iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** XII **Month of publication:** December 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 4 Issue XII, December 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An Ensemble Model for Teaching Assistant Evaluation using Classification Technique

Akhilesh Kumar Shrivas[1], Shikha Mishra[2]
Department of IT, Dr. C. V. Raman University

Abstract— Teaching Assistant Evaluation is very important for every education sector for academic improvement. To improve the performance of teaching skill, criteria is increasing day by day. Due large number of data and criteria, data mining is one of the important factor. In this research work, we have used various data mining based classification techniques for classifying teaching skill. We have proposed ensemble models (CART+CHAID and ANN+BayesNet) to improve the teaching performance, but archived highest testing accuracy as 63.61% in case of ensemble of ANN and Byes Net with 80-20% training-testing partition.
Keywords— Ensemble Model, Classification, Decision Tree.

I. INTRODUCTION

Teaching Assistants(TAs)[1] are individuals who assists professor or teachers with instructional responsibilities. Teaching assistants includes Graduates Teaching assistants(GTs) ,Undergraduate Teaching Assistants(UTAs) secondary school Teaching Assistants(TAs) and elementary school Teaching Assistants(TAs).Teaching Assistant Evaluation Program is developed in accordance with the Guidelines for Uniform Performance Standards and Evaluation adopted Teacher. Evaluation Program supports our district's student achievement goals and our schools. This system provides us effective instructional environment in the education centers. It is allowing for creativity and individual teacher initiative. The goal is to support the continuous growth and development of each teacher by monitoring, analyzing, and applying pertinent data compiled within a system of meaningful feedback to improve student academic progress and educator effectiveness. The main objective of this research work is to develop the robust model for teaching assistant evaluation. There are various authors have worked in the field of teaching assistant evaluation. Anuj Gupta et al. (2015)[5] have proposed classification techniques like J48 ,Decision Table, Multilayer Perceptron (MLP), Naive bayes and other algorithms for evaluation of teaching assistant .They got better accuracy (41.05%) in case of J48 and Naive bayes classifier. Thakaa Z. Mohammad et al.(2015)[10] have proposed an intelligent educational data mining classification model designed for teaching English for slow learner's students. The model is also called IEDM-SL. The main motive of IEDM_SL model is to identifying learning pattern and improve their performance. J. Yang et al. (2011) [1] have used SVM technique for Teaching Assistant Evaluation. They have trained and tested the SVM model with different kernel functions like linear, polynomial, radial basis function. Other authors have also used various classifications techniques to develop various models for evaluation of teaching assistant evaluation.

II. ARCHITECTURE OF PROPOSED SYSTEM

The architecture of proposed system as shown in figure1. The teaching assistant evaluation data set is categorized into three different data partitions like 70-30% , 75-25% and 80-20% training- testing data partitions. All mention data partitions are applied on different models like C4.5, CART,ANN, Bayesian Net, ensemble of CART and CHAID and ensemble of ANN and Bayesian Net. The performance of all classifier are calculated in terms of accuracy, sensitivity and specificity.

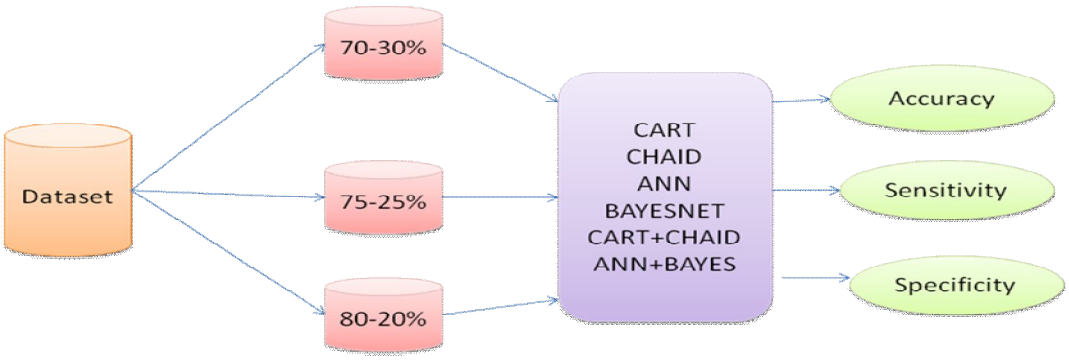


Figure 1: Architecture of proposed system

III. CLASSIFICATION TECHNIQUES

Classification is one of the important applications of data mining techniques for classification of data. Classification is also called supervised learning because each record associated with class. Classification process consist two steps: first step is trained the model using training samples of data set and second one is testing the trained model using testing sample of same data set. Figure 2 shows that generic architecture of training and testing the model.

