Intelligent Information Management, 2023, 15, 391-425
https://www.scirp.org/journal/iim
ISSN Online: 2160-5920
ISSN Print: 2160-5912

# MultiDMet: Designing a Hybrid Multidimensional Metrics Framework to Predictive Modeling for Performance Evaluation and Feature Selection

**Tesfay Gidey Hailu[1], Taye Abdulkadir Edris[2]**

[1]Department of Information and Communication Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
[2]Department of Computer Science, HILCOE School of Computer Science and Technology, Addis Ababa, Ethiopia
Email: tesfaygidey21@gmail.com, tesfaygidey22@ieee.com, tabdulkadir@gmail.com

**How to cite this paper:** Hailu, T.G. and Edris, T.A. (2023) MultiDMet: Designing a Hybrid Multidimensional Metrics Framework to Predictive Modeling for Performance Evaluation and Feature Selection. *Intelligent Information Management*, **15**, 391-425.
https://doi.org/10.4236/iim.2023.156019

**Received:** August 30, 2023
**Accepted:** November 21, 2023
**Published:** November 24, 2023



## Abstract

In a competitive digital age where data volumes are increasing with time, the ability to extract meaningful knowledge from high-dimensional data using machine learning (ML) and data mining (DM) techniques and making decisions based on the extracted knowledge is becoming increasingly important in all business domains. Nevertheless, high-dimensional data remains a major challenge for classification algorithms due to its high computational cost and storage requirements. The 2016 Demographic and Health Survey of Ethiopia (EDHS 2016) used as the data source for this study which is publicly available contains several features that may not be relevant to the prediction task. In this paper, we developed a hybrid multidimensional metrics framework for predictive modeling for both model performance evaluation and feature selection to overcome the feature selection challenges and select the best model among the available models in DM and ML. The proposed hybrid metrics were used to measure the efficiency of the predictive models. Experimental results show that the decision tree algorithm is the most efficient model. The higher score of $HMM\,(m,\ r) = 0.47$ illustrates the overall significant model that encompasses almost all the user's requirements, unlike the classical metrics that use a criterion to select the most appropriate model. On the other hand, the ANNs were found to be the most computationally intensive for our prediction task. Moreover, the type of data and the class size of the dataset (unbalanced data) have a significant impact on the efficiency of the model, especially on the computational cost, and the interpretability of the parameters of the model would be hampered. And the efficiency of the predictive

model could be improved with other feature selection algorithms (especially hybrid metrics) considering the experts of the knowledge domain, as the understanding of the business domain has a significant impact.

## Keywords

Predictive Modeling, Hybrid Metrics, Feature Selection, Model Selection, Algorithm Analysis, Machine Learning

## 1. Introduction

In today's digital age, we are experiencing an explosion in data volume, data variety, and data dimensions. With the increase of high dimensional data in the last decade, feature selection becomes a necessary step in the field of machine learning, data mining, pattern recognition and statistics as one of the solutions to overcome the curse of dimensionality. Feature selection refers to the process of identifying a subset of the most relevant features that provide representative results for the original set of features [1] [2] [3] [4] [5]. As data sets grow larger over time, the ability to extract knowledge hidden in these large data sets and make decisions based on the extracted knowledge is becoming increasingly important in all business domains. The application of data mining techniques is at the heart of the pattern recognition and knowledge extraction process in all domains [6] [7]. Nevertheless, high-dimensional data (HDD) poses a major challenge to classification algorithms for both machine learning (ML) and data mining (DM) models due to its high computational cost and storage requirements [8]. Hypothetically, a larger number of features imply higher discriminative power in classification. In practice, however, this is not always true because the collected features are often not all equally informative, as some of them may be interdependent. In addition, high-dimensional data can cause the problem of the "curse of dimensionality" [9].

Feature extraction and selection methods are used in isolation or in combination with the goal of improving performance such as estimation accuracy, visualization, and understandability of the learned knowledge [10] [11]. In general, features can be categorized as relevant, irrelevant, or redundant. The best subset is the one with the least number of dimensions that contribute the most to learning accuracy [12]. The Demographic and Health Survey of Ethiopia dataset (EDHS 2016), which was used as the data source for this study, contains detailed information on respondents' background characteristics that may not be relevant for predicting contraceptive use [13]. In general, health care is considered "information rich" but "knowledge poor" due to the lack of effective analytical tools to discover hidden relationships and trends in the data. Data mining is used in biomedical sciences and research because it is a rapidly developing technology that can extract useful knowledge, and scientific decision making for

diagnosis and treatment of diseases from the database is becoming increasingly important [14]. It can also improve the management of hospital information and promote the development of telemedicine and community medicine [15], and is becoming increasingly popular in the healthcare industry due to its excellent efficiency, including but not limited to the health insurance sector, which can use it to minimize fraud and abuse [16] [17] [18] [19]. The algorithms DM and ML search for meaningful patterns in raw datasets and help to make scientific decisions from the database.

On the other hand, both DM and ML are computationally expensive for large datasets because they may contain several features irrelevant to the analysis. Therefore, reducing the dimensionality can effectively reduce these costs. In summary, feature selection methods help to: reduce the dimensionality of the feature spaces (to limit the memory requirements and increase the speed of the algorithm), improve the runtime of the learning algorithms, improve the data quality, increase the performance, and understand the nature of the data and the process that generated the data or visualize the data easily [12]. Several feature selection algorithms have been proposed and studied for ML and DM applications and can be broadly classified into three categories: Filters, Wrappers, and the embedded methods. Filter methods use inherent properties of the data such as information-based measures, distances, or statistical information to evaluate the quality of a selected subset. Nevertheless, the accuracy results obtained may not be guaranteed [20] [21] [22]. The wrapper methods use the prediction accuracy of a predetermined learning algorithm to determine the relevance of the selected subsets, and the accuracy of the learning algorithms is usually high. However, the generality of the selected features is limited (tendency to overfit for small training sets) and is very computationally intensive for high-dimensional data [23] [24] [25] [26]. And the embedded methods include feature selection as part of the training process and are usually specific to certain learning algorithms such as (SVM, ANNs, and DT classifiers) and thus may be more efficient than the other two categories [27] [28] [29] [30].

In this work, we developed a hybrid multidimensional metrics framework for predictive modeling for performance evaluation and feature selection to overcome the challenges in feature selection and select the best model among the available models used in DM and ML. The contributions are fourfold:

1) We proposed a novel feature selection method based on multidimensional metrics, including but not limited to correlation test, chi-square test, expert testimony, established knowledge, etc.

2) We developed a hybrid multidimensional metric for model selection, including: Confusion matrix analysis, ROC curve analysis, statistical significance, practicality or applicability, computation time, simplicity of rule extraction, consistency with existing knowledge.

3) Compare the proposed metrics with the classical approach most commonly used in feature selection and model selection.

4) This research work is of importance to researchers and the scientific community as well as academia to show how these hybrid metrics could be used in DM and ML algorithms in a broader healthcare setting for prediction tasks based on high-dimensional data in similar and/or other platforms.

The remaining of this paper is organized as follows: related works is presented in Section II. Section III describes the framework and problem statement. Experimental results, discussions, and evaluation metrics are presented in Section IV. Conclusions and recommendations are provided in Section V.

## 2. Related Works

This section presents two major parts: 1) applications of predictive modeling in healthcare industry with the goal to handle the fundamental challenges encountered in both feature selection and model selection supported with real word dataset. 2) Presenting different research works specific to fertility and contraception methods use.

### 2.1. Applications of Predictive Models in Healthcare Industry

Healthcare seems to be "information rich" but "knowledge poor" due to lack of effective analysis tools to discover hidden relationships and trends in data. DM and ML are two effective techniques suitable for data analysis and finding hidden patterns that can be used for medical decision making [31]. And ML approaches can benefit the health care system through various approaches such as shortening treatment time, detecting disease causes and symptoms [32]. The application of data mining techniques has long been encouraged in healthcare. For example, health insurance fraud and abuse have led many health insurers to try to reduce their losses by using data mining tools to find and prosecute offenders [16] [33]. In the commercial world, data mining tools are mainly used for fraud detection. Data mining is actively used in diagnosis and treatment, healthcare resource management, customer relationship management, and fraud and anomaly detection [34]. Numerous healthcare companies, hospitals, and pharmaceutical manufacturing facilities are using data mining tools due to their excellent efficiency [34]. The predominant use of data mining and machine learning algorithms compared to traditional statistical methods in healthcare applications or predictive tasks could be mainly due to their promising accuracy and more reliable results, as they offer greater efficiency in processing large amounts of data, are more flexible, and can handle any type of data [35].

A surveillance system that uses data mining techniques to detect new and interesting patterns in infection control data has been implemented at the College of Alabama [36]. Data mining techniques have been implemented to examine reporting practices using International Classification of Diseases, 9$^{th}$ Revision, codes (risk factors). By reconstructing patient profiles, cluster and association analyzes can show how risk factors are reported [37]. To improve its ability to prospectively identify high-risk patients, American Heathway's uses predictive

modeling technology [38]. Data mining tools have been used for fertility demographic analysis to determine which attributes have the greatest impact on a country's fertility rate [39]. In summary, data mining in healthcare is used to evaluate the effectiveness of treatments, manage healthcare, manage customer relationships, and detect fraud and abuse, among other applications [39]. In addition, previous research has demonstrated the application of data mining techniques for predictive tasks, including HIV testing [35], cancer [40], heart disease [41], tuberculosis [42], kidney dialysis [43], diabetes [44], dengue fever [45], hepatitis C [46], and IVF [47].

The most commonly used algorithms for prediction tasks in DM and ML include Decision Trees, Naïve Bayes Classifier, Artificial Neural Networks, Support Vector Machines, and K-Nearest Neighbor [48], based on their accuracy performance. However, the efficiency of a model is not just about accuracy. Many clients have multiple lists of requirements that need to be considered when using a predictive model for its prediction task (flexible metrics). In an environment like healthcare, where data can accumulate over time and potentially take on characteristics of Big Data, it is extremely important to analyze the characteristics and identify latent relationships between the characteristics. However, high-dimensional data has become a real challenge for predictive tasks in DM and ML algorithms because the data may come from multiple sources, which in turn affects performance, is computationally intensive, and introduces the problem of overfitting. Therefore, in order to remove irrelevant and redundant features, feature selection is a necessary preprocessing step of the classification process, which serves to reduce computation time and improve learning accuracy, especially for high-dimensional datasets [49] [50].

## 2.2. Applications of Data Mining Models to Contraceptive Use

Contraceptive use is considered critical for protecting women's health and rights, influencing fertility and population growth, and promoting economic development, especially in much of sub-Saharan Africa [51]. Data mining consists of the application of algorithms to identify and analyze information to create patterns or models [52]. In a study conducted in southern Brazil, data mining was used to analyze the profile of contraceptive method use in a college population [53]. The study found that the results obtained with the generated rules were largely consistent with the literature and global epidemiology and revealed significant vulnerabilities in the college population [53]. The study validated its results based on accuracy, sensitivity, specificity, and area under the ROC curve and obtained higher or at least similar values compared to recent studies using the same methodology [54]. Another study was conducted to examine in detail how a particular data mining method called General Unary Hypotheses Automaton (GUHA) helps predict women's use of contraceptive methods based on knowledge of their demographic and socioeconomic characteristics [55].

It also used a data mining approach to analyze patterns of contraceptive use in India by comparing contraceptive use among groups of women with different demographic, economic, cultural, and social characteristics. The decision tree classification and regression algorithms were applied to identify women with different social, economic, cultural, and demographic characteristics who use different contraceptive methods and then analyze how the pattern of contraceptive methods differs among these groups [56] [57]. The study found that currently married, nonpregnant women aged 15 - 49 years in India can be classified into 13 mutually exclusive groups based on six characteristics of the women: surviving children, household standard of living, religion, women's schooling, husband's education, and place of residence. The observed differences in patterns of contraceptive use have important policy and programmatic implications related to universal access to family planning. Another study was conducted at the Pratama Hasanah Pekanbaru clinic to examine contraceptive use data using the C4.5 decision tree. The study found that it was necessary to evaluate the contraceptive use data collection to determine the pattern of contraceptive choice. Nine attributes (age, duration of use, menstrual cycle, recently married, recently delivered, breastfeeding, already having offspring, health problems, and more than four children) were used to determine the pattern of contraceptive choice, and the class label was contraceptive use. The classification model of the study had achieved 93.15% accuracy [58].

