A NOVEL APPROACH OF ENSEMBLE MODELS BY USING EDM

G.Ayyappan[1], K.SivaKumar[2]

Research Scholar, Department of Computer Science, Bharath University[1]
SIPS Technologies, Chennai, Tamil Nadu, India[2]
ayyappangmca@gmail.com[1], kadirsiva@gmail.com[2]

Abstract: Study of research progress in the academic domain is challenging for student's academic performance in Tamil Nadu higher education sectors. The data collected from various engineering colleges augment this issue for supporting the results in this direction. Here in this paper we address this issue positively with the help of the educational data mining. The Classification method as one of the major data mining methodologies can be applied effectively for this purpose. The main focus of this paper is to check the learning algorithms for classification such examples based on selected dataset for student's academic performance. The main intention in this context is to deal with available data set for high accuracy. For this purpose, **AdaboostM1, Bagging and Dagging** models are built using an open source mining Weka under supervised learning algorithms. It is necessary to reduce the error before constructing the final models and thus the varying the parameters and number of iterations for training is carried out.

KEYWORDS: Educational Data Mining, Ensemble classifiers, AdaboostM1, Bagging, Dagging.

1. INTRODUCTION

Data Mining can be defined as the process involved in extracting interesting, interpretable, useful and novel information from the data. In this paper addresses the problem of student academic data in research progress prediction based on engineering students. In our present study a novel meta-feature generation method in the context of meta-learning, this is based on rules that compare the performance of individual base learners in a one-against-one manner. Experimental results are based on a large collection of datasets and show that the proposed new techniques can improve the overall performance of meta-learning for algorithm ranking significantly [1].

Nikita Bhatt et al [12] discussed the different approaches of Meta learning based on dataset characteristics provides a system that automatically provides ranking of the classifiers by considering different characteristics of datasets and different characteristics of classifiers after the generation of the Meta Knowledge Base, Ranking is provided based on Adjusted Ration of Ratio (ARR) or accuracy or time that helps non-experts in algorithm selection task.

Artur Ferreira et al [5] presented an overview of boosting algorithms to build ensembles of classifiers. The basic boosting technique and its variants are addressed and compared for supervised learning. The extension of these techniques for semi-supervised learning is also addressed. For face detection, boosting algorithms have been the most effective of all those developed so far, achieving the best results.