A. *Decision Trees*

A decision tree [3] is a classification scheme and is used to be find description of several predefined classes and classify them. The decision tree consists of nodes that form a rooted tree, meaning it is a directed tree with a node called "root" that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves.

In a decision tree, each internal node splits the instance space into two or more subspaces according to a certain discrete function of the input attributes values. In the simplest and most frequent case, each test considers a single attribute, such that the instance space is partitioned according to the attribute's value. In the case of numeric attributes, the condition refers to a range. Each leaf is assigned to one class representing the most appropriate target value.

Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path. There are various decision tree algorithm as described below:

1) *Classification and Regression Tree (CART):* CART [3] stands for Classification and Regression Trees. It is [3] one of the popular of building trees in the machine learning community. CART builds a binary decision tree by splitting the records at each node.

2) *CHI-Squared Automatic Interaction Detection (CHAID):* CHAID[4] (Chi-square–Automatic–Interaction–Detection) was originally designed to handle nominal attributes only. CHAID attempts to stop growing the tree before over fitting occurs where as the algorithms generate a fully grown tree and than carry out pruning as post processing step.

3) *Bayesian Classification:* Bayesian classifiers [2] are statistical classifiers. They can predict class membership probabilities. Bayesian classification is based on Bayes theorem .A simple Bayesion classifier known as a native Bayesian classifier to be comparable in performance with decision tree and neutral network classifier. Bayesian classifiers have also exhibited high accuracy and speed when applied large databases. Bayesian belief network can also be used for classification.

4) *Artificial Neural Network (ANN):* An artificial neural network (ANN)[3] are analytic techniques modeled after the processes of learning in the cognitive functions of the brain and capable of predicting new observation from other observation. "neural network" (NN) is a mathematical model or computational model based on biological neural networks, in other words, it is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

5) *Ensemble Model:* An ensemble model is a combination of two or more models to avoid the drawbacks of individual models and to achieve high accuracy. The two models are combined by using high confidential wins scheme [15] where weights are weighted based on the confidence value of each prediction. Then the weights are summed and the value with highest total is again selected. The confidence for the final selection is the sum of the weights for the winning values divided by the number of models included in the ensemble model. If one model predicts no with a higher confidence than the two yes predictions combined, then no wins. In this research work, we have used voting scheme for ensemble of models.

IV. DATA SET

The data set is collected from UCI repository. This data set [6] is provided by the statistics department of the University of Wisconsin Madison . This data set consists of the evaluation of teaching performance of three category .It contains 151 instances , 5 attributes and 1 class with three categories. Table 1 shows that features and class of data set.

Table 1: Features of dataset with type of attribute

Attributes Name	Attributes Type	Attributes Possible Values
English_ speaker	Binary	1=English speaker, 2=non English speaker
Course_ instructor	Categorical	25 category
Course	Categorical	26 category
Regular_ semester	Binary	1=summer semester 2=Regular semester
Class _ size	real	
Class_ attributes		1=Low 2=Medium 3=High

The data processed in order to be used by Clementine is software which is mature data mining toolkit. It contains a collection of visualization tools and algorithms for data classification modelling.

V. RESULTS AND DISCUSSION

In this research work, we have analysed the various classification techniques for classifying performance of teaching assistant. This experiment is carried out using WEKA data mining tools in window environment. We have applied the teaching assistant evaluation data set into different models like CART, CHAID, ANN, Bayesian Net, ensemble of CART and CHAID and ensemble of ANN and Bayesian Net with different data partitions. The accuracy of different models with different partitions as shown in table 2. Table 2 shows that accuracy of individuals and ensemble models. The accuracy of individual models is not satisfactory with all partitions, but ensemble models gives better accuracy compare to individual's models. Partition of data play very important role for classification of data. The accuracy is less or equals of ensemble of CART and CHAID compare to individuals model, but we have achieved better accuracy 56.25%, 58.14% and 63.16% with 70-30%, 75-25% and 80-20% training-testing partition respectively in case of ensemble of ANN and BAYES Net. In case of ensemble of ANN and BayesNet , the accuracy of model is increasing when traing partition increases and testing partition decreases. The confusion matrix of best model (ANN+Bayes Net) as shown in table 3 with high, low and medium category of class. To check the robustness of model , other performance measures like sensitivity and specificity play very important role. Table4 shows that various performance measures like sensitivity, specificity and also accuracy with three categories of class as high ,low and medium performance of teachers.