Another study was also conducted using data mining classification algorithm to predict the duration of contraceptive use of productive couples by adopting the CRISP-DM process method [59]. And data mining techniques were employed to different experimentations using Demography and Health Survey of Indonesia (DHSI) in 2017. The result exemplified that the Adaboost data mining technique produced the best performance of contraceptive used prediction model, with the accuracy score of the classification model as 85.1%. Moreover, the application of data mining technique was also used to predict the likelihood of contraceptive method use among women aged 15 - 49 years old using DHS of Ethiopia 2005 despite the survey doesn't reflect or lacks to represent the current sociodemographic status of the population [60]. Experimental results of the study revealed that J48 decision tree performs better than Naïve Bayes. It has also been reported the model had achieved an accuracy of 82.85% to detect contraceptive method users correctly. Furthermore, in several DM and ML applications, it has been revealed that models are described as best fit model with few criteria mostly the performance accuracy, the higher accuracy of the model would best fit the data. However, accuracy of a model, would not give the complete picture of the problem domain as for possible health interventions may require by decision makers based on their policies' prioritization and hence several features might affect for the prediction of the outcome. This is why an important question arises in the scientific communities, when should we use a particular model to get a complete picture of the variable under study pertinent to

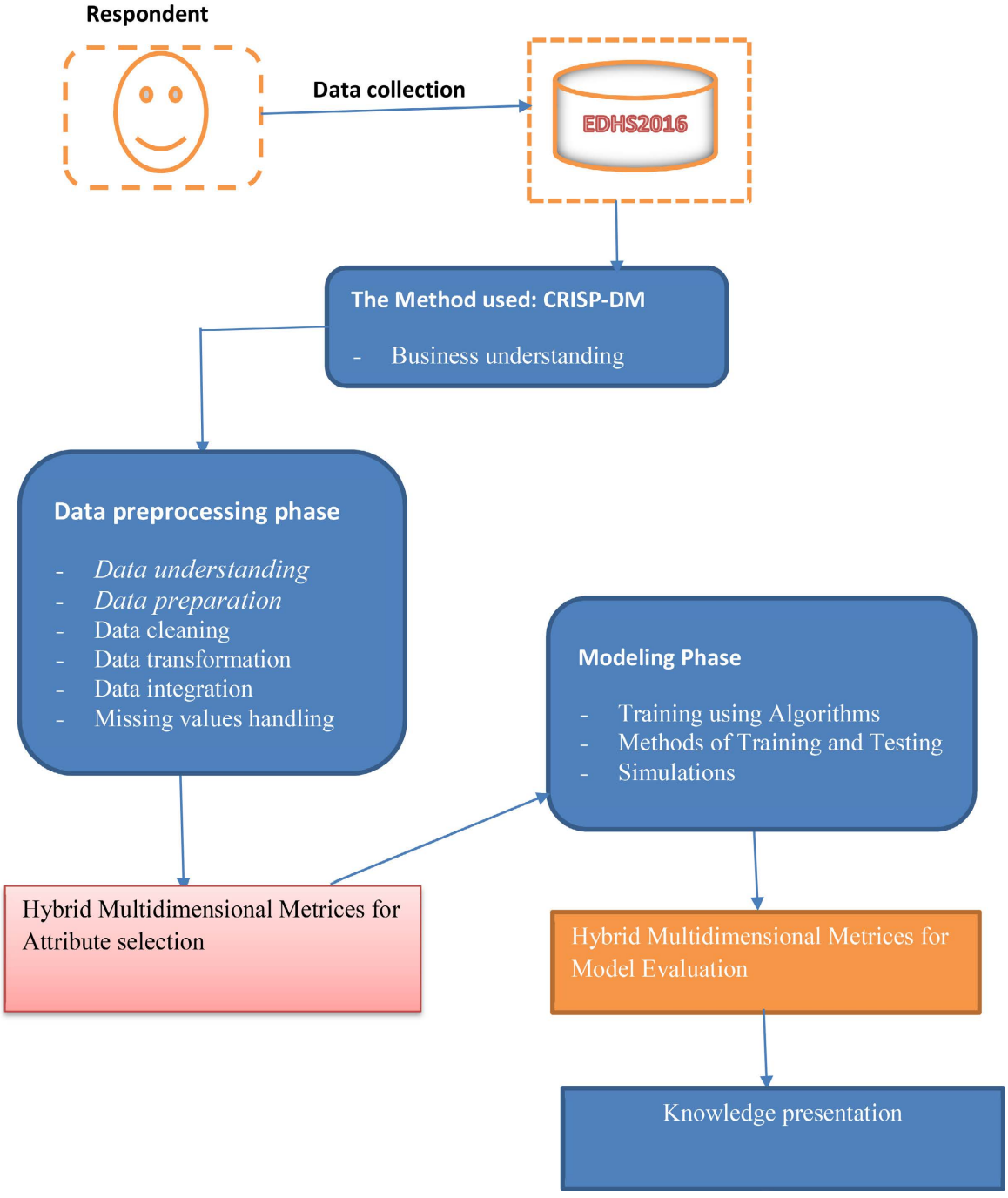
the target concept? To this end, we designed a novel approach which is more flexible for both feature selection and model selection criterion to be applied in predictive modeling that suit to our specific problem considering a hybrid multidimensional metrices of a client (based on user's requirements) into account unlike the classical approaches that mainly rely on unidimensional criteria to perform the tasks.

## 3. Problem Formulation and Framework

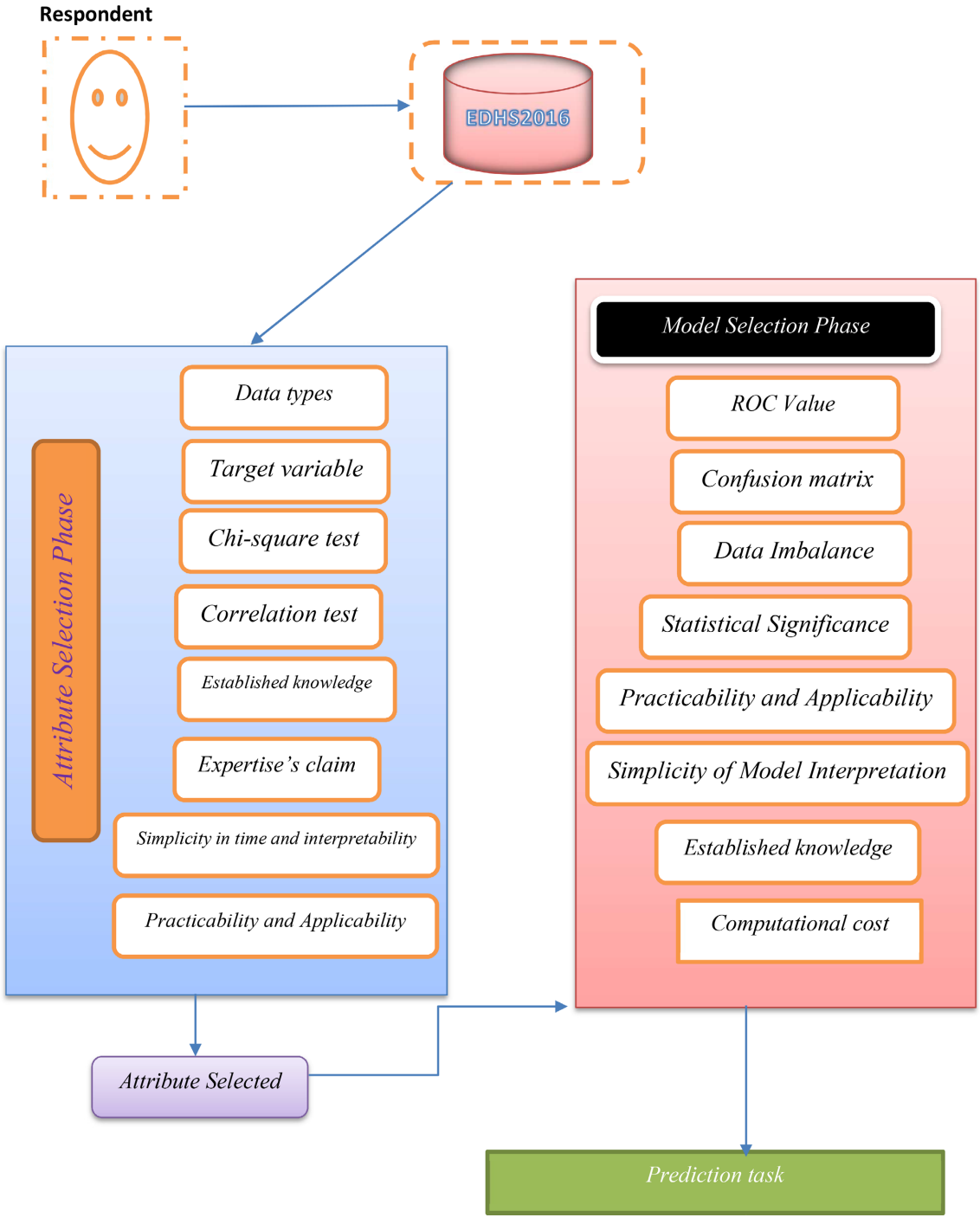
As the world grows in complexity, overwhelming us with the data it generates, data mining becomes the only hope for clarifying the patterns that underlie it [61]. However, both ML and DM processes requires high computational cost when dealing with high dimensional data as it may comprises several irrelevant features for the analysis. Therefore, feature selection is an essential technique used in DM and ML before any algorithms applied to train a classifier to avoid overfitting, improve model performance, provide faster and more cost-effective models. The selection of optimal features adds an extra layer of complexity in the modelling as instead of just finding optimal parameters for full set of features, first optimal feature subset is to be found and the model parameters are to be optimized [62]. Several feature selection algorithms have been proposed and studied for ML and DM applications and broadly classified into three: filter, wrapper and the embedded methods. The filter methods are suitable for high-dimensional dataset with good generalization and computational cost is also low as they are independent of learning classifiers when selecting the feature subset. Filter methods use inherent properties of the data such as information-based measures, distances, or statistical information to evaluate the quality of a selected subset. However, the accuracy results obtained might not be guaranteed [20] [21] [22]. The wrapper methods use the predictive accuracy of a predetermined learning algorithm to determine the relevance of the selected subsets, and the accuracy of the learning algorithms is usually high. However, the generality of the selected features is limited (tend to overfit on small training sets) and computationally expensive for high-dimensional data [23] [24] [25] [26]. And the embedded methods incorporate feature selection as a part of the training process and are usually specific to given learning algorithms of classifiers, and may be more efficient than the other two categories [27] [28] [29] as they are designed for specific classifiers [30].

To this end, there is a need for establishing a standard or framework more flexible as it comprises multidimensional metrics could help to view the complete picture of the prediction task. In line to this, different DM models were experimented to select the best fit model for the contraceptive dataset used in this paper. However, in the literature, there is no a such fixed guideline or rule to be adopted to pick the best model for the problem. Moreover, many predictive algorithms didn't perform well with large feature spaces as they could possibly irrelevant or redundant to the target variable. In this paper, we designed a hybrid

multidimensional metrices framework to predictive modeling for performance evaluation and feature selection to address the challenges encountered in feature selection and to pick the best model among the available models being used in DM and ML. CRISP-DM method was applied for this study for the reason that it has additional features on understanding the business perspective and its deployment [63]. It begins from understanding the business and ends with the deployment of the system. **Figure 1(a)** depicts the architecture of the KDD process




(a)




(b)

**Figure 1.** (a) The architecture of the KDD processes of the Proposed Hybrid Multidimensional Metrics for Feature Selection and Model Evaluations viewed entirely in one component. (b) Architecture of the Hybrid Multidimensional Metrices for Feature selection and Model selection in Predictive Modeling.

of the proposed hybrid multidimensional metrices for feature and model selection in predictive modeling comprises of six essential components: the method used, data preprocessing phase, feature selection, modeling and evaluation phase,

and the knowledge representation. Figure 1(b) demonstrates the architecture of the hybrid multidimensional metrices for feature and model selection in predictive modeling comprises of two major parts: feature selection and model selection phases. Their details are provided in the following sections.

### 3.1. Feature Selection

Several researches recently have studied both feature selection and clustering together with a single or unified criterion [64]. It has been stated that the importance of selecting of features in any data mining task; that the abundance of potential features constitutes a serious obstacle to the efficiency of most learning algorithms [65]. Popular methods such as k-nearest neighbor, C4.5, and back propagation are slowed down by the presence of many features, especially if most of these features are redundant and irrelevant to the learning task [65]. K-nearest neighbors (KNN) [66] is a simple and easy-to-implement supervised machine learning algorithm that can be used for both classification and regression problems. C4.5 algorithm [65] is a decision tree classifier used in data mining to generate a decision based on a certain sample of data, either univariate or multivariate predictors. Backpropagation is an algorithm used to train artificial neural networks by computing the gradient of the loss function with respect to the weights of the network. It works by performing a forward pass through the network, computing the output of each neuron, and then computing the error between the predicted output and the actual output. The error is then propagated backward through the network, and the weights of the network are adjusted to minimize the error. In this paper, we applied the proposed hybrid multidimensional criterion to aggregate features from a single or multiple source to create a target dataset which are pertinent to the data mining goals.

