

Integrated UnderOversampling Responsive based Fetal Health Prediction using Cardiotocographic Data

S. Sridevi^a, M. Shyamala Devi^b, Kanamukkala Vinod Kumar Reddy^c and Ramya Harika Dadi^d

^a Associate Professor, Department of Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu.

^b Professor, Department of Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu.

^c Third Year B.Tech Student, Department of Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu.

^d Third Year B.Tech Student, Department of Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu.

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: With the current enhancement of advancement towards pharmaceutical, particular ultrasound strategies are open to discover the fetal thriving. It is analyzed with assorted clinical parameters with 2-D imaging and other test. In any case, success want of fetal heart still remains an open issue due to unconstrained works out of the hatchling, the minor heart assess and lacking of information in fetal echocardiography. The machine learning techniques can discover out the classes of fetal heart rate which can be utilized for prior assessing. 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 NeighbourhoodCleaningRule, OneSidedSelection, RandomUnderSampler, TomekLinks, SmoteENN and SmoteTomek. 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 Random Forest classifier tends to retain 98% before and after feature scaling for the undersampling with NeighbourhoodCleaningRule, SmoteENN and SmoteTomek methods comparing to other methods.

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

1. Introduction

The sporadic changes of the baby must be observed through the clinical parameters in coordinate to initiate to the fetal thriving. The passing rate of the fetal can be controlled by foreseeing the changes interior the clinical parameters of the fetal success. With the mechanical progression, the ultrasound strategies are utilized for the evaluation of fetal wellbeing and other changes interior the required properties. The natural calculations can other than be utilized for the figure of any sicknesses interior the fetal thriving by translating the acumen gotten through the parameter changes. Fetal heart rate observing may be a method of checking the rate and cadence of the fetal beat. The normal fetal heart rate is between 120 and 160 beats per miniature. This rate may change as the hatchling reacts to conditions interior the uterus. The appraisal of fetal success has thought for different a long time. The appraisal of fetal thriving has had our competent thought for different a long time. As the upgrade of advancements for pre-birth symptomatic techniques has advanced, applications of such improvements have supported interior the colossal examination of fetal well-being. Fetal heart-rate checking remains the first shape of fetal evaluation for high-risk pregnancies. The extra examinations directed by the examination of ST and T-wave changes of the fetal electrocardiogram hold guarantee for moving forward the prescient respect of fetal heart-rate evaluations. Ultrasound has been priceless for assessment of fetal life systems, and the utilization of Doppler ultrasound has given data into fetal cardiovascular reactions to such conditions as intrauterine progression control and fetal slowness caused by red blood cell immunization. There is a tremendous scope for machine learning calculations in anticipating cardiovascular maladies or heart related illnesses. Each of the specified calculations have performed amazingly well in a few cases but ineffectively in a few other cases.

2. Background

This paper deals with the examination of cardiotocography for fetal state classification utilizing machine learning algorithms. The aim is to extend the cross examination remove and to expand the estimation of single

parameter to numerous parameters. An LC sensor is ordinarily developed from a winding inductor associated with a detecting capacitor, shaping a thunderous LC tank. The capacitor changes in reaction to the parameter of intrigued, coming about in a move in its resounding frequency LC detached remote sensor with one inductor is concerned, in arrange to screen different parameters, diverse touchy components are required to be coordinates within the same chip. The detached LC remote sensors with littler measure, longer interrogating distance and more capacities proceed to advance within the future [1]. This paper attempts to classify the Cardiotocogram Information for expectation of fetal dangers utilizing sacking outfit classifier. It discover fetal dangers based on CTG, machine learning procedures to form an expanding drift to deliver choice bolster frameworks in medication. Distinctive methods have carried out for the classification of the CTG data. The data taken from CTG is utilized for early distinguishing proof of a neurotic state and can offer assistance the obstetrician to predict. The issues and ruin some time recently happening a lasting disability to the hatchling. Three proficient obstetricians categorized the CTG information, by testing the status of the baby whether solid or undesirable. WEKA information mining apparatus were utilized for the execution of the classifiers. Execution of each classifier was assessed utilizing 10-fold cross approval and CTG dataset. The CTG information utilized for the classification and prepared numerous times by changing the specific parameters of each classifier to attain greatest classification performance [2].

This paper foresee heart disease utilizing machine learning. Heart maladies have risen as one of the foremost conspicuous cause of passing all around the world. In this way, attainable and precise forecast of heart related illnesses is exceptionally critical. Machine learning calculations have ended up exceptionally valuable, in later times, to foresee the nearness or nonappearance of heart related diseases accurately. Dimensionality Diminishment may be exceptionally vital step considered whereas building any show. Dimensionality Diminishment is by and large accomplished by two strategies -Include Extraction and Highlight Selection. SVM performed greatly well for most of the cases. Frameworks based on machine learning calculations and procedures have been exceptionally precise in anticipating the heart related diseases [3]. This paper investigate fetal wellbeing status utilizing an affiliation based classification. In wellbeing care space mining of valuable information from restorative information such as CTG could be a major challenge. By recording fetal heart rate and uterine withdrawals (UC) amid conveyance, CTG evaluates maternal and fetal well-being. The proposed framework is pointed at planning a show utilizing acquainted classification method to analyze fetal development to upgrade the quality of determination for pregnant females with lessening fetal development (DFM) [4].

This paper anticipate the fetal wellbeing status classification utilizing MOGA-CD based highlight determination approach. One of the multi-objective highlight determination strategies MOGA-CD is actualized in this paper to discover critical highlights for classifying fetal wellbeing into three bunches ordinary, suspicious and obsessive from CTG data. The prescient precision of a foreordained learning calculation is utilized by the wrapper show to test the exactness of the chosen features [5]. This article anticipate the diabetes in therapeutic in fetal wellbeing care utilizing machine learning. Healthcare industry contains exceptionally expansive and touchy information and has to be dealt with exceptionally carefully. Diabetes Mellitus is one of the developing amazingly lethal maladies all over the world. Restorative experts need a solid forecast framework to analyze Diabetes. Diverse machine learning strategies are valuable for looking at the information from differing viewpoints and synopsisizing it into valuable information. Information preprocessing could be a procedure of machine learning that comprises of converting raw information into a consistent or comprehensible feature. Naive Bayes could be the information mining classification method and it is utilized as a classifier. This classifier is utilized for likelihood expectation in the event that a test has a place to specific class [6]. This paper identify the heart beat and uterine withdrawals utilizing machine learning. Objective: Machine learning calculations within the healthcare space can move forward healthcare and clinical hone morally and dependably. The fetal heart rate (FHR) and the uterine compression (UC) movement are recorded by utilizing the method called Cardiotocography (CTG). It gives back for the obstetricians to get total physiological data approximately new-borns [7].