Table 2: Accuracy of model with different data partition

Model	70-30%	75-25%	80-20%
CART	47.92%	46.51%	50%
CHAID	41.67%	46.51%	50%
ANN	52.06%	53.49%	55.26%
Bayes Net	47.92%	51.16%	55.26%
CART +CHAID	50%	51.16%	55.26%
ANN+BAYES Net	56.25%	58.14%	63.16%

Table 3: Confusion matrix of best model (ANN+BayesNet)

Actual Vs. Predicted	70-30% Data partition			75-25% Data partition			80-20% Data partition		
	High	Low	Medium	High	Low	Medium	High	Low	Medium
High	9	1	6	9	1	5	9	1	3
Low	0	10	7	0	8	6	0	8	4
Medium	2	5	8	2	4	8	2	4	7

Table 4: Various performance measures

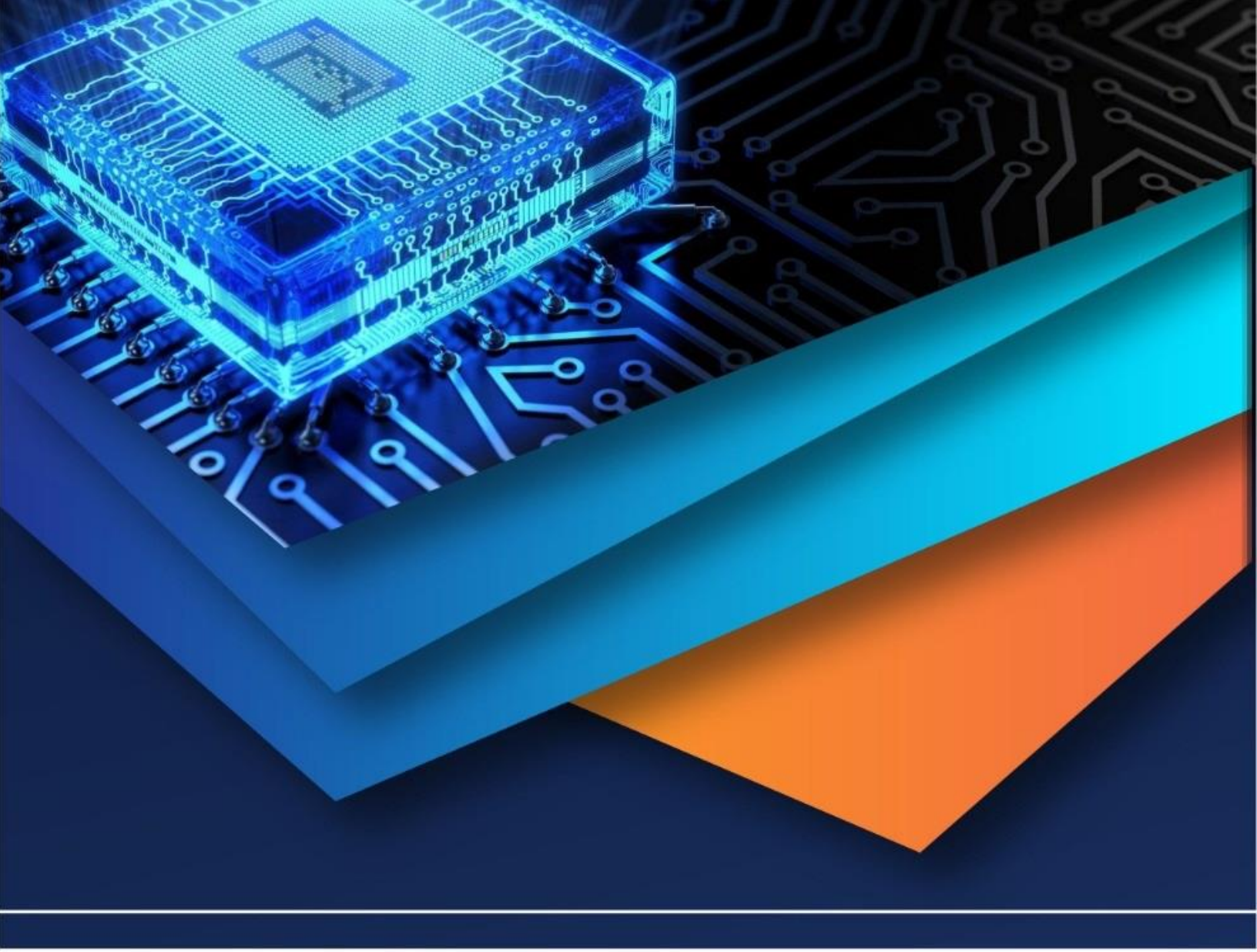
Categories of performance of teacher	70-30% Data Partition			75-25% Data Partition			80-20% Data Partition		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
High	56.25%	93.75%	81.25%	60%	92.85%	81.39%	69.23%	92%	84.21%
Low	58.82%	80.64%	72.91%	57.14%	82.75%	76.74%	66.66%	80.76%	76.31%
Medium	53.33%	68.29%	58.33%	57.14%	62.06%	60.46%	53.84%	72%	65.78%

