ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

DEVELOPMENT OF ENSEMBLE PREDICTIVE MODELS FOR CONTRACEPTIVE UPTAKE IN WOMEN

[1]Joda Shade Christiana, [2]Prof. O.O. Obe, [3]Prof. O.K. Boyinbode, [4]Prof B.N. Olagbuji

Department of Computer Science, Federal College of Agriculture Akure, Ondo State,

Department of Computer Science, Federal University of Technology, Akure, Ondo State, Department of Obsterics and Gynaecology college of Medicine Ekiti State University, Ado Ekiti

DOI: https://doi.org/10.5281/zenodo.14564997

Published Date: 28-December-2024

Abstract: **The purpose of this study is to improve the prediction of contraceptive uptake with the use of the stacked ensemble model. The main objectives include the identification of the relevant predictors associated with that uptake, development of ensemble designs for improving the accuracy and robustness of prediction, the implementation involved in these models, and performance evaluation. The proposed ensemble model encompasses four MLAs: Support Vector Machine, Random Forest, Decision Tree, and KNN, respectively, while Logistic Regression will be the meta learner. Due care was taken for the collection and preprocessing of data by discretization, normalization, and cleansing. Next, feature selection was done based on Information Gain and Chi-Square methods in order to focus only on the most relevant predictors. The ensemble model was implemented in Python and the performance of the model evaluated using Accuracy, Precision, Recall, and F1 score. The performance was able to give an accuracy of 0.8187, precision of 0.8223, recall of 0.8187, and F1 score of 0.8198-in effect, a balance in all the metrics. The ensemble model was much more robust and reliable compared to the single models, hence a tool in the prediction of uptake of contraceptives. Overall, the performance of the ensemble model suggests that sophisticated machine learning algorithms will significantly support healthcare providers with informed contraceptive recommendations to better the outcomes of family planning. Future research may investigate other predictors and longitudinal data further to improve model performance and generalizability across diverse healthcare domains.**

Keywords: **Contraceptive uptake, ensemble learning, machine learning, prediction model, stacked ensemble, Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, feature selection, accuracy, precision, recall, F1 score, healthcare data, family planning.**

1. INTRODUCTION

Unwanted pregnancies are among the most serious public health issues of our time. Contraceptives, whether through pills, implants, or IUDs, not only avert unintended pregnancy but also play a vital role in family planning and women's self-determination. Women's adoption of contraceptives is crucial for reproductive health and the best possible outcomes of family planning. However, this decision to use contraceptives is influenced by an interplay of individual knowledge, attitudes, access, societal norms, and cultural beliefs. Such diverse influencing factors make the prediction of contraceptive uptake particularly difficult. Conventionally, healthcare providers relied on consultations with their patients and some limited data in order to make recommendations. However, recent developments in artificial intelligence offer promising avenues for more accurate and data-driven predictions. Machine learning, a subfield of AI, offers a set of powerful tools for analyzing large amounts of healthcare data and finding patterns that relate to contraceptive uptake.

2. LITERATURE REVIEW

The integration of AI in medical practice has been driven by the need for accurate prediction of recommended contraceptives. AI utilizes advanced algorithms in drawing from the vast datasets of healthcare insights that help improve clinical decision-making. Machine learning techniques, and especially ensemble predictive models, have recently emerged as promising tools in tackling complexities in predicting contraceptive uptake. Though a host of studies have been undertaken on the factors that influence contraceptive use, works done in contraceptive prevalence using machine learning algorithms are still at their evolving stages: Iqramul Haq et al. (2022), Rosenfield (2012), and Ajayi et al. (2018).

Although much research has explored the domain of machine learning for the prediction of contraceptive uptake, including works by Iqramul Haq et al. (2022), Tesfaya et al. (2019), Kebede et al. (2023), and Yu-En (2022), none of those works have applied ensemble models to this task. An important opportunity is presented to extend this study by exploring some key aspects of ensemble learning for more accurate prediction performance. According to Oreski et al. (2012) and Shigeyuki et al. (2018), ensemble learning was introduced so as to boost the performance of prediction with strategic combination towards multiple models regarding computational intelligence problems. It has also been proven that in most cases, the ensemble model performs better when compared to solitary classifiers; therefore, important in predicting contraceptive uptake. First, it is outside any reasonable doubt that contraceptives are considered effective with regard to avoiding unintended pregnancies; at the same time, no single way can offer absolute assurance.

Of course, efficacy rates vary among methods, as do side effects, convenience, and STI prevention.

MOTIVATION FOR ENSEMBLE MODELS

A review of the literature on contraception and machine learning revealed some key insights. First, while very effective at preventing unintended pregnancy, no contraceptive method is perfect. Efficacy rates and user preferences for side effects, convenience, and STI prevention play a huge role in the choice of a woman's contraceptive method. These are hard to take into consideration using traditional methods of prediction. Ensemble models overcome this limitation. These ensembles combine the results of several machine learning models to give a more accurate and robust prediction, as stated by Oreski et al. (2012) and Shigeyuki et al. (2018). This is very important in predicting contraceptive uptake, given the complex interaction of factors that influence the decision of a woman.