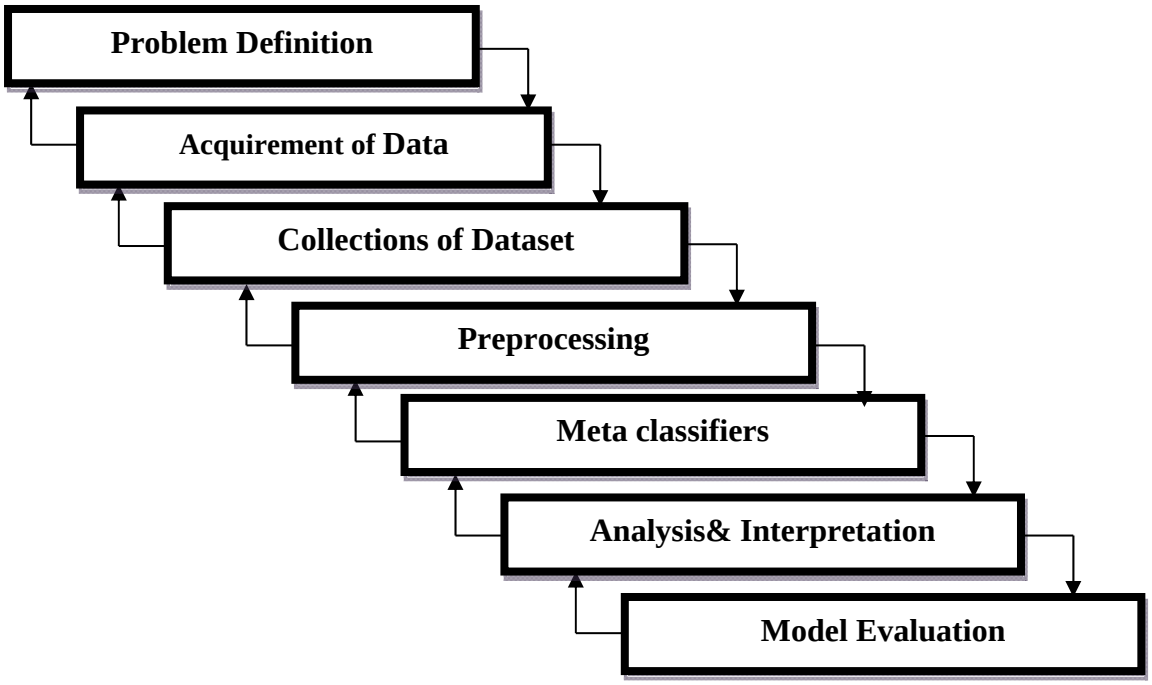

Figure 1. Architecture of Proposed System

2. RELATED WORKS

Data mining technology can help bridging knowledge gaps in higher educational system. Educational Mining is the application of Data mining techniques to educational data. M. Vranic et al have explained how data raining algorithms and techniques can be used by the academic community to potentially improve some aspects of the quality of education. One of the main concerns of any higher educational system is evaluating and enhancing the educational organization so as to improve the quality of their services and satisfy their customer's needs.

Hoda Waguih et al have investigated and justified the importance of data mining in evaluating student performance in a particular course in a real world higher education system through predicting the likelihood of success. The results of the experiment demonstrate how extracted knowledge may help in improving decision making processes. Higher education institutions have long been interested in predicting the paths of students and alumni, thus identifying which students join particular course programs, and which students require assistance in order to graduate. At the same time, institutions want to learn whether some students are more likely to transfer than others, and what groups of alumni are most likely to offer pledges.

Elena susnea et al discussed Data mining (DM) is useful for collecting and interpreting significant data from huge database. The education field offers several potential data sources for data mining applications. These applications can help both instructors and students in improving the learning process.

3. MATERIALS AND METHODS

The first step of our analysis was to reduce the high data dimensionality. For this purpose, we used Weka tool [4,6,8,9,13] for attribute selection based on various search methods16 made in the attribute space as shown in table 1. We used factors which are selected after preprocessing as new predictors.

Meta Classifier has showed spectacular success in reducing classification error from learned classifiers. These techniques develop a classifier in the form of a committee of classifiers. The committee members are applied to a classification task and their individual outputs combined to create a single classification. Meta learning approaches like AdaBoostM1, Bagging and Dagging, [2,3,10,11] Parameter Selection have received extensive attention. They are the recent methods for improving the predictive power of classifier learning systems.

