Jestr

Journal of Engineering Science and Technology Review 9 (5) (2016) 201- 207

Review Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Ensembled Rule Based Classification Algorithms for predicting Imbalanced Kidney Disease Data

Sai Prasad Potharaju* and M.Sreedevi

Dept of CSE , K L University, Guntur, Andhra Pradesh, India

Received 25 August 2016; Accepted 8 December 2016

Abstract

Imbalanced data is a type of data where there exists a difference in the ratio of classes. It occurs easily in real life of data analysis. In Data mining the functioning of learning algorithms caused by the imbalanced data. Most of the machine learning algorithms has a tendency to prejudice towards the class of majority in case of imbalanced data and hence those algorithms misjudge the minority class. Therefore, In this article we discuss a systematic way to address the imbalanced data classification problem by applying the rule based ensemble learning techniques like bagging, boosting, voting and stacking to build models, and then accelerates the performance of learning algorithms. In this research, we have preferred real data of chronic kidney disease which is collected from Appolo Hospitals, Tamil Nadu, India, to predict kidney disease of patients .The collected data is initially imbalanced. Firstly, the imbalanced data is balanced by applying SMOTE algorithm, which is an over sampling technique. Then applied various ensemble learning techniques to make better prediction. The incurred results showed that the model template chosen can minimize the problem of misclassification of imbalanced data efficaciously. But this model template cannot classify correctly when imbalanced rate of class increases i.e. in case of Big Data. For better result of imbalanced Big Data, new algorithmic plan of action has to be exploited which can be measured by using Hadoop framework and mapreduce programming model.

Keywords: Mining, Ensemble Learning, Health Informatics, Imbalanced Data sets, Rule Based Classification, SMOTE.

1. Introduction

Data mining [1] is a practical method, which has been put into service to retrieve the unfamiliar knowledge from a huge database. Depending on the intention, there are numerous categories of Data mining tasks. Classification is one of those categories .The main aim of classification is to gain knowledge of hidden patterns to make prediction about the class of some unknown data. Most of the standard machine learning algorithms for data classification can be applied very efficiently for classification precision if class labels are equally scattered. However, these standard algorithms show less or poor learning execution in case of classifying the imbalanced data that have variation in the class labels [2] .In order to get the accuracy of classification algorithms, one or more algorithms can be combined and can get the reasonable accuracy. The process of combining the multiple algorithms is called ensembling [3].The current research aims to predict the chances of getting kidney disease from given patient data set. The main drive of predictions in data mining of health care is to discover trends in patient data in order to make better their health [4].Chronic Kidney Disease (CKD) has become a superior cause of deaths recent days due to the alternation in regular life style of citizens.

Accelerative amount of obtainable information of patient health related details provide a goldmine that can be used to know the condition of various parts of body. The advantages of data mining have been creating a scope of research in health care informatics (HCI) [5]. Unveiled information can be useful to perceive how patient's body is responding with several medical test reports. HCI is an aggregation of computer and information science of health care .Latest studies in health care has aimed on predicting many diseases. Example of these includes prediction of liver cancer, prediction of heart disease [6], prediction of breast cancer [7], prediction of Dermatological disease, prediction of diabetes [8], prediction of hepatitis C virus (HCV). However, the scope of this study targets to examine the use of various ensemble learning methods with a collection of kidney disease data belongs to Tamil Nadu, Apollo Hospitals. Besides, the current study furnishes serious set of findings, from which more refined and precise classification models can be built. Anticipations from imbalanced data sets i.e. huge division in instances of various classes may not bring accurate results. To get better results, imbalanced (unequal) data set should be balanced (stabled) by applying different techniques. In this study over sampling algorithm i.e. SMOTE is employed on unequal data set to make it stable data set, then various ensemble learning techniques such as Boosting,Bagging,Voting and Stacking with Rule Based Algorithms Jrip,OneR,Ridor are used to produce the results .The remaining parts of this paper is formed as below. Section 2 presented related work which will set the path for the subsequent sections. Data mining stages and data collection and preparation to predict kidney disease is presented in section 3. Experimental setup and its corresponding results with various rule based induction with ensembling methods are presented in section 4. Section 5 contains results .At the end section 6 contains conclusion with future suggestions.