Yu-En's contribution shows an interesting case when machine learning algorithms are being applied for the prediction of contraceptive prevalence in Asian countries. This research deals with selecting the classification models through the means of ML techniques that best forecast that a certain woman would possibly use contraception in Thailand or Mongolia or Laos by their WRA status and more about not using contraception. For analyzing, five different machine-learning mechanisms are used, namely: KND, Decision tree, Ensembled random forest, Bayesian classifier combined logistic regression mode, and Generalized addition model. Of all, Random forest had the maximum (76.67%). However, this study points out some limitations, such as the data being cross-sectional and the limited variables analyzed.

Haq et al. (2022) investigated the contraceptive practice using a machine learning algorithm among ever-married women in Bangladesh. The study applied several ML models, namely Logistic Regression, Random Forest, Naïve Bayes, Least Absolute Shrinkage and Selection Operation-LASSO, Classification Trees, Adaboost, and Neural Network.

The results showed that the best model was the Neural Network, which had an accuracy of 79.34%. This study underlines the variability in predictive performance across different models and calls for further research to enhance model accuracy.

Another study, conducted by Haq (2017), investigated the sociodemographic factors affecting contraceptive prevalence among ever-married women in Bangladesh, based on data from three rounds of Demographic and Health Surveys. The factors were analyzed using binary logistic regression, and the significant determinants identified included age, education, and media exposure. This research provides valuable insights into the changing patterns of contraceptive use over two decades. Islam, 2017, through logistic regression analysis using data from the national surveys of Bangladesh and India, examined the currently used contraceptive, the method choice, and discontinuation. The findings indicated that Bangladeshi women had a significantly higher chance of using contraceptives compared to Indian women, though the rates of discontinuation among them were higher. Such findings indicate that socio-economic factors have a great influence on contraceptive behaviors; thus, there is a need for targeted interventions.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

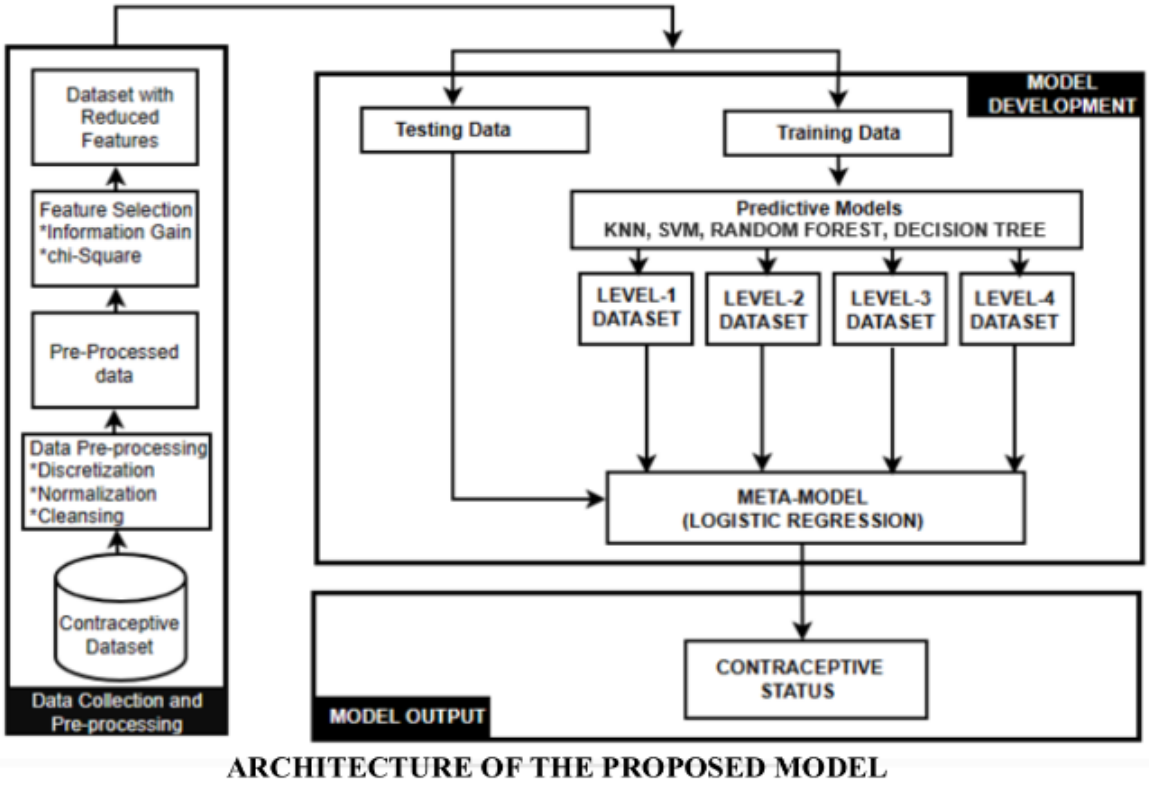
Kebede et al. aimed to predict and identify the predictors of contraceptive discontinuation using the Ethiopian Demographic and Health Survey 2016 dataset among reproductive-age women in Ethiopia. Several algorithms for classifications were presented in this study, such as logistic regression, Random Forest, KNN, artificial neural network, support vector machine, Naïve Bayes, extreme gradient boosting (XGBoost), and AdaBoost. The Random Forest was the best model, showing an accuracy of 0.68.