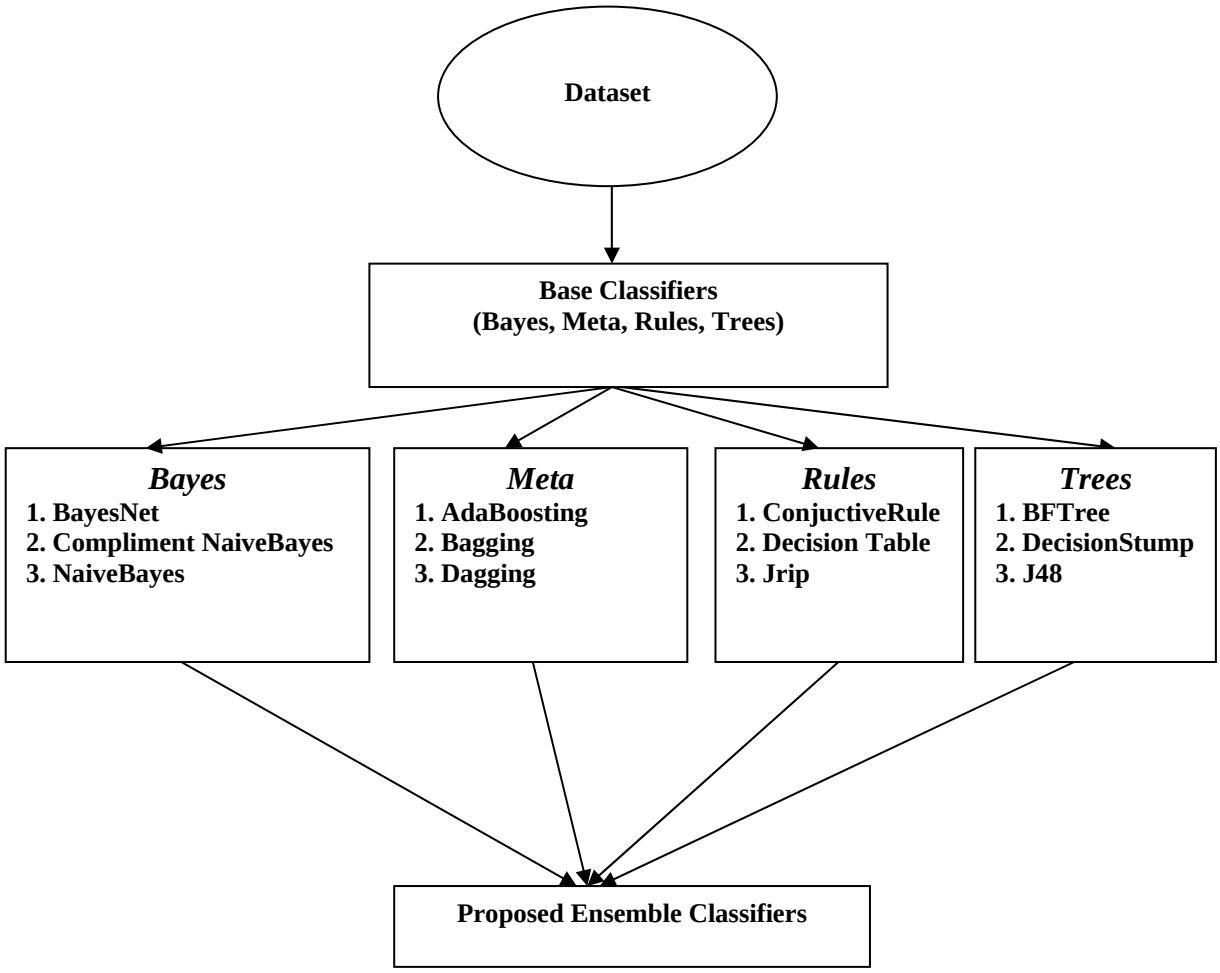

Figure 2 Flow to obtain a Strong Meta classifier.

The above diagram represents the flow of process of the proposed ensemble model by using several base classifiers, such as Bayes, Meta, Rules and Trees. In Bayes classifier has several methods are there but here consider only few of them like as BayesNet, Complement NaïveBayes and NaiveBayes. In Meta classifier has several methods are there but here consider only few of them like as AdaBoosting, Bagging and Dagging models. In Rules classifier has several methods are there but here consider only few of them like as Conjuctive Rule, Decision Table and JRip models.

Table1 Functions of classifiers

Meta Classifier name	Category	Functions
Meta	***Adaboost M1***	Class for boosting a nominal class classifier using the Adaboost M1 method.
	Bagging	Bag a classifier; works for regression too
	Dagging	It creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier.
Base	***BayesNet***	Numeric estimator precision values are chosen based on analysis of the training data
	Compliment NaiveBayes	Class for building and using a Complement class Naive Bayes classifier.
	NaiveBayes	Class for a Naive Bayes classifier using estimator classes.
	Adaboost M1	Class for boosting a nominal class classifier using the Adaboost M1 method.
	Bagging	Bag a classifier; works for regression too
	Dagging	It creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier.
	ConjuctiveRule	This class implements a single conjunctive rule learner that can predict for numeric and nominal class labels.
	Decision Table	Class for building and using a simple decision table majority classifier.
	JRip	This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER),
	BFTree	Class for building a best-first decision tree classifier.
	DecisionStump	Usually used in conjunction with a boosting algorithm.
	J48	For generating a pruned or unpruned C4.5 decision tree.

4. EXPERIMENTAL AND RESULT ANALYSIS

In this section, we test the implementation efficiency of various methods and compare with whole dataset and the selected attributes. Weka tool is used to construct classification models.

The datasets for these experiments are collected from various engineering colleges in tamil nadu.The original data format has been slightly modified and extended in order to get relational format.

The dataset of higher education student academic performance describes a set for selected attributes for Best first search method in the range as shown in the table 1. The output is categorized into Boys and Grils classes. The output class is denoting the possible category of infection affected. Number of Instances in this database is 7435.