**Hybrid Multidimensional Metrics for Feature Selection**

Table 1 below illustrates a multidimensional metrics designed for feature selection to predictive modeling. The proposed metrics considered multiple dimensions of the feature whether to retain it for further analysis or not including the following criterion but not limited to: identifying data types of a feature, labeling target feature as either categorical or continuous, propose a statistical measure computed to test the relationship between the two variables for retaining or removing the feature, chi-square test computed to test the independence between the two features, consistency to the established knowledge, expertise's claim, simplicity in time and interpretability and practicability and applicability of the features are required. Given a national dataset or high dimensional data $D_B$ with the attributes of $X_k\left(k=1,2,\cdots,N\right)$ and detailed information was collected on background characteristics of the respondents based on a nationally representative sample that provides estimates at the national and regional levels. In this paper, the features would be therefore selected based on the proposed hybrid metrics for feature selection pertinent to the data mining goals. Therefore, a new target dataset was prepared for predictive task purpose and the re-

sponse feature is the contraceptive methods use ($CU$) which is a binary outcome. And **Algorithm 1** below depicts the pseudo code for feature selection.

### 3.2. Missing Value Handling

Missing values and their problems play important role in the data cleaning

**Table 1.** A Multidimensional metrices designed for Feature selection to predictive modeling.

| *Criterion* | *Measures* | *List of Features* |
|---|---|---|
| Data types | Categorical or continuous | $A_1$ |
| Target feature | Categorical or continuous | $A_1$ |
| Correlation test | Continuous | . |
| Chi-square test | Categorical | |
| Established knowledge | (Positive, Negative, Neutral, Unstudied) | . |
| Expertise's claim | (State in scientific manner) | . |
| Simplicity in time and interpretability | Meaningful and clarity | |
| Applicability and Practicality | Model's direct impact on the domain | $A_k$ |

**Algorithm 1.** A hybrid multidimensional metrics for feature selection.

**Input:** 1) Load national database $D_B$; 2) Response feature $CU$; 3) The number of respondents $N$;

**Output:** Target dataset $T_{Atr}$

1. $T_{Atr} = [\ ]$
2. **for** $k = 1, 2, \cdots, n$ **do**
3. Data types: Identify as categorical or continuous
4. Target variable: Identify as categorical or continuous
5. Apply data transformation when appropriate
6. **for** continuous variables **do**
7. Compute Correlation test
8. **end for**
9. **for** categorical variables **do**
10. Compute Chi-square test using Equation (5)
11. **end for**
12. Established knowledge: Identify as Positive, Negative, Neutral, Unstudied
13. Expertise's claim: State in scientific manner
14. Simplicity in time and interpretability: Identify as yes or no
15. Practicability and Applicability: Identify as yes or no
16. **end for**
17. Obtain the Attribute vector of the k$^{th}$ respondent
18. return $T_{Atr}$

process. Several methods have been proposed so as to process missing data in datasets and avoid problems caused by it. When the dataset is small or the number of missing fields is large, not all records with a missing field can be deleted from the sample. Moreover, the fact that a value is missing may be significant itself. A widely applied approach is used to calculate a substitute value for missing fields, for example, the median or mean of a variable [67]. In this paper, for the categorical variable, the missing values were replaced by the modal value of the variable [68]. All features with five percent missing values (5%) selected for further analysis and otherwise discarded from analysis. In this paper, WEKA preprocessing techniques such as replace missing value (using the most frequent (modal) value methods) was used to handle missing values.

### 3.3. Data Transformation and Reduction

Data mining often requires data integration or the merging of data from multiple data sources [69]. In data transformation; the collected attributes were transformed into forms which are appropriate for data mining tools. The process of data transformation included feature construction, where new features were constructed and added from the given set of features to help the mining process [64] [70]. In order to make the analysis procedures manageable and cost-effective the data needed to be reduced. Data reduction techniques include data discretization which is one of data transformation methods used to reduce the number of values for a given continuous attribute by dividing the range of the feature into intervals [63] [64]. In this paper, some features were discretized to reduce the unlike values of the features to obtain knowledge (pattern) and to make the dataset suitable for data mining tools. Almost all the selected features have been transformed from their original state in such a way that could be easily understandable and interpretable. For instance, a feature of ethnicity had 46 distinct values but later converted into ten distinct categories as: Afar, Guragie, Tigrean, Amara, Somalie, Sidama, Nuwer, Welaita, Oromo and Others.

### 3.4. Methods of Training and Testing

In data mining predictive models, the classifiers rely on being trained before they can reliably be used on new data [71]. The more instances the classifier is exposed to during the training phase, the more reliable it will be as it has more experience. However, once trained, we would like to test the classifier too, so that we are confident that it works successfully. It has been also stated that, in order to predict the performance of a classifier on new data, we need to assess its error rate on an independent test set that played no part in the formation of the classifier [72]. The standard way of predicting the error rate of a learning technique is to use stratified 10-fold cross-validation. The data is divided randomly into 10 parts in which the class is represented in approximately the same proportions as in the full dataset. Each part is held out in turn and the learning scheme trained on the remaining nine-tenths; then its error rate is calculated on

the holdout set. Thus, the learning procedure is executed a total of 10 times on different training sets. Finally, the 10 error estimates are averaged to yield an overall error estimate [72]. The WEKA 3.7 tool provides a test options to test on the same set the classifier is trained on (use training set), to test on a user-specified test data (Supplied test set), test on k-fold cross validation, and to train on a percentage of the data and test on the remainder (percentage split). In this paper, 10-fold cross validation was used for the prediction task.

### 3.5. Methods of Analysis and Evaluation of the Models

The output of several experiments of the classification models were analyzed and evaluated in terms of the details of the hybrid multidimensional metrics listed below. The complexity of each model in terms of the number of trees and leaves had also been evaluated. Furthermore, the models were evaluated using F-measures to test their statistical significance at 5% level of significance to be used for prediction purposes. In this paper, we designed a hybrid multidimensional criterion for model selection.

### 3.6. Methods of Training and Testing

**1) The ROC Curve**

ROC (Receiver Operating Characteristic): ROC curves are a useful tool for comparing classification models [73]. The performance of the classifiers with different parameters was also compared by examining their ROC curve. The ROC curve shows the trade-off between the true positive rate (*i.e.*, true contraceptive user) and the false positive rate (false contraceptive user) for a given model. Moreover, models can be compared with respect to their speed, robustness, scalability, and interpretability which may have an influence on the model [52]. Besides, the ROC curve is a two-dimensional plane; the vertical axis (Y-axis which denotes the sensitivity) represents the true contraception user rate (TCUR) and the horizontal axis (X-axis which denotes 1-specificity) represents the false-contraception user rate (FCUR).

**2) The Confusion Matrix**

Previous studies on data mining and machine learning techniques revealed that, a confusion matrix was often used to measure performance of the models in terms of accuracy, sensitivity and specificity it achieved as depicted in **Table 2** below. The confusion matrix is a matrix representation of the classification results. In a two-class prediction problem the upper left cell denotes the number of samples classified as true while they are true (*i.e.*, true users), and lower right cell denotes the number of samples classified as false while they were actually false (*i.e.*, true false or true not users). The other two cells (lower left cell and upper right cell) denote the number of samples misclassified. Particularly, the lower left cell denotes the number of samples classified as false while they were actually true (*i.e.*, false negative or false non-users), and the upper right cell denotes the number of samples classified as true while they were actually false (*i.e.*, false

**Table 2.** Summary of two-class prediction problem.

| | | *Predicted value for contraceptive use* | | |
|---|---|---|---|---|
| *Actual value of current contraceptive use* | | *No* | *Yes* | *Total* |
| | *No* | TN | FP | TN + FP |
| | *Yes* | FN | TP | FN + TP |
| | Total | TN + FN | FP + TP | Grand = TP + FP + TN + FN |

positive or false contraceptive users). Once the confusion matrixes were constructed, the accuracy, sensitivity and specificity of each model was calculated using the respective formulas presented below. In summary, there are three measures for model performance evaluations, namely: -*accuracy*, *sensitivity and specificity*.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

a) Data imbalance problem: Handling and test the effect of data imbalance for the Target variable.

b) Model's Statistical Significance: Measured using F test or paired test the overall significance of the model.

c) Practicability and Applicability: This is significance of the model from its direct impact on the institution or determined by the manager. This is quite different from statistical significance.

d) Simplicity of Model Interpretation: Model's clarity of rules extraction from user's side.

e) Established knowledge: identified as: Positive, Negative, Neutral (borderline significance), Unstudied.

f) Computational cost: Algorithm's simplicity in terms of time and space.

g) Hybrid Multidimensional Metrics for Model Selection: Considers multidimensional scenarios.

**3) Hybrid Multidimensional Metrics for Model Selection**

Given a national dataset or high dimensional database $D_B$ with the features of $X_k\left(k = 1, 2, \cdots, N\right)$ detailed information was collected on background characteristics of the respondents. Given that the features were selected based on the proposed hybrid metrics as described in **Table 2** for feature selection pertinent to the data mining goals, a new target dataset was prepared for predictive task and the response feature is the contraceptive methods use (*CU*) which is a binary outcome. Now, suppose there are n requirements received from the user or organization where the model intended to be used by and we define: $RI_i = \left\{R_i\right\}_{i=1}^{n}$ to

denote the set of user's requirements for each model, where $\{R_i\}_{i=1}^{n} \in \{0,1\}$. The user's requirement indicators are a binary outcome when the value is 1, the corresponding user requirement is selected otherwise unselected. And suppose the list of models to be compared against the user's requirements are given as: $M_k^{i,R} = \left[M_1^{i,R}, M_2^{i,R}, \cdots, M_k^{i,R}\right]^{\mathrm{T}}$. And we proposed a hybrid multidimensional metrices used to compute the overall significance of the model taking both the effects of the user's requirements and their corresponding weights of their importance basically assigned based on the user's requirements and defined as:

$$HMM(m,r) = \frac{\sum_{r=1}^{R}\sum_{m=1}^{k} w_i \cdot RI_i}{M} \tag{4}$$

The higher $HMM(m,r)$ indicates the overall significant model that comprises almost all requirements of the user unlike the classical metrics that used one criterion to pick the best fit model. **Algorithm 2** illustrates the pseudo code for the hybrid multidimensional metrics of model selection which is provided below.

**Algorithm 2.** Proposed hybrid multidimensional metrics for model selection.

**Input:** 1. Target dataset $T_{Atr}$; 2) Response variable $CU$; 3) The number of respondents $N$;

**Output:** 1. Model selected $M_k$; 2. The metrics $HMM(m,r)$; 3) Knowledge representation: $CU$

1. Start
2. **for** $m = 1,2,\cdots,k$ **do**
3. Compute ROC Values for each model
4. Assign weight for each model based on user's requirements
5. Compute Confusion matrix for each model
6. Assign weight for each model based on user's requirements
7. Test the effect of Data Imbalance problem for each model
8. Test Statistical Significance using F-test
9. Practicability and Applicability: Identify as yes or no
10. Simplicity of Model Interpretation: Identify as yes or no
11. Established knowledge: Identify as Positive, Negative, Neutral, Unstudied
12. Computational cost: identify as High, Moderate, and Low
13. Compute the hybrid multidimensional metrics using equation (4)
14. Find $m,r$ s.t $HMM_{mr} = \max\left(HMM(m,r)\right)$
15. **end for**
16. return $M_k$; $HMM(m,r)$; $CU$

Table 3 below described the criteria and measures of the hybrid multidimensional metrics designed for model selection of predictive modeling.