The study faced challenges such as interpreting model parameters and relying on retrospective data, indicating areas for improvement in future research.

These case studies illustrate the different ways in which machine learning can be used to predict contraceptive use and discontinuation. They also bring out the potential of ensemble models in improving predictive accuracy and dealing with the multi-faceted nature of contraceptive decision-making. This study, therefore, will be informed by addressing the shortcomings identified in the literature review. We develop and evaluate different ensemble models that are focused on the prediction of the uptake of contraceptives among women. By using the strengths of ensemble learning besides considering diverse factors of choice in contraceptive use, we hope to contribute toward an improvement of the predictive accuracy that thus will enable healthcare providers to tailor contraception recommendations and identify women at elevated risk for unintended pregnancy.

3. METHODOLOGY

Figure 1: The Architecture of the proposed model. It is in two basic stages: the first stage, which is Data collection and Preprocessing; the second section is the model development Stage. Further explanation will be done for these.



ARCHITECTURE OF THE PROPOSED MODEL

1. DATA COLLECTION AND PREPROCESSING

This information was collected from Kaggle.com and the Federal Medical Centre, Owo, Ondo State, Nigeria. Contraceptive databases are obtained containing identical form of clinical instances concerning the same patient. Relevant identified predictors of contraceptive uptake come forward to this procedure. The method involves scrutinizing variables of the datasets being identified in related works with what a gynecologist examines for confirmation in an opinion statement. It combines domain expertise with empirical knowledge regarding the procedures in machine learning applied in literature. The data is preprocessed by discretization, normalization, and cleansing to be fit for machine learning processing.

Data Discretization was performed by converting nominal or literal attributes into numeric or integer values. For a nominal attribute A with n unique categories {A_1,A_2,…,A_n }, you define a mapping function f that maps each category to a unique integer:

$$f(A_i) = i \; for \; i = 1,2, \dots, n$$

Where:

A_i represents each unique category in attribute A.

$f(A_i)$ assigns a unique integer i to each A_i.

This will normalize the data so that objects of these datasets are similar in features. Normalization is a strategy that rescales the range into a common range, usually within 0 and 1. Normalization is used to scale numeric columns in a dataset to a common scale for data preprocessing without distorting differences in the ranges of values. This is very helpful, especially in the case of employing machine learning algorithms, since it is helpful to improve the performance or stability of training the model. Normalization in the context of data preprocessing means bringing numerical values into a common scale without affecting differences in ranges. That is particularly useful in the usage of machine learning algorithms and helps to improve performance and training stability for a model.

The formula given:

$$x_{i_{new}} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

normalizes the value x_i to a new range between 0 and 1. Here:

- $x_{i_{new}}$ is the normalized value.
- x_i is the original value of the data point.
- x_{min} is the minimum value in the dataset.
- x_{max} is the maximum value in the dataset.

This transformation ensures that the normalized data will have a minimum value of 0 and a maximum value of 1.

It cleanses the data by removing irregular features that may affect the performance of the proposed model. In this case, instances with a large amount of missing values are removed.

The process is further implemented through the identification and discarding of unrelated features to contraceptives by using a filter-based feature selection method. The reason chosen for two Filter-based feature selection algorithms in this work is that they give a considerable difference in performance yet lightweight and requiring lesser computational complexity, hence suitable for large data-sets.

Feature Selection

The complexity could be further removed by salient feature extraction using some feature selection algorithms, which determines and selects the attributes related to the target class. Filter-based feature selection methods used include Information Gain and Chi-Square. Both are used to identify the most relevant features for variables of contraceptive.