This paper points with the special issue to assist progress the logical inquire about inside the wide field of machine learning in restorative imaging. Cell discovery and classification are regularly performed sequentially and independently by machine learning or profound learning. Tune et al., in any case, propose a synchronized profound auto-encoder arrange for synchronous location and classification of cells in bone marrow histology pictures. The proposed arrange employments a single design to identify the positions of cells and classifiers [8]. This paper foresee the fetal illness utilizing machine learning over enormous data. The examination precision is decreased when the quality of therapeutic information in fragmented. Besides, diverse districts show interesting characteristics of certain territorial infections, which may debilitate the forecast of infection episodes. In any case, those existing work generally considered organized information. Machine learning strategies, approaches, and devices that can offer assistance settling expository and prescient hitches in a variety of therapeutic zones [9]. This paper anticipate the heart illness utilizing machine learning and information analytics. The assessment of a person's hazard for coronary heart illness is vital for numerous angles of wellbeing advancement and clinical pharmaceutical. A hazard expectation show may be gotten through multivariate regression. By utilizing distinctive sorts of information mining and machine learning methods to foresee the event of heart malady have summarized. It decides the expectation execution of each calculation and apply the proposed framework for the zone it required. Utilizes more significant include determination strategies to move forward the precise execution of algorithms

[10]. This paper gives the overview of machine learning calculations for malady diagnostic. It deals with the examination of high-dimensional and multimodal bio-medical information, machine learning offers a commendable approach for making classy and programmed calculations. This study paper gives the comparative examination of diverse machine learning calculations for conclusion of distinctive maladies such as heart illness, diabetes disease [11].

3. Proposed Work

The CTG Cardiographic Fetal heart rate dataset with 36 independent variables and 1 dependent variable has been used for implementation. Fig. 1 shows the overall workflow of this work

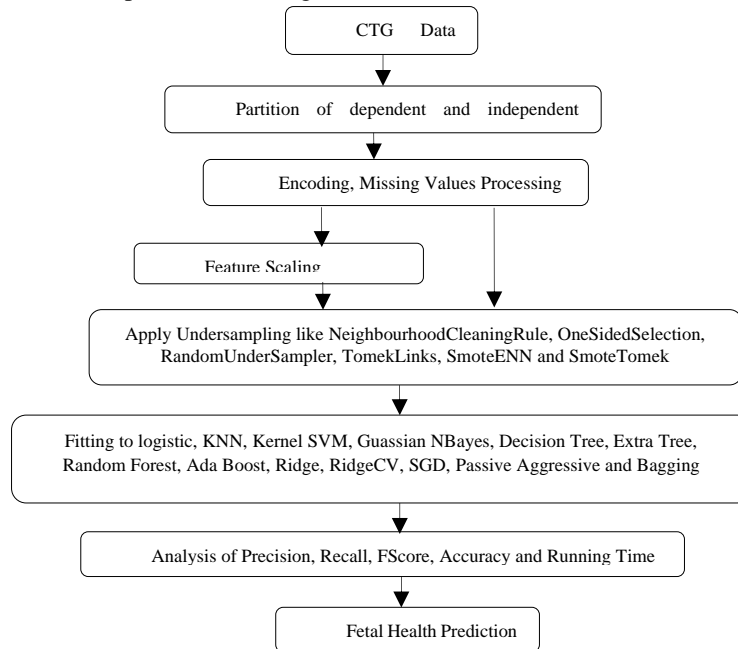


Fig.1. Overall workflow of the system.

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 NeighbourhoodCleaningRule, OneSidedSelection, RandomUnderSampler, TomekLinks, SmoteENN and SmoteTomek.
- (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.

4. Exploratory Data Analysis

The CTG dataset extricated from the UCI machine learning store is utilized for usage. The dataset comprises of 2127 patients information with 21 autonomous highlights (baseline value, accelerations, fetal movement, Uterine contractions, light decelerations, severe decelerations, prolonged 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". 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.

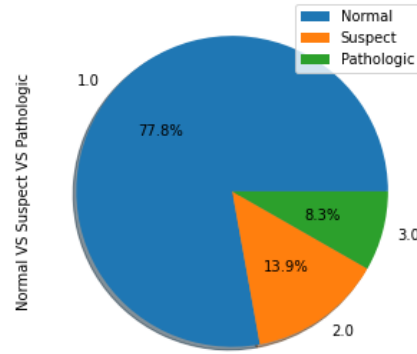


Fig.2. Target feature analysis of the dataset

5. 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 the running time comparison is shown in Figure. 3 - 4.