* E-mail address: psaiprasadcse@gmail.com

2. Related Work

In machine learning problems, in recent days it has become very popular to use multiple classifiers instead of one single classifier. As an advantage, model that can be induced will be more reliable and sophisticated to classify the instances if multiple classifiers are combined, instead of one classifier. Ensemble learning is a type of learning in which, multiple finite number of classifiers are get trained for the same classification task and thus it can get better accuracy.

There are various ensembling techniques, which includes Bagging, Boosting, Voting and Stacking. Bagging and Boosting are standalone classification algorithms, but they use multiple classification tasks into one. In Voting [9], different combinations of probability estimates for classification are available.

Stacking [10] is another procedure of combining several classifiers, that brings in the concept of a meta learner. The methodology of stacking is as follows:

1. Partition the training set into two separate sets.
2. Train multiple base learners on the initial (first) part.
3. Test the base learners on the second part.
4. Using the predictions from above step as the inputs, and the correct responses as the outputs, train a higher level learner.

In [11], a type of selection scheme based on accuracy and diversity is explored to achieve improved classification performance. In the article [12], The strength of the (LDB) Linear Discriminant Boosting algorithm is tested by churn prediction investigation on data set of real bank customer churn data set. The algorithm found to improve the accuracy, and output is tested with other algorithms, such as support vector machines (SVM), artificial neural networks (ANN), decision trees, and classical Adaboost algorithm. In research article [13], worked on prediction of kidney disease using different rule based and tree based algorithms using SMOTE .In [14] ,the authors demonstrated an improvement of the precision of classification algorithm findings. Two different strategies bagging and boosting are used to increase the precision in their article. The paper describes a set of experiments with bagging and boosting methods. The applications of those techniques targets at classification algorithms generating decision trees. In[15], The authors presented integrating base models via stacking of model ,majority voting sets with varying diversity, as well as a re sampling/boosting integration technique called RUSBoost to predicting disorders that occurs in blood transfusion.

In [16], Analysis of Various Data Mining algorithms to predict the Heart Disease were presented. In [17], the authors examined accuracy and efficiency using RBF-Radial Basis Function, BPA-Back Propagation Algorithm, SVM. The objective of authors study was to propose the tool for kidney stone detection, to minimize the diagnosis time and to increase the accuracy and efficiency.

In [18], the researchers introduced the framework to minimize the effort and cost of choosing patients for clinical examinations. Two mining techniques were used in order to get hidden knowledge in the way of decision rules. Patient can be chosen based on the prediction results and the most significant parameters discovered in this paper. In [19], it was discussed about various clustering methods that can be applied to Big Data Mining, which is current research area. In the research article [20], authors introduced an ANFIS-Adaptive Neuro Fuzzy Inference System for to detect the CKD-Chronic Kinney Disease based on real medical data. In [21], CKD prediction is presented by the authors using Naive Bayes (NB) classifier and SVM. From the findings of SVM and NB, it is concluded that the accuracy of the SVM is accelerated than the NB classifier algorithm. But, comparison of SVM and NB algorithm is done on imbalanced data set. In [22], researchers applied ID3 technique to predict the Information Technology performance students. The investigation suggests that, Computer Science student's performance is better than students of in different branches. In article [23], authors compared different classification strategies to predict the dropout of course registration. The popular WEKA tool is used to compare those techniques and it is found that NBTree generated the significant level of accuracy. In paper [24], the authors investigated the behaviour of C4.5, CART and ID3 tree methods to predict the first year student's performance. In this article, among the various models ID3 is identified as most strong model for their purpose. In [22, 23, 24] various mining techniques are applied for different tasks. All those techniques are applied in this study to predict the chronic kidney disease. The research article [2], it was introduced the over sampling technique called SMOTE, by which difficulties of imbalanced data can be solved.

3. Methodology

In this part of the article, the framework designed in the current study is presented. Fig.1 shows processing steps of data mining, which includes data collection or gathering, preparation of data (data cleaning), sdata analysis and generation of model. There are multiple factors which lead to kidney disease, which are listed in Table 1.

3.1 Data Collection and Preparation

To evaluate the proposed approach of this study, the data is collected from Apollo Hospitals, Tamil Nadu, India . Table 1. contains 24 (Twenty four) characteristics used in this study.