Information Gain is a probabilistic model for feature selection based on a nominal feature, X, and target class, Y, by determining individually the probabilities of values from the dataset. In the selection method of Information Gain,

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

Set of attribute with the highest accuracy is selected and returned

Information Gain (IG) for attribute(s) x in the dataset is given by:

$$IG(X) = H(Y) - H(Y|X)$$

Where H(Y) is Entropy of Y

H(Y/X) is Entropy of Y given X.

$$H(Y) = -\sum_{i=1}^{n} p(y_i) log_2 p(y_i)$$

$$H(Y/X) = -\sum_{i=1}^{n} p\left(\frac{y_i}{x_i}\right) log_2\, p\left(\frac{y_i}{x_i}\right)$$

The larger the value of Information Gain, the more relevant the feature.

n : is the number of instances in the contraceptive dataset

$P(y_i)$: is the probability of occurrence of class label value of instance I

$P(y_i/x_i)$: is the probability of the class label value of instance i will occur given the occurrence of attribute value x of the instance i.

Chi–Square is then used as a feature selection algorithm.

- To use chi-square (χ^2) for feature selection, we calculate χ^2 between each feature and the target. and select the desired number of features with the best χ^2 scores.
- The intuition is that if a feature is independent to the target, it is uninformative for classifying observations.
- Let the feature have x attribute values and the output have y class labels. Chi- square score is given by:

$$\chi^2 = \sum_{i=1}^{x}\sum_{j=1}^{y} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where:

- O_i is *the number of observation of class i*
- E_i is *the number of expected observations in class i if there was no relationship between the feature and the target.*

The attributes with larger chi-square values are then selected for the classification.

- For each of the feature in the dataset, the χ^2 is calculated and then ordered in descending order according to the χ^2 values.
- The higher the value, the more dependent the class label is on the feature and the higher the importance the feature has on determining the class label.
- For each feature, a contingency table is created with x rows and y columns.
- Each cell (i, j) denotes the number of rows having attribute feature as i and class label as y – ***the observed frequency***.
- To calculate **the expected frequency** for each cell, first the proportion of the feature value in the total dataset is calculated and then it is multiplied by the total number of the current class label.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

2. THE MODEL DEVELOPMENT STAGE

It consists of the training and development of predictive base models using four different supervised machine learning algorithms, which were later on stacked into what they generally termed stacked generalization. Then it listed all the supervised algorithms to represent themselves. Support vector machine as the first base model, Random Forest, now as base model II; it is followed by Decision Tree as base model III; ending with KNN as the base model IV. Each model is applied individually to an equal number of objects during the training phase.

– Base Model I (SVM)

- **Support Vector Machine** SVM algorithm predicts the contraceptive uptake in women by plotting predicting attributes in multidimensional hyperplane and classifies the classes optimally by creating the margin between two data clusters.

Let:

- x be a feature vector (i.e., the input of the SVM). $X \in R^n$, where n is the dimension of the feature vector.
- R^n Real Vector of n dimension
- y be the class (i.e., the output of the SVM). $y \in \{-1,1\}$, i.e. the classification task is binary.
- w and b be the parameters of the SVM: we need to learn them using the training set.
- $(x^{(i)}, y^{(i)})$ be the i^{th} sample in the dataset. Let's assume we have N samples in the training set.
- The class y is determined as follows:

$$y^{(i)} = \begin{cases} -1 & if\ w^T x^{(i)} + b \leq -1 \\ 1 & if\ w^T x^{(i)} + b \geq 1 \end{cases}$$

which can be more concisely written as $y^{(i)}(w^T x^{(i)} + b) \geq 1$

The SVM aims at satisfying **two requirements**:

(i) The SVM should maximize the distance between the two decision boundaries. Mathematically, this means we want to maximize the distance between the hyperplane defined by

$w^T x + b = -1$ and the hyperplane defined by $w^T x + b = 1$

Where $w^T = weight\ vector$ (distance between hyperplane and attribute)

b = hyperplane bias

x = data feature vector

(ii) The SVM should also correctly classify all $x^{(i)}$, which means

$$y^{(i)}\left(w^T x^{(i)} + b\right) \geq 1, \forall i \in \{1, \ldots, N\}$$

- Once the lagrange multipliers $\alpha^{(i)}$ are learned, one can **predict the class** of a new sample with the feature vector x^{test} as follows:

$$y^{test} = sign(w^T \phi(x^{test}) + b$$

$$= sign(\sum_{i=1}^{N} \alpha^{(i)} y^{(i)} \phi(x^{(i)})^T \phi(x^{test}) + b)$$

ϕ represents nonlinear feature mapping function

– Base Model II (Random forest)

Random forest is a tree based classification algorithm.