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

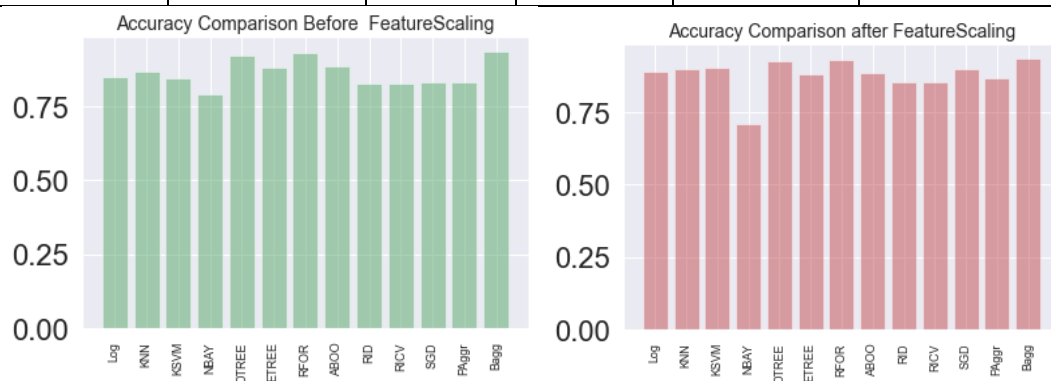


Fig.3. Accuracy analysis of raw dataset before and after feature scaling

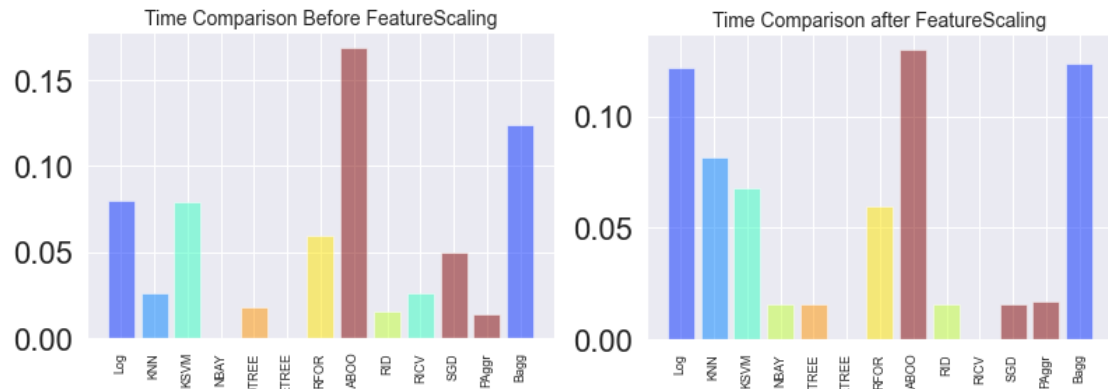


Fig.4. Response time analysis of raw dataset before and after feature scaling

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

6. Undersampling Results and Performance Analysis

The raw data set is subjected to undersampling methods NeighbourhoodCleaningRule, OneSidedSelection, RandomUnderSampler, TomekLinks, SmoteENN and SmoteTomek. The resampled dataset distribution after undersampling is shown in Fig.5. The raw data set is subjected to undersampling method namely NeighbourhoodCleaningRule 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 Fig. 6 - 7.

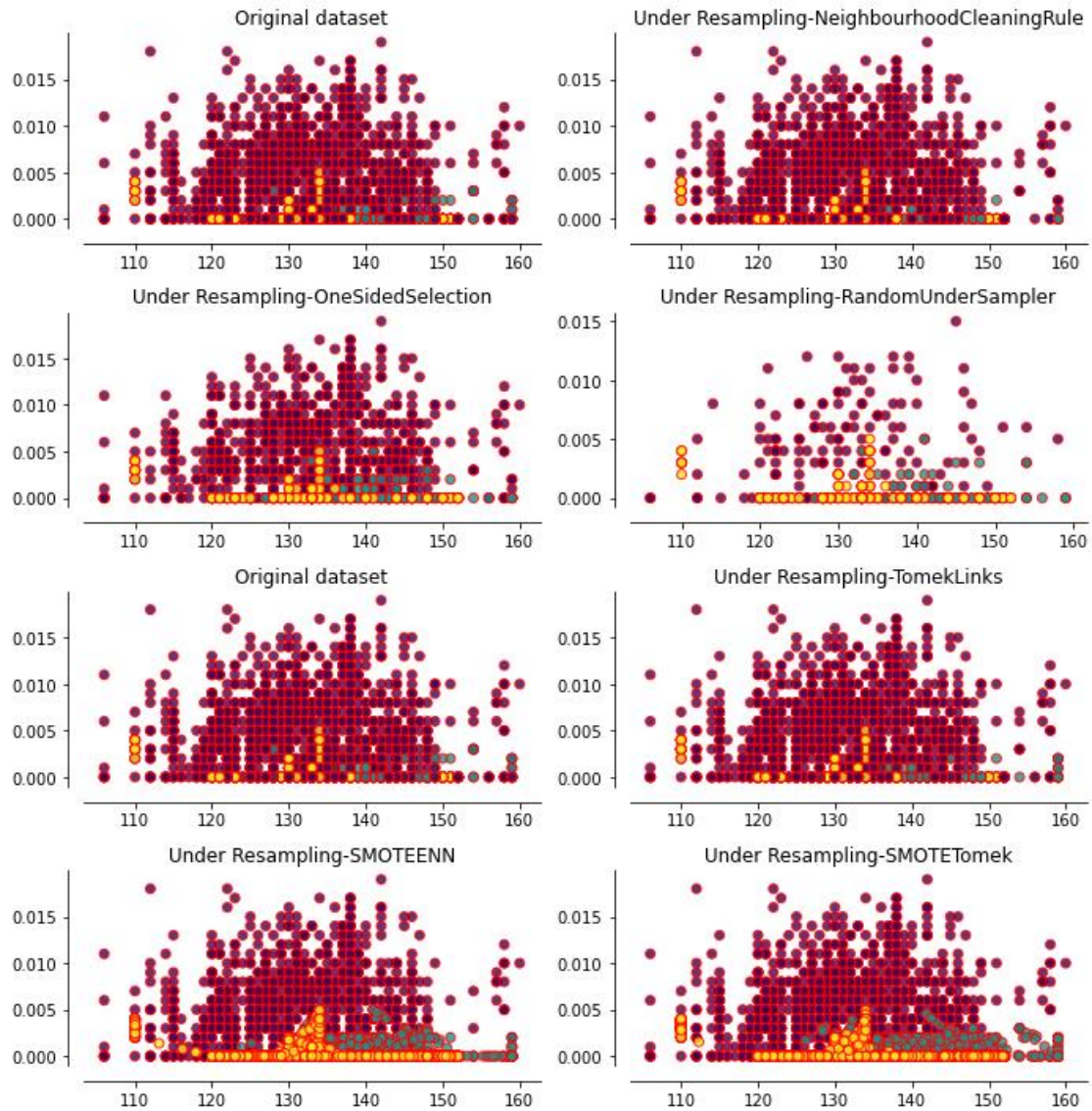


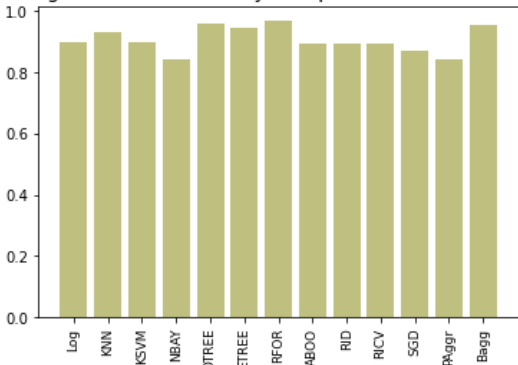
Fig.5. Data distribution after undersampling methods