Table 1. Description of Investigated Attributes

Attribute	Type of Attribute
Age	numerical
Albumin	nominal
Blood Pressure	numerical
Sugar	nominal
Red Blood Cells	nominal
Pus Cell	nominal
Pus Cell clumps	nominal
Bacteria	nominal
Blood Glucose	numerical
Blood Ure	numerical
Serum	numerical
Sodium	numerical
Potassium	numerical
Hemoglobin	numerical
Packed Cell Volume	numerical
White Blood Cell Count	numerical
Red Blood Cell Count	numerical
Hypertension	nominal
Diabetes Mellitus	nominal
Coronary Artery Disease	nominal
Appetite	nominal
Pedal Edema	nominal
Anemia	nominal
Class (cdk or not cdk)	nominal

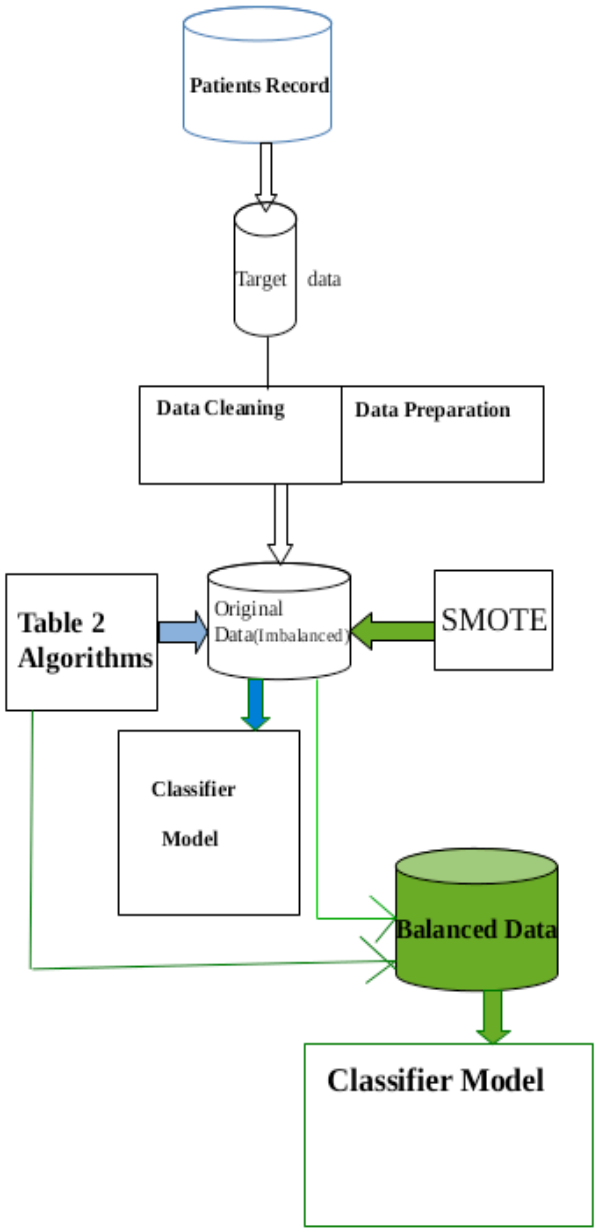

Fig.1. Processing Stages

At first, collective data may come across loss of value problem or missing value problem .There may be chance of having unknown or incorrect attribute values in the process of data preparation stage. In this research, any record with such faulty instances have been removed from the original data set, after this stage, complete data set is having 400 records. For numerical features, mean of the feature value is considered to head off bias in the learning process.

4. Evaluation

Several type of rule based induction techniques like Jrip, OneR, Ridor are applied to construct a classifier model. These models are analyzed with ensemble learning techniques and that can be interpretable. The algorithms used in this study are listed in Table 2.