- It is an ensemble algorithm which combines multiple algorithms
- **It creates a set of decision trees from a random sample of the training set.**
- It repeats the process with multiple random samples and makes a final decision based on majority voting.
- **The Random forest algorithm is effective in handling missing values but it is prone to overfitting.**
- Appropriate parameter tuning can be applied to avoid overfitting. The algorithm for Random forest

Training data set: Θ

Let D be a training set $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$
Let $h = h_1(x), h_2(x), \ldots, h_k(x)$, *an ensemble of weak classifiers*
If each h_k *is a decision tree, the parameters of the tree are defined as* $\Theta = (\theta_{k1}, \theta_{k2}, \ldots, \theta_{kp})$
Each decision tree k leads to a classifier $h_k(X) = h(X|\Theta_k)$
Final Classification $f(x)$ = *Majority of* $h_k(X)$

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

– Base Model III (Decision Tree)

Decision tree is a flow chart like tree structure. It can be binary or non-binary. In a decision tree each non-leaf node denotes a test on an attribute, every branch is an outcome of a test and each leaf contains a class label. The top node in tree is called root node. (Chandan *et al.*, 2019)

The most common algorithm of decision tree is ID3 (Iterative Dichotomiser 3).

- There are two mathematical tools needed to complete ID3 algorithms. a) Entropy: it is used to measure important of information relative to its size.

$$H(a) = - p_+ \log_2 p_+ - p\log_2 p$$

where, a is training set b) Information Gain:

It is the difference between original information required and the new requirement

$$Gain\ (a,b) = H(a) - \sum_{i\in\ value\ (b)} \frac{|\Delta b_i|}{|\Delta b|} H(b_i)$$

Δbi – *Possible Value of b*
Δb – *Set example of (x)*
bi – Subset where Xb = I
$i\in\ value$ – partition of the training set data

– Base Model IV (KNN)

- **K-Nearest Neighbour** classify an unlabeled instance (tuple) in the contraceptive test dataset by assigning them to the class of the most similar labeled instances in the contraceptives training dataset.
- The goal of KNN is to;
  - Use distance function to compute the distance between the given instance and all the instances in the training dataset
  - Rank the computed distances of all the training instances with the given instance in ascending order
  - For K = N, use simple majority vote to determine instance that has highest number of class label among the top N ranked instances.
  - Classify the given instance class categories as the majority class categories among the top n ranked instances

– Base Model I (KNN)

- KNN is based on Euclidean distance between the training set and the testing set.
- Given that p_i is the instance to be classified ranging from 1 to n, p_t is the total number of instances in contraceptive dataset ranging from 1 to k with same number of features.
- The Euclidean distance between p_i and p_t can be defined below:

$$dist(p_i, p_t) = \sqrt{\sum_{k=1}^{n} (p_{ik} - p_{tk})^2}$$

where

- p_i is the current instance of the dataset $(i = 1, 2, \ldots, n)$
- p_t is the given instance in the test dataset being classified
- n is the total number of instances in the training dataset
- k is the total number of features in the entire dataset apart from the class label.

- Diagnosis of the input instance was based on the closest n neighbor according to equation

These base models were then used to construct the meta-level dataset. A stacked ensemble predictive model was adopted where the meta-model would serve as the combiner model. Stacked ensemble was chosen because it performed best among several ensemble methods. The architecture provides an illustration of the research process to be conducted. Logistic Regression was used for meta-level learning.

Stacked Generalization Approach

The stacking algorithm is an ensemble learning class that describes the combining of multiple classifiers in a hierarchical manner. Level-0 classifiers will form attributes in a new training set, level-1 data, which keeps the original class labels. Selection of the classifiers was based on the strengths and weaknesses of the individual classifiers revealed in the reviewed related works.

Stacking is a general method of ensembling different machine learning algorithms and can be thought of as a method for collectively using several models to estimate their generalizing biases with respect to a certain learning set and filter those biases out.

– Stacked Generalization Approach

- Defining x_n as the n^{th} sample of N sample-training data and y_n as the class label of the sample, raw data set is constructed by,

$$D = \{(y_n, x_n) \quad n = 1, \ldots, N\}$$

- The raw CP data set, firstly, was divided into *j(10)* parts by random selection in order to avoid over-fitting.
- In the 10-fold cross-validation technique, D_j is defined as test set, and $D^{(-j)} = D - D_j$ as training set.
- k^{th} classifier among K classifiers on the layer-0 is applied on the $D^{(-j)}$ training set for the construction of model $M_k^{(-j)}$.
- If model $M_k^{(-j)}$ is used to classify an instance x in D_j (test set). Let $v_k^{(-j)}(x_n)$ denote the probability of the i^{th} output class (prediction of label x)

$$z_{kn} = v_k^{(-j)}(x_n)$$

- The class probabilities in equation 17 are assembled with the original class to constitute Level-1 dataset.