## 4. Results and Discussions

### 4.1. Chi-Square Test Analysis

In this paper, we used EDHS 2016 as a source of data applied to contraceptive use meeting to address the main challenges of both feature and model selection encountered in predictive modeling task. A chi-square test ( $\chi^2_{cal}$ ) was used to test the association between each feature with the contraceptive use with the purpose to retain it in the model or not for further analysis of the prediction task (Table 4). Suppose we have a population consisting of observations having two attributes or qualitative characteristics say A and B. If the attributes are independent then the probability of possessing both A and B is P (A) * P (B). Where: P (A) is the probability that a member has attribute A and P (B) is the probability that a member has attribute B. The chi-square procedure test is used to test the hypothesis of independency of two attributes. For instance, we may be interested whether the contraceptive user or non-user is independent of marital status or not. The chi-square statistics is given by:

$$\chi^2_{cal} = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left[o_{ij} - e_{ij}\right]^2}{e_{ij}} \sim \chi^2_{(r-1)\times(c-1)} \tag{5}$$

where: $o_{ij}$ = observed number of units in category $i$ of A and category $j$ of B and $e_{ij} = \dfrac{R_i \times C_j}{n}$ = Expected number of units in category $i$ of A and category $j$ of B with $R_i = i^{\text{th}}$ row total and $R_i = j^{\text{th}}$ column total. Accordingly, the socio demographic factors: marital status, religion, wealth index, region, place of residence, ethnicity and highest education level were found to be significantly associated with contraceptive methods use (P-value < 0.000). However, the missing

Table 3. A hybrid multidimensional metrics designed for model selection to predictive modeling.

| Criterion | Measures | List of Models |
|---|---|---|
| ROC value | Categorical or continuous | $M_1$ |
| Confusion matrix | Accuracy, Specificity and Sensitivity | $M_2$ |
| Data Imbalance Problem | Compare the differences | . |
| Statistical Significance | F-test | . |
| Practicability and Applicability | Model's direct impact on the domain | . |
| Simplicity of Model Interpretation | Meaningful and clarity | |
| Established knowledge | (Positive, Negative, Neutral, Unstudied) | . |
| Computational cost | In time | $M_k$ |

**Table 4.** Statistical association of socio-demographic attributes related to Contraception use using Chi-square test, EDHS 2016.

| No | Features | Category | Contraception Use | | P-value |
|---|---|---|---|---|---|
| | | | Yes | No | |
| 1 | Marital status | Divorced | 122 | 756 | 0.000 |
| | | Married | 2887 | 6715 | |
| | | Living with partner | 93 | 129 | |
| | | No longer living with partner | 52 | 200 | |
| | | Never in union | 132 | 4146 | |
| | | Widowed | 26 | 425 | |
| 2 | Religion | Catholic | 22 | 69 | 0.000 |
| | | Protestant | 670 | 2144 | |
| | | Orthodox | 1845 | 4568 | |
| | | Muslim | 761 | 5448 | |
| | | Traditional | 4 | 84 | |
| | | Others | 10 | 62 | |
| 3 | Highest level of Education | No education | 5686 | 1347 | 0.000 |
| | | Primary | 4040 | 1173 | |
| | | Secondary | 1782 | 456 | |
| | | Higher | 863 | 336 | |
| 4 | Wealth Index combined | Poorest | 3562 | 332 | 0.000 |
| | | Poorer | 1610 | 436 | |
| | | Middle | 1502 | 500 | |
| | | Richer | 1498 | 544 | |
| | | Richest | 4199 | 1500 | |
| 5 | Husbands Education level | No education | 901 | 3530 | 0.000 |
| | | Primary | 1141 | 1913 | |
| | | Secondary | 494 | 732 | |
| | | Higher | 423 | 597 | |
| | | Don't know | 21 | 72 | |
| | | Missing | 5527 | 332 | |
| 6 | Ethnicity | Afar | 13 | 934 | 0.000 |
| | | Guragie | 153 | 502 | |
| | | Tigrean | 469 | 1436 | |
| | | Amara | 1186 | 2502 | |
| | | Oromo | 743 | 2868 | |

**Continued**

| | | |
|---|---|---|
| Welaita | 71 | 251 |
| Sidama | 144 | 211 |
| Nuwer | 4 | 280 |
| Somalie | 20 | 1443 |
| Others | 509 | 1944 |

values of features that exceeds 5% were discarded from further analysis. For instance, husbands' education levels were discarded from analysis (Table 4). Table 4 indicates that for an attribute of marital status with six levels; respondents were asked about contraception use and the difference between the categories of marital status was tested using P-value. The association between marital status and contraception use were found to be significant (P-value < 0.000). One can also understand that the effect of marital status at every levels of category on contraception use is different. Hence, marital status would be included as potential predictor in the model.

Table 5 illustrates that the statistical association of knowledge features related to contraception use were also assessed. Accordingly, media exposure, contraception intent use, heard family planning, ever heard of AIDS, ever heard of STI, recent sexual activity and knowledge ovulatory cycle were found to be significantly associated with contraceptive method use (P-value < 0.000). And HIV transmitted during pregnancy was discarded from analysis (Table 5). An attribute of media exposure with two levels (*Yes* or *No*); respondents were asked about contraception use. And the differences between the users and non-users of media exposure against contraceptive use were found to be statistically significant. Hence, media exposure was included as potential predictor to train the model.

## 4.2. Feature Pattern Analysis

The pattern analysis was done to understand the effect of each feature at every level of category related to contraception use. Of the study participants, almost 22% were in the age group of 15 to 19 years and of the total participants only 1.35% was reported to be contraceptive users (Figure 2). Similarly, 18% were in the age group of 20 to 24 years and of the total participants only 4.5% were contraceptive users. Besides, 18% were in the age group of 25 to 29 years and of which 30% have been reported as contraceptive users. It has also been reported that both age groups 25 to 29 and 30 to 34 years had the higher proportions of contraceptive users among other age groups. The pattern indicates that participants both in the age groups of 15 to 19 and 40 to 49 years of age the proportion of contraceptive uses among these groups got declined (Figure 2). The two lines are not parallel hence it indicates there are variations on contraceptive users among the different age groups of the respondents.

**Table 5.** Statistical association of knowledge attributes related to Contraception use using Chi-square test, EDHS 2016.

| No | Attributes | Category | Contraception Use | | P-value |
|---|---|---|---|---|---|
| | | | No | Yes | |
| 1 | Media exposure | No | 6715 | 1412 | 0.000 |
| | | Yes | 5656 | 1900 | |
| 2 | Contraception Use Intention | Doesn't Intend to use | 6708 | 0 | 0.000 |
| | | Non-user intends to use later | 5663 | 0 | |
| | | Using modern method | 0 | 3217 | |
| | | Using traditional method | 0 | 95 | |
| 3 | Knowledge of Ovulatory Cycle | After period ended: 2 | 3085 | 936 | 0.000 |
| | | At any time: 5 | 2557 | 530 | |
| | | Before period begins: 4 | 879 | 274 | |
| | | Middle of cycle: 3 | 2694 | 1005 | |
| | | During her period: 1 | 374 | 108 | |
| | | Don't know: 8 | 2782 | 459 | |
| 4 | Heard Family planning | No | 8223 | 1788 | 0.000 |
| | | Yes | 4148 | 1524 | |
| 5 | Ever heard of STI | No | 1170 | 75 | 0.000 |
| | | Yes | 11201 | 3237 | |
| 6 | Recent sexual activity | Active last 4 weeks | 4832 | 2723 | 0.000 |
| | | Never had sex | 3709 | 12 | |
| | | No active: No postpartum abstinence | 2937 | 517 | |
| | | No active: Postpartum abstinence | 893 | 60 | |
| 7 | Ever heard of AIDS | No | 1233 | 81 | 0.000 |
| | | Yes | 11138 | 3231 | |

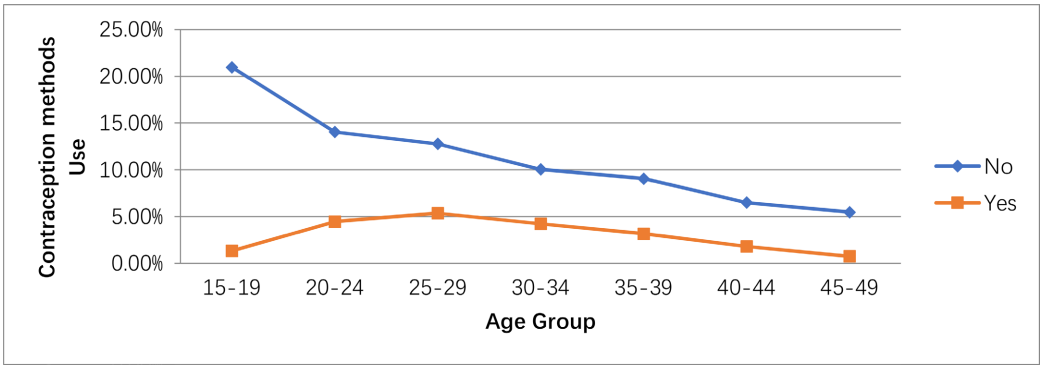



**Figure 2.** Patterns of Contraception methods use by age group, EDHS 2016.

Of the study participants, who have been asked whether contraceptive methods used in the survey, 12.06%, 11.79%, 11.63%, 10.96% and 10.72% were found to be from Oromiya, SNNP, Addis Ababa, Amhara, and Tigray regions respectively (Figure 3).

Among the study participants who were higher in their educational status (7.64%), only 2.14% has been reported that as contraceptive users. However, participants with no education (45%) have reported the least proportion (8.59%) of contraception methods use. One can see the gap for contraceptive use from the graph below for the participants with no education is huge. The pattern for contraception use gets decrease as educational level get decrease (Figure 4).

Among the study participants who were married (51%), only 18% has been reported that as contraceptive users. However, participants who never been in union (27%) have reported the least proportion (0.84%) of contraception methods use (Figure 5).

Of the study participants, 65% were found to be from rural residents and only 13% of rural residents reported as contraceptive users. On the other side, only

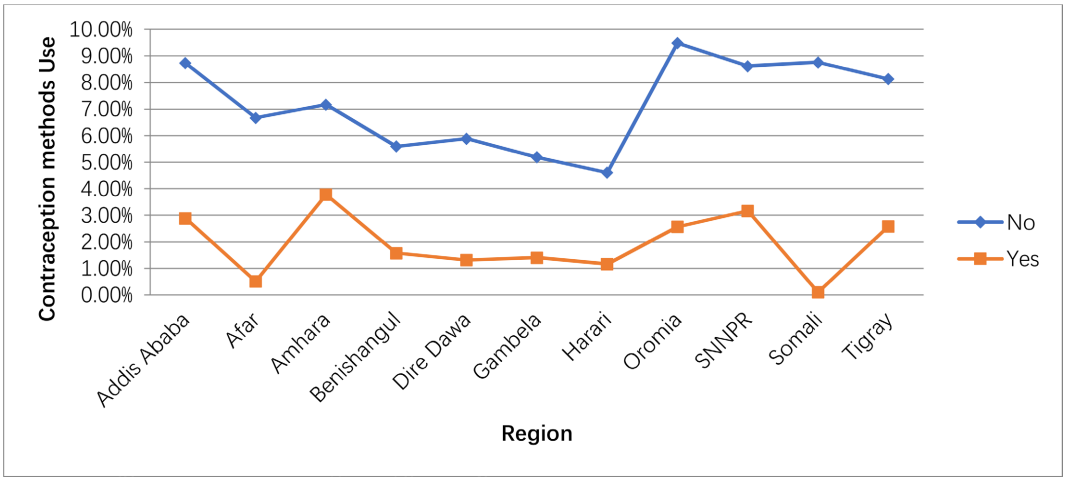



Figure 3. Patterns of contraception methods use by region, EDHS 2016.

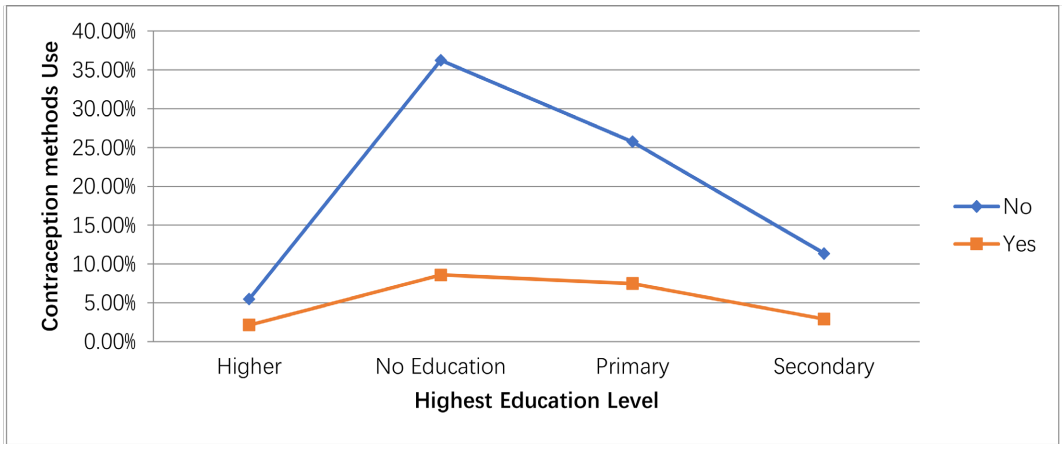



Figure 4. Patterns of contraception methods use by highest education level, EDHS 2016.

8.5%of the urban residents were reported as contraceptive users (Figure 6).

Among the study participants, Muslims and Orthodox constituted 40% and 41% respectively of which only 5% and 12% reported as contraceptive users. However, participants who are Catholic and traditional religion followers have shown unique pattern unlike the huge gap which is observed among other religions (Figure 7).

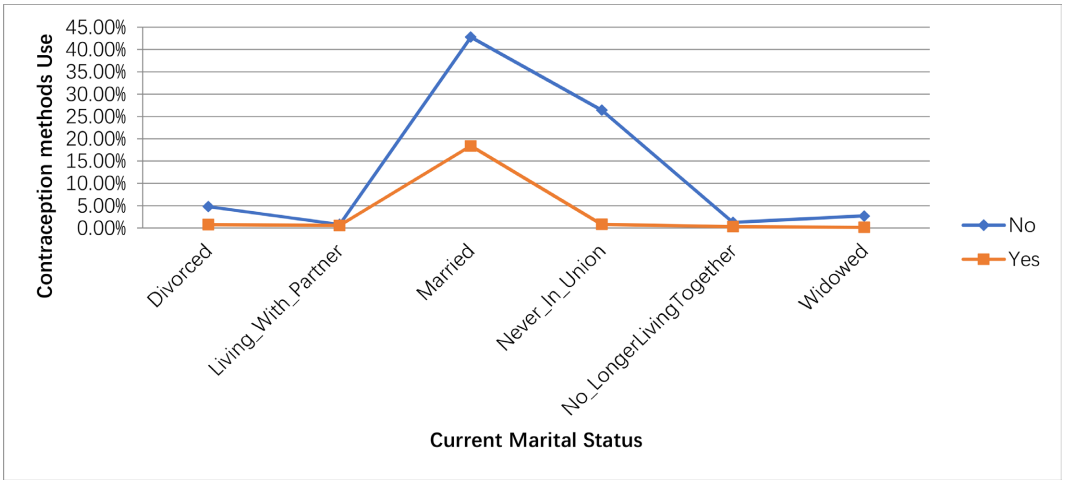



Figure 5. Patterns of contraception methods use by current marital status, EDHS 2016.

