Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6, July, 2021: 7730-7749

Research Article

Undersampling Aware Learning based Fetal Health Prediction using Cardiotocographic Data

M. Shyamala Devi[1*], S. Sridevi[2] , D.Umanandhini[3] , A. Peter Soosai Anandaraj[4] ,
Sudheer Kumar Gupta[5], Bhumireddy Sidhartha[6]

Abstract

With the current improvement of development towards pharmaceutical, distinctive ultrasound methodologies are open to find the fetal prosperity. It is analyzed with diverse clinical parameters with 2-D imaging and other test. In any case, prosperity desire of fetal heart still remains an open issue due to unconstrained works out of the hatchling, the minor heart appraise and inadequate of data in fetal echocardiography. The machine learning strategies can find out the classes of fetal heart rate which can be utilized for earlier evaluating. With this background, we have utilized Cardiotocographic Fetal heart rate dataset removed from UCI Machine Learning Store for predicting the fetal heart rate health classes. The Prediction of fetal health rate are achieved in six ways. Firstly, the data set is preprocessed with Feature Scaling and missing values. Secondly, exploratory data investigation is done and the dispersion of target feature is visualized. Thirdly, the raw data set is fitted to all the classifiers and the performance is analysed before and after feature scaling. Fourth, the raw data set is subjected to undersampling methods like ClusterCentroids, RepeatedENN, AllKNN, CondensedNearestNeighbour, EditedNearestNeighbours, InstanceHardnessThreshold and NearMiss. Fifth, the undersampled dataset by above mentioned methods are fitted to all the classifiers and the performance is analyzed before and after feature scaling. Sixth, performance analysis is done using metrics like Precision, Recall, F-score, Accuracy and running time. The execution is done using python language under Spyder platform with Anaconda Navigator. Experimental results shows that the Decision Tree classifier tends to retain 98% before and after feature scaling for the underrsampling with EditedNearestNeighbours, RepeatedENN and AllKNN methods.

Keywords: *Machine learning, scaling, undersampling, precision, accuracy, classification*

[1,2,3,4,5,6] Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology

shyamaladevim@veltech.edu.in

Introduction

The irregular changes of the fetus must be watched through the clinical parameters in orchestrate to induce to the fetal prosperity. The passing rate of the fetal can be controlled by anticipating the changes inside the clinical parameters of the fetal prosperity. With the mechanical advancement, the ultrasound techniques are utilized for the assessment of fetal health and other changes inside the specified properties. The innate calculations can besides be utilized for the figure of any illnesses inside the fetal prosperity by decoding the discernment gotten through the parameter changes. Fetal heart rate watching may be a technique of checking the rate and cadence of the fetal beat. The typical fetal heart rate is between 120 and 160 beats per minute. This rate may alter as the fetus responds to conditions inside the uterus. The assessment of fetal prosperity has had our capable thought for various a long time. As the enhancement of developments for pre-birth symptomatic strategies has progressed, applications of such developments have backed inside the huge examination of fetal well-being. Fetal heart-rate checking remains the foremost shape of fetal assessment for high-risk pregnancies. The additional examinations overseen by the examination of ST and T-wave changes of the fetal electrocardiogram hold ensure for moving forward the prescient regard of fetal heart-rate assessments . Ultrasound has been invaluable for evaluation of fetal life frameworks, and the utilization of Doppler ultrasound has given information into fetal cardiovascular responses to such conditions as intrauterine advancement control and fetal slightness caused by reddish blood cell immunization.

Related works

Background

This paper employments non-parametric Bayesian techniques to classify the fetal heart rate. They have utilized non-parametric Bayesian Procedure and SVM-based methodology to classify the FHR and result are compared to find the execution of both procedures. Bayesian methodology have predominant execution than SVM-based procedure [1]. This paper find any natural contaminations and cardiac quirks on fetal heart utilizing full convolution neural organize (FCN). They are utilizing FCN to recognize the range of heart and examination the any inconsistencies utilizing FCN. They have concluded that FCN-16 illustrate has lower botch than other traces whereas classifying [2]. This paper endeavor to accomplish productive fetal acidosis disclosure to help pros in the midst of movement. They are utilizing sparse-SVM to choose and classify highlights and calculate execution. They have concluded that classification done by modified assurance is far off predominant than clinical sharpen[3]. The Classification and regression tree (CART) to distinguish high-risk in the midst of pregnancy. They found that accuracy gotten utilizing entropy calculation and GINI list is 88.87% and 90.12% independently[4]. This paper classify the fetal heart rate utilizing convolution neural organize (CNN) to make a neural organize which can classify heart rate subsequently. They found that accuracy is tall of CNN than of SVM and MLP. CNN is an profitable for fetal heart checking system[5]. They have utilized classification based on affiliation (CBA) to classify fetal prosperity status. They have found that the accuracy of appear was extended after utilizing highlight assurance methodology. They found out that subjective forest and XGBoost have extraordinary execution for classifying fetal prosperity status[6].

In this paper, they are advancing to find out the irregular changes to recognize sinusoidal heart rate. Unpredictable forest is utilized for the data preprocessing and sinusoidal plan is calculated utilizing soft theoretic procedure . The conclusion is that given methodology can be considered a gold standard for the discontinuous change revelation [7] . In this paper they are coming to to find within the occasion that there are any inconvenience happening in fetal and a overwhelming system utilizing significant learning is sketched out. Exploratory appear rot is utilized to break down any one-dimensional timing hail S(t) into Inborn Mode Work (IMF) with assorted frequencies. LSMT is utilized to classify the data. As a result, it is showed up that significant slanting has most precision for recognizing fetal heart inconvenience [8]. In this paper, they are pointing to anticipate fetal risk utilizing machine learning for shirking of any adolescent passing. At to start with for dataset derivation least reiteration most prominent relevance methodology is utilized . And after that the calculations like navie bayes, choice trees, irregular timberland, back vector machines are associated for classification of heart rate[9]. In this paper detail heart rate is classified utilizing two methodologies and the predominant procedure is found out between them. To utilize hybrid k-means, the incorporate extraction is done utilizing k infers clustering. CTG dataset is recorded utilizing calculation and is compared to SVM. Cross breed K-means and Bolster Vector Machine is compared and accuracy of cross breed K-means was 90.64% though ordinary exactness for SVM was 76.72 which shows up Half breed K-means has predominant exactness [10].

In this paper, they are utilizing 2D ultrasound to choose and degree the hatchling heart rate. Number of Pixel from the picture is taken care of and gives the information around fetal heart condition as each zone of the heart or parcel of heart pictures are arranged utilizing 2D ultrasound [11]. In this paper, they classify and compare CTG data system utilizing managed SVM and choice tree to encourage which have best execution and exactness. The CTG dataset were arranged utilizing coordinated SVM and choice tree [12]. In this paper, they have utilize SVM as classifier to execute the f-score. They have utilized f-score procedure to sort the highlights. F-score have unimaginable precision in anticipating fetal status[13]. In this paper, they are centered on a formative multi objective generic algorithm (MOGA) by utilize of which the basic components causing fetal passing is removed with offer help of cardiotocographic examination . Along these methods, it is found out that execution of any classifier is boosted on the off chance that genuine incorporate assurance is done [14]. In this paper, they display the function of data-driven entropy profiling to distinguish fetal arrhythmia normally. The fetal QRS extraction method is utilized to remove fetal heart rate from the data set and after that entropy highlights are associated for profiling of the data set. The proposed methodology talks to strong entropy assess that donate predominant execution than existing methodology [15-16].

Proposed Work

The CTG Cardiotocographic Fetal heart rate dataset with 36 independent variables and 1 dependent variable has been used for implementation [17-20]. The prediction of fetal health is done with the following contributions.