Logistic Regression (Meta Model)

- It's a classifier that predicts the outcome of a categorical dependent variable from a set of predictor or independent variables.
- **It calculates the probability of a discrete outcome based on the given data.**
- **Predictive attribute are referred to as independent variables while target attribute is referred to as response variable.**
- Let $X = \{x_1, x_2, \ldots, x_m\}$ be the vector of independent variables
- Let Y be the response variable with possible outcome 0 or 1.
- Let p be the probability that $Y = 1$ i.e. $p = P(Y = 1)$
- In logistic regression, it is assumed that there is a linear relationship between X and the logit transform of p.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

Then we have

$$\left(\frac{p(x)}{1-p(x)}\right) = e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m}$$

Let $Y = e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m}$, then $\left(\frac{p(x)}{1-p(x)}\right) = Y$

Consequently, $p(x) = Y(1-p(x))$

$$p(x) = Y - Y(p(x)) = \frac{Y}{(1-Y)} = \frac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m}}{1-e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m}}$$

The equation of a sigmoid function:

$$p(x) = \frac{e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m}}$$

$$= \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_mx_m)}}$$

IMPLEMENTATION

Implementation of the proposed ensemble models will be done in a Python environment. Standard metrics to be considered are: precision, sensitivity, and Accuracy. This model is performed in an Jupyter notebook 6.4.8 version on windows 8 operating system. The CPU is an Intel core i5 processor, equipped with a 4GB RAM.

Dataset

We trained the model on the dataset obtained from Kaggle.com. and dataset obtained from from a female questionnaire from a medical research center and evaluation. After feature extraction, feature reduction was done in order to reduce time complexity and excessive features in the storage memory. In this paper, 70% of the respondents are used for the purpose of model tuning, and the remaining 30% are used to check.

MODEL PERFORMANCE,

Attribute Information:

The following table shows that the attribute information of the data is Wife's age, numerical; Wife's education, categorical: 1 = low, 2, 3, 4 = high; Husband's education categorical - 1= low, 2, 3, 4 = high, Number of children ever born continuous, Wife's religion

bnary 0 = Non-Islam, 1 = slam, Wife now working? binary 0 = Yes, 1 = No, Husband's occupation categorical 1, 2, 3, 4 Standard-of-living index categorical 1 = low, 2, 3, 4 = high

Media exposure-binary 0=Good, 1=Not good, Contraceptive method used-class attribute 1 = No-use, 2 = Long-term, 3 = Short-term. Below is a sample of the dataset.

	Wife Age	Wife Education	Husband Education	Children	Wife religion	Wife working	Husband Occupation	SOLI	Media Exposure	Contraceptive Method
0	24	2	3	3	1	1	2	3	0	1
1	45	1	3	10	1	1	3	4	0	1
2	43	2	3	7	1	1	3	4	0	1
3	42	3	2	9	1	1	3	3	0	1
4	36	3	3	8	1	1	3	2	0	1

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

FEATURE IMPORTANCE

	features	scores
0	Wife Age	128.920246
1	Wife Education	41.898666
2	Husband Education	7.827384
3	Children	43.577809
4	Wife religion	3.640215
5	Wife working	1.103385
6	Husband Occupation	14.649733
7	SOLI	16.901136
8	Media Exposure	24.610799

So, three most important features for our dataset are Wife Age, Wife Education, and Children. The two least relevant features are Wife Religion and Wife working features. We can make decisions with these facts.

RESULT EVALUATION AND ANALYSIS

Training of the whole stacked architecture is performed after training of individual classifiers using the cross validation technique.

Two parameters analyze the result of the classification.

a) Accuracy: Precision, recall, and f1-score are the metrics applied for the measurement of accuracy on the input dataset.

b) Confusion Matrix: The confusion matrix is able to show the Performance of the model.

A confusion matrix, sometimes referred to as an error matrix, is a mapping between actual and predicted class instances, where rows where rows represent the actual class and columns represent the predicted class.

In this research, we used a confusion matrix and a classification report provided by Given by the Sklearn library, we have all the performance evaluations inside the classification report. matrix and the accuracy score.

METRICS USED IN CLASSIFICATION

Explanation of the metrics involved in Classification Report:

Precision Score

TP – True Positives

FP - False Positives

Precision: Exactness of the positive predictions.

Precision = TP/(TP + FP)

Recall Score

FN - False Negatives

Recall (aka sensitivity or true positive rate): Proportion of positives That were

Correctly identified.

Recall = TP/(TP+FN)

F1 score

F1 Score (aka F-Score or F-Measure): A helpful metric when looking to compare two Classifiers. F1 Score incorporates precision and recall. It is generated by The harmonic mean of precision and recall is calculated.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 12, Issue 4, pp: (64-76), Month: October - December 2024, Available at: www.researchpublish.com

F1 = 2 x (precision x recall)/(precision + recall)

Confusion Matrix

Confusion Matrix This allows you to look at all the particular misclassified examples yourself, and do any other analysis you want. Here is a tabular result in diagram.