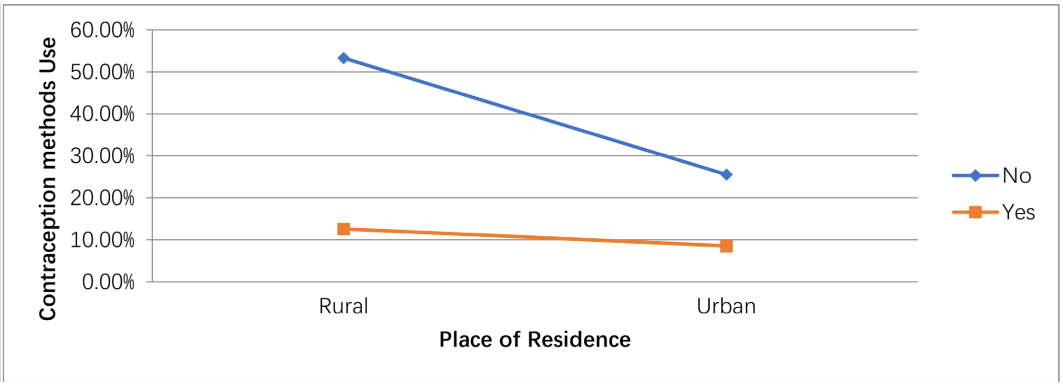



Figure 6. Patterns of contraception methods use by place of residence, EDHS 2016.

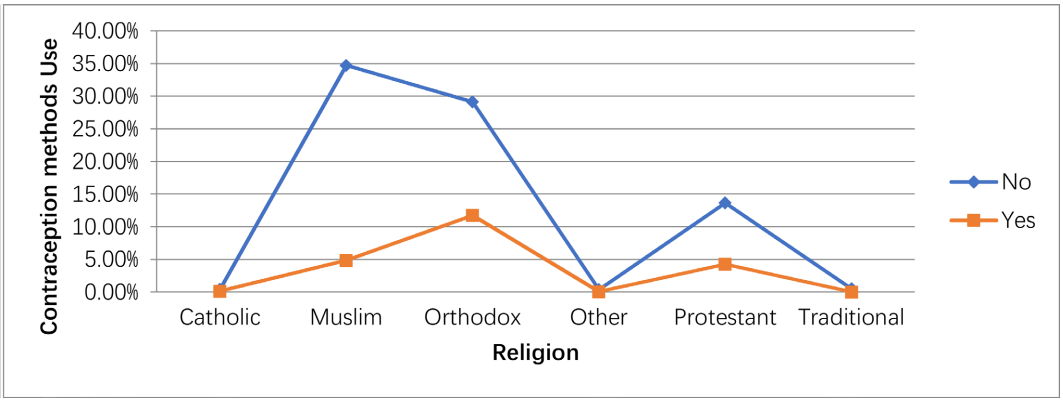



Figure 7. Patterns of Contraception methods use by place of residence, EDHS 2016.

### 4.3. Experimentation

The classifiers were used 15,683 instances for training the predictive models applied to contraceptive users. Different data mining algorithms such as: decision tree (J48, random tree, and random forest), Naïve Bayes, and artificial neural network (ANNs) algorithms were used to train the classifiers. Five of the classifiers were trained with two scenarios and with varying testing parameters. The performances of the data mining models were evaluated using 10 k cross validation test option as it is the standard for controlling a bias. Two scenarios were considered with respect to the attribute selections adopted to train the models. These are the classical and the proposed approaches.

1) In classical approach, we used both selection feature and search methods algorithms from the available Weka packages. Accordingly, five attributes (Ethnicity, knowledge any method, current marital status, recent sexual activity and ever been tested for HIV) have got selected using classifier subset evaluator algorithm, and both bestFirst and GreedyStepwise search methods.

2) In the second approach, we applied the hybrid multidimensional metrics approach for the feature selection and accordingly "18 selected features" (socio-demographic determinants, knowledge related to contraception use, knowledge related to AIDS and/or STI, exposure to mass-media, and knowledge on family planning) were used in all experimentations. The current contraception methods use (CCMU) is a binary outcome which is the response variable of the study. List of the features used for this study are presented as shown below in Table 6.

The efficiency of the predictive models was evaluated based on the proposed hybrid multidimensional metrics for model selection as can be shown in Table 7 below. These performance measures are used or designed to be used to fulfil the user's requirements.

Figure 8 depicts that the Artificial Neural Network (ANNs) takes a sample of features (individual inputs $p_1, p_2, \cdots, p_R$) to build the predictive modeling for contraceptive use for demonstration purpose. Each individual feature is weighted by the corresponding elements $w_{1,1}, w_{1,2}, \cdots, w_{1,R}$ of the *weight matrix* ***W***. The ANNs predictive model has been trained with: 1) no hidden layers, 2) two hidden neurons, and 3) with two layers hidden neurons if improvement of prediction power could gain. However, the results for the three cases of layer configurations using the ANNs were found to be similar. Therefore, we recommend the ANNs to model the contraceptive use with no hidden layers for simplicity of model interpretation purpose.

### 4.4. Comparison Analysis of the Classifiers

#### 1) The ROC Curve Analysis

The ROC value for the data mining algorithm of Naïve Bayes used for modeling of contraception use was found to be 85.1%. The ROC curve analyses for the Naïve Bayes displayed below showed that the curve moves sharply up from zero

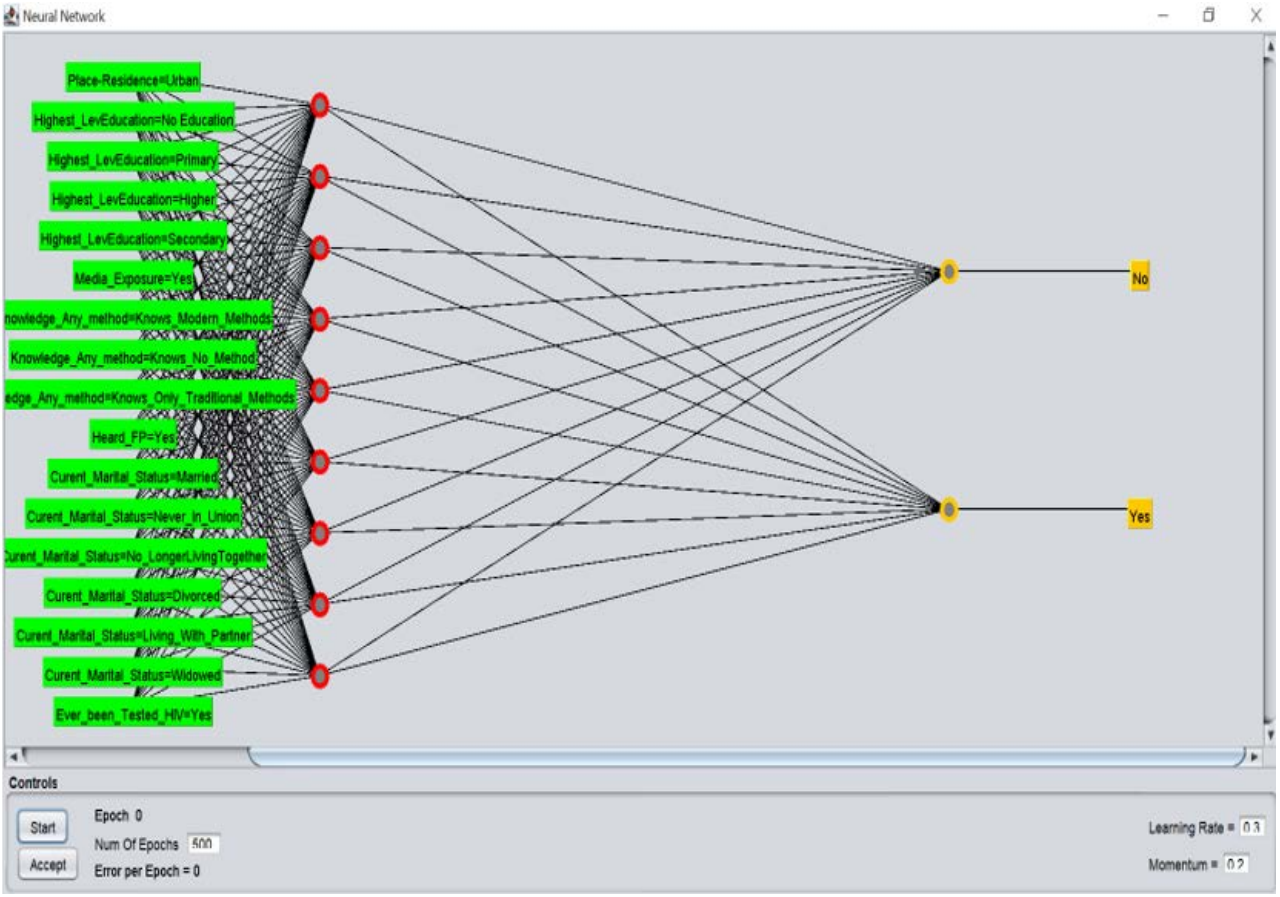
showing that there are higher true tested than false tested rates. Then the curve starts to become more horizontal as it encounters less true tested and more false

**Table 6.** List of possible attributes for predicting the model for contraceptive use, EDHS 2016.

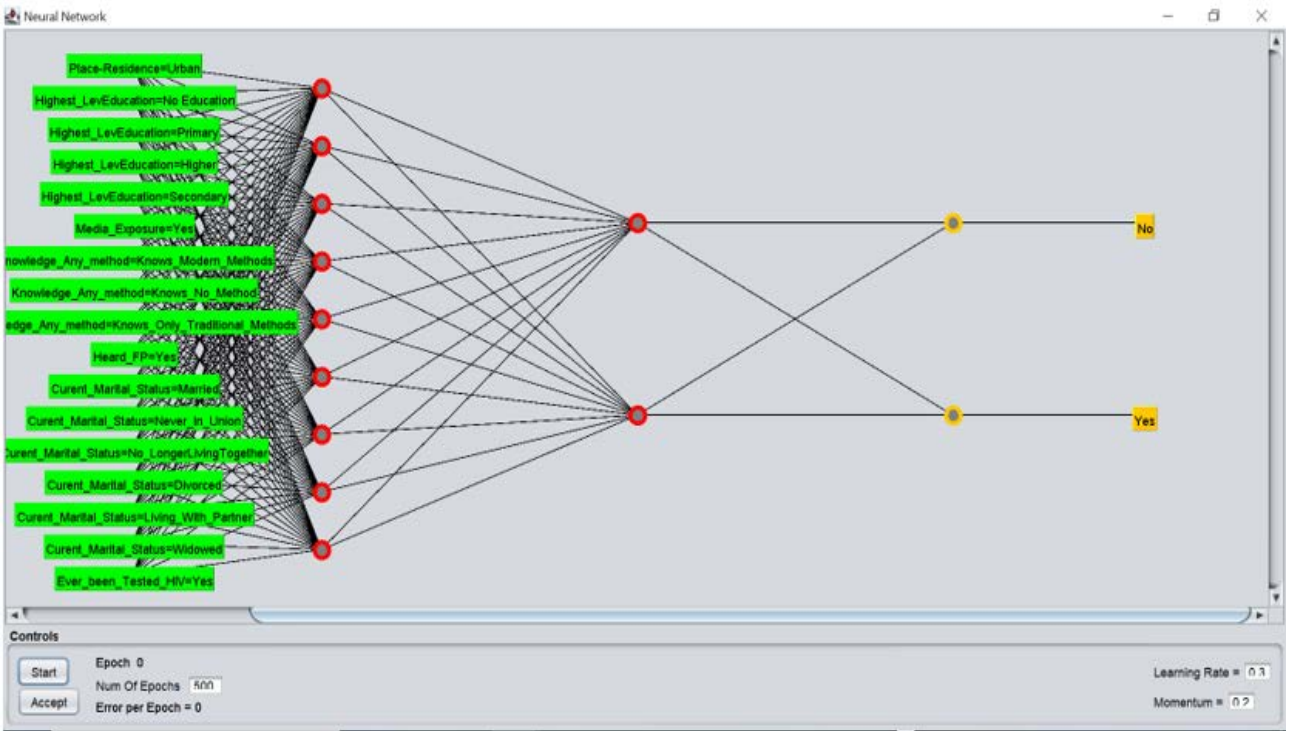
| Rank | Attributes | Contribution of each attribute to the model | Data type | Distinct values | % of missing values |
|---|---|---|---|---|---|
| 1 | Recent_Sexual_Activity | 0.12728 | Tex | 4 | 0 |
| 2 | Curent_Marital_Status | 0.08209 | Tex | 6 | 0 |
| 3 | Ethnicity | 0.06039 | Numeric | 46 | 0 |
| 4 | Num_Living_Children | 0.05981 | Tex | 4 | 0 |
| 5 | Ever_been_Tested_HIV | 0.04436 | Tex | 2 | 0 |
| 6 | AgeGroup | 0.04222 | Tex | 9 | 0 |
| 7 | Region | 0.04219 | Tex | 11 | 0 |
| 8 | WI_Combined | 0.02728 | Tex | 5 | 0 |
| 9 | Religion | 0.02639 | Tex | 6 | 0 |
| 10 | Desire_For_More_Children | 0.02579 | Tex | 3 | 0 |
| 11 | Knowledge_Any_method | 0.01613 | Tex | 3 | 0 |
| 12 | Ever_Heard_AIDS | 0.01125 | Tex | 2 | 0 |
| 13 | Ever_Heard_STI | 0.01086 | Tex | 2 | 0 |
| 14 | Knowledge_Ovulatory_Cycle | 0.01041 | Tex | 6 | 0 |
| 15 | Heard_FP | 0.00793 | Tex | 2 | 0 |
| 16 | Media_Exposure | 0.00654 | Tex | 2 | 0 |
| 17 | Place of residence | 0.00328 | Tex | 2 | 0 |
| 18 | Highest_LevEducation | 0.00255 | Tex | 4 | 0 |