Table 2. Algorithms Used in this Study

Category	Algorithms
Rule Based	Jrip, OneR, Ridor
Ensemble (AdaBoost)	AdaBoost+Jrip, AdaBoost+OneR, AdaBoost+Ridor
Ensemble (Bagging)	Bagging+Jrip, Bagging+OneR, Bagging+Ridor
Ensemble (Voting)	Jrip+OneR, OneR+Ridor, Ridor+Jrip, Jrip+OneR+Ridor
Ensemble (Stacking)	Jrip+OneR, OneR+Ridor, Ridor+Jrip, Jrip+OneR+Ridor

For evaluation of various methods performance, k-fold cross validation is considered. Data set of 400 records are divided into training and test sets, in the ratio of 66% to 34% .This process is continued for 10 times.(k =10). For every iteration of these ten settings, the training data set is applied to the learning technique with a listed classification method. After that, the respective test data set is employed to measure the accuracy of listed classifier model. Average accuracy of those ten different runs considered as classification performance.

By contrast, the trouble of imbalanced data that visible in the considered data set has not been taken into consideration. Specifically, the amount of cdk class records (record leading to Improper Functioning of Kidney) and not cdk class (record with Proper Functioning of Kidney) are presented equally. After the pre processing of data set, total cdk class instances present in the data set are 150, and not cdk are 250. Because of this variation in data set, the algorithm tends to take a side of majority class category So, the result obtained may not be accurate.

This type of problems in data set are minimized by employing a SMOTE [2] method, which is an over sampling technique. This technique is considered in order to balance the imbalanced data set. With the outcome of this, it increments the minority class instances, with synthetic instances using the concept of k-nearest neighbours(KNN). As for the observations shown in the next section, the value of k is set to 5. After applying the imbalanced data set total 1170 record were created, out of those, 600 records are belongs to not cdk class and 570 records belong to cdk class after 3 samplings.

5. Results

After completion of analysis by different algorithms using ensembling techniques, different computing factors like Precision,Recall,F-Measure,ROC Area are examined .All listed classification techniques in Table 2 are investigated in the first experiment to obtain classifier models from the original training sets(imbalanced data set) . The findings of those classifier models are gain below**.** Note In the below tables(From Table 3 to Table 12) Column 1 is Algorithm, Column 2 is TP Rate, Column 3 is, FP Rate ,Column 4 is Precision , Column 5 is Recall, Column 6 is F-Measure, Column 7 is ROC Area ,Column 8 is Class, Column 9 is Accuracy.

Table 3. Classification Result of Imbalanced Data set Without Ensembling

1	**2**	3	4	5	6	7	8	9
JRIP	0.953 0.972	0.047 0.047	0.953 0.972	0.953 0.972	0.953 0.972	0.961 0.961	Not ckd ckd	96.5
OneR	0.927 0.961	0.084 0.073	0.869 0.954	0.927 0.916	0.897 0.935	0.921 0.921	Notckd ckd	92
RidoR	0.98 0.964	0.036 0.02	0.942 0.988	0.98 0.964	0.961 0.976	0.972 0.972	No tckd ckd	97

Table 4. Classification Result of Balanced Data set Without Ensembling

1	2	3	4	5	6	7	8	9
JRIP	1 0.993	0.007 0	0.993 1	1 0.993	0.997 0.996	0.961 0.961	Not ckd ckd	99.65
OneR	0.973 0.918	0.082 0.082	0.926 0.926	0.973 0.973	0.949 0.949	0.921 0.921	Notckd ckd	94.6
RidoR	0.988 0.997	0.023 0.002	0.979 0.998	0.998 0.977	0.988 0.988	0.972 0.972	No tckd ckd	98.8

From the above tables it is cleared that ensembling technique on balanced data set performing better.

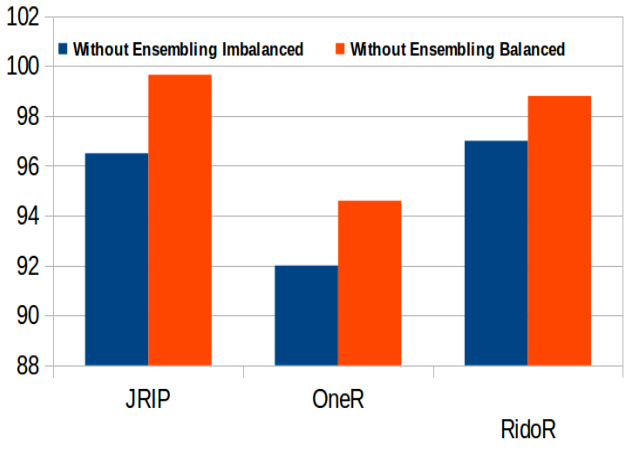

Fig. 2. Comparision of Table 3. And Table 4

The result obtained by Adaboost on Imbalanced and Balanced data set is shown in Table 5. And Table 6. These results indicating ,Adaboost on balanced data set accelerating classification performance.