BUILD AND EVALUATION OF EACH CLASSIFICATION

```
Accuracy score for KNeighboursClassifier is :0.8603796693202694
F1_score score for KNeighboursClassifier is :0.8606761966067022
Confusion Matrix for KNeighboursClassifier is :[[554  18  47]
 [ 13 331  51]
 [ 40  59 520]]
              precision    recall  f1-score   support

           1       0.91      0.89      0.90       619
           2       0.81      0.84      0.82       395
           3       0.84      0.84      0.84       619

    accuracy                           0.86      1633
   macro avg       0.86      0.86      0.86      1633
weighted avg       0.86      0.86      0.86      1633
```

```
Accuracy score for SVC Classifier is :0.5646050214329456
F1_score score for SVC Classifier is :0.5499936602781909
Confusion Matrix for SVC Classifier is :[[364  33 222]
 [ 63 101 231]
 [101  61 457]]
              precision    recall  f1-score   support

           1       0.69      0.59      0.63       619
           2       0.52      0.26      0.34       395
           3       0.50      0.74      0.60       619

    accuracy                           0.56      1633
   macro avg       0.57      0.53      0.52      1633
weighted avg       0.58      0.56      0.55      1633
```

```
Accuracy score for RandomForestClassifier is :0.7544396815676668
F1_score score for RandomForestClassifier is :0.7547253607495458
Confusion_matrix for RandomForestClassifier is :[[478  27 114]
 [ 42 270  83]
 [ 62  73 484]]
              precision    recall  f1-score   support

           1       0.82      0.77      0.80       619
           2       0.73      0.68      0.71       395
           3       0.71      0.78      0.74       619

    accuracy                           0.75      1633
   macro avg       0.75      0.75      0.75      1633
weighted avg       0.76      0.75      0.75      1633
```

```
Accuracy score for  DecisionTreeClassifier is :0.6962645437844458
F1_score score for  DecisionTreeClassifier is :0.6959639570785962
Confusion Matrix for  DecisionTreeClassifier is :[[478  57  84]
 [ 51 241 103]
 [103  98 418]]
              precision    recall  f1-score   support

           1       0.76      0.77      0.76       619
           2       0.61      0.61      0.61       395
           3       0.69      0.68      0.68       619

    accuracy                           0.70      1633
   macro avg       0.69      0.69      0.69      1633
weighted avg       0.70      0.70      0.70      1633
```

RESULT OF THE STACK ENSEMBLE MODEL

Accuracy	0.8187385180649112
Precision	0.8222512740324512
Recall	0.8187385180649112
F1 Score	0.8198219216235713

Overall, the model appears to perform reasonably well, with accuracy, precision, recall, and F1 score all hovering around 0.82, indicating a good balance between true positive and false positive predictions.

4. DISCUSSION OF RESULTS

Comparison Table: Performance Metrics of Classification Models

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
K-Nearest Neighbors	**0.86**	**0.86**	**0.86**	**0.86**
Support Vector Machine	**0.56**	**0.58**	**0.56**	**0.55**
Random Forest	**0.75**	**0.76**	**0.75**	**0.75**
Decision Tree	**0.70**	**0.70**	**0.70**	**0.70**
Ensemble	**0.8187**	**0.8223**	**0.8187**	**0.8198**

ANALYSIS:

1. Accuracy:

The K-Nearest Neighbors classifier provides the best accuracy at 0.86.

- The Ensemble method also presents a high accuracy of 0.8187, outperforming the Random Forest classifier with 0.75, Decision Tree with 0.70, and Support Vector Machine (SVM) with 0.56.

2. Precision

- KNN again leads with a precision of 0.86.

It is followed by the Ensemble method with an almost similar accuracy of 0.8223, far above the Random Forest classifier's 0.76, the Decision Tree classifier's 0.70, and the SVM classifier's 0.58.

3. Recall:

- KNN keeps the maximum recall of 0.86.

- The Ensemble method has a recall of 0.8187, outperforming Random Forest at 0.75, Decision Tree at 0.70, and SVM at 0.56.

4. F1 Score:

- KNN has the highest F1 score, which is 0.86. It reaches an Ensemble F1 score that is much stronger compared to the rest: 0.8198, beating Random Forest's score of 0.75, Decision Tree with a score of 0.70, and that of SVM with 0.55.

5. DISCUSSION

The ensemble model performs balanced with respect to all the metrics: though for KNN classifiers, performing better compared to the ensemble approach in accuracy, precision, recall, or F1 score, the results brought by the ensemble method are generally far more consistent and reliable.

- Support Vector Machine: This performs the worst on all the metrics, which means that this SVM probably is not suitable for either this particular task or dataset.

- Random Forest and Decision Tree: Both models present an average performance, with Random Forest performing better in all metrics when compared to the Decision Tree; however, both are outperformed by the ensemble method.

- Ensemble Model: While KNN has slightly higher scores, the performance of the ensemble model is close behind and far better than the rest of the models like SVM, Random Forest, and Decision Tree classifiers. This may be due to the fact that ensembling different models leverages their strengths for more robust and accurate predictions.