**Table 7.** Summarization of various experimentations applying with different testing parameters.

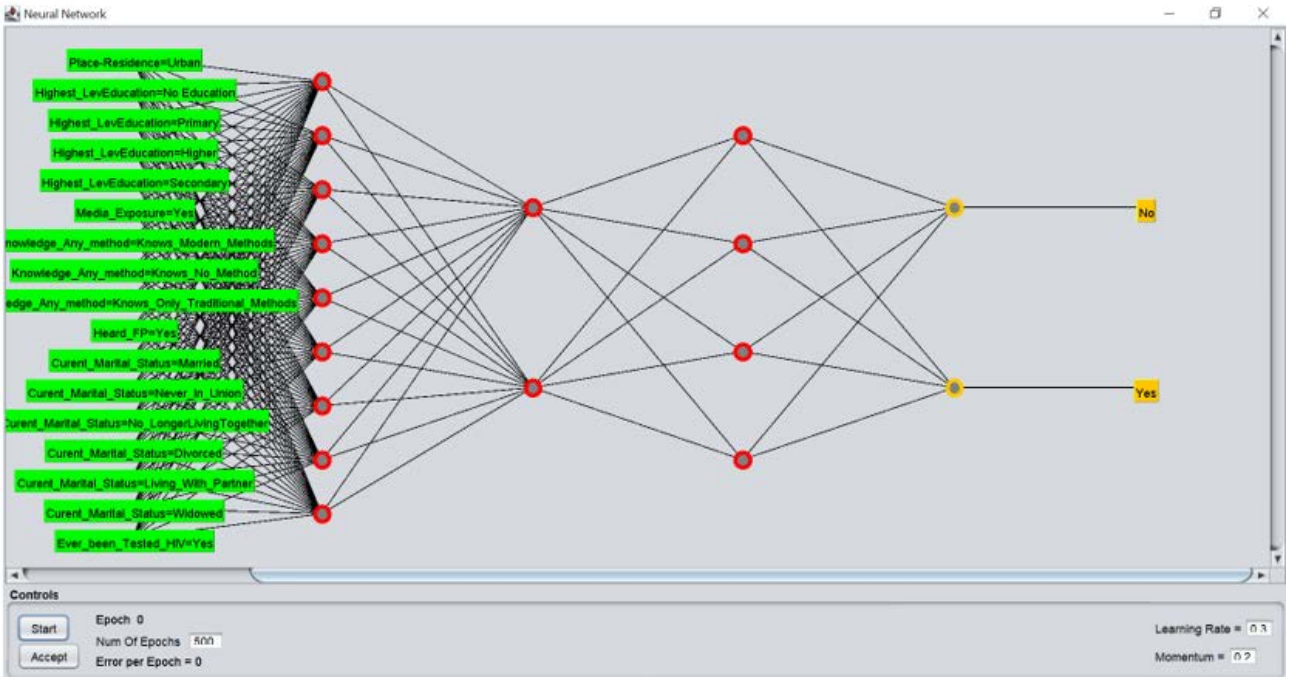
| S.No | Experimentation of models | Testing options | No. of attribute | Selection Attributes |
|---|---|---|---|---|
| **Scenario 1** | Naïve base | Training<br>Cross validation<br>Percentile | 5 | CfsSubsetEval: +BestFirst:<br>CfsSubsetEval: +GreedyStepwise |
| | Decision tree (J48) | | 5 | |
| | Random tree | | 5 | |
| | Random forest | | 5 | |
| | Artificial Neural Networks | | 5 | |
| **Scenario 2** | Naïve base | Training<br>Cross validation<br>Percentile | 18 | Proposed approach |
| | Decision tree (J48) | | 18 | |
| | Random tree | | 18 | |
| | Random forest | | 18 | |
| | Artificial Neural Networks | | 18 | |


Neural Network
Place-Residence=Urban
Highest_LevEducation=No Education
Highest_LevEducation=Primary
Highest_LevEducation=Higher
Highest_LevEducation=Secondary
Media_Exposure=Yes
nowledge_Any_method=Knows_Modern_Methods
Knowledge_Any_method=Knows_No_Method
edge_Any_method=Knows_Only_Traditional_Methods
Heard_FP=Yes
Curent_Marital_Status=Married
Curent_Marital_Status=Never_in_Union
Curent_Marital_Status=No_LongerLivingTogether
Curent_Marital_Status=Divorced
Curent_Marital_Status=Living_With_Partner
Curent_Marital_Status=Widowed
Ever_been_Tested_HIV=Yes
No
Yes
Controls
Start
Accept
Epoch 0
Num Of Epochs 500
Error per Epoch = 0
Learning Rate = 0.3
Momentum = 0.2

(a)


Neural Network
Place-Residence=Urban
Highest_LevEducation=No Education
Highest_LevEducation=Primary
Highest_LevEducation=Higher
Highest_LevEducation=Secondary
Media_Exposure=Yes
nowledge_Any_method=Knows_Modern_Methods
Knowledge_Any_method=Knows_No_Method
edge_Any_method=Knows_Only_Traditional_Methods
Heard_FP=Yes
Curent_Marital_Status=Married
Curent_Marital_Status=Never_in_Union
Curent_Marital_Status=No_LongerLivingTogether
Curent_Marital_Status=Divorced
Curent_Marital_Status=Living_With_Partner
Curent_Marital_Status=Widowed
Ever_been_Tested_HIV=Yes
No
Yes
Controls
Start
Accept
Epoch 0
Num Of Epochs 500
Error per Epoch = 0
Learning Rate = 0.3
Momentum = 0.2

(b)




(c)

**Figure 8.** Results for Neural Network (Multilayer Perception) classifier with some sample attributes, (a) with no hidden layers, (b) two hidden neurons, and (c) with two layers hidden neurons.

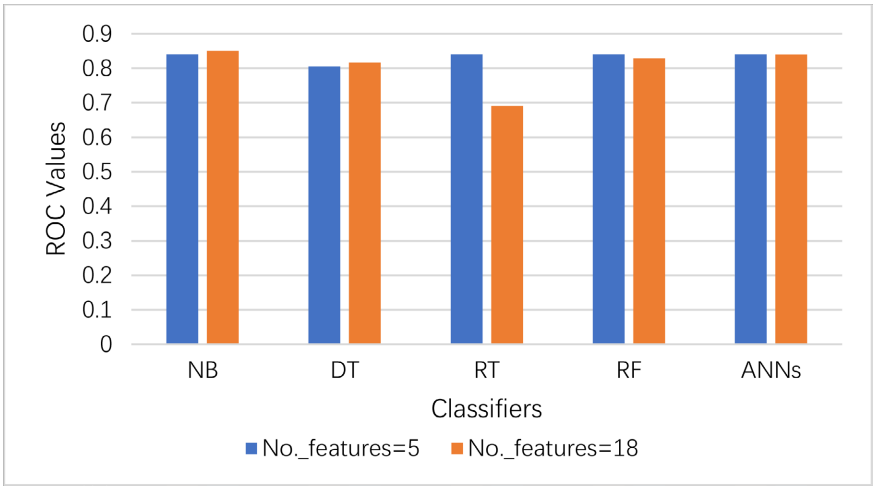



**Figure 9.** Analysis of ROC values for the classifiers for both scenarios.

tested rates. The area under the curve for the naïve Bayes model was found to be 85.2% (Figure 9).

**2) The Confusion Matrix**

Intensive experimentations with different testing parameter options (training test, cross validation and percentage) were done but comparison was done using cross validation (CV) test options only as it is a standard for controlling the bias. Accordingly, the results for Naïve Bayes with CV test option achieved an accuracy of 79.85%, a sensitivity of 58.78% and specificity of 85.49% were demon-

strated. But, the Naïve Bayes classifier achieved the minimum cost for time computation in second. Similarly, ANNs (Multiple Perceptron) classifier scored an accuracy of 80.24%, a sensitivity of 44.89% and specificity of 89.70% respectively associated with maximum cost for time computation in seconds. Moreover, the results for decision trees with algorithms of (J48, RT and RF) achieved accuracy better than the above-mentioned classifiers (NB and ANNs) as seen in (Table 8). If we simply see the performance of the model in terms of accuracy it achieved, one can observe that the decision tree of scenario two (J48) is the best model predictor (Table 9). However, we need to check whether these performance measures achieved by each model have a statistical significance at 5% level of significance for further analysis and for feature prediction purpose. And this objective of statistical testing model significance would be achieved using F-test in DM models (Table 10). The complete set of results used for comparison of each model performance was prepared in a tabular format (Table 10 and Table 11).

**3) Model Evaluation for Data Imbalance Problem**

**a) Data Imbalance Case**

In this paper, the receiver operator characteristics curve analysis (ROC curve)

Table 8. Comparison of performance of different Classifiers, scenario 1 (n = 5).

| Evaluation criteria's | Naïve Bayes | | Decision tree (J48) | | Decision tree (random tree) | | Decision tree (forest) | | Neural networks | | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Confusion matrix | 10,576 | 1795 | 11,676 | 695 | 11,586 | 785 | 11,586 | 785 | 11,097 | 1274 | No |
| | 1365 | 1947 | 2320 | 992 | 2235 | 1077 | 2232 | 1080 | 1825 | 1487 | Yes |
| Accuracy (%) | 79.85% | | 80.77% | | 80.74% | | 80.76% | | 80.24% | | |
| Sensitivity (%) | 58.78% | | 29.95% | | 32.51% | | 32.60% | | 44.89% | | |
| Specificity (%) | 85.49% | | 94.38% | | 93.65% | | 93.65% | | 89.70% | | |
| ROC (%) | 84.1% | | 80.5% | | 84.1% | | 84.4% | | 84.1% | | |
| Computations time in seconds | 0.01 | | 0.02 | | 0.09 | | 0.7 | | 61.16 | | |

Table 9. Comparison of performance of different Classifiers, scenario 2 (n = 18).

| Evaluation criteria's | Naïve Bayes | | Decision tree (J48) | | Decision tree (random tree) | | Decision tree (forest) | | Neural networks | | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Confusion matrix | 9680 | 2691 | 11,416 | 955 | 10,732 | 1639 | 11,376 | 995 | 10,964 | 1407 | No |
| | 868 | 2444 | 1810 | 1502 | 1845 | 1467 | 1847 | 1465 | 1680 | 1632 | Yes |
| Accuracy (%) | 77.30% | | 82.36% | | 77.78% | | 81.87% | | 80.32% | | |
| Sensitivity (%) | 73.79% | | 45.35% | | 44.29% | | 44.23% | | 49.27% | | |
| Specificity (%) | 78.25% | | 92.28% | | 86.75% | | 91.96% | | 88.63% | | |
| ROC (%) | 85.1% | | 81.7% | | 69.1% | | 85.5% | | 84.0% | | |
| Computations time in seconds | 0.00 | | 0.24 | | 0.07 | | 3.78 | | 518.2 | | |

was also used to measure the performance of the models. All the four classifiers using imbalanced data case have achieved ROC values much more than 81% except the random tree with 69.1%. If we simply see the performance of the model in terms of accuracy it achieved, one can observe that the decision tree (J48) is the best model predictor of the other two (Table 8 and Table 9). However, a paired two-tailed comparison was done using paired corrected test option to measure the difference of performances among the models in predicting the contraceptive use at 5% level of significance for further analysis and for future prediction purpose (Table 10). This objective of testing model significance would be achieved using F-measure in data mining models. The four data mining models (Decision tree (J48), Decision tree (random tree), decision tree (random forest) and Neural networks (MLP)) were compared against to the "Naïve Bayes" model given for the same number of inputs. Hence, all the models used in this paper are efficient enough (prediction power exceeds 77%) to predict the contraceptive methods use among women since all the models achieved the same F-measures. Unlike statistical value that uses P-value for measuring significance of an interest, WEKA uses three symbols ((v/ /*)) for measuring the differences of the models and represented as (v------the difference in performance of the models is considered as victory (better difference), / /-------There is no difference, *-----The difference in performance among the models for prediction is poorer).