Table 3. Classifier performance of NeighbourhoodCleaningRule dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.89	0.90	0.89	0.90	0.06
KNN	0.93	0.93	0.93	0.93	0.03
KSVM	0.89	0.90	0.89	0.90	0.05
GNBayes	0.90	0.84	0.86	0.84	0.00
DTree	0.96	0.96	0.96	0.96	0.02
ETree	0.95	0.94	0.94	0.94	0.00
RForest	0.97	0.97	0.97	0.97	0.06
AdaBoost	0.90	0.90	0.89	0.90	0.16
Ridge	0.89	0.89	0.88	0.89	0.01
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
---------	------	------	------	------	------

Neighbourhood Accuracy Comparison Before Scaling



Neighbourhood Accuracy Comparison after Scaling

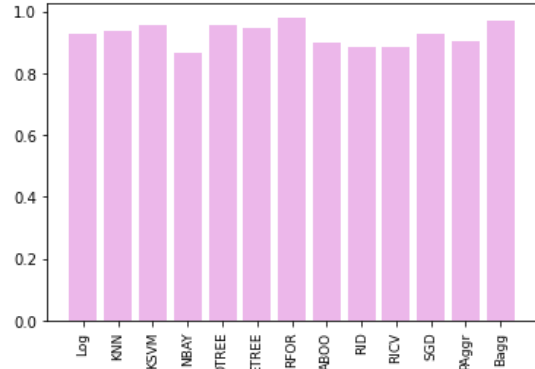
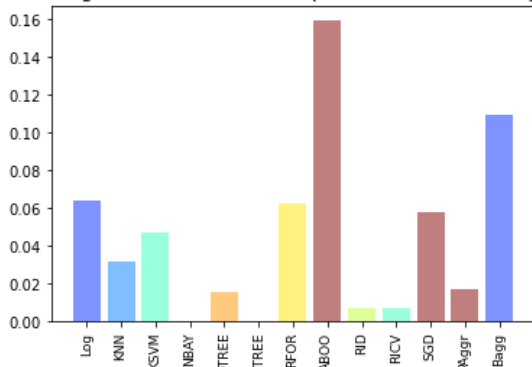


Fig.6. Accuracy analysis of NeighbourhoodCleaningRule dataset before and after feature scaling

Neighbourhood Time Comparison Before Scaling



Neighbourhood Time Comparison after Scaling

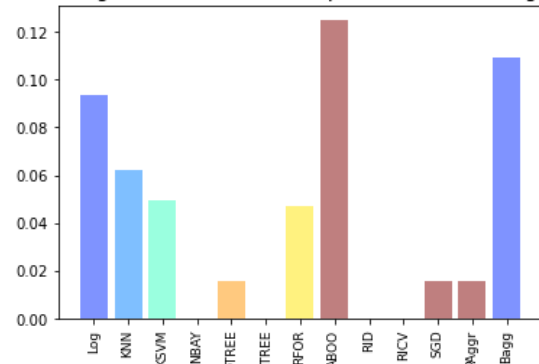


Fig.6. Response time analysis of NeighbourhoodCleaningRule dataset before and after feature scaling

Table 4. Classifier performance of NeighbourhoodCleaningRule 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	0.95	0.95	0.95	0.95	0.05
GNBayes	0.91	0.87	0.88	0.87	0.00
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	0.90	0.90	0.89	0.90	0.12
Ridge	0.87	0.89	0.87	0.89	0.00
RidgeCV	0.87	0.89	0.87	0.89	0.00
SGD	0.93	0.93	0.93	0.93	0.02
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 OneSidedSelection 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 Fig. 8 - 9.

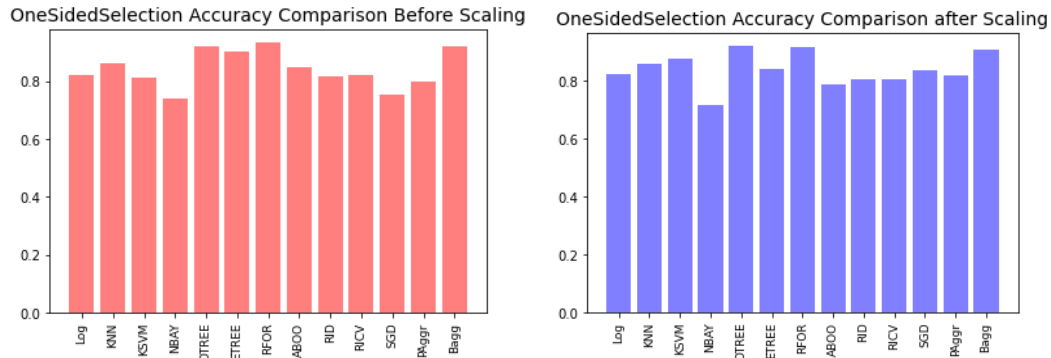


Fig.8. Accuracy analysis of OneSidedSelection dataset before and after feature scaling

Table 5. Classifier performance of OneSidedSelection dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.81	0.82	0.82	0.82	0.07
KNN	0.86	0.86	0.86	0.86	0.03
KSVM	0.80	0.81	0.80	0.81	0.05
GNBayes	0.82	0.74	0.76	0.74	0.00
DTree	0.92	0.92	0.92	0.92	0.01
ETree	0.90	0.90	0.90	0.90	0.00
RForest	0.93	0.93	0.93	0.93	0.06
AdaBoost	0.86	0.85	0.84	0.85	0.13
Ridge	0.81	0.82	0.80	0.82	0.01
RidgeCV	0.82	0.82	0.81	0.82	0.01
SGD	0.69	0.75	0.71	0.75	0.03
PAGgress	0.80	0.80	0.80	0.80	0.02
Bagging	0.92	0.92	0.92	0.92	0.12