Table 5. Classification Result of Ensemble Adaboost for Imbalanced Data set

1	**2**	3	4	5	6	7	8	9
JRIP	1 0.984	0.016 0	0.974 1	1 0.98	0.987 0.992	1 1	Not ckd ckd	99
OneR	0.973 0.976	0.024 0.027	0.961 0.984	0.97 0.97	0.967 0.98	0.99 0.99	Notckd ckd	97.5
RidoR	1 0.992	0.008 0	0.987 1	1 0.99	0.993 0.996	1 1	Not ckd ckd	99.5

Table 6 . Classification Result of Ensemble Adaboost for Balanced Data set

1	**2**	3	4	5	6	7	8	9
JRIP	1 0.993	0.007 0	0.993 1	1 0.993	0.997 0.996	0.997 0.996	Not ckd ckd	99.65
OneR	0.997 0.979	0.021 0.003	0.98 0.996	0.997 0.979	0.988 0.988	0.999 0.999	Not ckd ckd	98.80
RidoR	1 0.996	0.004 0	0.997 1	1 0.996	0.998 0.998	1 1	Not ckd ckd	99.82

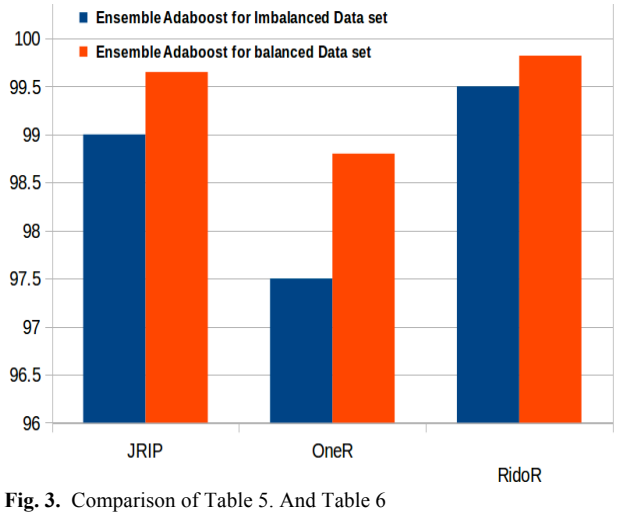

Fig. 3. Comparison of Table 5. And Table 6

The result obtained by Bagging on Imbalanced and Balanced data set is shown in Table 7. And Table 8. These results indicating ,Bagging technique on balanced data set accelerating classification performance.

Table 7. Classification Result of Ensemble Bagging for Imbalanced Data set

1	2	3	4	5	6	7	8	9
JRIP	0.96	0.012	0.98	0.96	0.97	0.998	Not ckd	97.75
	0.988	0.04	0.976	0.988	0.982	0.998	ckd	
OneR	0.933	0.084	0.87	0.933	0.9	0.948	Not ckd	92.25
	0.916	0.067	0.958	0.916	0.937	0.948	ckd	
RidoR	1	0.016	0.974	1	0.987	1	Not ckd	99
	0.984	0	1	0.984	0.992	1	ckd	

Table 8. Classification Result of Ensemble Bagging for Balanced Data set

1	2	3	4	5	6	7	8	9
JRIP	1	0.005	0.995	1	0.998	1	Not ckd	99.74
	0.995	0	1	0.995	0.997	1	ckd	
OneR	0.975	0.082	0.926	0.975	0.95	0.978	Not ckd	94.70
	0.918	0.025	0.972	0.918	0.944	0.978	ckd	
RidoR	1	0.012	0.988	1	0.994	0.997	Not ckd	99.40
	0.988	0	1	0.988	0.994	0.997	ckd	

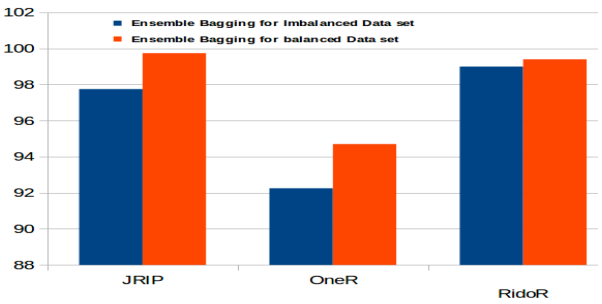

Fig.4. Comparision of Table 7. And Table 8

The result obtained by Voting on Imbalanced and Balanced data set is shown in Table 9. And Table 10. These results indicating, Voting technique on balanced data set accelerating classification performance. RIP and OneR , Ridor and OneR ,JRIP and RidoR,

JRIP ,OneR and RidoR algorithms are combined in this evaluation.

Each combination of algorithms are performing better than
its individuals.

Table 9. Classification Result of Ensemble Voting for Imbalanced Data set

1	2	3	4	5	6	7	8	9
JRIP + OneR	0.97	0.11	0.912	0.97	0.94	0.996	Not ckd	93.27
	0.888	0.03	0.961	0.888	0.923	0.996	ckd	
Ridor+OneR	0.99	0.14	0.892	0.99	0.938	0.991	Not ckd	92.90
	0.856	0.01	0.986	0.856	0.916	0.991	ckd	
JRIP+RidoR	0.987	0.04	0.964	0.987	0.975	0.99	Not ckd	97.27
	0.956	0.01	0.984	0.956	0.97	0.99	ckd	
JRIP+ OneR + RidoR	0.967	0.02	0.96	0.967	0.963	0.996	Not ckd	97.25
	0.976	0.03	0.98	0.976	0.978	0.996	ckd	

Table 10. Classification Result of Ensemble Voting for Balanced Data set

1	2	3	4	5	6	7	8	9
JRIP+ OneR	0.985	0.086	0.923	0.985	0.953	0.997	Not ckd	95.04
	0.914	0.015	0.983	0.914	0.947	0.997	ckd	
Ridor+ OneR	1	0.1	0.913	1	0.955	0.996	Not ckd	95.12
	0.9	0	1	0.9	0.947	0.996	ckd	
JRIP+ RidoR	1	0.023	0.979	1	0.989	0.996	Not ckd	98.88
	0.977	0	1	0.977	0.988	0.996	ckd	
JRIP+ OneR+ RidoR	1	0.011	0.99	1	0.995	0.999	Not ckd	99.48
	0.989	0	1	0.989	0.995	0.999	ckd	

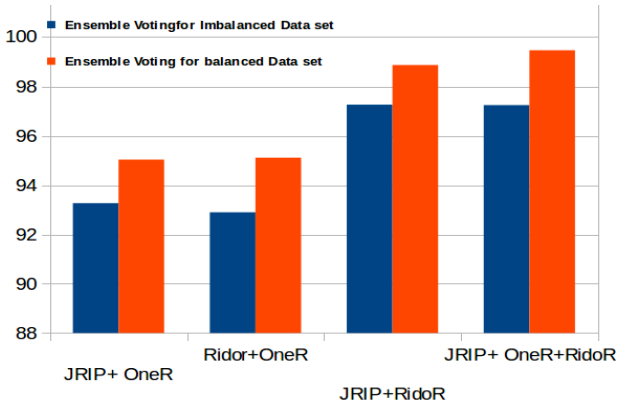

Fig. 5 . Comparison of Table 9. And Table 10.

The result obtained by Stacking on Imbalanced and Balanced data set is shown in Table 11. And Table 12. These results indicating ,Voting technique on balanced data set accelerating classification performance. JRIP and OneR ,Ridor and OneR ,JRIP and RidoR,

JRIP, OneR and RidoR algorithms are combined in this evaluation.

Each combination of algorithms are performing better than its individuals.