**b) Handling the Problem of Imbalance Data**

The percentage of contraceptive methods use class data size consists about 21% of the respondents was reported as contraceptive users. This class size was considered to be unbalanced data which might be a bias to evaluate the classifier methods. An equal amount of both contraceptive users and non-users was taken randomly using WEKA 3.7.7 pre-processing option to balance these two classes to avoid dominance one over the other. And the overall significance of this balanced data should be compared with the above unbalanced data if there are differences on the models based on the performance measures used for the purpose of prediction. The original sample size was 15,683 but after the data imbalance problem was adjusted the new resample size would become 6586. On other word, the following below experimental results are re-run by considering equal amount of both contraceptive users and non-users. Table 11 illustrates that, after the adjustment of data imbalance, we evaluated if there exist effect due to the

Table 10. Model Evaluation for the classifiers, Paired corrected Tester-measure, Confidence: 0.05 (*two tailed*), *for* imbalance data.

| Dataset | (1) Naïve Bayes | (2) Decision tree (J48) | (3) Rando m tree | (4) Random forest tree | (5) Neural networks |
|---|---|---|---|---|---|
| DataSet_CPR_2018_19_ Model: F-measures | 0.84 | 0.89 | 0.88 | 0.88 | 0.87 |
| | (v/ /*) | \|(0/1/0) | (0/1/0) | (0/1/0) | (0/1/0) |

Table 11. Comparison of performance of different classifiers, for balanced data.

| Evaluation criteria's | Naïve Bayes | | Decision tree (J48) | | Decision tree (random tree) | | Decision tree (forest) | | Neural networks | | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Confusion matrix | 2176 | 1136 | 2325 | 987 | 2475 | 837 | 2405 | 907 | 2477 | 835 | No |
| | 391 | 2921 | 454 | 2858 | 964 | 2348 | 477 | 2835 | 782 | 2530 | Yes |
| Accuracy (%) | 76.94% | | 78.24% | | 72.81% | | 79.10% | | 75.58% | | |
| Sensitivity (%) | 65.70% | | 70.19% | | 74.72% | | 72.61% | | 76.38% | | |
| Specificity (%) | 88.19% | | 86.29% | | 70.89% | | 85.59% | | 74.78% | | |
| ROC (%) | 84.80% | | 81.70% | | 74.80% | | 86.70% | | 84.20% | | |
| Computations time in seconds | 0.0 | | 0.13 | | 0.07 | | 0.19 | | 260.59 | | |

Table 12. Model Evaluation for the classifiers, paired corrected Tester-measure, Confidence: 0.05 (*two tailed*); after adjusting the data imbalance problem.

| Dataset | (1) Naïve Bayes | (2) Decision tree (J48) | (3) Rando m tree | (4) Random forest tree | (5) Neural networks |
|---|---|---|---|---|---|
| DataSet_CPR_2018_19_Model: F-measures | 0.74 | 0.76 v | 0.73 | 0.77 v | 0.76 |
| | (v/ /*) | \|(0/1/0) | (0/1/0) | (0/1/0) | (0/1/0) |

imbalance of target variable using the same measures. The models used to predict with unbalanced data achieved slightly higher in overall performance than the models with balanced target dataset this is due to as possibly one target have got chance to dominate over the other target. Despite the slight differences observed due the imbalance of data, all the four classifiers have ROC values much more than 81% and with an improved ROC value of 74.80% for random tree. This indicates that given the features as input, the classifiers are efficient to predict the true contraceptive method users (more than 81% of ROC value) being an individual is contraception user or not. Besides, if we simply see the performance of the model in terms of accuracy it achieved, one can observe that the decision tree (random tree algorism) is the best model predictor of the other two (Table 11). However, we need to check whether these performances of measures achieved by each model have a statistical significance at 5% level of significance for further analysis and for future prediction purpose (Table 12).

Table 12 depicts a paired two-tailed comparison was done (paired corrected tested) to measure the difference of performance among the models in predicting the contraception use by the women at 5% level of significance [after adjusting the data imbalance problem]. Four data mining models (Decision tree (J48), Decision tree (random tree), Decision tree (random forest) and Neural networks (MLP)) were compared against to the Naïve Bayes model given for the same number of inputs. But, there were statistically significant differences between the

decision tree models (both J48 and random forest algorithms) and the Naïve Bayes model used for prediction to contraception methods use (Table 12). Moreover, the difference in performances of the models used for prediction using the decision tree models were considered as victory (significantly different) as compared to the naïve Bayes model. Nevertheless, all the models used in this paper are efficient enough (prediction power exceeds 77%) to predict the contraceptive methods use among women.

**4) Hybrid Multidimensional Metrics for Model Selection**

A hybrid multidimensional metrices was used to compute the overall significance of the model taking both the effects of the user's requirements and their corresponding weights of their importance basically assigned based on the user's requirements and defined as in Equation (4). The higher $HMM(m,r)$ indicates the overall significant model that comprises almost all requirements of the user unlike the classical metrics that used one criterion to pick the best fit model (Table 13). Accordingly, decision tree (J48) was found be the best fit model for the prediction task based on the hybrid metrics criterion. On the other side, the ANNs was found to be the most computationally expensive for our prediction task.

## 5. Conclusions

Two scenarios were considered with respect to both feature and model selections adopted to train the models: the classical approach employed the most commonly used algorithms for feature selection. However, this approach has been criticized for its weak side on drawing the complete picture of the prediction

Table 13. Hybrid multidimensional metrics criterion for final model selection.

| Metrics | Requirement's indicator | Classifier's weight score | | | | |
|---|---|---|---|---|---|---|
| | | NB | DT | RT | RF | ANNs |
| Roc values | 1 | 0.15 | 0.15 | 0.15 | 2/5 | 0.15 |
| Accuracy | 1 | 0.15 | 0.15 | 0.15 | 2/5 | 0.15 |
| data imbalance problem handled | 1 | 0.15 | 0.15 | 0.15 | 2/5 | 0.15 |
| statistical significance | 1 | 0.13 | 0.305 | 0.13 | 0.305 | 0.13 |
| practicability and applicability of the model | 1 | 0.15 | 2/5 | 0.15 | 0.15 | 0.15 |
| simplicity of model interpretation | 1 | 0.15 | 2/5 | 0.15 | 0.15 | 0.15 |
| consistency to the established knowledge | 1 | 0.15 | 2/5 | 0.15 | 0.15 | 0.15 |
| algorithm's simplicity in terms of time and space | 1 | 0.15 | 2/5 | 0.15 | 0.15 | 0.15 |
| | $HMM(m,r)$: | 0.236 | 0.47 | 0.25 | 0.42 | 0.236 |

task. Therefore, we proposed the hybrid multidimensional metrics for both feature and model selections would be an efficient approach in comprising the entire requirements of the user. Experimental results have revealed that all the predictive models used for this study except random tree were able to predict whether an individual was being contraceptive user or not given that the socio-demographic determinants, knowledge related to contraception use, knowledge related to AIDS and/or STI, exposure to mass-media, and knowledge on family planning as inputs with predictive power of more than 81%. Slight differences were also observed due the imbalance of data and the classifiers have ROC values much more than 81% after adjusting data imbalance problem. However, there was statistically significant difference between the decision tree models and the Naïve Bayes model used for prediction to contraceptive use (after adjusting for imbalance data problem). In conclusion, decision tree (J48) was found to be the best fit model for the prediction task based on the hybrid metrics criterion as the higher score of $HMM(m,r)=0.47$ indicates the overall significant model that comprises almost all requirements of the user unlike the classical metrics that rely on one criterion to pick the best fit model which lacks practicality or several characteristics of the model. On the other side, the ANNs was found to be the most computationally expensive for our prediction task. Specifically, this paper concluded that:

- Efficiency of predictive model could be better measured based on multidimensional criterion of the performance measures as this approach is more flexible to entertain user's requirements.
- Decision tree (J48) is the most efficient model (with a score of $HMM(m,r)=0.47$) and found statistically as victory model for the balanced data.
- The nature of data and the class size of the dataset (balanced or imbalanced data) have negative impact on the efficiency or prediction power of the model.

Following recommendations are forwarded to the academia, scientific communities and healthcare industries for future work in both feature and model selection scenarios could be considered in similar and/or different platforms of prediction tasks:

- Efficiency of a model is a multi-dimensional phenomenon. Hence, different model selection criteria as more flexible as hybrid metrics can be applied including scalability of the model, accuracy and specificity of the model, computational time cost, simplicity of the model, and others.
- Efficiency of predictive model could be improved through more flexible feature selection algorithms (specifically flexible hybrid metrics) considering the knowledge domain experts into account as understanding the business domain affects significantly.
- The class size or problem of imbalance data need to be handled hence an equal amount of both targets could be taken to minimize a bias that could be introduced to the model.

- Data transformation techniques, specifically on continuous features, need to be addressed to make the features suitable for the prediction task and to make the analysis procedures manageable and cost-effective.

## Acknowledgements

The authors would like to acknowledge the MEASURE Demographic and Health Survey (DHS) authority that they have authorized us to access all the necessary dataset and documents which we needed for this work.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Molina, L.C., Belanche, L. and Nebot, A. (2022) Feature Selection Algorithms: A Survey and Experimental Evaluation. 2002 *IEEE International Conference on Data Mining*, Maebashi City, 9-12 December 2002, 306-313.

[2] Liu, H., Motoda, H. and Yu, L. (2004) Selective Sampling Approach to Active Feature Selection. *Artificial Intelligence*, **159**, 49-74. https://doi.org/10.1016/j.artint.2004.05.009

[3] Chandrashekar, G. and Sahin, F. (2014) A Survey on Feature Selection Methods. *Computers & Electrical Engineering*, **40**, 16-28. https://doi.org/10.1016/j.compeleceng.2013.11.024

[4] Nakariyakul, S. (2014) Suboptimal Branch and Bound Algorithms for Feature Subset Selection: A Comparative Study. *Computers & Electrical Engineering*, **45**, 62-70. https://doi.org/10.1016/j.patrec.2014.03.002

[5] Sheikhpour, R., Sarram, M.A., Gharaghani, S. and Chahooki, M.A.Z. (2017) A Survey on Semi-Supervised Feature Selection Methods. *Pattern Recognition*, **64**, 141-158. https://doi.org/10.1016/j.patcog.2016.11.003

[6] Agrawal, R. and Psaila, G. (1995) Active Data Mining. *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, Montreal, 20-21 August 1995, 3-8.

[7] Liao, S.H., Chu, P.H. and Hsiao, P.Y. (2012) Data Mining Techniques and Applications—A Decade Review from 2000 to 2011. *Expert Systems with Applications*, **39**, 11303-11311. https://doi.org/10.1016/j.eswa.2012.02.063

[8] Janecek, A.G.K., Gansterer, G.F., *et al.* (2008) On the Relationship between Feature Selection and Classification Accuracy. *New challenges for feature selection in data mining and knowledge discovery*, Antwerp, 15 September 2008, 90-105.

[9] Bellman, R. (1961) Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton. https://doi.org/10.1515/9781400874668

[10] Chumerin, N. and Van Hulle, M.M. (2006) Comparison of Two Feature Extraction Methods Based on Maximization of Mutual Information. 2006 16*th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, Maynooth, 6-8 September 2006, 343-348. https://doi.org/10.1109/MLSP.2006.275572

[11] Motoda, H. and Liu, H. (2002) Feature Selection, Extraction and Construction. Springer, New York.

[12] Ladla, L. and Deepa, T. (2011) Feature Selection Methods and Algorithms. *International Journal on Computer Science and Engineering* (*IJCSE*), **3**, 1787-1797.

[13] (2021) Ethiopia and Demographic and Health Survey 2016 [FR328]. https://dhsprogram.com/

[14] Joshi, S., Deepa Shenoy, P., Vibhudendra Simha, G.G., Venugopal, K.R. and Patnaik, L.M. (2010) Classification of Neurodegenerative Disorders Based on Major Risk Factors EmployingMachine Learning Techniques. *International Journal of Engineering and Technology*, **2**, 350-355. https://doi.org/10.7763/IJET.2010.V2.146

[15] Kunwar, V., Chandel, K., Sabitha, A.S. and Bansal, A. (2016) Chronic Kidney Disease Analysis Using Data Mining Classification Techniques. 2016 6*th International Conference—Cloud System and Big Data Engineering* (*Confluence*), Noida, 14-15 January 2016, 300-305. https://doi.org/10.1109/CONFLUENCE.2016.7508132

[16] Bellazzi, R. and Zupan, B. (2008) Predictive Data Mining in Clinical Medicine: Current Issues and Guidelines. *International Journal of Medical Informatics*, **77**, 81-97. https://doi.org/10.1016/j.ijmedinf.2006.11.006

[17] Ge, Z.Q., Song, Z.H., Ding, S.X. and Huang, B. (2017) Data Mining and Analytics in the Process Industry: The Role of Machine Learning. *IEEE Access*, **5**, 20590-20616.

