	0.8262	0.8263	0.6748

The RFE-SVM outperforms the RFE-NB in all performance metrices. However, differences in precision are almost negligible implying that the two models demonstrate a good trade-off between precision and recall, indicating their overall effective performances.

4. Discussions

In this research study, Recursive Feature Elimination was applied to identify the most relevant features for predicting modern contraceptive usage. Eight significant features, including age, marital status, residence, education level, wealth status, parity, family planning awareness, and the number of births were found to be the most important features and were thus selected for training a Naive Bayes classification model. The results showed promising accuracy, with a classification accuracy of 0.8117 (0.792, 0.8303), indicating the model's ability to make accurate predictions. The model had a sensitivity score of 0.7899 as the model's effectiveness in correctly identifying positive instances of modern contraceptive use, while specificity was at 0.8327, highlighting its ability to correctly classify negative instances. These findings suggested the potential of this approach for aiding decisions relating to family planning. The performance of the comparative model, RFE-SVM with an RBF kernel, results showed an accuracy of 0.838 (0.8196, 0.8553). The sensitivity was 0.8034 while the specificity was 0.8699. The RFE-SVM model performed better than the RFE-NB model. It was noted, however, that the RFE-NB model had an advantage over the RFE-SVM as the computational time here was low. It still achieved a higher accuracy even with its assumption that the features are independent.

5. Conclusion

This research's main objective was to formulate and implement a cross-validated RFE-NB classifier on modern contraceptive data and compare its performance to that of RFE-SVM. The findings of stratified cross-validation with 10 folds and 5 repetitions RFENB showed that the optimal number of features was 8. It was also noted that the determinants of modern contraception were age, marital status, residence, education level, wealth status, parity, and family planning awareness. The number of births a woman has was observed to be the most important determinant of modern contraception. There was a consensus that awareness of family planning has an influence on contraception uptake. The F1, precision, and recall all indicated that the binary classification by the RFE-NB algorithm works well. When comparisons were made with RFE-SVM, the two models performed well in classifications with almost similar output. The classification model exhibits strong accuracy, sensitivity, and specificity, making it a valuable tool for accurate prediction and classification tasks. The high Kappa value further underscores its reliability in distinguishing between different classes.

References

[1] Cleland, J. G., Ndugwa, R. P., & Zulu, E. M. (2011). Family planning in subsaharan africa: Progress or stagnation? *Bulletin of the World Health Organization*, *89* (2), 137–143.

[2] Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC genetics*, *19* (1), 1–6.

[3] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, *6* (2), 94.

[4] Leung, K. M., et al. (2007). Naive bayesian classifier. *Polytechnic University Department of Computer Science/Finance and Risk Engineering*, *2007*, 123– 156.

[5] Loucks, J., Davenport, T., & Schatsky, D. (2018). State of ai in the enterprise. *Deloitte Insights Report*.

[6] Marston, L. (2010). *Introductory statistics for health and nursing using spss*. Sage Publications.

[7] Nasien, D., Yuhaniz, S. S., & Haron, H. (2010). Statistical learning theory and support vector machines. *2010 Second International Conference on Computer Research and Development*, 760–764.

[8] Negash, W. D., Eshetu, H. B., & Asmamaw, D. B. (2022). Predictors of modern contraceptive use among reproductive age women in high fertility countries in sub-saharan africa: Evidence from demographic and health surveys. *BMC Women's Health*, *22* (1), 1–10.

[9] Qteat, H., & Awad, M. (2021). Using hybrid model of particle swarm optimization and multi-layer perceptron neural networks for classification of diabete. *Int. J. Intell. Eng. Syst*, *14* (3), 11–22.

[10] Rish, I., et al. (2001). An empirical study of the naive bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, *3* (22), 41–46.

[11] Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naıve bayes. *Encyclopedia of machine learning*, *15*, 713–714.