(i) Firstly, the data set is preprocessed with Feature Scaling and missing values.
(ii) Secondly, exploratory data investigation is done and the dispersion of target feature is visualized.
(iii) Thirdly, the raw data set is fitted to all the classifiers and the performance is analysed before and after feature scaling.
(iv) Fourth, the raw data set is subjected to undersampling methods like ClusterCentroids, RepeatedENN, AllKNN, CondensedNearestNeighbour, EditedNearestNeighbours, InstanceHardnessThreshold and NearMiss.
(v) Fifth, the undersampled dataset by above mentioned methods are fitted to all the classifiers and the performance is analyzed before and after feature scaling.
(vi) Sixth, performance analysis is done using metrics like Precision, Recall, F-score, Accuracy and running time. Fig. 1 shows the overall workflow of this work

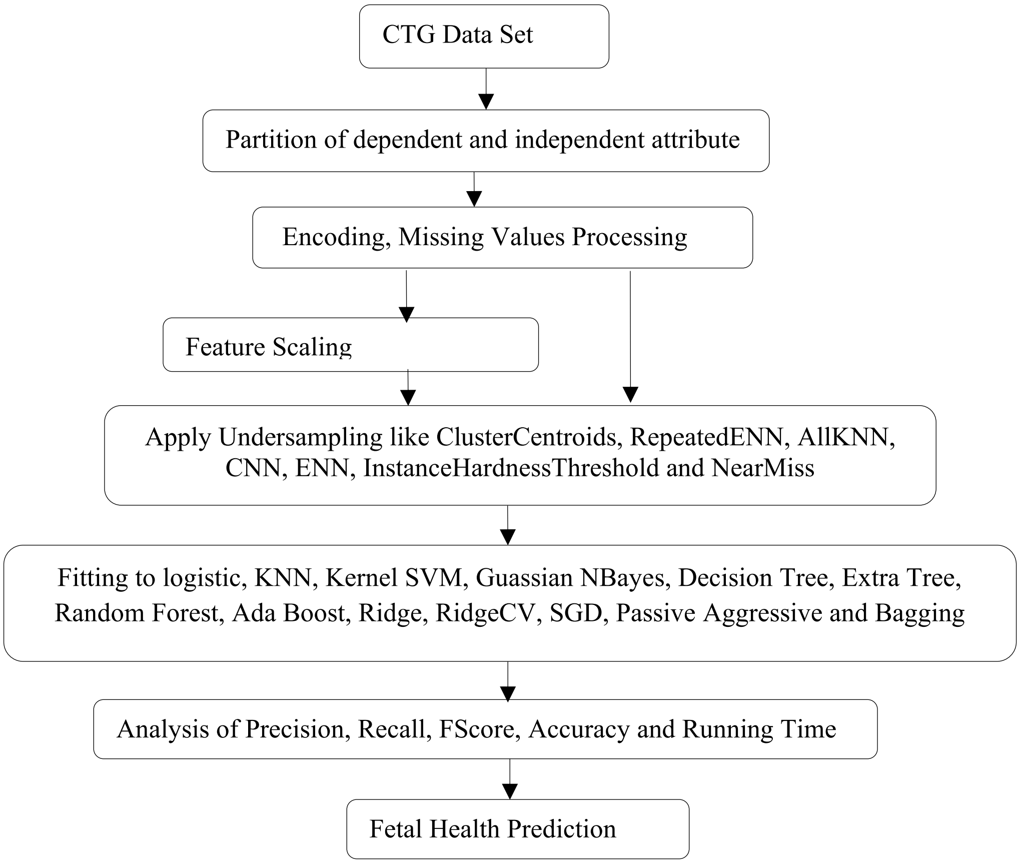

Figure 1. Overall Workflow of the system

Exploratory Data Analysis

The CTG dataset extricated from the UCI machine learning store is utilized for usage [21]. The dataset comprises of 2127 patients information with 21 autonomous highlights (baseline value, accelerations, fetal movement, Uterine contractions, light decelerations, severe decelerations, prolongued decelerations, abnormal short term variability, mean value of short term variability, percentage of time with abnormal long term variability, mean value of long term variability, histogram width, histogram min, histogram max, histogram number of peaks, histogram number of zeroes, histogram mode, histogram mean, histogram median, histogram variance, histogram tendency) and 1 Target “Fetal Health” [22]. The code is implemented with python under Anaconda Navigator with Spyder IDE. The data set is splitted with 80:20 for training and testing dataset. Fig.2. shows the target feature analysis and found to be non-sampled [23-24].

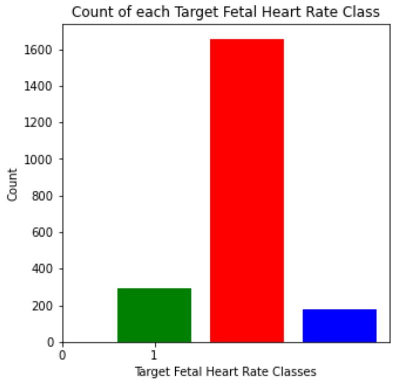

Figure 2. Target Feature Analysis of Dataset

Implementation and Discussion

The raw data set is fitted to all the classifier like logistic regression, KNN, Kernel SVM, Decision Tree, Random Forest, Ada Boost, Ridge, RidgeCV, SGD, Passive Aggressive and Bagging classifier with and without the presence of feature scaling and performance is shown in Table 1 and Table 2, the accuracy and running time comparison is shown in Figure. 3 - 4.

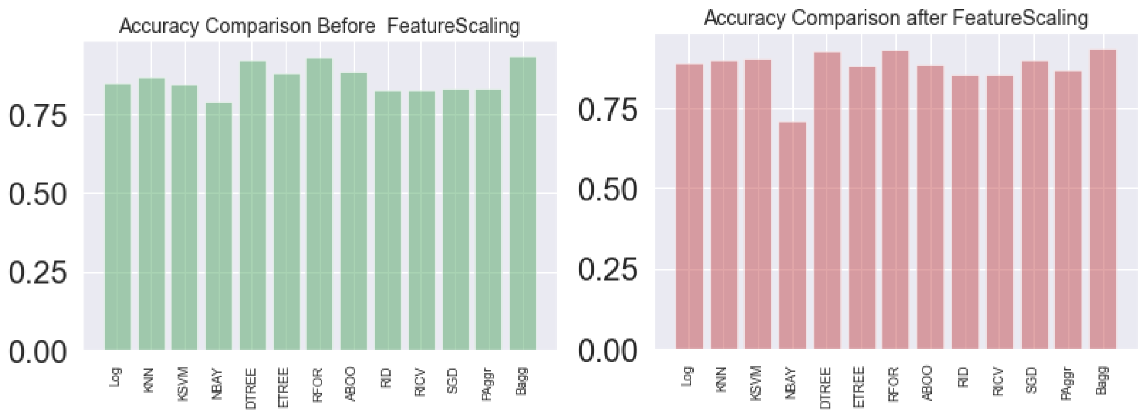

Figure 3. Accuracy analysis of raw dataset before and after feature scaling

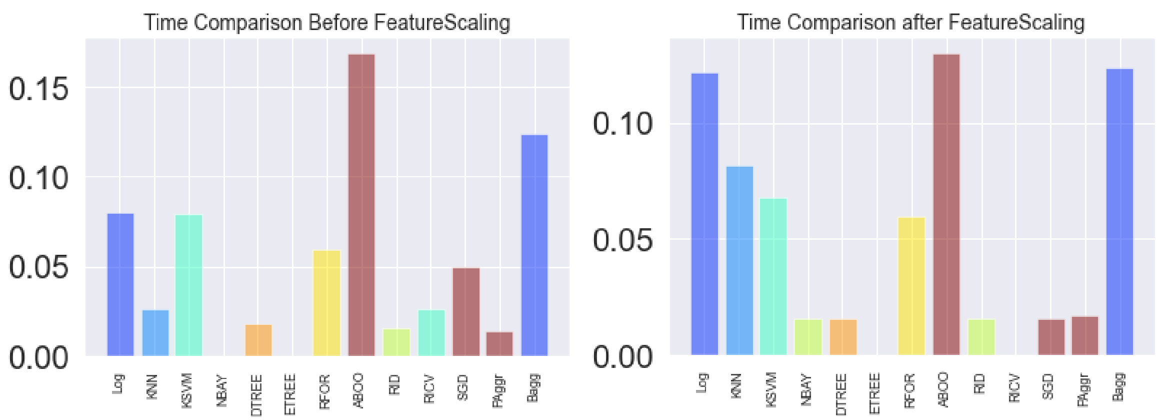

*Figure 4.*Response time analysis of raw dataset before and after feature scaling