[18] Roberts, A. (2005) AI32: Guide to Weka. http://lia.deis.unibo.it/Courses/SistInt/Lucidi/lab01-small_weka_guide.pdf

[19] Han, J. and Kamber, M. (2006) Data Mining: Concepts and Techniques. 2nd Edition, Morgan Kaufmann Publishers, San Francisco.

[20] Brown, G., Pocock, A., Zhao, M.J. and LujÆn, M. (2012) Conditional Likelihood Maximization: A Unifying Framework for Information Theoretic Feature Selection. *Journal of Machine Learning Research*, **13**, 27-66.

[21] Souza, J. (2004) Feature Selection with a General Hybrid Algorithm. Ph.D. Thesis, University of Ottawa, Ottawa.

[22] Yu, L. and Liu, H. (2004) Efficient Feature Selection via Analysis of Relevance and Redundancy. *Journal of Machine Learning Research*, **5**, 1205-1224.

[23] Kohavi, R. and John, G.H. (1997) Wrappers for Feature Subset Selection. *Artificial Intelligence*, **97**, 273-324. https://doi.org/10.1016/S0004-3702(97)00043-X

[24] Nakariyakul, S., Liu, Z.P. and Chen, L. (2012) Detecting Thermophilic Proteins through Selecting Amino Acid and Dipeptide Composition Features. *Amino Acids*, **42**, 1947-1953. https://doi.org/10.1007/s00726-011-0923-1

[25] Dash, M. and Liu, H. (1997) Feature Selection for Classification. *Intelligent Data Analysis*, **1**, 131-156. https://doi.org/10.3233/IDA-1997-1302

[26] Das, S. (2001) Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection. *Proceedings of the Eighteenth International Conference on Machine Learning*, San Francisco, 28 June-1 July 2001, 74-81.

[27] Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002) Gene Selection for Cancer Classification Using Support Vector Machines. *Machine Learning*, **46**, 389-422. https://doi.org/10.1023/A:1012487302797

[28] Neumann, J., Schnörr, C. and Steidl, G. (2005) Combined SVM-Based Feature Selection and Classification. *Machine Learning*, **61**, 129-150. https://doi.org/10.1007/s10994-005-1505-9

[29] Guyon, I. and Elisseeff, A. (2003) An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, **3**, 1157-1182.

[30] Mitchell, T.M. (1982) Generalization as Search. *Artificial Intelligence*, **18**, 203-226.

https://doi.org/10.1016/0004-3702(82)90040-6

[31] Shafique, U., Majeed, F., Qaiser, H. and Mustafa, I.U. (2015) Data Mining in Healthcare for Heart Diseases. *International Journal of Innovation and Applied Studies*, **10**, 1312-1322.

[32] Soni, J., Ansari, U., Sharma, D. and Soni, S. (2011) Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications*, **17**, 43-48. https://doi.org/10.5120/2237-2860

[33] Koh, H.C. and Tan, G. (2011) Data Mining Applications in Healthcare. *Journal of Healthcare Information Management*, **19**, 64-72.

[34] Mahindrakar, P. and Hanumanthappa, M. (2013) Data Mining in Healthcare: A Survey of Techniques and Algorithms with Its Limitations and Challenges. *International Journal of Engineering Research and Applications*, **3**, 937-941.

[35] Hailu, T. (2015) Comparing Data Mining Techniques in HIV Testing Prediction. *Intelligent Information Management*, **7**, 153-180. https://doi.org/10.4236/iim.2015.73014

[36] Brosette, S.E., Spragre, A.P., Jones, W.T. and Moser, S.A. (2000) A Data Mining System for Infection Control Surveillance. *Methods of Information in Medicine*, **39**, 303-310. https://doi.org/10.1055/s-0038-1634449

[37] Sandhya, J., Deepa Shenoy, P., Venugopal, K.R. and Patnaik, A. (2010) Classification and Treatment of Different Stages of Alzheimer's Disease Using Various Machine Learning Methods. *International Journal of Bioinformatics Research*, **2**, 44-52. https://doi.org/10.9735/0975-3087.2.1.44-52

[38] Giudici, P. (2003) Applied Data Mining: Statistical Methods for Business and Industry. John Wiley, New York.

[39] Kharya, S. (2012) Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease. *International Journal of Computer Science, Engineering and Information Technology* (*IJCSEIT*), **2**, 55-66. https://doi.org/10.5121/ijcseit.2012.2206

[40] Sundar, N.A., Latha, P.P. and Chandra, M.R. (2012) Performance Analysis of Classification Data Mining Techniques over Heart Disease Database. *International Journal of Engineering Science & Advanced Technology*, **2**, 470-478.

[41] Obenshain, M.K. (2004) Application of Data Mining Techniques to Healthcare Data. *Infection Control and Hospital Epidemiology*, **25**, 690-695. https://doi.org/10.1086/502460

[42] Maniya, H., Hasan, M. and Patel, K.P. (2011) Comparative Study of Naïve Bayes Classifier and KNN for Tuberculosis. *International Conference on Web Services Computing* (*ICWSC*), **2**, 22-26.

[43] Kusiak, A., Dixon, B. and Shah, S. (2005) Predicting Survival Time for Kidney Dialysis Patients: A Data Mining Approach. *Computers in Biology and Medicine*, **35**, 311-327. https://doi.org/10.1016/j.compbiomed.2004.02.004

[44] Shetty, D., Rit, K., Shaikh, S. and Patil, N. (2017) Diabetes Disease Prediction Using Data Mining. 2017 *International Conference on Innovations in Information, Embedded and Communication Systems* (*ICIIECS*), Coimbatore, 17-18 March 2017, 1-5. https://doi.org/10.1109/ICIIECS.2017.8276012

[45] Rahim, N.F., Taib, S.M. and Abidin, A.I.Z. (2017) Dengue Fatality Prediction Using Data Mining. *Journal of Fundamental and Applied Sciences*, **9**, 671-683. https://doi.org/10.4314/jfas.v9i6s.52

[46] Uhmn, S., Kim, D.H., Cho, S.W., Cheong, J.Y. and Kim, J. (2007) Chronic Hepatitis

Classification Using SNP Data and Data Mining Techniques. 2007 *Frontiers in the Convergence of Bioscience and Information Technologies*, Jeju, 11-13 October 2007, 81-86. https://doi.org/10.1109/FBIT.2007.64

[47] Passmore, L., Goodside, J., Hamel, L., Gonzalez, L., Silberstein, T.A.L.I. and Trimarchi, J.A.M.E.S. (2003) Assessing Decision Tree Models for Clinical *in-vitro* Fertilization Data. Dept. of Computer Science and Statistics University of Rhode Island, Technical Report TR03-296.

[48] Dwivedi, A., Rehman, K., Ghosh, M. and Raman, R. (2018) Data Mining Algorithms in Healthcare. *International Journal of Computer Applications*, **180**, 26-31. https://doi.org/10.5120/ijca2018916901

[49] Guyon, I., Gunn, S., Nikravesh, M. and Zadeh, L.A. (2008) Feature Extraction: Foundations and Applications. Springer, Berlin.

[50] Xue, B., Zhang, M. and Browne, W.N. (2013) Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach. *IEEE Transactions on Cybernetics*, **43**, 1656-1671. https://doi.org/10.1109/TSMCB.2012.2227469

[51] Caldwell, J.C. and Caldwell, P. (2002) Africa: The New Family Planning Frontier. *Studies in Family Planning*, **33**, 76-86. https://doi.org/10.1111/j.1728-4465.2002.00076.x

[52] Fayyad, U. (1997) Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases. Proceedings. *Ninth International Conference on Scientific and Statistical Database Management*, Olympia, 11-13 August 1997, 2-11.

[53] Chaurasia, A.R. (2014) Contraceptive Use in India: A Data Mining Approach. *International Journal of Population Research*, **2014**, Article ID: 821436. https://doi.org/10.1155/2014/821436

[54] Han, J. and Kamber, M. (2006) Data Mining Concepts and Techniques. Morgan Kaufmann Publishers, Burlington.

[55] Berry, M.J. and Linoff, G. (1997) Data Mining Techniques: For Marketing, Sales and Customer Support. Wiley, New York.

[56] Parr Rud, O. (2001) Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management. Wiley, New York.

[57] Azevedo, A. and Santos, M.F. (2008) KDD, SEMMA and CRISP-DM: A Parallel Overview. *IADIS European Conference on Data Mining* 2008, Amsterdam, 24-26 July 2008, 182-185.

[58] Suryani, D., Labellapansa, A. and Marsela, E. (2018) Accuracy of Algorithm C4.5 to Study Data Mining against Selection of Contraception. In: Saian, R. and Abbas, M., Eds., *Proceedings of the Second International Conference on the Future of ASEAN* (*ICoFA*) 2017—*Volume* 2, Springer, Singapore, 955-962. https://doi.org/10.1007/978-981-10-8471-3_95

[59] Dwi Fajar Maulana, Y., Ruldeviyani, Y. and Indra Sensuse, D. (2020) Data Mining Classification Approach to Predict the Duration of Contraceptive Use. 2020 *Fifth International Conference on Informatics and Computing* (*ICIC*), Gorontalo, 3-4 November 2020, 1-6. https://doi.org/10.1109/ICIC50835.2020.9288568

[60] Hailemariam, T., Gebregiorgis, A., Meshesha, M. and Mekonnen, W. (2017) Application of Data Mining to Predict the Likelihood of Contraceptive Method Use among Women Aged 15-49 Case of 2005 Demographic Health Survey Data Collected by Central Statistics Agency, Addis Ababa, Ethiopia. *Journal of Health & Medical Informatics*, **8**, Article ID: 1000274. https://doi.org/10.4172/2157-7420.1000274

[61] Witten, I.H., Frank, E. and Hall, M.A. (2011) Data Mining Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publisher, Burlington.

[62] Daelemans, W., Hoste, V., Meulder, F.D. and Naudts, B. (2003) Combined Optimization of Feature Selection and Algorithm Parameter Interaction in Machine Learning of Language. In: Lavrač, N., Gamberger, D., Blockeel, H. and Todorovski, L., Eds., *Machine Learning*: *ECML* 2003, Springer, Berlin, 84-95. https://doi.org/10.1007/978-3-540-39857-8_10

[63] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000) CRISP-DM 1.0: Step by Step Data Mining Guide. https://books.google.com/books/about/CRISP_DM_1_0.html?id=po7FtgAACAAJ

[64] Liu, H. and Motoda, H. (1998) Feature Selection for Knowledge Discovery and Data Mining. Springer, New York. https://doi.org/10.1007/978-1-4615-5689-3

[65] Brachman, R.J. and Anand, T. (1996) The Process of Knowledge Discovery in Databases. *Advances in Knowledge Discovery and Data Mining*, AAAI Press/the MIT Press, Menlo Park, 37-57. https://dl.acm.org/doi/10.5555/257938.257944

[66] Yahia, M.E. and Ibrahim, B.A. (2003) K-Nearest Neighbor and C4.5 Algorithms as Data Mining Methods: Advantages and Difficulties. *ACS/IEEE International Conference on Computer Systems and Applications*, 2003. *Book of Abstracts*, Tunis, 14-18 July 2003, 103-109. https://doi.org/10.1109/AICCSA.2003.1227535

[67] Bach, M.P. and Ćosić, D. (2008) Data Mining Usage in Health Care Management: Literature Survey and Decision Tree Application. *Medicinski Glasnik*, **5**, 57-64.

[68] Chakrabarti, S., Cox, E., Frank, E., Hartmut, G.R., Han, J., Jiang, X., Kamber, M. and Witten, I. (2009) Data Mining: Know It All. Morga Kaufmann Publishers, Burlington.

[69] Famili, A. and Turney, P. (1997) Data Preprocessing and Intelligent Data Analysis. Institute of Information Technology, National Research Council Canada.

[70] Lu, H., Setiono, R. and Liu, H. (1996) Effective Data Mining Using Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*, **8**, 957-961. https://doi.org/10.1109/69.553163

[71] Witten, I.H. and Frank, E. (2005) Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition, Morgan Kaufmann Publishers, San Francisco.

[72] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) Classification and Regresion Trees.

[73] Velickov, S. and Solomatine, D. (2000) Predictive Data Mining: Practical Examples: Artificial Intelligence in Civil Engineering. 2*nd Joint Workshop on Applied AI in Civil Engineering, Cottbus, Germany*, Cottbus, March 2000, 1-17.