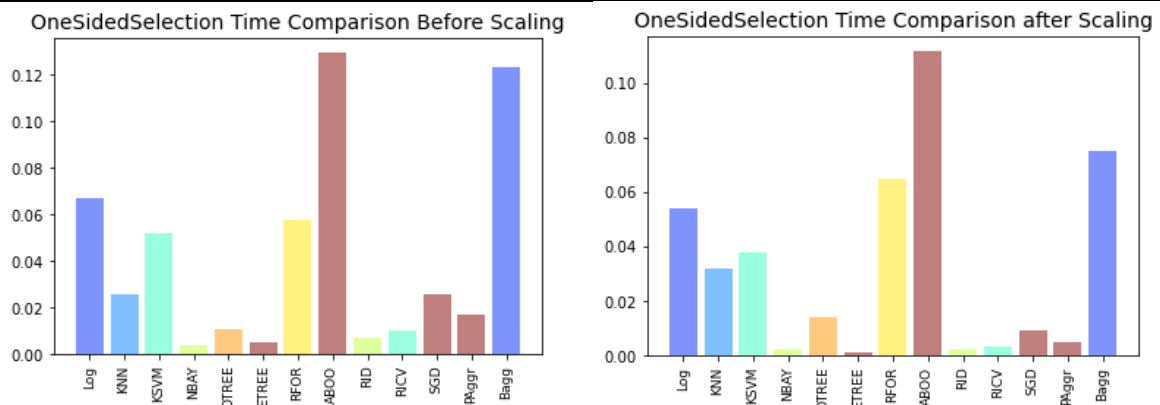


Fig.9. Response time analysis of OneSidedSelection dataset before and after feature scaling

Table 6. Classifier performance of OneSidedSelection dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.84	0.83	0.83	0.83	0.05
KNN	0.86	0.86	0.85	0.86	0.03
KSVM	0.89	0.88	0.88	0.88	0.04
GNBayes	0.81	0.72	0.74	0.72	0.00

DTree	0.92	0.92	0.92	0.92	0.01
ETree	0.86	0.84	0.85	0.84	0.00
RForest	0.92	0.92	0.92	0.92	0.06
AdaBoost	0.80	0.79	0.78	0.79	0.11
Ridge	0.80	0.81	0.80	0.81	0.00
RidgeCV	0.80	0.80	0.79	0.80	0.00
SGD	0.85	0.84	0.84	0.84	0.01
PAGgress	0.82	0.82	0.82	0.82	0.00
Bagging	0.91	0.91	0.91	0.91	0.07

The raw data set is subjected to undersampling method namely RandomUnderSampler 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 Fig. 10 - 11.

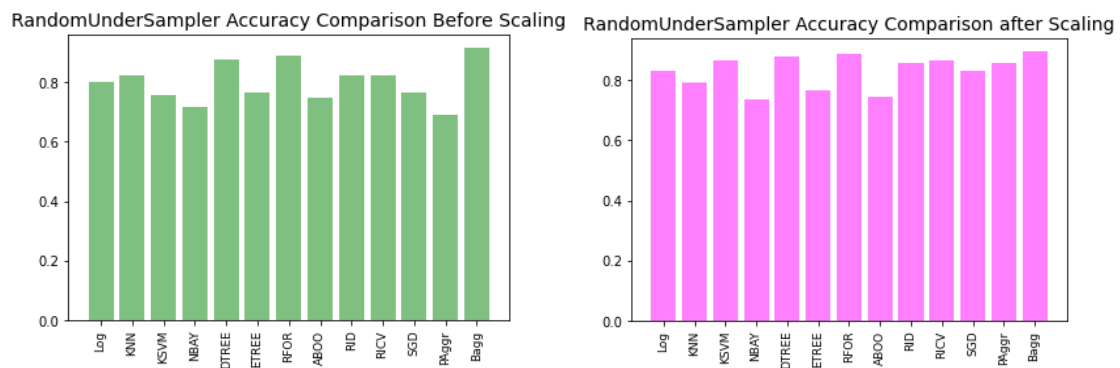


Fig.10. Accuracy analysis of RandomUnderSampler dataset before and after feature scaling

Table 7. Classifier performance of RandomUnderSampler dataset before scaling

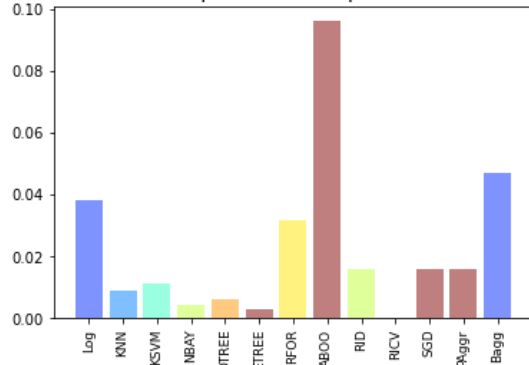
Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.80	0.80	0.80	0.80	0.04
KNN	0.83	0.82	0.82	0.82	0.01
KSVM	0.76	0.75	0.76	0.75	0.01
GNBayes	0.74	0.72	0.72	0.72	0.00
DTree	0.88	0.88	0.88	0.88	0.01
ETree	0.76	0.76	0.76	0.76	0.00
RForest	0.89	0.89	0.89	0.89	0.03
AdaBoost	0.78	0.75	0.75	0.75	0.10
Ridge	0.83	0.82	0.82	0.82	0.02
RidgeCV	0.83	0.82	0.82	0.82	0.00
SGD	0.79	0.76	0.77	0.76	0.02
PAGgress	0.68	0.69	0.65	0.69	0.02
Bagging	0.92	0.92	0.92	0.92	0.05

Table 8. Classifier performance of RandomUnderSampler dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.83	0.83	0.83	0.83	0.03
KNN	0.79	0.79	0.79	0.79	0.02

KSVM	0.87	0.87	0.87	0.87	0.00
GNBayes	0.76	0.74	0.74	0.74	0.00
DTree	0.88	0.88	0.88	0.88	0.00
ETree	0.76	0.76	0.76	0.76	0.00
RForest	0.89	0.89	0.89	0.89	0.03
AdaBoost	0.78	0.75	0.75	0.75	0.10
Ridge	0.87	0.86	0.86	0.86	0.00
RidgeCV	0.88	0.87	0.87	0.87	0.00
SGD	0.84	0.83	0.83	0.83	0.01
PAGgress	0.87	0.86	0.86	0.86	0.01
Bagging	0.90	0.90	0.90	0.90	0.06

RandomUnderSampler Time Comparison Before Scaling



RandomUnderSampler Time Comparison after Scaling

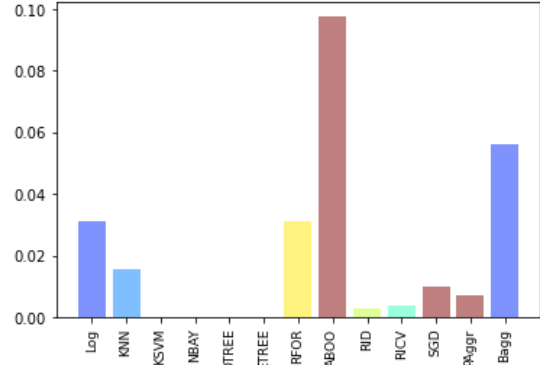


Fig.11. Response time analysis of RandomUnderSampler dataset before and after feature scaling

The raw data set is subjected to undersampling method namely TomekLinks 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 Fig. 12 - 13.

Table 9. Classifier performance of TomekLinks dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.86	0.88	0.86	0.88	0.07
KNN	0.88	0.89	0.88	0.89	0.05
KSVM	0.86	0.87	0.86	0.87	0.08
GNBayes	0.88	0.83	0.85	0.83	0.00
DTree	0.94	0.94	0.94	0.94	0.02
ETree	0.89	0.89	0.89	0.89	0.00
RForest	0.93	0.93	0.93	0.93	0.05
AdaBoost	0.87	0.86	0.86	0.86	0.15
Ridge	0.84	0.86	0.84	0.86	0.01
RidgeCV	0.84	0.86	0.84	0.86	0.01
SGD	0.80	0.83	0.79	0.83	0.05
PAGgress	0.88	0.74	0.78	0.74	0.01
Bagging	0.94	0.94	0.94	0.94	0.12

Table 10. Classifier performance of TomekLinks dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
------------	-----------	--------	--------	----------	-------------------

Logistic	0.91	0.91	0.91	0.91	0.08
KNN	0.91	0.91	0.91	0.91	0.06
KSVM	0.93	0.93	0.93	0.93	0.05
GNBayes	0.90	0.84	0.86	0.84	0.00
DTree	0.94	0.93	0.94	0.93	0.02
ETree	0.91	0.91	0.91	0.91	0.02
RForest	0.95	0.95	0.95	0.95	0.06
AdaBoost	0.90	0.90	0.90	0.90	0.14
Ridge	0.88	0.89	0.88	0.89	0.00
RidgeCV	0.88	0.89	0.88	0.89	0.00
SGD	0.92	0.91	0.91	0.91	0.02
PAGgress	0.92	0.89	0.90	0.89	0.00
Bagging	0.95	0.95	0.95	0.95	0.13

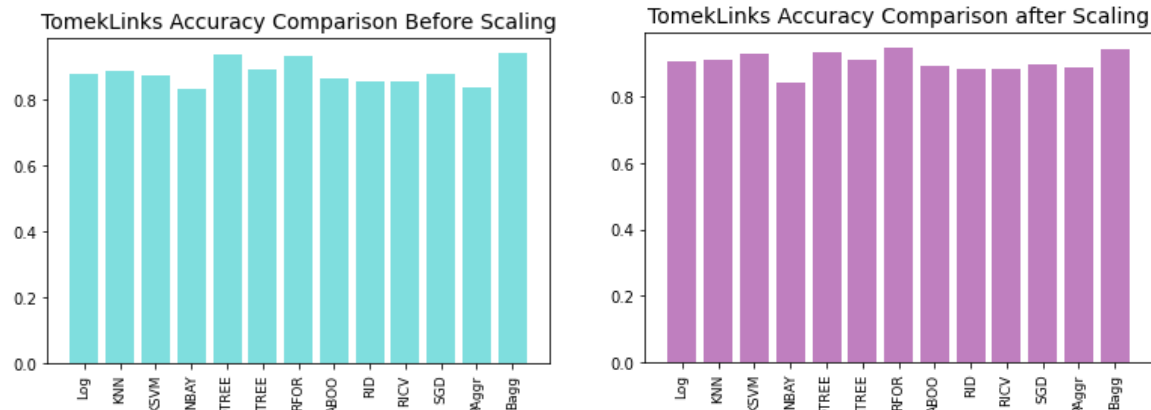


Fig.12. Accuracy analysis of TomekLinks dataset before and after feature scaling

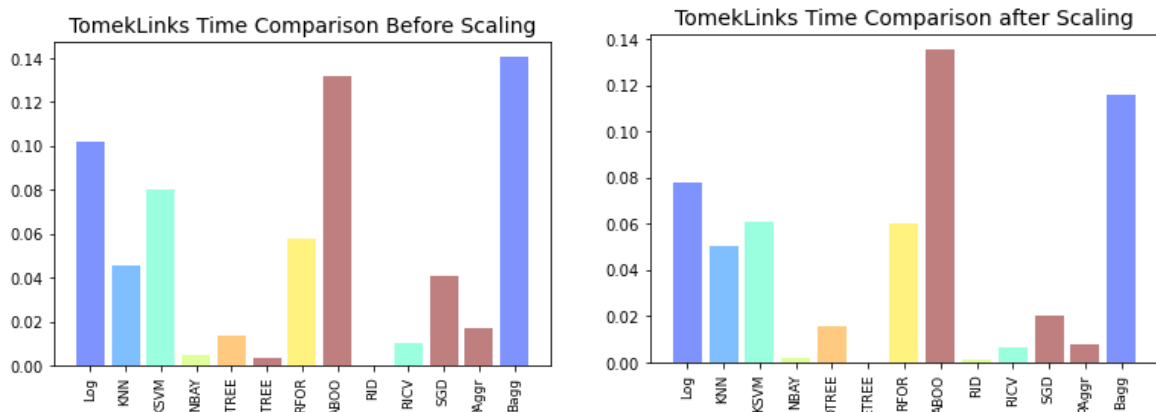


Fig.13. Response time analysis of TomekLinks dataset before and after feature scaling

The raw data set is subjected to undersampling method namely SmoteENN 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 Fig. 14 - 15.

