Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Oversampling Response Stretch based Fetal Health Prediction using Cardiotocographic Data

M. Shyamala Devi[1], S. Sridevi[2], Kalyan kumar Bonala[3], Ramya Harika Dadi[4], Kanamukkala Vinod Kumar Reddy[5]

[1]Professor, [2]Associate Professor, Department of Computer Science & Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

[3, 4, 5] Third Year B.Tech Students, Department of Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India .Email Id: shyamaladevim@veltech.edu.in

ABSTRACT

With the current development of innovation towards medication, different ultrasound strategies are accessible to discover the fetal wellbeing. It is analyzed with different clinical parameters with 2-D imaging and other test. However, wellbeing expectation of fetal heart still remains an open issue due to unconstrained exercises of the hatchling, the minor heart estimate and insufficiency of information in fetal echocardiography. The machine learning methods can discover out the classes of fetal heart rate which can be utilized for prior estimating. With this outline, we have utilized Cardiotocographic Fetal heart rate dataset extricated from UCI Machine Learning Store for foreseeing the fetal heart rate wellbeing classes. The Prediction of fetal health rate are accomplished in six ways. Firstly, the data set is preprocessed with Feature Scaling and lost values. Secondly, exploratory data analysis is done and the distribution 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 oversampling methods like Random Oversampler, SMOTE, Borderline SMOTE, KMeansSMOTE, SVMSMOTE and ADASYN. Fifth, the oversampled 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 and Decision Tree classifier tends to retain 93% and 92% accuracy respectively before and after feature scaling. The Random Oversampled dataset shows that the Random Forest and Decision Tree classifier tends to retain 99% and 98% accuracy respectively before and after feature scaling. The SMOTE, Borderline SMOTE, KMeansSMOTE, SVMSMOTE and ADASYN resampled dataset shows that the Random Forest and Decision Tree classifier tends to retain 97% and 96% accuracy respectively before and after feature scaling.

Keywords: Machine learning; Classification; Oversampling; Accuracy; Precision; SMOTE

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Introduction

The intermittent changes of the embryo must be observed through the clinical parameters in arrange to get to the fetal wellbeing. The passing rate of the fetal can be controlled by predicting the changes within the clinical parameters of the fetal wellbeing. With the mechanical development, the ultrasound strategies are utilized for the evaluation of fetal health and other changes within the required properties. The hereditary calculations can moreover be used for the forecast of any maladies within the fetal wellbeing by deciphering the perception gotten through the parameter changes. Fetal heart rate observing may be a strategy of checking the rate and cadence of the fetal pulse. The normal fetal heart rate is between 120 and 160 beats per diminutive. This rate may change as the embryo reacts to conditions within the uterus. The evaluation of fetal wellbeing has possessed our proficient consideration for numerous a long time. As the improvement of innovations for pre-birth symptomatic procedures has advanced, applications of such innovations have supported within the large appraisal of fetal well-being. Fetal heart-rate checking remains the most shape of fetal evaluation for high-risk pregnancies. The extra appraisals managed by the examination of ST and T-wave changes of the fetal electrocardiogram hold guarantee for moving forward the prescient esteem of fetal heart-rate evaluations. Ultrasound has been priceless for appraisal of fetal life systems, and the utilization of Doppler ultrasound has given knowledge into fetal cardiovascular reactions to such conditions as intrauterine development confinement and fetal frailty caused by ruddy blood cell immunization.

Literature Review

This paper employments non-parametric Bayesian strategies to classify the fetal heart rate. They have utilized non-parametric Bayesian Strategy and SVM-based strategy to classify the FHR and result are compared to discover the execution of both strategies. Bayesian strategy have superior execution than SVM-based strategy [1]. This paper discover any intrinsic infections and cardiac peculiarities on fetal heart utilizing full convolution neural organize (FCN). They are utilizing FCN to distinguish the area of heart and examination the any irregularities utilizing FCN. They have concluded that FCN-16 demonstrate has lower mistake than other outlines while classifying [2]. This paper attempt to attain efficient fetal acidosis discovery to assist specialists amid conveyance. They are utilizing sparse-SVM to decide and classify highlights and calculate execution. They have concluded that classification done by programmed determination is distant superior than clinical hone [3]. The Classification and regression tree (CART) to identify high-risk amid pregnancy. They found that precision gotten utilizing entropy calculation and GINI list is 88.87% and 90.12% separately [4]. This paper classify the fetal heart rate utilizing convolution neural arrange (CNN) to create a neural arrange which can classify heart rate consequently. They found that precision is tall of CNN than of SVM and MLP. CNN is an productive for fetal heart checking framework [5]. They have utilized classification based on association (CBA) to classify fetal wellbeing status. They have found that the precision of show was expanded after utilizing highlight determination strategy. They found out that arbitrary woodland and XGBoost have great execution for classifying fetal wellbeing status [6].

In this paper, they are progressing to discover out the intermittent changes to identify sinusoidal heart rate. Irregular woodland is utilized for the information preprocessing and sinusoidal design is calculated utilizing fluffy theoretic strategy. The conclusion is that given strategy can be considered a gold standard for the intermittent alter discovery [7]. In this paper

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

they are reaching to discover in the event that there are any trouble happening in fetal and a predominant framework utilizing profound learning is outlined. Experimental show decay is utilized to break down any one-dimensional timing flag S(t) into Intrinsic Mode Function (IMF) with diverse frequencies. LSMT is utilized to classify the information. As a result, it is appeared that profound inclining has most exactness for recognizing fetal heart trouble [8].

In this paper, they are aiming to foresee fetal hazard utilizing machine learning for avoidance of any juvenile passing. At to begin with for dataset deduction minimum repetition greatest pertinence strategy is utilized. And after that the calculations like navie bayes, decision trees, random forest, support vector machines are connected for classification of heart rate[9]. In this paper detail heart rate is classified utilizing two strategies and the prevalent strategy is found out between them. To utilize crossover k-means, the include extraction is done utilizing k implies clustering. CTG dataset is recorded utilizing calculation and is compared to SVM. Hybrid K-means and Support Vector Machine is compared and precision of hybrid K-means was 90.64% whereas normal accuracy for SVM was 76.72 which appears Half breed K-means has superior accuracy [10]. In this paper, they are utilizing 2D ultrasound to decide and degree the hatchling heart rate. They prepare pictures to calculate FHR. Number of Pixel from the picture is handled and gives the data approximately fetal heart condition as each zone of the heart or portion of heart pictures are prepared utilizing 2D ultrasound. They found out, alter chamber zone is way better than the changes within the ventricle region with exactness more than 90% [11]. In this paper, they classify and compare CTG information framework utilizing administered SVM and choice tree to urge which have best execution and precision. The CTG dataset were prepared utilizing directed SVM and choice tree. They found out exactness for support vector machine was 97.93% and decision tree was 97.41%. So, they concluded t that the SVM classifier was able of recognizing Ordinary, Suspicious and Pathologic condition, from the nature of CTG information with exceptionally great exactness [12].

In this paper, they have utilize SVM as classifier to execute the f-score. They have utilized f-score strategy to sort the highlights. F-score strategy is utilized for include determination whereas SVM will classify the highlight. F-score have incredible exactness in foreseeing fetal status [13]. In this paper, they are centered on a developmental multi objective generic algorithm (MOGA) by utilize of which the critical components causing fetal passing is extricated with offer assistance of cardiotocographic examination.

The preferred choice is done utilizing Bland Calculation which is heuristic look calculation which incorporates the forms of encoding, Wellness work assessment, Determination by means of roulette wheel mechanism, Hybrid, Change, Multi-objective optimization, Store support. Subsequently it is found out that execution of any classifier is boosted on the off chance that legitimate include determination is done [14]. In this paper, they present the use of data-driven entropy profiling to identify fetal arrhythmia naturally. The fetal QRS extraction procedure is utilized to extricate fetal heart rate from the information set and after that entropy highlights are connected for profiling of the information set. The proposed strategy speaks to solid entropy appraise that give prevalent execution than existing strategy [15, 16, 17].

Proposed Work

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

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

(i) Firstly, the data set is preprocessed with Feature Scaling and lost values.
(ii) Secondly, exploratory data analysis is done and the distribution 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 oversampling methods like Random Oversampler, Synthetic Minority Oversampling Technique SMOTE, Borderline SMOTE, KMeansSMOTE, SVMSMOTE and ADASYN.
(v) Fifth, the oversampled 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.

The overall workflow is shown in Figure. 1.

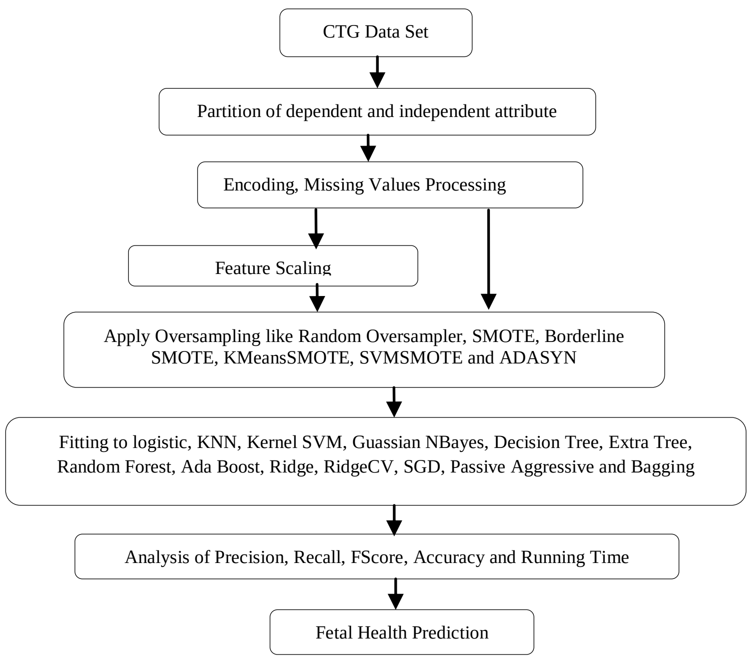

Figure 1. Overall Work Flow

Results and Discussion

The CTG dataset extracted from the UCI machine learning repository is used for implementation. The dataset consists of 2127 patients data with 21 independent features (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”. 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.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Target Feature Analysis

The target "Fealth health" is available as non-sampled as we can visualize in target feature distribution and is shown in Figure. 2.

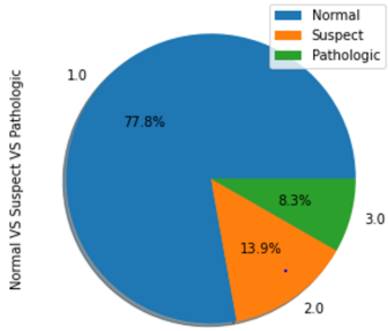

Figure 2. Target Feature Distribution

The relation of each feature is depicted in the correlation of the dataset and is shown in Figure. 3.

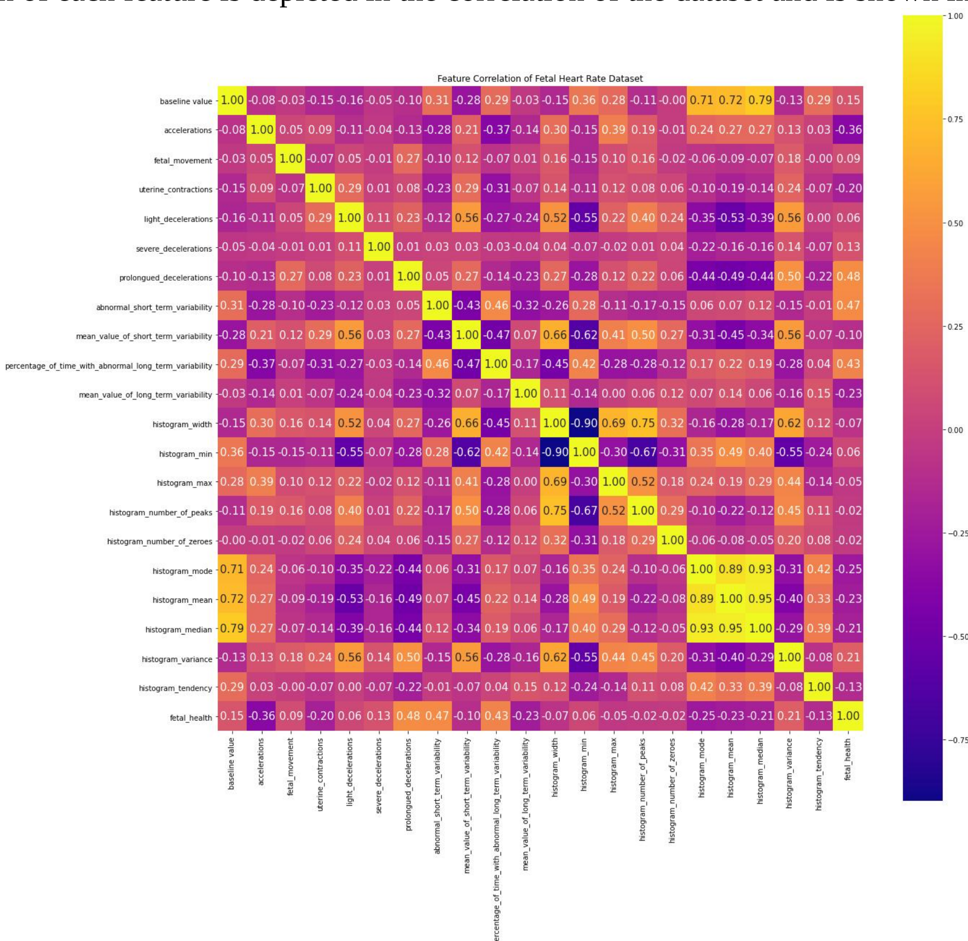

Figure 3. Target Feature Distribution

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Raw Dataset with Classifier Performance Analysis

The raw data set is fitted to all the classifier like logistic regression, KNN, Kernel SVM, Decision Tree, Random Forest, Ada Boost, Ridge, Ridge CV, 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. 4 - 5.

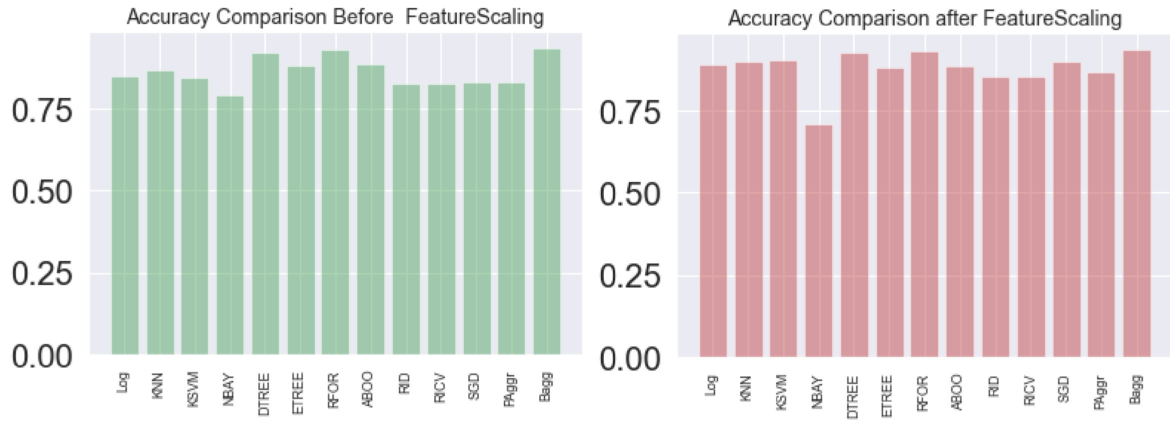

Figure 4. Raw Dataset Accuracy Comparison

Table 1. Classifier Performance before Feature 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 after Feature 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

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
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

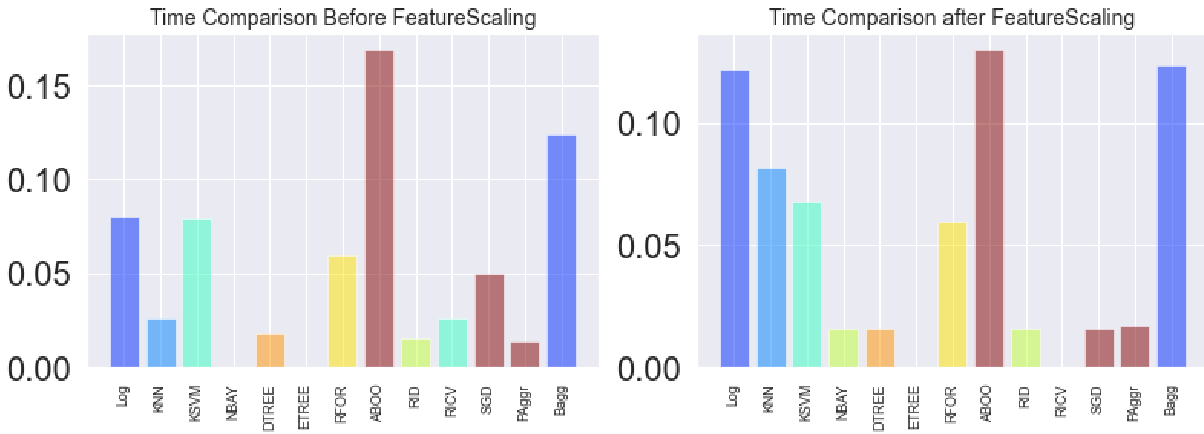

Figure 5. Raw Dataset Running Time Comparison

Oversampling with Dataset

The raw data set is subjected to oversampling methods like Random Oversampler, SMOTE, Borderline SMOTE, KMeansSMOTE, SVMSMOTE and ADASYN. The resampled dataset distribution after oversampling is shown in Figure. 6-7.

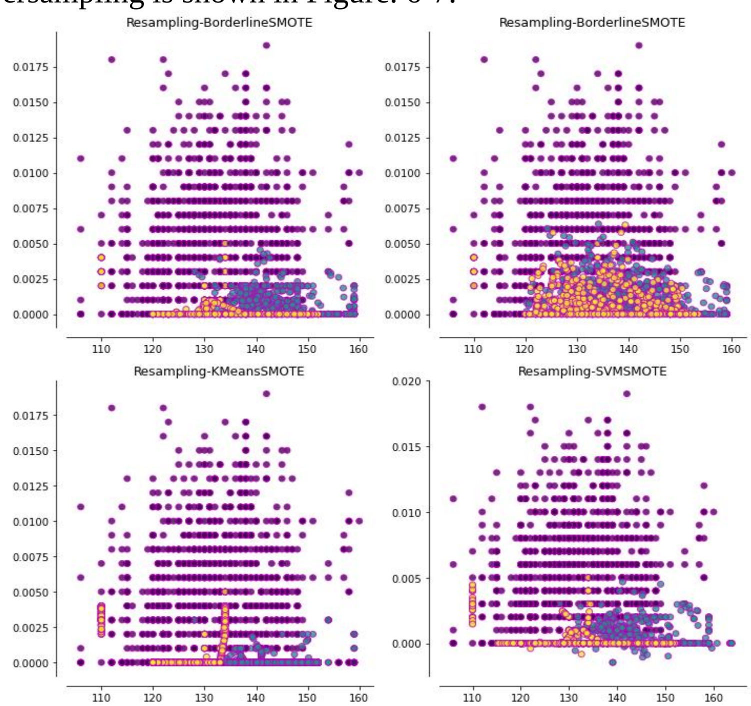

Figure 6. Dataset Oversampling with Random, SMOTE and ADASYN

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

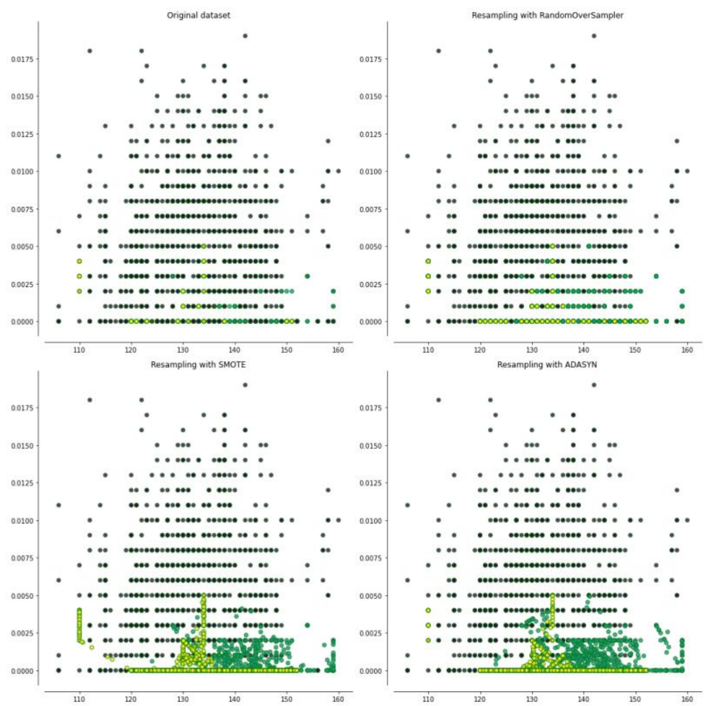

Figure 7. Dataset Oversampling with Borderline, KMeans and SVMSMOTE

Random Oversampler Classifier Performance Analysis

The raw data set is subjected to oversampling method namely Random Oversampler 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. 8 - 9.

Table 3. Random Oversampling Performance before Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.81	0.81	0.81	0.81	0.18
KNN	0.90	0.90	0.90	0.90	0.09
KSVM	0.84	0.84	0.84	0.84	0.46
GNBayes	0.79	0.77	0.77	0.77	0.00
DTree	0.99	0.99	0.99	0.99	0.03
ETree	0.97	0.97	0.97	0.97	0.00
RForest	0.98	0.98	0.98	0.98	0.08
AdaBoost	0.89	0.89	0.89	0.89	0.24
Ridge	0.84	0.83	0.84	0.83	0.00
RidgeCV	0.84	0.84	0.84	0.84	0.02
SGD	0.84	0.84	0.84	0.84	0.13
PAggress	0.76	0.77	0.76	0.77	0.02
Bagging	0.99	0.99	0.99	0.99	0.20

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

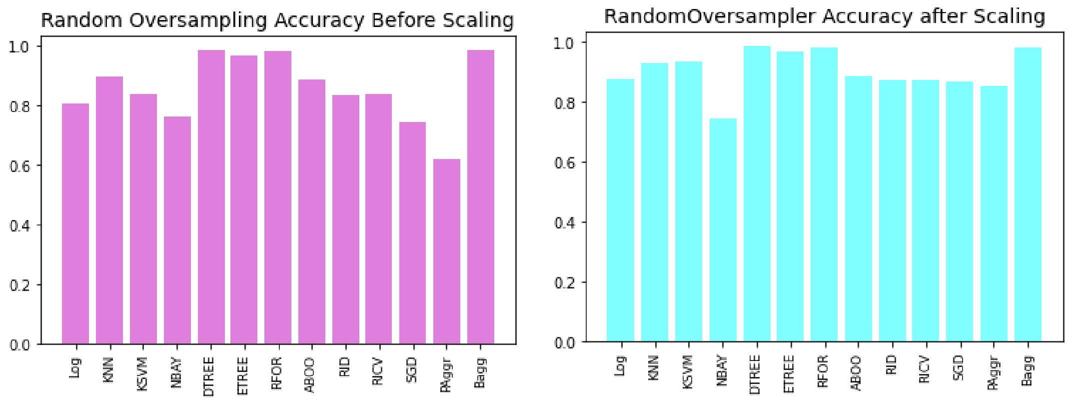

Figure 8. Random Oversampling Accuracy Comparison

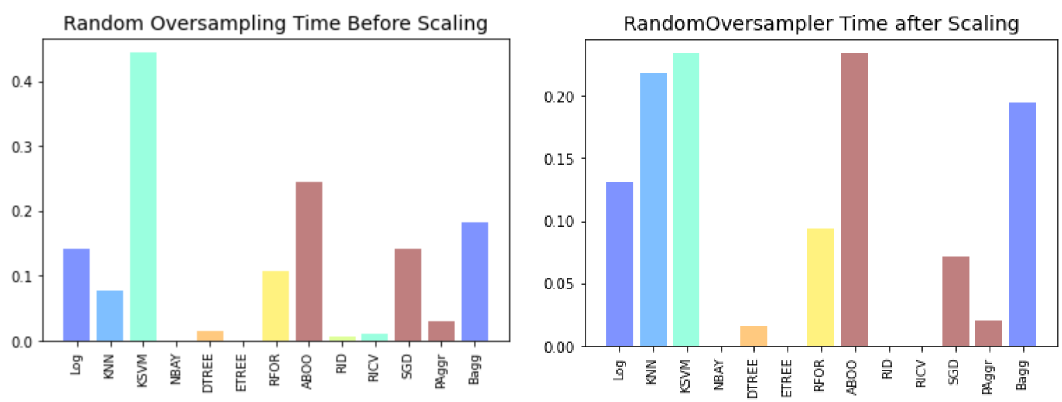

Figure 9. Random Oversampling Running Time Comparison

Table 4. Random Oversampling Classifier Performance after Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.88	0.88	0.88	0.88	0.19
KNN	0.94	0.93	0.93	0.93	0.23
KSVM	0.94	0.93	0.93	0.93	0.25
GNBayes	0.83	0.75	0.75	0.75	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	0.98	0.98	0.98	0.98	0.09
AdaBoost	0.89	0.89	0.89	0.89	0.24
Ridge	0.88	0.87	0.88	0.87	0.02
RidgeCV	0.88	0.87	0.88	0.87	0.01
SGD	0.86	0.86	0.86	0.86	0.07
PAggress	0.82	0.79	0.79	0.79	0.02
Bagging	0.98	0.98	0.98	0.98	0.19

SMOTE Oversampling Classifier Performance Analysis

The raw data set is subjected to oversampling method namely Synthetic Minority Oversampling Technique SMOTE and the resampled dataset is fitted to all the classifiers with and without the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

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. 10 - 11.

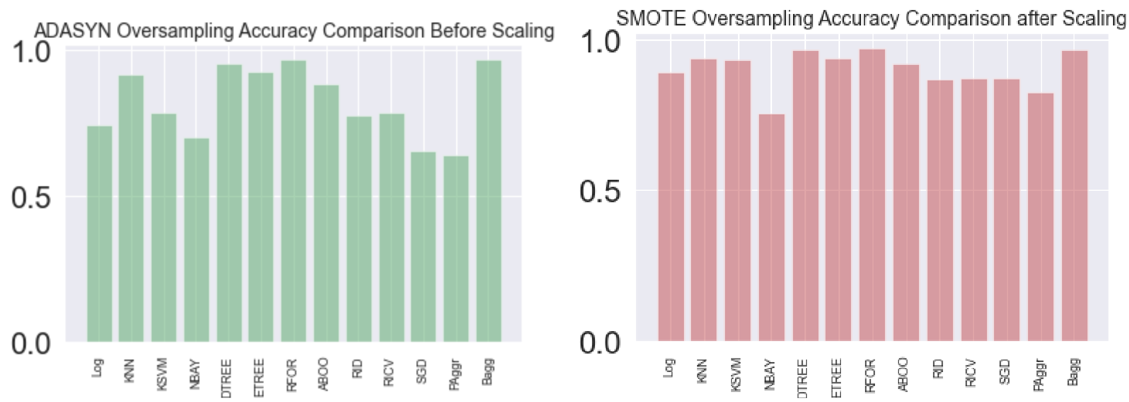

Figure 10. SMOTE Oversampling Accuracy Comparison

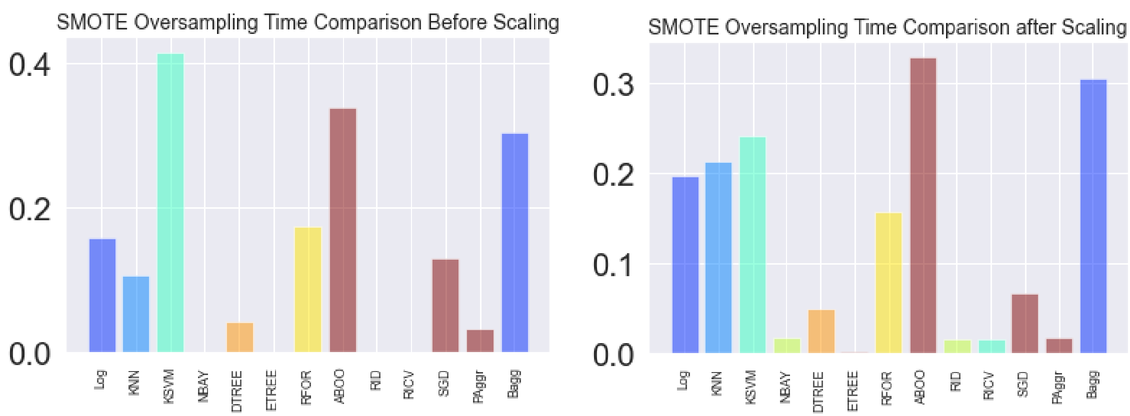

Figure 11. SMOTE Oversampling Running Time Comparison

Table 5. SMOTE Oversampling Performance before Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.80	0.81	0.81	0.81	0.16
KNN	0.92	0.92	0.92	0.92	0.11
KSVM	0.85	0.85	0.85	0.85	0.41
GNBayes	0.79	0.76	0.77	0.76	0.00
DTree	0.96	0.96	0.96	0.96	0.04
ETree	0.95	0.95	0.95	0.95	0.00
RForest	0.98	0.98	0.98	0.98	0.17
AdaBoost	0.92	0.92	0.92	0.92	0.34
Ridge	0.84	0.84	0.84	0.84	0.00
RidgeCV	0.85	0.85	0.85	0.85	0.00
SGD	0.79	0.77	0.77	0.77	0.13
PAggress	0.77	0.66	0.58	0.66	0.03
Bagging	0.97	0.97	0.97	0.97	0.30

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Table 6. SMOTE Oversampling Classifier Performance after Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.89	0.89	0.89	0.89	0.20
KNN	0.94	0.94	0.94	0.94	0.21
KSVM	0.94	0.93	0.93	0.93	0.24
GNBayes	0.83	0.75	0.76	0.75	0.02
DTree	0.97	0.97	0.97	0.97	0.05
ETree	0.94	0.94	0.94	0.94	0.00
RForest	0.97	0.97	0.97	0.97	0.16
AdaBoost	0.92	0.92	0.92	0.92	0.33
Ridge	0.88	0.87	0.87	0.87	0.02
RidgeCV	0.88	0.87	0.87	0.87	0.02
SGD	0.88	0.87	0.87	0.87	0.07
PAggress	0.82	0.83	0.82	0.83	0.02
Bagging	0.96	0.96	0.96	0.96	0.31

Borderline SMOTE Oversampling Classifier Performance Analysis

The raw data set is subjected to oversampling method namely Borderline SMOTE 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. 12 - 13.

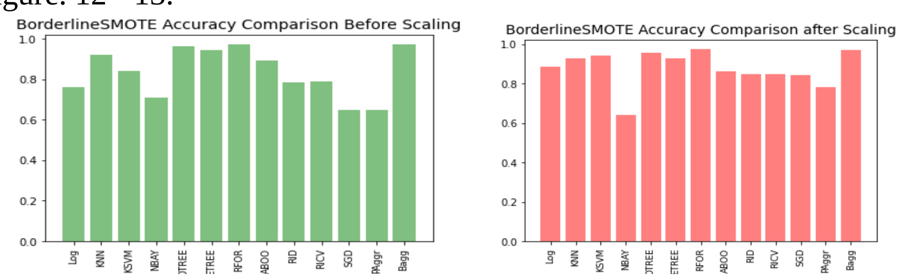

Figure 12. Borderline SMOTE Oversampling Accuracy Comparison

Table 7. Borderline SMOTE Oversampling Performance before Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.76	0.76	0.76	0.76	0.17
KNN	0.93	0.92	0.92	0.92	0.09
KSVM	0.85	0.84	0.84	0.84	0.53
GNBayes	0.77	0.71	0.71	0.71	0.01
DTree	0.97	0.96	0.96	0.96	0.06
ETree	0.95	0.95	0.95	0.95	0.00
RForest	0.97	0.97	0.97	0.97	0.16
AdaBoost	0.89	0.89	0.89	0.89	0.33
Ridge	0.78	0.78	0.78	0.78	0.02
RidgeCV	0.79	0.79	0.79	0.79	0.02
SGD	0.69	0.65	0.59	0.65	0.18
PAggress	0.44	0.65	0.53	0.65	0.02

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Bagging	0.97	0.97	0.97	0.97	0.31

Table 8. Borderline SMOTE Classifier Performance after Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.89	0.89	0.89	0.89	0.16
KNN	0.93	0.93	0.93	0.93	0.22
KSVM	0.94	0.94	0.94	0.94	0.24
GNBayes	0.78	0.64	0.64	0.64	0.00
DTree	0.96	0.96	0.96	0.96	0.05
ETree	0.93	0.93	0.93	0.93	0.00
RForest	0.98	0.97	0.97	0.97	0.16
AdaBoost	0.87	0.86	0.86	0.86	0.33
Ridge	0.86	0.85	0.85	0.85	0.00
RidgeCV	0.86	0.85	0.85	0.85	0.00
SGD	0.86	0.84	0.85	0.84	0.06
PAggress	0.78	0.78	0.77	0.78	0.02
Bagging	0.97	0.97	0.97	0.97	0.33

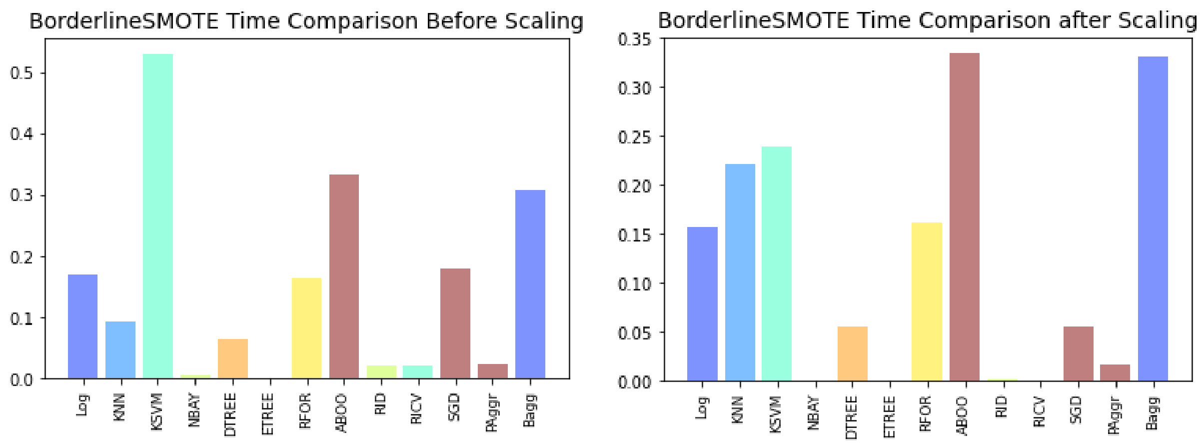

Figure 13. Borderline SMOTE Oversampling Running Time Comparison

KMeans SMOTE Oversampling Classifier Performance Analysis

The raw data set is subjected to oversampling method namely KMeans SMOTE and the resampled dataset is fitted to all classifiers with and without the presence of feature scaling and performance is shown in Table 9 and Table 10, the accuracy and time comparison is shown in Figure. 14 - 15.

Table 9. KMeans SMOTE Classifier Performance before Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.94	0.94	0.94	0.94	0.17
KNN	0.95	0.95	0.95	0.95	0.12
KSVM	0.94	0.94	0.94	0.94	0.20
GNBayes	0.91	0.91	0.91	0.91	0.00

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

DTree	0.96	0.96	0.96	0.96	0.03
ETree	0.96	0.96	0.96	0.96	0.00
RForest	0.97	0.97	0.97	0.97	0.14
AdaBoost	0.94	0.93	0.93	0.93	0.32
Ridge	0.93	0.93	0.93	0.93	0.02
RidgeCV	0.93	0.93	0.93	0.93	0.02
SGD	0.92	0.92	0.92	0.92	0.10
PAggress	0.93	0.93	0.93	0.93	0.05
Bagging	0.97	0.97	0.97	0.97	0.27

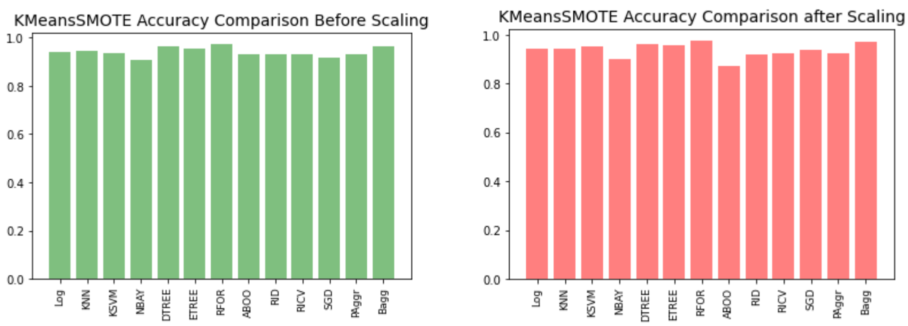

Figure 14. KMeans SMOTE Oversampling Accuracy Comparison

Table 10. KMeans SMOTE Classifier Performance after Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.95	0.95	0.95	0.95	0.15
KNN	0.95	0.95	0.95	0.95	0.20
KSVM	0.96	0.95	0.95	0.95	0.14
GNBayes	0.92	0.90	0.90	0.90	0.00
DTree	0.96	0.96	0.96	0.96	0.04
ETree	0.96	0.96	0.96	0.96	0.00
RForest	0.98	0.98	0.98	0.98	0.13
AdaBoost	0.89	0.88	0.88	0.88	0.30
Ridge	0.93	0.92	0.92	0.92	0.00
RidgeCV	0.93	0.92	0.92	0.92	0.00
SGD	0.94	0.94	0.94	0.94	0.08
PAggress	0.93	0.92	0.92	0.92	0.02
Bagging	0.97	0.97	0.97	0.97	0.28

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

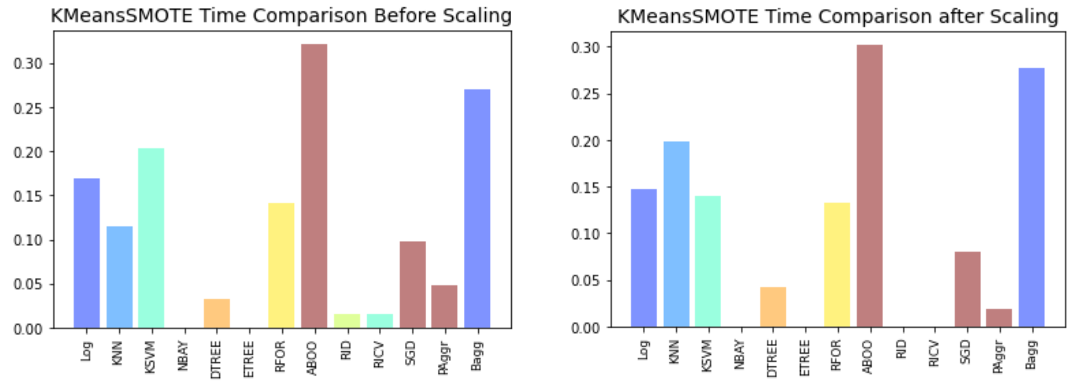

Figure 15. KMeans SMOTE Oversampling Running Time Comparison

SVM SMOTE Oversampling Classifier Performance Analysis

The raw data set is subjected to oversampling method namely SVMSMOTE 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. 12 - 13.

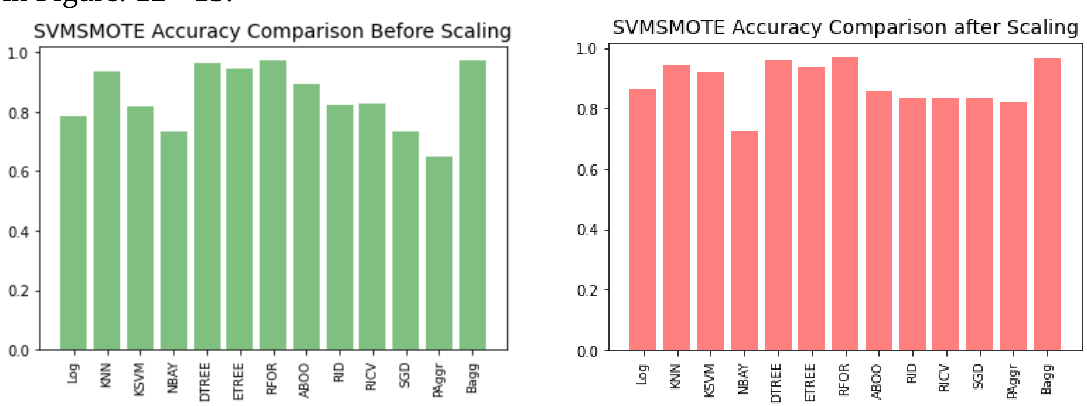

Figure 12. SVMSMOTE Oversampling Accuracy Comparison

Table 11. SVMSMOTE Classifier Performance before Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.78	0.78	0.78	0.78	0.20
KNN	0.94	0.93	0.93	0.93	0.12
KSVM	0.83	0.82	0.82	0.82	0.57
GNBayes	0.78	0.73	0.74	0.73	0.01
DTree	0.97	0.97	0.97	0.97	0.10
ETree	0.95	0.95	0.95	0.95	0.00
RForest	0.97	0.97	0.97	0.97	0.22
AdaBoost	0.90	0.90	0.90	0.90	0.41
Ridge	0.83	0.82	0.83	0.82	0.01
RidgeCV	0.83	0.83	0.83	0.83	0.01
SGD	0.81	0.74	0.74	0.74	0.14
PAggress	0.70	0.65	0.63	0.65	0.02

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 5, 2021, Pages. 1448 - 1464
Received 15 April 2021; Accepted 05 May 2021.

Bagging	0.97	0.97	0.97	0.97	0.30

Table 12. SVMSMOTE Classifier Performance after Feature Scaling

Classifier	Precision	Recall	FScore	Accuracy	Running Time (ms)
Logistic	0.87	0.86	0.86	0.86	0.16
KNN	0.94	0.94	0.94	0.94	0.24
KSVM	0.92	0.92	0.92	0.92	0.27
GNBayes	0.81	0.73	0.73	0.73	0.00
DTree	0.96	0.96	0.96	0.96	0.05
ETree	0.94	0.94	0.94	0.94	0.02
RForest	0.97	0.97	0.97	0.97	0.17
AdaBoost	0.86	0.86	0.86	0.86	0.32
Ridge	0.85	0.84	0.84	0.84	0.02
RidgeCV	0.85	0.84	0.84	0.84	0.02
SGD	0.84	0.83	0.84	0.83	0.10
PAggress	0.82	0.82	0.82	0.82	0.02
Bagging	0.97	0.96	0.96	0.96	0.32

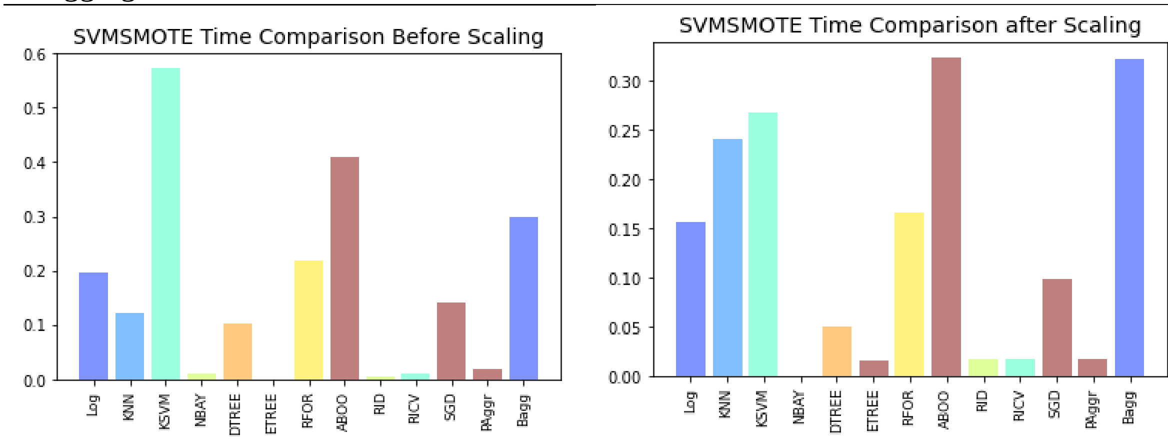

Figure 13. SVMSMOTE Oversampling Running Time Comparison

Conclusion

An attempt is done to analyze the performance of non sampled data with sampled data. The CTG dataset used in this paper found to have nonsampled data with Normal, Suspect and Pathologic. This paper attempt to perform oversampling with Random Oversampler, SMOTE, Borderline SMOTE, KMeansSMOTE, SVMSMOTE and ADASYN to resample the data. Experimental results shows that the Random Forest and Decision Tree classifier tends to retain 93% and 92% accuracy respectively before and after feature scaling. The Random Oversampled dataset shows that the Random Forest and Decision Tree classifier tends to retain 99% and 98% accuracy respectively before and after feature scaling. The SMOTE, Borderline SMOTE, KMeansSMOTE, SVMSMOTE and ADASYN resampled dataset shows that the Random Forest and Decision Tree classifier tends to retain 97% and 96% accuracy respectively before and after feature scaling.

References

[1] Jianqiang Li., Luxiang Huang., Zhixia Shen., Yifan Zhang., Min Fang, Bing Li., Xianghua Fu., Qingguo Zhao., & Huihui Wang. (2019, April). 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, March). 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, May). 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, June). 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, June). 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, March). 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, September). 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, May) 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. 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, February). 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, December). Determination of Fetus Heart Rate Based 2D Ultrasound. Proceedings of the International Conference on Information and Communication Technology Convergence. 5.

[12] Jagannathan, 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, August). 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, July). 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, K., Udhayakumar., Chandan Karmakar, K., Fiona Brownfoot, C & Marimuthu Palaniswami. (2020, August). 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] M Ramkumar, M Mathankumar, A Manjunathan, "Classification of Electrocardiogram Cardiac Arrhythmia Signals Using Genetic Algorithm - Support Vector Machines", Bioscience Biotechnology Research Communications (BBRC), Special Issue Vol.13 (11), pp.138-146, 2020.

[17] M Ramkumar, C Ganesh Babu, A Manjunathan, S Udhayanan, M Mathankumar, R Sarath Kumar, "A Graphical User Interface based Heart Rate Monitoring process and Detection of PQRST peaks from ECG Signal", International Conference on Inventive Computation and Information Technologies, Lecture Notes in Networks and Systems, pp. 445-461, 2020.