Table 1

Classifier performance of the raw dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.84	0.85	0.84	0.85	0.08
KNN	0.86	0.87	0.86	0.87	0.03
KSVM	0.83	0.84	0.83	0.84	0.08
GNBayes	0.86	0.79	0.81	0.79	0.00
DTree	0.92	0.92	0.92	0.92	0.02
ETree	0.88	0.88	0.88	0.88	0.00
RForest	0.93	0.93	0.93	0.93	0.06
AdaBoost	0.88	0.88	0.88	0.88	0.17
Ridge	0.82	0.83	0.81	0.83	0.02
RidgeCV	0.82	0.83	0.81	0.83	0.03
SGD	0.84	0.83	0.83	0.83	0.05
PAggress	0.80	0.83	0.80	0.83	0.01
Bagging	0.93	0.94	0.93	0.94	0.12

Table 2

Classifier performance of the raw dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.89	0.89	0.89	0.89	0.12
KNN	0.90	0.90	0.89	0.90	0.08
KSVM	0.90	0.90	0.90	0.90	0.07
GNBayes	0.86	0.71	0.75	0.71	0.02
DTree	0.92	0.92	0.92	0.92	0.02
ETree	0.88	0.88	0.88	0.88	0.00
RForest	0.93	0.93	0.93	0.93	0.06
AdaBoost	0.88	0.88	0.88	0.88	0.13
Ridge	0.84	0.85	0.84	0.85	0.02
RidgeCV	0.84	0.85	0.84	0.85	0.00
SGD	0.89	0.90	0.89	0.90	0.02
PAggress	0.88	0.87	0.87	0.87	0.02
Bagging	0.93	0.94	0.93	0.94	0.12

Undersampling Results and Performance Analysis

The raw data set is subjected to undersampling methods ClusterCentroids, CondensedNearestNeighbour, AllKNN, RepeatedENN, EditedNearestNeighbours, InstanceHardnessThreshold and NearMiss. The resampled dataset distribution after undersampling is shown in Figure.5. The raw data set is subjected to undersampling method namely ClusterCentroids and the resampled dataset is fitted to all the classifiers with and without the presence of feature scaling and performance is shown in Table 3 and Table 4, the accuracy and the running time comparison is shown in Figure. 6 - 7.

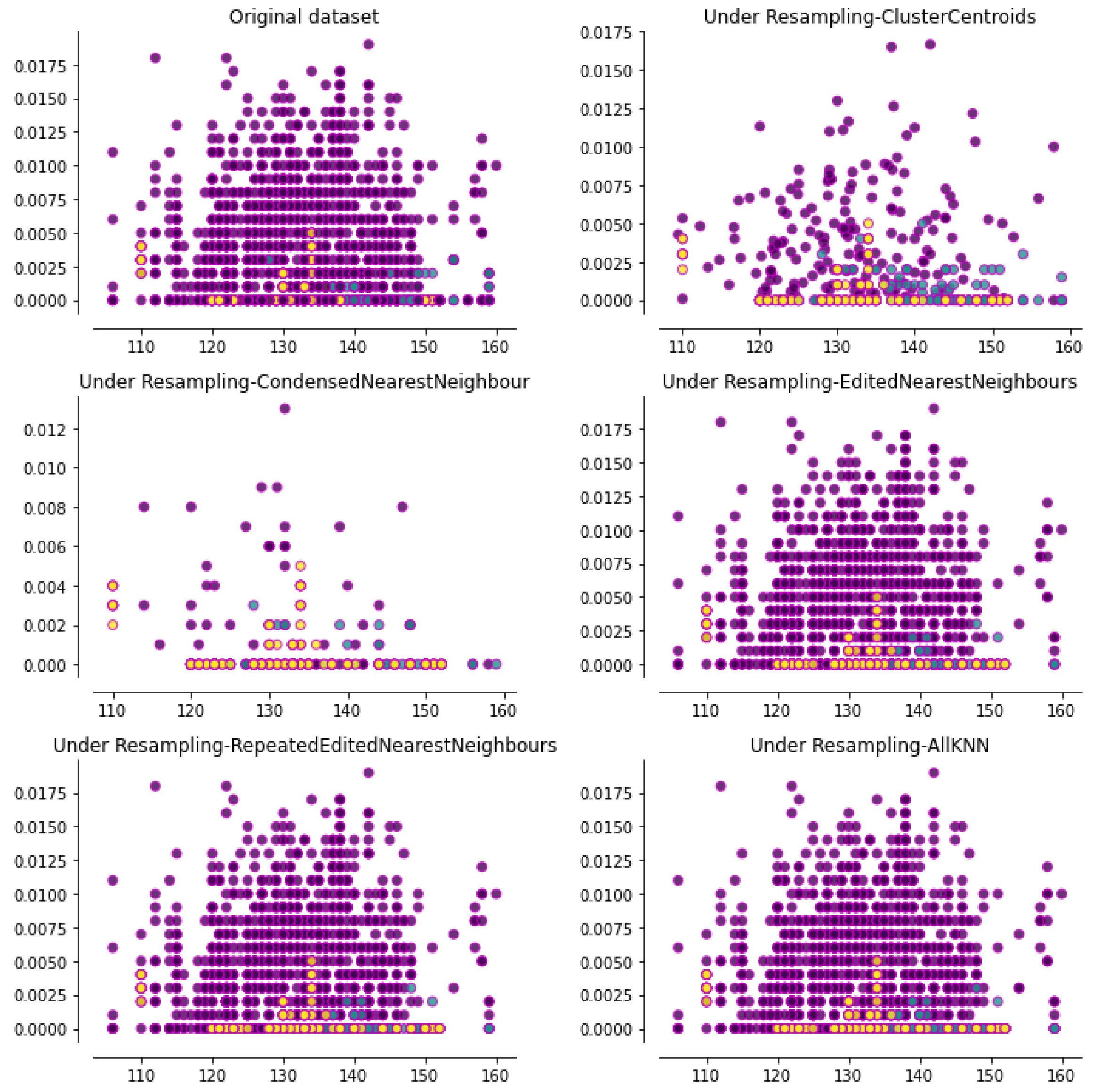

Figure 5. Data distribution after undersampling methods

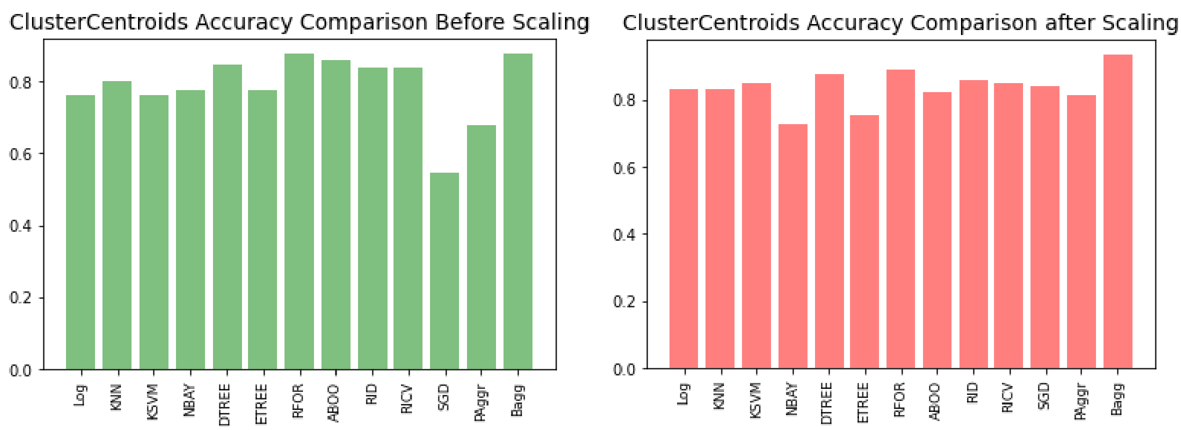

Figure 6. Accuracy analysis of ClusterCentroids dataset before and after feature scaling

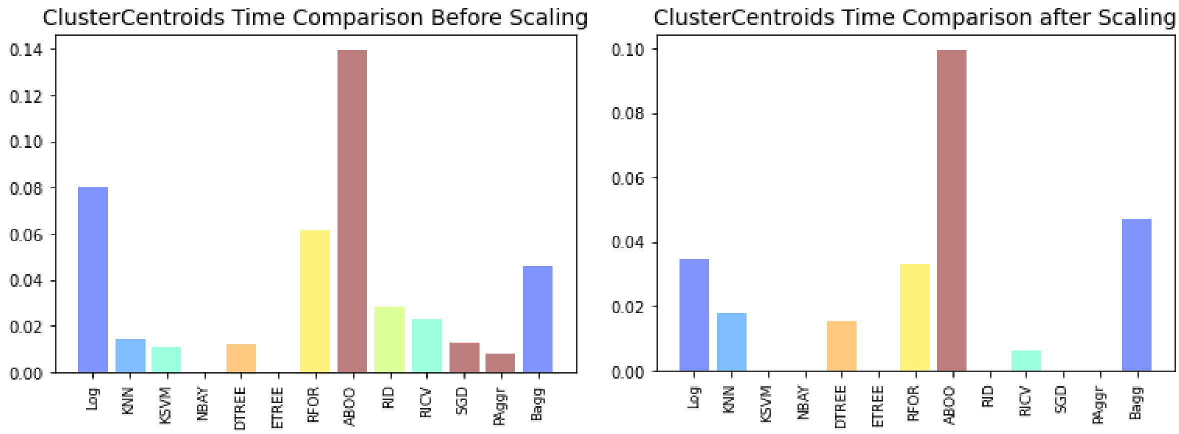

Figure 7. Response time analysis of raw dataset before and after feature scaling

Table 3

Classifier performance of Cluster Centroids dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.76	0.76	0.76	0.76	0.08
KNN	0.80	0.80	0.80	0.80	0.01
KSVM	0.76	0.76	0.76	0.76	0.01
GNBayes	0.79	0.77	0.78	0.77	0.00
DTree	0.85	0.85	0.85	0.85	0.01
ETree	0.78	0.77	0.77	0.77	0.00
RForest	0.88	0.88	0.88	0.88	0.06
AdaBoost	0.86	0.86	0.86	0.86	0.14
Ridge	0.84	0.84	0.84	0.84	0.03
RidgeCV	0.84	0.84	0.84	0.84	0.02
SGD	0.71	0.55	0.54	0.55	0.01
PAggress	0.69	0.68	0.68	0.68	0.01
Bagging	0.88	0.88	0.88	0.88	0.05

Table 4

Classifier performance of Cluster Centroids dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.83	0.83	0.83	0.83	0.03
KNN	0.84	0.83	0.83	0.83	0.02
KSVM	0.85	0.85	0.85	0.85	0.00
GNBayes	0.83	0.73	0.73	0.73	0.00
DTree	0.88	0.88	0.88	0.88	0.02
ETree	0.75	0.75	0.75	0.75	0.00
RForest	0.89	0.89	0.89	0.89	0.03
AdaBoost	0.85	0.82	0.82	0.82	0.10
Ridge	0.87	0.86	0.86	0.86	0.00
RidgeCV	0.85	0.85	0.85	0.85	0.01
SGD	0.84	0.84	0.84	0.84	0.00
PAggress	0.81	0.81	0.81	0.81	0.00
Bagging	0.93	0.93	0.93	0.93	0.05

The raw data set is subjected to undersampling method namely CondensedNearestNeighbour and the resampled dataset is fitted to all the classifiers with and without the presence of feature scaling and performance is shown in Table 5 and Table 6, the accuracy and the running time comparison is shown in Figure. 8 - 9.

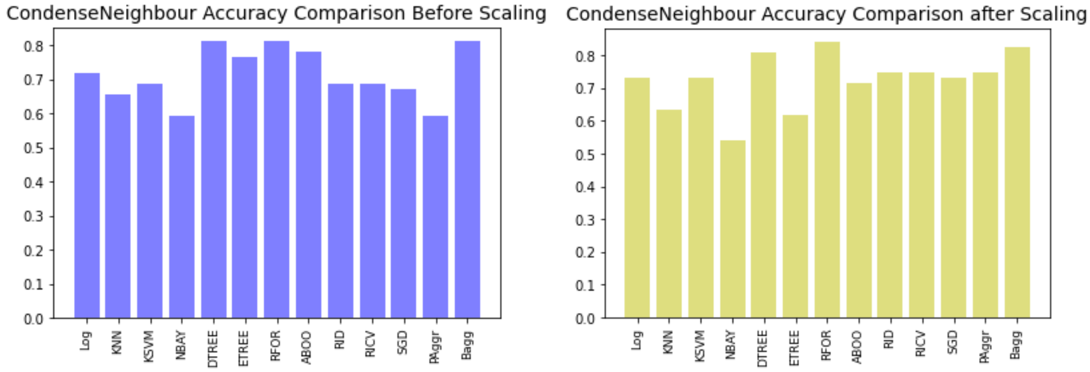

Figure 8. Accuracy analysis of CondensedNearestNeighbour dataset before and after scaling

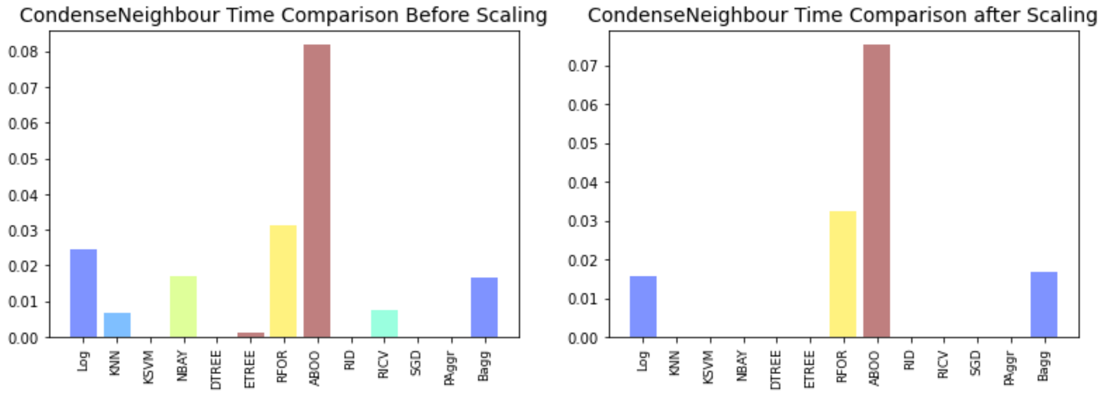

Figure 9. Response time analysis of CondensedNearestNeighbour before and after scaling

Table 5

Classifier performance of CondensedNearestNeighbour dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.73	0.72	0.71	0.72	0.02
KNN	0.65	0.66	0.65	0.66	0.01
KSVM	0.73	0.69	0.63	0.69	0.00
GNBayes	0.71	0.59	0.62	0.59	0.02
DTree	0.79	0.81	0.79	0.81	0.00
ETree	0.75	0.77	0.75	0.77	0.00
RForest	0.82	0.81	0.82	0.81	0.03
AdaBoost	0.84	0.78	0.79	0.78	0.08
Ridge	0.68	0.69	0.67	0.69	0.00
RidgeCV	0.67	0.69	0.67	0.69	0.01
SGD	0.69	0.67	0.64	0.67	0.00
PAggress	0.45	0.59	0.51	0.59	0.00
Bagging	0.82	0.81	0.81	0.81	0.02

Table 6

Classifier performance of CondensedNearestNeighbour dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.72	0.73	0.72	0.73	0.02
KNN	0.66	0.63	0.63	0.63	0.00
KSVM	0.77	0.73	0.72	0.73	0.00
GNBayes	0.72	0.54	0.58	0.54	0.00
DTree	0.79	0.81	0.80	0.81	0.00
ETree	0.60	0.62	0.61	0.62	0.00
RForest	0.83	0.84	0.83	0.84	0.03
AdaBoost	0.74	0.71	0.72	0.71	0.08
Ridge	0.72	0.75	0.73	0.75	0.00
RidgeCV	0.72	0.75	0.73	0.75	0.00
SGD	0.73	0.73	0.73	0.73	0.00
PAggress	0.74	0.75	0.74	0.75	0.00
Bagging	0.81	0.83	0.82	0.83	0.02

The raw data set is subjected to undersampling method namely AllKNN and the resampled dataset is fitted to all the classifiers with and without the presence of feature scaling and performance is shown in Table 7 and Table 8, the accuracy and the running time comparison is shown in Figure. 10 - 11.

Table 7

Classifier performance of ALLKNN dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.93	0.93	0.93	0.93	0.07
KNN	0.96	0.96	0.96	0.96	0.03
KSVM	0.92	0.92	0.92	0.92	0.07
GNBayes	0.92	0.84	0.86	0.84	0.00
DTree	0.98	0.98	0.98	0.98	0.00
ETree	0.97	0.97	0.97	0.97	0.01
RForest	0.98	0.98	0.98	0.98	0.06
AdaBoost	0.86	0.86	0.86	0.86	0.16
Ridge	0.91	0.91	0.91	0.91	0.01
RidgeCV	0.91	0.92	0.91	0.92	0.01
SGD	0.90	0.91	0.89	0.91	0.04
PAggress	0.86	0.91	0.89	0.91	0.02
Bagging	0.98	0.98	0.98	0.98	0.12

Table 8

Classifier performance of ALLKNN dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.95	0.94	0.94	0.94	0.08
KNN	0.94	0.94	0.94	0.94	0.08
KSVM	0.96	0.95	0.95	0.95	0.06
GNBayes	0.92	0.89	0.90	0.89	0.00
DTree	0.97	0.97	0.97	0.97	0.00
ETree	0.95	0.95	0.95	0.95	0.00
RForest	0.98	0.98	0.98	0.98	0.05
AdaBoost	0.91	0.91	0.90	0.91	0.14
Ridge	0.92	0.93	0.92	0.93	0.00
RidgeCV	0.92	0.92	0.91	0.92	0.00
SGD	0.94	0.93	0.93	0.93	0.02
PAggress	0.94	0.94	0.94	0.94	0.01
Bagging	0.97	0.97	0.97	0.97	0.07

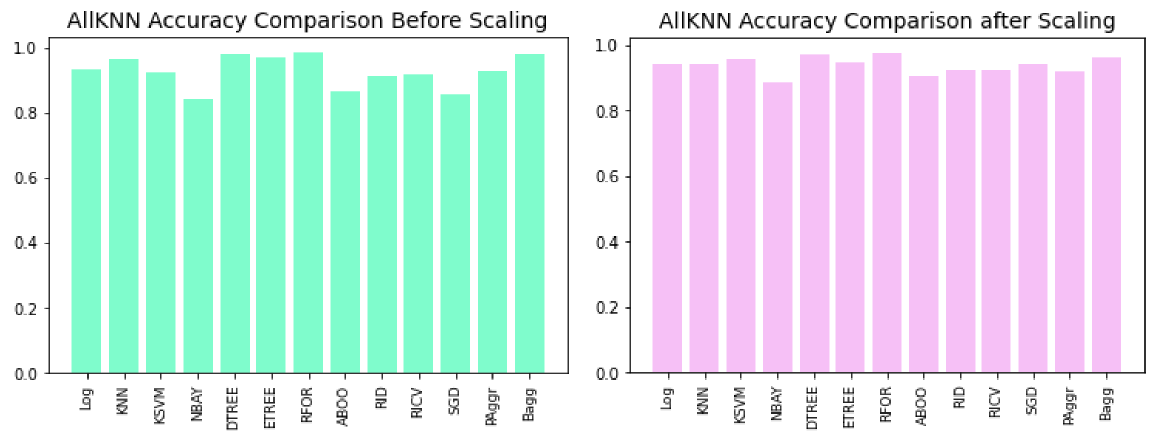

Figure 10. Accuracy analysis of AllKNN dataset before and after feature scaling

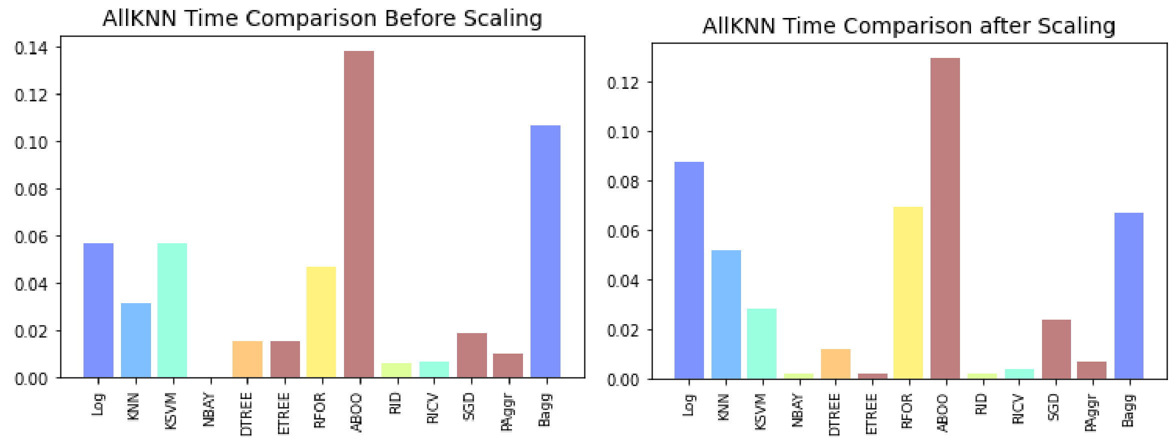

Figure 11. Response time analysis of AllKNN dataset before and after feature scaling

The raw data set is subjected to undersampling method namely RepeatedENN and the resampled dataset is fitted to all the classifiers with and without the presence of feature scaling and performance is shown in Table 9 and Table 10, the accuracy and the running time comparison is shown in Figure. 12 - 13.

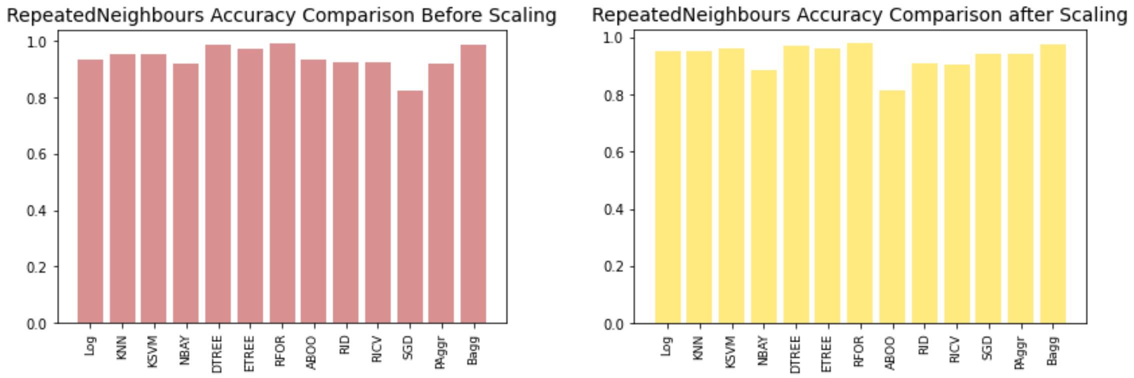

Figure 12. Accuracy analysis of RepeatedENN dataset before and after feature scaling

Table 9

Classifier performance of RepeatedENN dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.93	0.93	0.93	0.93	0.07
KNN	0.95	0.95	0.95	0.95	0.03
KSVM	0.95	0.95	0.95	0.95	0.05
GNBayes	0.94	0.92	0.93	0.92	0.00
DTree	0.99	0.99	0.99	0.99	0.02
ETree	0.97	0.97	0.97	0.97	0.00
RForest	0.99	0.99	0.99	0.99	0.06
AdaBoost	0.94	0.93	0.92	0.93	0.15
Ridge	0.92	0.92	0.91	0.92	0.00
RidgeCV	0.92	0.92	0.91	0.92	0.01
SGD	0.87	0.82	0.83	0.82	0.03
PAggress	0.92	0.92	0.91	0.92	0.01
Bagging	0.99	0.99	0.99	0.99	0.09

Table 10

Classifier performance of RepeatedENN dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.95	0.95	0.95	0.95	0.05
KNN	0.95	0.95	0.95	0.95	0.03
KSVM	0.96	0.96	0.96	0.96	0.03
GNBayes	0.91	0.89	0.89	0.89	0.00
DTree	0.97	0.97	0.97	0.97	0.01
ETree	0.96	0.96	0.96	0.96	0.02
RForest	0.98	0.98	0.98	0.98	0.06
AdaBoost	0.90	0.81	0.84	0.81	0.14
Ridge	0.90	0.91	0.89	0.91	0.00
RidgeCV	0.89	0.90	0.89	0.90	0.00
SGD	0.94	0.94	0.94	0.94	0.02
PAggress	0.94	0.94	0.94	0.94	0.01
Bagging	0.98	0.97	0.97	0.97	0.07

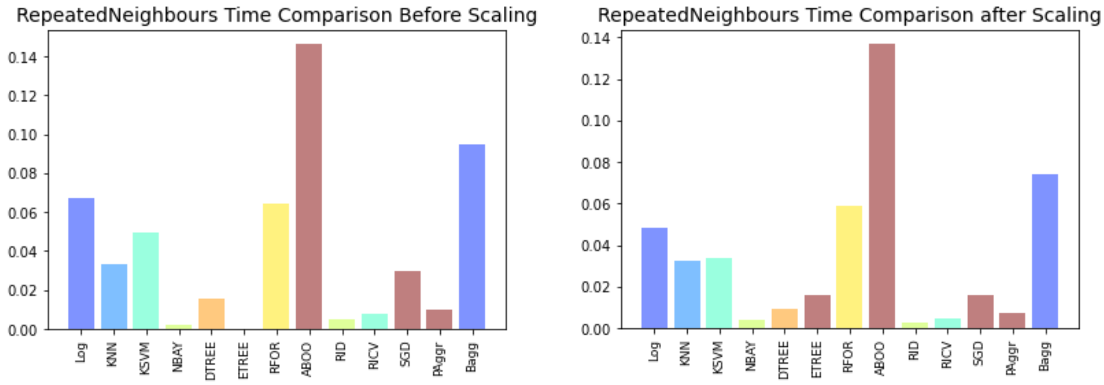

Figure 13. Response time analysis of RepeatedENN dataset before and after feature scaling

The raw data set is subjected to undersampling method namely EditedNearestNeighbours and the resampled dataset is fitted to all the classifiers with and without the presence of feature scaling and performance is shown in Table 11 and Table 12, the accuracy and the running time comparison is shown in Figure. 14 - 15.

Table 11

Classifier performance of EditedNearestNeighbours dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.93	0.93	0.93	0.93	0.07
KNN	0.95	0.95	0.95	0.95	0.03
KSVM	0.95	0.95	0.95	0.95	0.03
GNBayes	0.94	0.92	0.93	0.92	0.00
DTree	0.99	0.99	0.99	0.99	0.02
ETree	0.97	0.97	0.97	0.97	0.00
RForest	0.99	0.99	0.99	0.99	0.08
AdaBoost	0.94	0.93	0.92	0.93	0.14
Ridge	0.92	0.92	0.91	0.92	0.01
RidgeCV	0.92	0.92	0.91	0.92	0.01
SGD	0.92	0.92	0.91	0.92	0.03
PAggress	0.94	0.93	0.93	0.93	0.01
Bagging	0.99	0.99	0.99	0.99	0.12

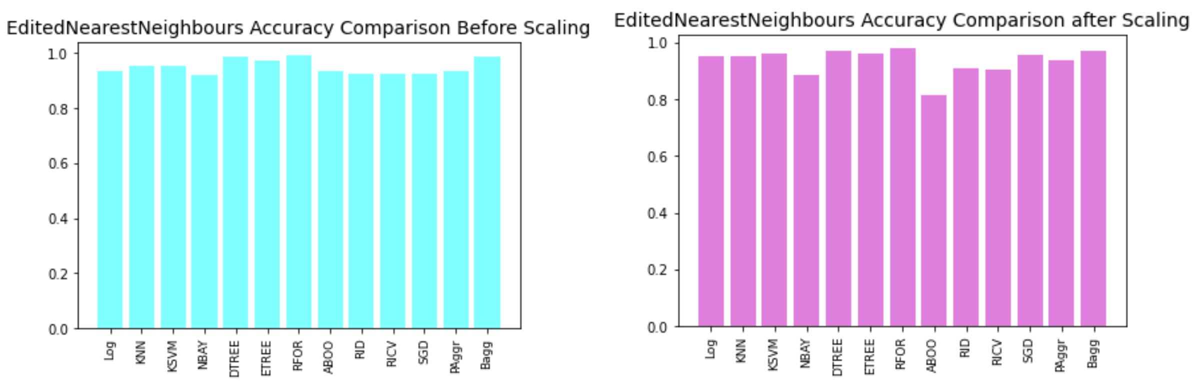

Figure 14. Accuracy analysis of EditedNearestNeighbours dataset before and after scaling

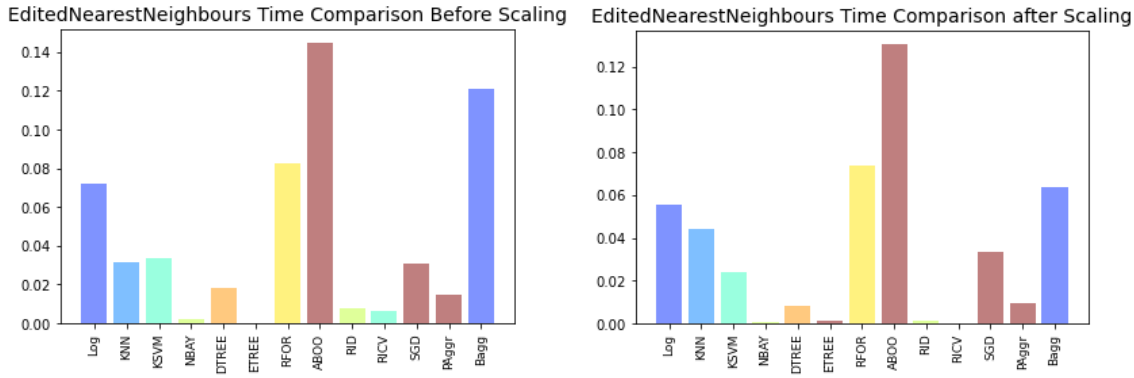

Figure 15. Response time analysis of EditedNearestNeighbours dataset before and after scaling

Table 12

Classifier performance of EditedNearestNeighbours dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.95	0.95	0.95	0.95	0.06
KNN	0.95	0.95	0.95	0.95	0.04
KSVM	0.96	0.96	0.96	0.96	0.02
GNBayes	0.91	0.89	0.89	0.89	0.00
DTree	0.97	0.97	0.97	0.97	0.01
ETree	0.96	0.96	0.96	0.96	0.00
RForest	0.98	0.98	0.98	0.98	0.07
AdaBoost	0.90	0.81	0.84	0.81	0.13
Ridge	0.90	0.91	0.89	0.91	0.00
RidgeCV	0.89	0.90	0.89	0.90	0.00
SGD	0.96	0.96	0.95	0.96	0.03
PAggress	0.93	0.94	0.93	0.94	0.01
Bagging	0.97	0.97	0.97	0.97	0.06

The raw data set is subjected to undersampling method namely InstanceHardnessThreshold and the resampled dataset is fitted to all the classifiers with and without the presence of feature scaling and performance is shown in Table 13 and Table 14, the accuracy and the running time comparison is shown in Figure. 16 - 17.

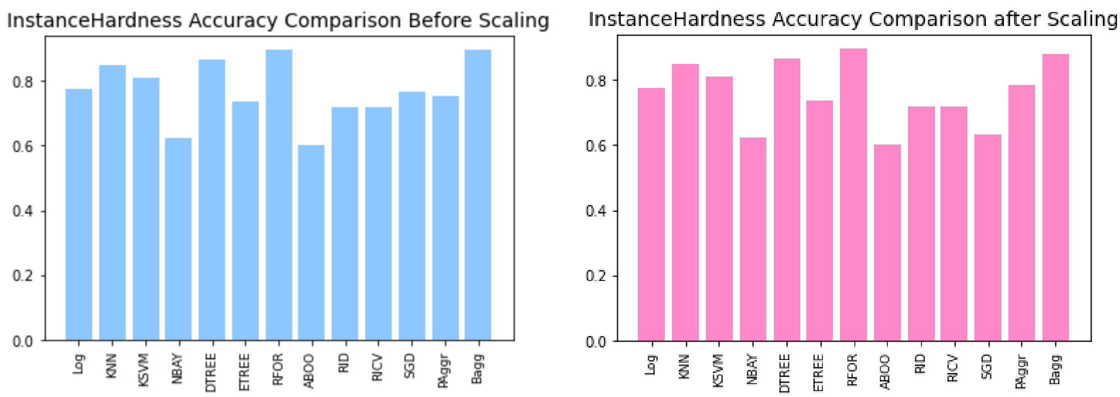

Figure 16. Accuracy analysis of InstanceHardnessThreshold dataset before and after scaling

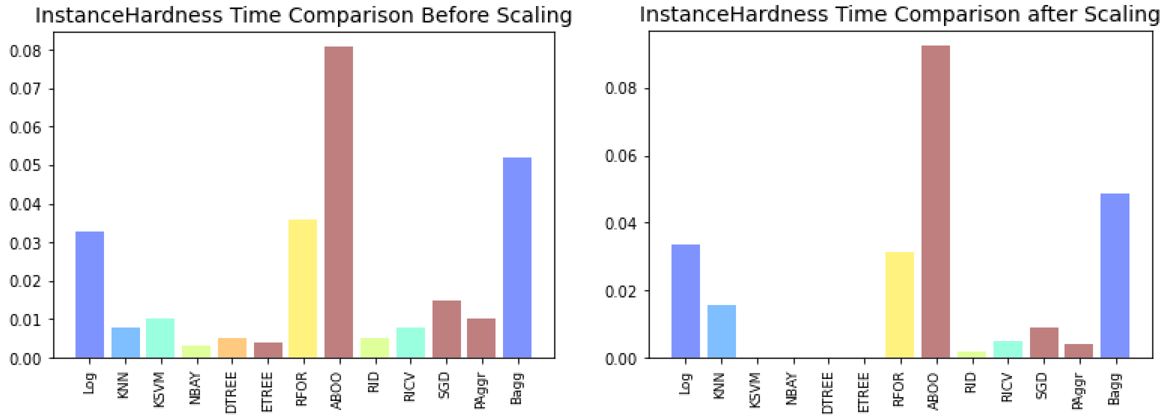

Figure 17. Response time analysis of InstanceHardnessThreshold before and after scaling

Table 13

Classifier performance of InstanceHardnessThreshold dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.89	0.90	0.89	0.90	0.06
KNN	0.80	0.87	0.82	0.87	0.06
KSVM	0.81	0.84	0.82	0.84	0.02
GNBayes	0.90	0.84	0.86	0.84	0.00
DTree	0.96	0.96	0.96	0.96	0.02
ETree	0.90	0.84	0.86	0.84	0.00
RForest	0.96	0.96	0.96	0.96	0.02
AdaBoost	0.95	0.94	0.94	0.94	0.00
Ridge	0.97	0.97	0.97	0.97	0.06
RidgeCV	0.89	0.90	0.89	0.90	0.01
SGD	0.80	0.87	0.82	0.87	0.06
PAggress	0.81	0.84	0.82	0.84	0.02
Bagging	0.95	0.95	0.95	0.95	0.11

Table 14

Classifier performance of InstanceHardnessThreshold dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.93	0.93	0.93	0.93	0.09
KNN	0.94	0.94	0.93	0.94	0.06
KSVM	0.90	0.90	0.90	0.90	0.02
GNBayes	0.97	0.97	0.97	0.97	0.11
DTree	0.95	0.95	0.95	0.95	0.02
ETree	0.94	0.94	0.94	0.94	0.00
RForest	0.98	0.98	0.98	0.98	0.05
AdaBoost	0.95	0.95	0.95	0.95	0.05
Ridge	0.91	0.87	0.88	0.87	0.00
RidgeCV	0.95	0.95	0.95	0.95	0.02
SGD	0.94	0.94	0.94	0.94	0.00
PAggress	0.90	0.90	0.90	0.90	0.02
Bagging	0.97	0.97	0.97	0.97	0.11

The raw data set is subjected to undersampling method namely NearMiss and the resampled dataset is fitted to all the classifiers with and without the presence of feature scaling and performance is shown in Table 15 and Table 16, the accuracy and the running time comparison is shown in Figure. 18 - 19.

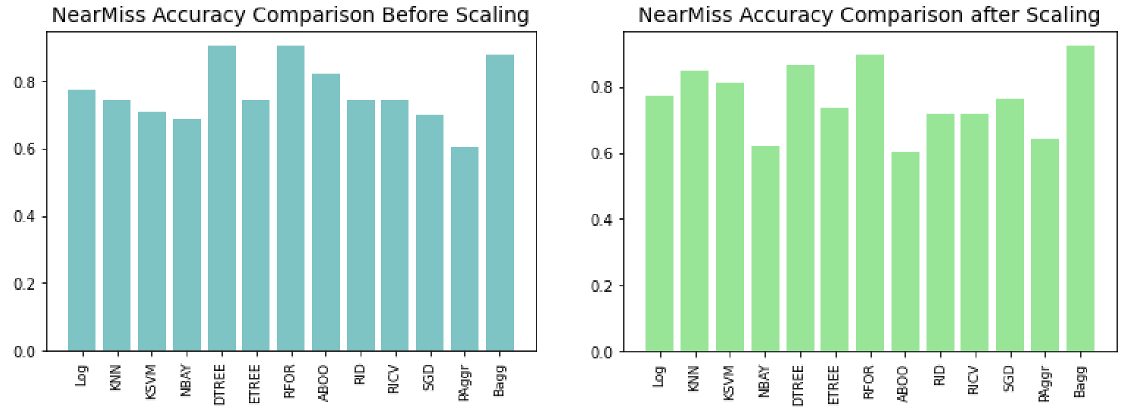

Figure 18. Accuracy analysis of NearMiss dataset before and after feature scaling

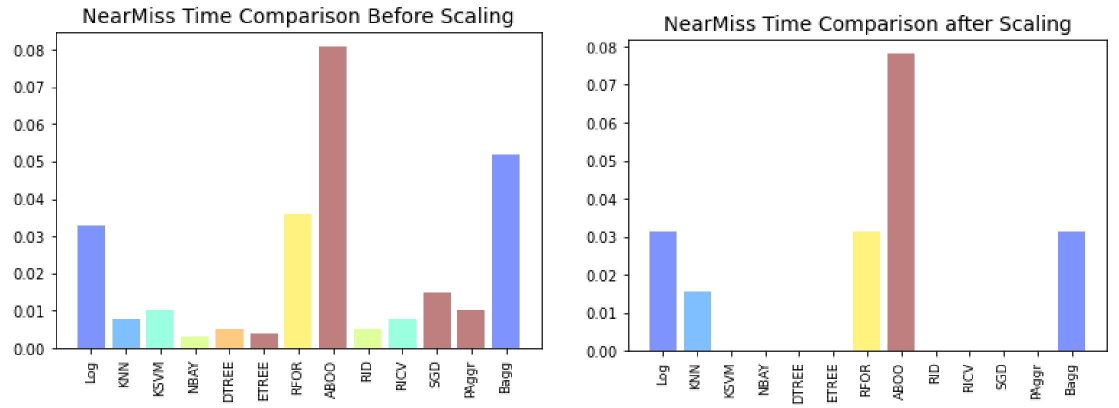

Figure 19. Response time analysis of NearMiss dataset before and after feature scaling

Table 15

Classifier performance of NearMiss dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.78	0.77	0.77	0.77	0.03
KNN	0.77	0.75	0.75	0.75	0.01
KSVM	0.76	0.71	0.71	0.71	0.01
GNBayes	0.71	0.69	0.68	0.69	0.00
DTree	0.91	0.91	0.91	0.91	0.01
ETree	0.76	0.75	0.75	0.75	0.00
RForest	0.91	0.91	0.91	0.91	0.04
AdaBoost	0.83	0.82	0.82	0.82	0.08
Ridge	0.75	0.75	0.75	0.75	0.00
RidgeCV	0.75	0.75	0.75	0.75	0.01
SGD	0.83	0.70	0.68	0.70	0.01
PAggress	0.43	0.60	0.50	0.60	0.01
Bagging	0.88	0.88	0.88	0.88	0.05

Table 16

Classifier performance of NearMiss dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.78	0.77	0.78	0.77	0.03
KNN	0.86	0.85	0.85	0.85	0.02
KSVM	0.82	0.81	0.81	0.81	0.00
GNBayes	0.78	0.62	0.62	0.62	0.00
DTree	0.87	0.87	0.87	0.87	0.00
ETree	0.74	0.74	0.74	0.74	0.00
RForest	0.90	0.90	0.90	0.90	0.03
AdaBoost	0.71	0.60	0.60	0.60	0.09
Ridge	0.73	0.72	0.72	0.72	0.00
RidgeCV	0.73	0.72	0.72	0.72	0.00
SGD	0.59	0.63	0.59	0.63	0.01
PAggress	0.80	0.78	0.79	0.78	0.00
Bagging	0.88	0.88	0.88	0.88	0.05

Conclusion

An endeavor is done to analyze the execution of non sampled target features with tested information. The CTG dataset utilized in this paper found to have nonsampled information with Ordinary, Suspect and Pathologic. This paper endeavor to perform undersampling with ClusterCentroids, CondensedNearestNeighbour, AllKNN, RepeatedENN, EditedNearestNeighbours, InstanceHardnessThreshold and NearMiss methods. Experimental results shows that the Decision Tree classifier tends to retain 98% before and after feature scaling for the undersampling with EditedNearestNeighbours, RepeatedENN and AllKNN methods.

References

1. Jianqiang Li., Luxiang Huang., Zhixia Shen., Yifan Zhang., Min Fang., Bing Li., Xianghua Fu., Qingguo Zhao., Huihui Wang. (2019). Automatic Classification of Fetal Heart Rate Based on Convolutional Neural Network. *IEEE Internet of Things Journal*, 6(2).

2. Ramla, M., Sangeetha, S., Nickolas, S. (2019). Fetal Health State Monitoring Using Decision Tree Classifier from Cardiotocography Measurements. *Proceedings of the Second International Conference on Intelligent Computing and Control Systems*, 11.

3. Jiri Spilka., Jordan Frecon., Roberto Leonarduzzi., Nelly Pustelnik., Patrice Abry., Muriel Doret. (2017). Sparse Support Vector Machine for Intrapartum Fetal Heart Rate Classification. *IEEE Journal of Biomedical And Health Informatics*, 21(3).

4. Vaanathi Sundaresan., Christopher Bridge, P., Christos Ioannou., Alison Noble, J. (2017). Automated Characterization of The Fetal Heart In Ultrasound Images Using Fully Convolutional Neural Networks. *Proceedings of the IEEE 14th International Symposium on Biomedical Imaging*.

5. Kezi Yu, J., Gerald Quirk., Petar, M. (2017). Fetal Heart Rate Classification By Non-Parametric Bayesian Methods. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*.

6. Jayashree Piri., Puspanjali Mohapatra. (2020). Exploring Fetal Health Status Using an Association Based Classification Approach *Proceedings of the International Conference on Information Technology*.

7. Sahana Das., Himadri Mukherjee., Santosh, K. C., Chanchal Kumar Saha., Kaushik Roy. (2020). Periodic Change Detection in Fetal Heart Rate Using Cardiotocograph. *Proceedings of the International Symposium on Computer-Based Medical Systems*.

8. Weisheng Gao., Yao Lu (2020) Fetal Heart Baseline Extraction And Classification based on Deep Learning. *Proceedings of the International Conference on Information Technology and Computer Application*

9. Jayashree, J., Harsha, T., Anil Kumar, C., Vijayashree, J. (2020). Enhanced Optimal Feature Selection Techniques for Fetal Risk Prediction using Machine Learning Algorithms. *International Journal of Engineering and Advanced Technology*, 9 (3).

10. Nurul Chamidah., Ito Wasito. (2016). Fetal State Classification from Cardiotocography Based on Feature Extraction Using Hybrid K-Means and Support Vector Machine *International Conference on Advanced Computer Science and Information Systems*, 25.

11. David Hareva, H., Dewi Sartika., Samuel Lukas., Lazarusli, A., Suryasari. (2016). Determination of Fetus Heart Rate Based 2D Ultrasound *Proceedings of the International Conference on Information and Communication Technology Convergence*, 5.

12. Jaganathan, D. (2017). Cardiotocography - A Comparative Study between Support Vector Machine and Decision Tree Algorithms. *International Journal of Trend in Research and Development*, 4(1).

13. Nina Sevani., Indra Hermawan., Wisnu Jatmiko, (2019). Feature Selection based on F-score for Enhancing CTG Data Classification. *Proceedings of the IEEE International Conference on Cybernetics and Computational Intelligence*.

14. Jayashree Piri., Puspanjali Mohapatra., Raghunath Dey. (2020) Fetal Health Status Classification Using MOGA - CD Based Feature Selection Approach. *Proceedings of the*

IEEE International Conference on Electronics, Computing and Communication Technologies.

15. Emerson Keenan., Radhagayrathi., Udhayakumar., Chandan Karmakar, K., Fiona Brownfoot, C., Marimuthu Palaniswami. (2020). Entropy Profiling for Detection of Fetal Arrhythmias in Short Length Fetal Heart Rate Recordings. *Proceedings of the International Conference of the IEEE Engineering in Medicine & Biology Society*.
16. Karthick, R., and M. Sundararajan. “SPIDER-based out-of-order execution scheme for HtMPSOC.” International Journal of Advanced Intelligence paradigms 19.1 (2021): 28-41. https://doi.org/10.1504/IJAIP.2021.114581
17. Sabarish, P., et al. “An Energy Efficient Microwave Based Wireless Solar Power Transmission System.” IOP Conference Series: Materials Science and Engineering. Vol. 937.No. 1.IOP Publishing, 2020. doi:10.1088/1757-899X/937/1/012013
18. Vijayalakshmi, S., et al. “Implementation of a new Bi-Directional Switch multilevel Inverter for the reduction of harmonics.” IOP Conference Series: Materials Science and Engineering. Vol. 937.No. 1.IOP Publishing, 2020. doi:10.1088/1757-899X/937/1/012026
19. P. Sabarish, R. Karthick, A. Sindhu, N. Sathiyanathan, Investigation on performance of solar photovoltaic fed hybrid semi impedance source converters, Materials Today: Proceedings, 2020, https://doi.org/10.1016/j.matpr.2020.08.390
20. Karthick, R., and M. Sundararajan. “A novel 3-D-IC test architecture-a review.” International Journal of Engineering and Technology (UAE) 7.1.1 (2018): 582-586.
21. Karthick, R., and M. Sundararajan. “Design and implementation of low power testing using advanced razor based processor.” International Journal of Applied Engineering Research 12.17 (2017): 6384-6390.
22. Suresh, HelinaRajini, et al. “Suppression of four wave mixing effect in DWDM system.” Materials Today: Proceedings (2021). https://doi.org/10.1016/j.matpr.2020.11.545
23. M Ramkumar, C Ganesh Babu, K Vinoth Kumar, D Hepsiba, A Manjunathan, R Sarath Kumar, “ECG Cardiac arrhythmias Classification using DWT, ICA and MLP Neural Network”, Journal of Physics: Conference Series, vol.1831, issue.1, pp.012015, 2021
24. M Ramkumar, C Ganesh Babu, A Manjunathan, S Udhayanan, M Mathankumar,”A Graphical User Interface Based Heart Rate Monitoring Process and Detection of PQRST Peaks from ECG Signal” Lecture Notes in Networks and Systems, 2021, 173 LNNS, pp. 481–496