Table 2: List of Attribute and their Data Type

S.No	Attribute Name	Data Type
1	Register_Number	VarChar
2	Student_Name	Character
3	Residence	VarChar
4	Medium_of_study	Character
5	Sex	Character
6	Mathematics-I	VarChar
7	Technical_English-I	VarChar
8	Engineering_Physics-I	VarChar
9	Engineering_Chemistry-I	VarChar
10	Computer_Programming	VarChar
11	Engineering_Graphics	VarChar
12	Attendance_Percentage	Numeric
13	Grade	VarChar

The above table contains the 13 attributes. Such as Register_Number, Student_Name, Residence, Medium_of_study, Sex, Mathematics-I, Technical_English-I, Engineering_Physics-I, Engineering_Chemistry-I, Computer_Programming, Engineering_Graphics, Attendance_Percentage,Grade. Based on these attributes compute the ensemble model which is apply for our proposed system.

Table 3: Various Classifiers accuracies

Base Classifiers \ Meta Classifiers		AdaBoostM1	Bagging	Dagging
Bayes	**BayesNet**	**72.18%**	71.18%	69.3%
	Compliment NaiveBayes	52.96%	52.95%	53.02%
	NaiveBayes	69.98%	70.1%	70.21%
Meta	**Adaboost M1**	51.38%	51.38%	54.21%
	Bagging	71.65%	**72.1%**	71.5%
	Dagging	70.66%	71.38%	67.28%
Rules	**ConjuctiveRule**	51.38%	51.38%	60.43%
	Decision Table	71.65%	**72.1%**	70.41%
	JRip	70.66%	71.06%	70.28%
Trees	**BFTree**	72.03%	**72.08%**	71.83%
	DecisionStump	51.38%	51.38%	61.53%
	J48	71.73%	71.73%	68.41%

The above table represents, Every proposed model of the ensemble models like as Meta with Bayes,Meta with Meta, Meta with Rules, Meta with Trees. In Meta with Bayes ensemble model, the AdaBoostM1with BayesNet parameter had 72.18% accuracy level rest of others less than this model while computing the meta with Bayes classifiers. In Meta with Meta ensemble model, the Bagging with Bagging parameter had 72.10% accuracy level rest of others less than this model while computing the Meta with Meta classifiers. In Meta with Rules ensemble model, the Bagging with Decision Table parameter had 72.10% accuracy level rest of others less than this model while computing the Meta with Rules classifiers. In Meta with Trees ensemble model, the Bagging with BFTree parameter had 72.08% accuracy level rest of others less than this model while computing the Meta with Trees classifiers.