Table 11. Classification Result of Ensemble Stacking for Imbalanced Data set

1	2	3	4	5	6	7	8	9
JRIP + OneR	0.927 0.972	0.028 0.073	0.952 0.957	0.927 0.972	0.939 0.964	0.955 0.955	Not ckd ckd	95.5
Ridor + OneR	0.973 0.984	0.016 0.027	0.973 0.984	0.973 0.984	0.973 0.984	0.991 0.991	Not ckd ckd	98
JRIP + RidoR	0.98 0.988	0.012 0.02	0.98 0.988	0.98 0.988	0.98 0.988	0.993 0.993	No tckd ckd	98.5
JRIP + OneR + RidoR	0.987 0.988	0.012 0.013	0.98 0.992	0.987 0.988	0.983 0.99	0.992 0.992	Not ckd ckd	98.75

Table 12. Classification Result of Ensemble Stacking for Balanced Data set

1	**2**	3	4	5	6	7	8	9
JRIP + OneR	1 0.989	0.011 0	0.99 1	1 0.989	0.995 0.995	0.996 0.996	Not ckd ckd	99.48
Ridor + OneR	1 0.981	0.019 0	0.982 1	1 0.981	0.991 0.99	0.993 0.993	Not ckd ckd	99.05
JRIP + RidoR	1 0.989	0.011 0	0.99 1	1 0.989	0.995 0.995	0.996 0.996	Not ckd ckd	99.48
JRIP + OneR + RidoR	1 0.989	0.011 0	0.99 1	1 0.989	0.995 0.995	0.996 0.996	Not ckd ckd	99.48

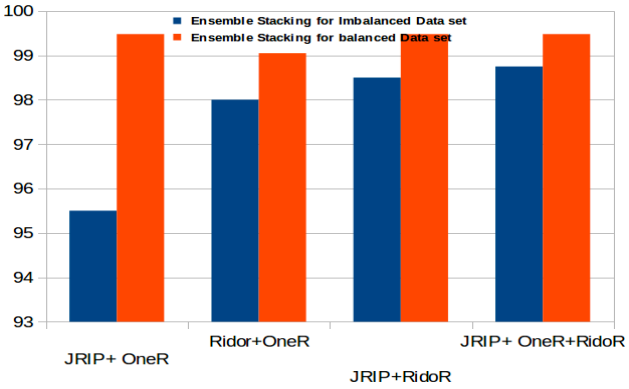

Fig.6. Comparision of Table 11. And Table 12.

It may be improvised with the help of SVM (Support Vector Machine) which is an advanced classification strategy. But, it might not be easy to understandable by a normal user.

6. Conclusion

After analysis of algorithms in Table 2, several parameters like Precision,Recall,F-Measure,ROC Area are compared .This study has produced a research work on the development of analytical framework for kidney functioning prediction. The proposed method abides the basics of data mining procedure collection of data, data cleaning, create the required classifier model, and analyze the results. Firstly, data has been chosen and aggregated to bring the target data set. For the stage of data cleaning, traditional techniques are applied; normalization technique is applied to diminish bias and errors identified in the data. Finally, various rule induction models with ensembling techniques like Boosting, Bagging, Stacking and Voting are explored for the classification purpose. The accuracy obtained by Ensembling models are performed better than rule based models without ensembled. The accuracy obtained is better with the use of SMOTE in case of imbalanced data. It has been observed that OneR algorithm performance is increased after ensembling with Jrip and RidoR in the case of imbalanced and balanced data. Further, SMOTE and ensembling techniques can be applied for Big Data analysis using Hadoop framework with the help of mapreduce programming model with new algorithmic approach, which is our future work.

References

1 J. Han and M. Kamber, Data mining: concepts and techniques, 3rd ed. Burlington, MA: Elsevier, 2011.

2. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.

3. M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp. 463–484, Jul. 2012.

4. A. K. Sen, S. B. Patel, and D. D. Shukla, "A data mining technique for prediction of coronary heart disease using neuro-fuzzy integrated approach two level," International Journal Of Engineering And Computer Science ISSN, pp. 2319–7242, 2013.

5. M. Herland, T. M. Khoshgoftaar, and R. Wald, "A review of data mining using big data in health informatics," Journal of Big Data, vol. 1, no. 1, p. 1, 2014.

6. Hlaudi Daniel Masethe, and Mosima Anna Masethe,"Prediction of Heart Disease usin Classification Algorithms,"Proceedings of the World Congress on Engineering and Computer Science 2014, Vol II , 22-24 October, 2014.