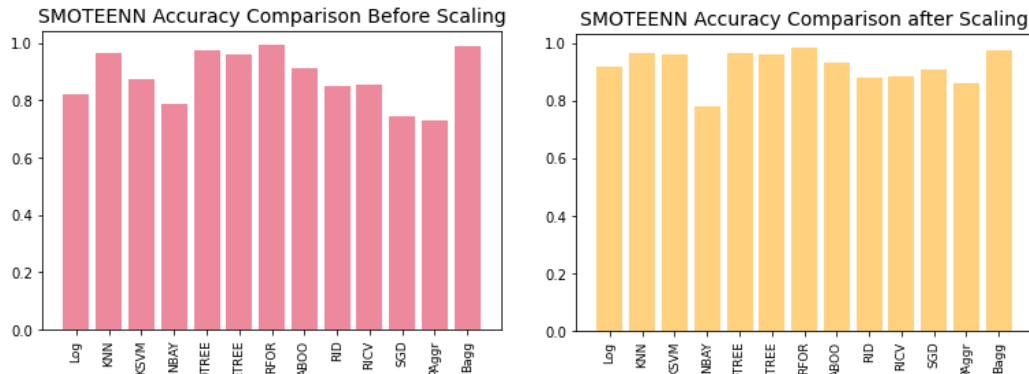


Fig.14. Accuracy analysis of SmoteENN dataset before and after feature scaling

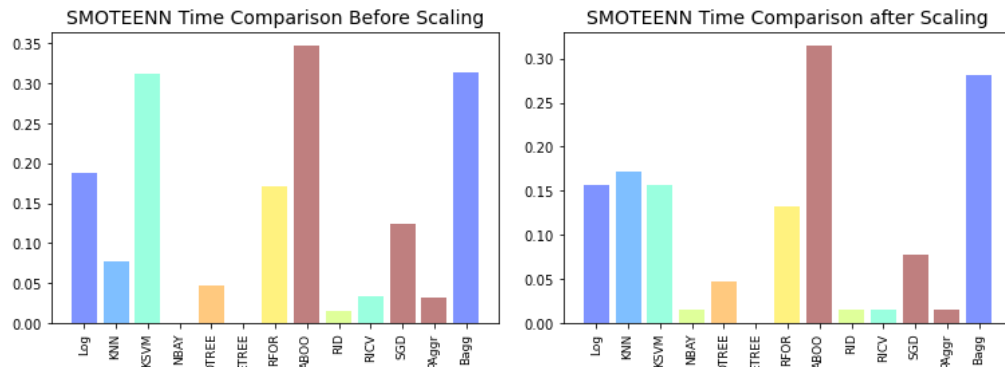


Fig.15. Response time analysis of SmoteENN dataset before and after feature scaling

Table 11. Classifier performance of SmoteENN dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.82	0.82	0.82	0.82	0.19
KNN	0.97	0.96	0.96	0.96	0.08
KSVM	0.87	0.87	0.87	0.87	0.31
GNBayes	0.81	0.79	0.79	0.79	0.00
DTree	0.98	0.98	0.98	0.98	0.05
ETree	0.96	0.96	0.96	0.96	0.00
RForest	0.99	0.99	0.99	0.99	0.17
AdaBoost	0.92	0.91	0.91	0.91	0.35
Ridge	0.85	0.85	0.85	0.85	0.02
RidgeCV	0.86	0.86	0.86	0.86	0.03
SGD	0.77	0.75	0.71	0.75	0.12
PAggress	0.82	0.73	0.74	0.73	0.03
Bagging	0.99	0.99	0.99	0.99	0.31

Table 12. Classifier performance of SmoteENN dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.92	0.92	0.92	0.92	0.16
KNN	0.97	0.97	0.97	0.97	0.17
KSVM	0.96	0.96	0.96	0.96	0.16
GNBayes	0.81	0.78	0.77	0.78	0.02
DTree	0.97	0.97	0.97	0.97	0.05

ETree	0.96	0.96	0.96	0.96	0.00
RForest	0.99	0.98	0.98	0.98	0.13
AdaBoost	0.93	0.93	0.93	0.93	0.31
Ridge	0.89	0.88	0.88	0.88	0.02
RidgeCV	0.89	0.88	0.88	0.88	0.02
SGD	0.91	0.91	0.91	0.91	0.08
PAggress	0.86	0.86	0.86	0.86	0.02
Bagging	0.98	0.98	0.98	0.98	0.28

The raw data set is subjected to undersampling method namely SmoteTomek 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 Fig. 16 - 17.

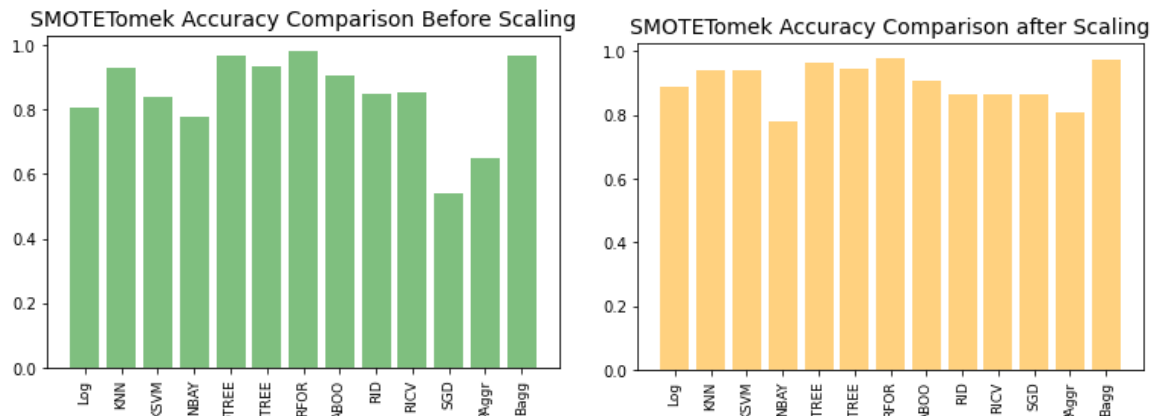


Fig.16. Accuracy analysis SmoteTomek dataset before and after feature scaling

Table 13. Classifier performance of SmoteTomek dataset before scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.80	0.80	0.80	0.80	0.18
KNN	0.93	0.93	0.93	0.93	0.10
KSVM	0.84	0.84	0.84	0.84	0.47
GNBayes	0.80	0.78	0.78	0.78	0.01
DTree	0.96	0.96	0.96	0.96	0.06
ETree	0.93	0.93	0.93	0.93	0.00
RForest	0.98	0.98	0.98	0.98	0.18
AdaBoost	0.91	0.91	0.90	0.91	0.35
Ridge	0.85	0.85	0.85	0.85	0.00
RidgeCV	0.85	0.85	0.85	0.85	0.01
SGD	0.80	0.54	0.47	0.54	0.15
PAggress	0.71	0.65	0.56	0.65	0.03
Bagging	0.97	0.97	0.97	0.97	0.30