While the KNN classifier represents the highest performance individually, the ensemble method presents a reliable balanced approach; by reducing the risk of model overfitting, this increases the robustness of predictions. This will make it a strong candidate for practical application in predicting contraceptive uptake.

6. CONCLUSION

This study highlights how a stacked ensemble model is efficient in predicting the uptake of contraception. The combination of four different supervised machine learning algorithms-namely, Support Vector Machine, Random Forest, Decision Tree, and K-Nearest Neighbors-into one predictive system allows several strengths of each algorithm to be utilized while making up for their respective individual limitations. The stacked generalization approach has significantly improved the performance-0.8187 accuracy, 0.8223 precision, 0.8187 recall, and an F1 score of 0.8198 by using Logistic Regression as the meta-learner. These results reflect a superior balance and robustness compared to the individual models.

Extensive data collection and preprocessing, including discretization, normalization, and cleansing, were done to ensure the quality and consistency of the input data. Feature selection using Information Gain and Chi-Square further refined the dataset, focusing on the most relevant predictors of contraceptive uptake.

The Python implementation, which was validated by standard metrics such as precision, recall, and accuracy, also proved the model's efficiency.

While the KNN classifier showed marginally higher individual metrics of performance, the balanced ensemble results across all metrics under study point out its robustness and reliability. This advanced machine learning technique holds great promise for dealing with complex challenges in healthcare; the resultant models provide more accurate predictions that help health providers make informed recommendations, which may be useful for resultant better family planning and reduction of unwanted pregnancies.

This, in itself, calls for further future studies that shall apply an ensemble approach in other health domains, adding other predictors with the hope of increasing the performance of the models and considering longitudinal data for uncovering temporal patterns of interest in contraceptive use. In each refined and extended model that comes along, the power of using artificial intelligence for improving patient outcomes by supporting public health becomes even stronger.

REFERENCES

[1] Ajayi, O. I., Onifade, T. I., Owolabi, O. D., & Omisakin, A. (2018). Factors influencing contraceptive use among women in developing countries: a review of literature. International journal of women's health, 10(7), 321. [Ajayi et al., 2018]

[2] Dilsizian, V., & Siegel, E. (2014). Machine learning in medicine. Applying machine learning to clinical data. Academic Press. [Dilsizian & Siegel, 2014]

[3] Gorucu, M., Yalcin, O., & Colak, I. (2020). Machine learning for prediction of contraceptive use in Turkey. Studies in health technology and informatics, 271, 1231-1235. [Gorucu et al., 2020]

[4] Haq, I., Hossain, I., Rahman, M. M., Methun, M. I., Talukder, A., Habib, M. J., & Hossain, M. S. (2022). Machine learning algorithm-based contraceptive practice among ever-married women in Bangladesh: A hierarchical machine learning classification approach. [Haq et al., 2022]

[5] Haq, I. U. (2017). Sociodemographic factors influencing contraceptive use among ever-married women in Bangladesh: A comparative analysis of three Bangladesh Demographic and Health Surveys (BDHS) 1993-2014. International quarterly of community health education, 38(2), 142-152. [Haq, 2017]

[6] Islam, A. K. M. S. (2017). Contraceptive use, method choice and discontinuation in Bangladesh and India: A comparative analysis using nationally representative surveys. PloS one, 12(1), e0169223. [Islam, 2017]

[7] Iqbal, Y. U. (2022). A machine learning approach to predict contraceptive use in Asian countries. [Yu-En, 2022]

[8] Kebede, A., Gebremariam, Z. A., & Derese, A. (2023). Prediction of contraceptive discontinuation among reproductive-age women in Ethiopia using machine learning approach. BMC medical informatics and decision making, 23(1), 1-12. [Kebede et al., 2023]

[9] Oreski, D. P., Oreski, C. M., & Oreski, J. C. (2012). Ensemble machine learning for customer relationship management. John Wiley & Sons. [Oreski et al., 2012]

[10] Rosenfield, A. (2012). Patient factors influencing contraceptive choice. Obstetrics and gynecology clinics of North America, 39(2), 301-312. [Rosenfield, 2012]

[11] Sedgh, G., Singh, S., Appleby, M., Ashford, T., Blumstein, S., Gangadhar, B., ... & Zhao, Y. (2020). Abortion incidence worldwide. The Lancet, 394(10194), 1603-1616. [Sedgh et al., 2020]

[12] Shigeyuki, M., Bao, T., & Nakashima, T. (2018, August). Ensemble learning for machine learning beginners. In 2018 International Conference on Systems, Man, and Cybernetics (SMC) (pp. 2239-2244). IEEE. [Shigeyuki et al., 2018]

[13] World Health Organization. (2021). Family planning/Contraception. [World Health Organization, 2021]