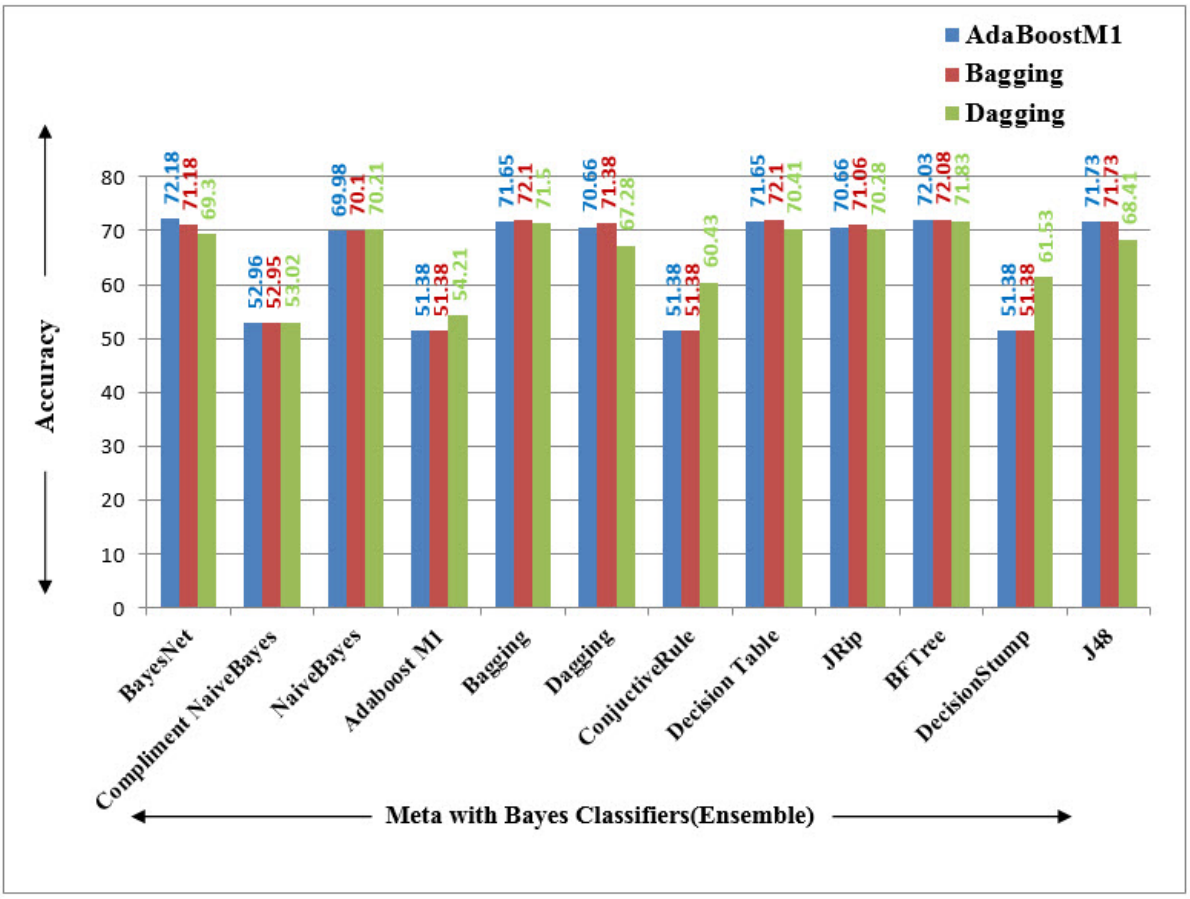

Figure 3 Comparison of meta classifier algorithms for accuracy.

The above figure clearly represents the highest accuracy 72.18% had AdaBoostM1 with BayesNet Parameter only. The rest of other ensemble models have less than the AdaBoostM1 with BayesNet model.

4. CONCLUSION

The above results clearly demonstrate the high accuracy produces the Meta classifications with base classifiers in various analysis. It produces the best accuracy result in ensemble model of AdaBoostM1 with BayesNet Classifiers. I contribute my research works based on student academic performance analysis for higher education.

5. References:

Journal article

[1] Ankit Desai and P M Jadav. Article: An Empirical Evaluation of Ada Boost Extensions for Cost-Sensitive Classification. International Journal of Computer Applications 44(13):34-41, April 2012.

[2] Abdullah Wahbeh H, Mohammed Al-Kabi., "Comparative Assessment of the Performance of Three WEKA Text Classifiers Applied to ArabicText", Vol. 21, No. 1, pp. 15- 28, 2012.

[3] Nikita Bhatt, Amit Thakkar, Amit Ganatra., "A Survey & Current Research Challenges in Meta Learning Approaches based on Dataset Characteristics", Volume-2, Issue-1, March 2012.

[4] Quan Sun, Pfahringer, "Pairwise meta-rules for better meta-learning-based algorithm ranking Machine learning", Springer US, Machine Learning, 93(1):141-161, 2013.

[5] Shilpa Dhanjibhai Serasiya, Neeraj Chaudhary., “Simulation of Various Classifications results using WEKA”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-3, August 2012.
[6] Shaidah Jusoh, Hejab Alfawareh M., “Techniques, Applications and Challenging Issues in Text Mining”, Vol. 9, Issue 6, No 2, November2012.
[7] Tao Wang, Zhenxing Qin, Zhi Jin and Shichao Zhang , "Handling overfitting in test cost-sensitive decision tree learning by feature selection, smoothing and pruning", The journal of systems and software, 2010.

Books

[1] Data mining - Concept and Techniques by Han & Kamber. Data mining: concepts and techniques. The Morgan Kaufmann series in data management systems, ISBN- 1558609016, 9781558609013, publisher morgan, 2006.

Online:

[1] http://weka.sourceforge.net/doc.dev/ [online December’2017]
[2] http://www.ijpbs.net/cms/php/upload/2938_pdf.pdf [online December’2017]
[3] https://fenix.tecnico.ulisboa.pt/downloadFile/282093452003810/Boosting%20-%20Ferreira%20and%20Figueiredo%202013.pdf [online December’2017]
[4] https://aminer.org/data [online December’2017]