Table 14. Classifier performance of SmoteTomek dataset after scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.89	0.89	0.89	0.89	0.14
KNN	0.95	0.94	0.94	0.94	0.22

KSVM	0.94	0.94	0.94	0.94	0.22
GNBayes	0.82	0.78	0.78	0.78	0.00
DTree	0.96	0.96	0.96	0.96	0.05
ETree	0.94	0.94	0.94	0.94	0.01
RForest	0.98	0.98	0.98	0.98	0.16
AdaBoost	0.91	0.91	0.91	0.91	0.33
Ridge	0.88	0.86	0.87	0.86	0.00
RidgeCV	0.88	0.86	0.87	0.86	0.01
SGD	0.87	0.87	0.87	0.87	0.09
PAGgress	0.80	0.81	0.80	0.81	0.02
Bagging	0.97	0.97	0.97	0.97	0.31

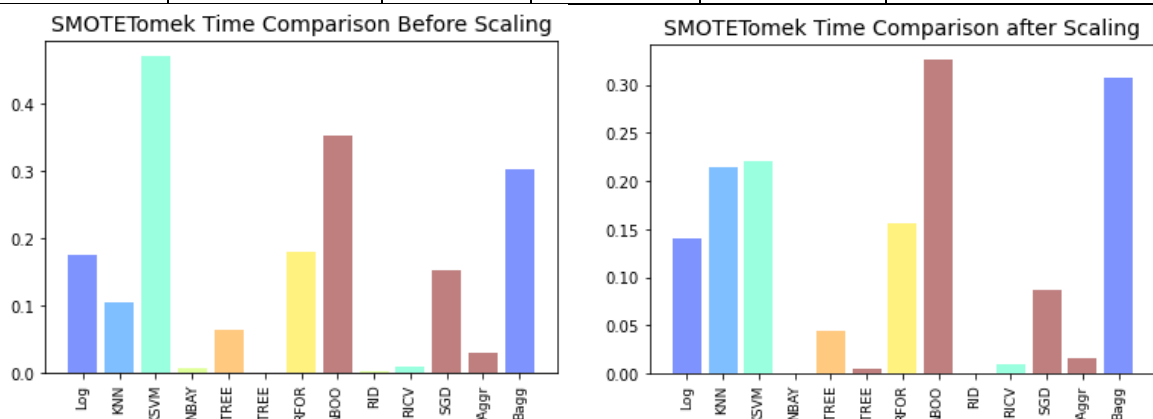


Fig.17. Response time analysis of SmoteTomek dataset before and after feature scaling

7. Conclusion

An endeavor is done to analyze the execution of non inspected target highlights with trained CTG data. The CTG dataset utilized in this paper found to have nonsampled data with Conventional, Suspect and Pathologic. This paper endeavor to perform undersampling with NeighbourhoodCleaningRule, OneSidedSelection, RandomUnderSampler, TomekLinks, SmoteENN and SmoteTomek methods. Experimental results shows that the Random Forest classifier tends to retain 98% before and after feature scaling for the undersampling with NeighbourhoodCleaningRule, SmoteENN and SmoteTomek methods comparing to other methods.

References

- Qing-An Huang, Lei Dong, and Li-Feng Wang, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms," in *Journal of Micro Electromechanical Systems*, vol. 25, no. 5, October 2016.
- Abdulhamit Subasia, Bayader Kadasaa, and Emir Kremicb, "Classification of the Cardiotocogram Data for Anticipation of Fetal Risks using Bagging Ensemble Classifier," in *Elsevier, Procedia Computer Science*, vol. 168, pp. 34–39, 2020
- V.V. Ramalingam, Ayantan Dandapath, and M Karthik Raja, "Heart disease prediction using machine learning techniques," in *International Journal of Engineering & Technology*, March 2018.
- Jayashree Piri, and Puspanjali Mohapatra, "Exploring Fetal Health Status Using an Association Based Classification," in *Proceedings of the International Conference on Information Technology*, 2019.
- Jayashree Piri, and Puspanjali Mohapatra, "Fetal Health Status Classification Using MOGA -CD Based Feature Selection Approach," in *Proceedings of the IEEE International conference on Electronics, Computing and Communication technologies*, 2020.
- Uswa Ali Zia, and Naeem Khan, "Predicting Diabetes in Medical Datasets Using Machine Learning Techniques," in *Journal of Scientific & Engineering Research*, vol. 8, no. 5, 2017.
- S. Vairachilai, Mondaddula Nivedh Vishnu Vardhana Reddy, and T. Krishnan, "Machine Learning Approach for Fetal Heartbeat and Uterine Contractions Monitoring," in *Journal of Advanced Research in Dynamical & Control Systems*, vol. 12, no. 8, 2020.
- Hayit Greenspan, Bram van Ginneken, and M. Ronald, "Machine Learning in Medical Imaging," in *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, 2019.

S. Vinitha, S. Sweetlin, H. Vinusha, and S. Sajini, "An Disease Prediction Using Machine Learning Over Big Data," in *International Journal (CSEIJ)*, vol.8, no.1, 2018.

M. Marimuthu, K. S. Hariesh, and K. Madhankumar, "Heart Disease Prediction using Machine Learning and Data Analytics Approach," in *International Journal of Computer Applications*, vol. 181, no. 18, September 2018

Meherwar Fatima, and Maruf Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic," in *Journal of Intelligent Learning Systems and Applications*, vol. 9, pp. 1-16, 2017.