7. AA.Safavi,N.M.Parandesh, and M.Salehi,"Predicting breast cancer survivability using data mining techniques, School of Electrical and Computer Engineering," Doi:10.1109/ICSTE.2010.5608818,2010.

8. X.-H. Meng, Y.-X. Huang, D.-P. Rao, Q. Zhang, and Q. Liu, "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors," The Kaohsiung Journal of Medical Sciences, vol. 29, no. 2, pp. 93–99, Feb. 2013.

9. Ludmila I. Kuncheva," Combining Pattern Classifiers: Methods and Algorithms," John Wiley and Sons, Inc,2004.
10. A. Ulaş, M. Semerci, O. T. Yıldız, and E. Alpaydın, "Incremental construction of classifier and discriminant ensembles," Information Sciences, vol. 179, no. 9, pp. 1298–1318, Apr. 2009.
11. L. Yang, "Classifiers selection for ensemble learning based on accuracy and diversity," Procedia Engineering, vol. 15, pp. 4266–4270, 2011.
12. Y. Xie and X. Li, "Churn prediction with linear discriminant boosting algorithm," in 2008 International Conference on Machine Learning and Cybernetics, 2008, vol. 1, pp. 228–233.
13. Sai Prasad Potharaju,M. Sreedevi,"An Improved prediction of Kidney disease using SMOTE,"IJST, Vol 9(31), DOI: 10.17485/ijst/2016/v9i31/95634,August,2016.
14. Kristina Machova, Miroslav Puszta , Frantisek Barcak , and Peter Bednar, "A Comparison of the Bagging and the Boosting Methods Using the Decision Trees Classifiers ",ComSIS, Vol. 3, No. 2, December 2006.
15. D. Murphree et al., "Ensemble learning approaches to predicting complications of blood transfusion," in 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 7222–7225.
16. Kodali Lohita, Adusumilli Amitha Sree, Doreti Poojitha, T. Renuga Devi, and A. Umamakeswari,"Performance Analysis of Various Data Mining Techniques in the Prediction of Heart Disease ",Indian Journal of Science and Technology,Vol 8,Issue 35, doi:10.17485/ijst/2015/v8i35/87458,2015.
17. Abhishek, Gour Sundar Mitra Thakur, and Dolly Gupta," Proposing Efficient Neural Network Training Model for Kidney Stone Diagnosis," International Journal of Computer Science and Information Technologies, Vol. 3 (3) , pp no 3900-3904,2013.
18. A. Kusiak, B. Dixon, and S. Shah, "Predicting survival time for kidney dialysis patients: a data mining approach," Computers in Biology and Medicine, vol. 35, no. 4, pp. 311–327, May 2005.
19. T. Sajana, C.M. Sheela Rani, and K. V. Narayana.,"A Survey on Clustering Techniques for Big Data Mining ,"IJST,Vol 9(3),Doi:10.17485/ijst/2016/v9i3/75971,2015.
20. J. Norouzi, A. Yadollahpour, S. A. Mirbagheri, M. M. Mazdeh, and S. A. Hosseini, "Predicting Renal Failure Progression in Chronic Kidney Disease Using Integrated Intelligent Fuzzy Expert System," Computational and Mathematical Methods in Medicine, vol. 2016, pp. 1–9, 2016.
21. V. S and D. S, "Data Mining Classification Algorithms for Kidney Disease Prediction," International Journal on Cybernetics & Informatics, vol. 4, no. 4, pp. 13–25, Aug. 2015.
22. Kumar .V, and Singh.S ,"Classification of students data using data mining techniques for training & placement department in technical education, "Proceedings of International Conference on Computer Science and Networks, pp. 121–126,2013.
23. P. Pumpuang, A. Srivihok, and P. Praneetpolgrang, "Comparisons of classifier algorithms: bayesian network, C4. 5, decision forest and NBTree for course registration planning model of undergraduate students," in Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on, 2008, pp. 3647–3651.
24. Bunkar K, Singh U.K, Pandy B,and Bunkar.R,"Data mining: prediction for performance improvement of graduate students using classification," Int Proceedings of Ninth International Conference on Wireless and Optical Communications Networks,pp. 1–5,2012.