VI. CONCLUSIONS AND FUTURE WORK

An evaluation of teaching assistant is very important in every academic sector for analysis of teacher's performance. In this research work, we have used data mining based classification techniques to evaluate the performance of teachers as low, high and medium. The proposed ensemble of ANN and Bayes net gives better accuracy and recommended best model to evaluate the performance of teachers. In future, we will extend the criteria and features of data set and applied the feature selection and optimization technique to develop computationally efficient model for classification teacher performance.

REFERENCES

[1] J. Yang, H. Jiang and Honglei Zhang, "Teaching Assistant Evaluation Based on Support Vector Machines with Parameters Optimization", Information Technology Journal, Vol.No. 10,PP.No. 2140-2146 , 2011.
[2] Jiawei Han and Micheline Kamber,"Data Mining: Concept and techniques" Simon Fraser University,2005.
[3] Bharat Bhushan Agarwal and Sumit Prakash, "Data Mining and Data Warehousing" I.F.T.M. University Moradabad (U.P.) UDM-9324-28,2009.
[4] A.K.Pujari
[5] A.Gupta, "Classification Of Complex UCI Datasets Using Machine Learning And Evolutionary Algorithms" , International journal of Scientific and technology research,Vol.No.4 , PP. No.94 , 2015.
[6] E.Georgiana Dragomir, "Teaching Performance Evaluation Using Supervised Machine Learning Techniques",The 5th International Conference on Virtual Learning (ICVL) ,PP.NO.390-394, 2010.
[7] A.A.B. Subramanian, Spramala, B.Rajalakshmi, Ramasamy Rajaram "Improving Decision Tree Performance by Exception Handling", International Journal of Automation and Computing 7(3), PP.No. 372-380 ,2010.
[8] S. Agarwal, G. N. Pandey, and M. D. Tiwari ,Data Mining in Education: Data Classification and Decision Tree Approach,Vol.No.2 , pp. 140-144. , 2012.
[9] P. Kaur , M. Singh and G. Singh Josan , "Classification and Prediction Based Data Mining Algorithms to predict slow learners in Education Sector",3rd International Conference On Recent Trends in computing , PP. No. 500-508, 2015.
[10] T. Z. Mohammad, A.M.Mahmoud, El-Sayed M. El-Horbart, Mohamed I.Roushdy and Abdel-Badeeh M. Salem, "An Intelligent Educational Data Mining Classification Model for Teaching English for Slow Learner Students" , IPASJ International Journal of Computer Science ,Vol.No.2,PP. No.6-15,2014 .
[11] L.Jena and N. K. Kamila , "Distributed Data Mining Classification Algorithms for Prediction of Chronic- Kidney-Disease", International Journal of Emerging Research in Management &Technology ,Vol.No.-4,PP.No.110-118 Issue-11, ISSN: 2278-9359 ,2015.
[12] Yu-Shan S. and Wei-Yin L.(2011) QUEST Classification tree (version 1.9.2),http://www.start.wisc.edu/~loh/quest.html.(Browsing date: 25th April 2012).
[13] S. Mardikyan and B.badur ,"Analyzing Teaching Performance Of Instructors Using Data Mining Techniques ",Informatics In Education,2011vilnius university ,PP No.245-257,2011.
[14] B. Tomasz, R. Tahir and P. Jens Myrup , "A Method for Classification of Network Traffic based on C5.0 Machine Learning Algorithm", ICNC, Maui, HI ,2012.
[15] A. M . Elsayad, Predicting the severity of breast masses with ensemble of Bayesian classifiers, Journal of Computer Science, 6(5), pp. 576-584, 2010.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY