Pedagogue Performance Assessment (PPA) using Data mining Techniques

Ashily M Baby[1]

[1]Graduate Student,
Dept. of Computer Science and Engg., St. Joseph's College of Engineering and Technology, Palai, Kerala, India
E-mail: ashily77@gmail.com

Abstract. Educational Data Mining (EDM) is a developing domain that exploring pedagogical data by applying different machine learning techniques. It can be considered as a research field which provides intrinsic knowledge of teaching and learning process for effective education. The main objective of any educational institution is to provide quality education to their students. One way to achieve highest level of quality in higher education system is by discovering knowledge that predicts pedagogues performance. This paper presents an efficient system model for evaluation and prediction of pedagogue performance in higher institutions of learning using data mining technologies. To achieve the objectives of this work, three classification algorithms like decision tree algorithms(C4.5), support vector machines (SMO) and artificial neural networks (MLP) are chosen to build classifier models on a dataset. The classifier system was tested successfully using case study data from St. Joseph college of Engineering and Technology, Palai. The data consists of the feedbacks that got from the students. The aim of this work is to demonstrate the capability of EDM in illuminating the criteria of fruitful instructor performance as perceived by the students. The result shows the accuracy of classifier models that predicts the performance of pedagogues. Decision tree J48 have higher accuracy of 94.37% than SMO and MLP.

Keywords: Artificial neural networks, classification algorithms, decision trees, performance evaluation, support vector machines.

1. Introduction
One of the biggest problem of higher education organizations are the large amount of data and how to use them to improve quality of academic programs and services and the managerial decisions. A variety of "formal and informal" procedures based on "qualitative and quantitative" methods is utilized by advanced education institutions to tackle issues, which keep them far from accomplishing their quality goals. Techniques utilized in higher education for quality reasons are mainly based on predefined questions and charts to investigate the information. In expansion, these strategies do not have the capacity to uncover useful hidden data. Covered up information in substantial datasets is best analyzed with data mining methods. Data mining can be characterized as the automated process of removing useful knowledge and information that are obscure from large datasets. One of the common

problems in higher education is the evaluation of pedagogues performances in a subject. The most widely applied tool to evaluate the pedagogues performance in a subject is through surveying students responses about the subject and its pedagogue through a questionnaire. Students are the only source of information about the learning environment, and they are the only ones who can rate the quality, the effectiveness, and the satisfaction of course content, method of instruction, textbook, and homework. Student evaluations are used primarily to improve course and teaching quality as well as a part of the evaluation process for staff assessment systems. The aim of this work is to demonstrate the capability of EDM in illuminating the criteria of fruitful instructor performance as perceived by the students. In this study, three classification techniques-decision tree algorithms, support vector machines (SMO) and artificial neural networks (MLP) are chosen to build classifier models on a dataset composed of the feedbacks of students to a subject evaluation questionnaire and the performances of these models are compared.

2. Related Works
Broad literature survey of the EDM research field are given by Romero and Ventura [1]. In this paper, they overviewed the use of data mining to traditional educational systems, surely understood learning content administration frameworks , specific web-based courses and versatile and keen web-based educational systems. Another broad survey was done by Baker and Yacef[2]. They investigated the past points of interest and current trends in the field of Educational Data Mining (EDM), thought about the methodological profile of research in the early years of EDM, and talked about trends and shifts in the examination led by this group. Another broad survey was done by Rajni and Malaya [3], concentrated on parts, research trends of EDM featuring its related tools, instructive outcomes and methods. They additionally featured the challenges in EDM. Another broad survey by Varun and Arupama [4] inspected the use of data mining procedures in higher instructive foundation to separate valuable data from big data sets and gave analytical tool to view and utilize this data for decision making processes by taking genuine illustrations. Chin-Chia Hsu and Tao Huang [5] directed an investigation on the utilization of data mining innovation to assess student's academic achievement through numerous channels of enrolment like joint enlistment enrolment, athletic enrolment and application enrolment. A comparative report was done by Osofisan and Olamiti [6] where they examined the academic background in relationship with the performance of students in a computer science programme in a Nigerian university. Their investigation demonstrated that the review got from senior secondary school examination (SSCE) in mathematics is the highest determinant of students' performance using the C4.5 learning algorithm in building the model of the student's performance.

3. Proposed Methodology
The proposed system aims at using the classification method of data mining for the prediction of pedagogues performance. The prediction model was developed using the Classification methods of the Data mining technique. The Neural Network data mining technique (the multilayer perceptron algorithm), Support Vector Machine (SMO), Decision Trees methods specifically the C4.5 algorithms were used and their performances were compared to each other.

3.1. Data Acquisition

The dataset used in this project was collected from St. Joseph College of Engineering and Technology, Palai. Student evaluation data has 13 variables including the name of faculty, the subject name, year, semester and questionnaire . Response values of these questions are of the form {1; 2; 3} where 1; 2; 3 represents the answers "Good", "Satisfactory", "Not satisfactory", for Q1 and 1; 2; 3 represents "Loud and clear", "Reasonably loud", "Too low" for Q2 and 1;2;3 represents "Very clear", "Somewhat clear", "Not clear" for Q3 and 1;2;3 represents "Exeecellent", "Reasonably good", "Occasional disorder in class" for Q4 and 1;2;3 represents "Too fast", "Good", "Too slow" for Q5 and 1;2;3 represents "Yes", "To some extend", "No" for Q6 and 1;2;3 represents "Very easy to understand" , "Somewhat easy to understand", "Difficult to understand" for Q7 and 1;2;3 represents "As a role model", "Good", "Satisfactory" for Q8. The last variable is dichotomous variable measured on a nominal scale in the
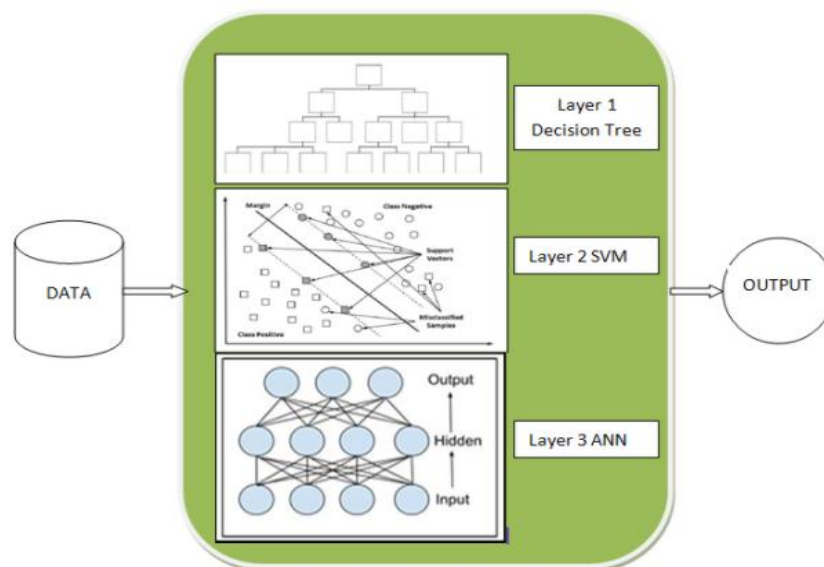
form of {1; 2} where 1 stands for "Good" and 2 for "Poor". Details of the variables used in this study are shown in Table I. Variable Q9 is the class variable that is to be predicted. The value "Good" for this variable is taken as positive class label whereas "Poor" is assumed to be negative.

Table 1. Performance Summary of Models

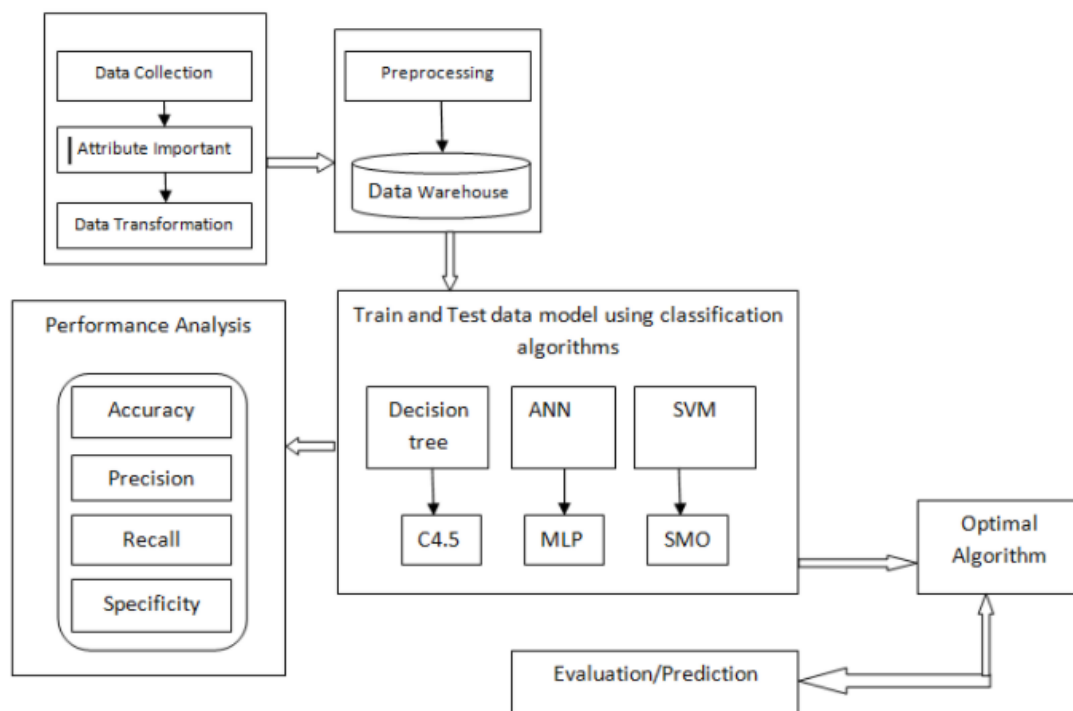| Variable	Description	Possible Values
Q1	Knowledge of the pedagogue	1,2,3
Q2	Pedagogue's voice	1,2,3
Q3	Writing and sketeches	1,2,3
Q4	Control and command over class	1,2,3,
Q5	Speed of presentation	1,2,3,
Q6	Pedagogue encourages questioning	1,2,3
Q7	Thw way the subject taught by the pedagogue	1,2,3
Q8	Rating of the pedagogue	1,2,3
Q9	Overall performance	1,2

3.2. System Design

The tasks involve machine learning and classification algorithms; hence in the research work a three-layered classifier system in figure 1 was designed to achieve the objective of the work. Layer 1 consists of an Artificial Neural Networks (MLP), layer 2 is made up of Decision Trees classifiers and layer 3 made up of Support Vector Machine (SMO). These classifiers have been selected because of their performances in various domains. They have both been successfully applied to a variety of real-world classification tasks in industry, business, science and education with good performances. The Neural Networks is known for its predictive accuracy, ability and aptitude to learn and remember. Decision Tree classifiers are considered "white box" classification model as they can provide explanation for their models and can be used directly for decision making. These abilities and aptitude are best suited and of good requirement for any effective and efficient intelligent system.

Figure 1. Three Layered Classifier System

3.3. Analysis Proposed Model

The goal of the proposed system it to aid higher institution management in determining teachers'
performance and recommend necessary actions to be taken on individual teacher based on the
prediction from the intelligent evaluation system. The proposed system framework includes three
components: The first and the second components take care of data acquisition and storage,
responsible for storing teachers' data, gathered from different data sources proposed in a data
warehouse. The third component is model building, responsible for obtaining knowledge about the
teachers, through appropriates classification models. Different classification algorithms are proposed
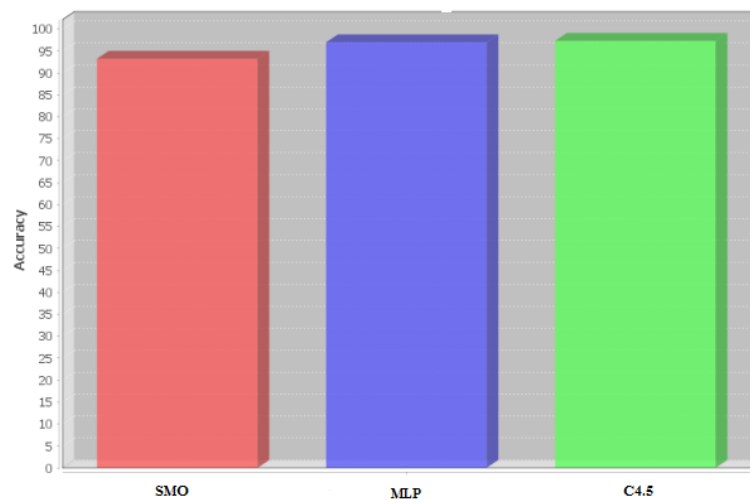in search for the best model with high predictive accuracy.

Figure 2. Architecture of Proposed System

4. Results and Discussion

In this study, three classification models are generated: decision tree algorithms (C4.5), Support
Vector Machine(SMO) and Artificial Neural Networks(MLP). The performances of these models are
evaluated on the test data in terms of accuracy, precision, recall, and specificity. Accuracy values,
which assess the effectiveness of the models, are all at least approximately 90%, C4.5 have 94.37%,
SMO have 90.85% and MLP have 92.96%. C4.5 classifier is the best in performance according to
accuracy followed by MLP, and SVM is the worst. Precision, which assesses the predictive power,
again indicated C4.5 as the best classifier; however, MLP and SVM also show high predictive power.
According to MCC and F-measure C4.5 is best. As a result, from the SVM classifier, and from the
artificial neural network classifier, the C4.5 is comparatively better. Ultimately, C4.5 can be
considered as the outstanding classifier among all according to the given performance measures. The
weighted averages of the models were compared using different performance measures like True
Positive Rate, False positive rate, Precision, Recall, MCC, F-Measure, ROC area, PRC area.

Table 2. Performance Summery of Models

Algorithm	TP Rate	FP Rate	Precision	Recall	MCC	F-Measure	ROC Area	PRC Area
J48(C4.5)	1.000	0.667	0.986	1.000	0.993	0.573	0.667	0.986
SVM(SMO)	1.000	1.000	0.980	1.000	0.990	0.000	0.500	0.980
ANN(MLP)	1.000	1.000	0.980	1.000	0.990	0.000	0.900	0.998

Figure 3. Accuracy Graph of Classification Algorithms

C4.5 algorithm predicts better than the SVM and MLP algorithms since its accuracy is the highest compared to others. The results obtained from the analysis demonstrated a slight higher performance of model built with decision tree (C4.5 algorithm) over neural network (MLP algorithm). Both C4.5 and MLP algorithms results show great superiority over SVM algorithm in terms of performance. Also the rules generated makes C4.5 decision tree algorithm clearer and understandable. The figure 3 shows the accuracy graph of classification algorithms, the C4.5 and MLP have much accuracy than the SMO.

5. Conclusion

Data mining techniques are applied in higher education more and more to give insights to educational and administrative problems in order to increase the managerial effectiveness. Based on student's perception instructor performances are found and these can help instructors to improve their performances. The aim of this project is to show the potential of EDM in enlightening the criteria or measures of effective instructor performance as perceived by the students. The Decision tree J48(C4.5) is the best algorithm suitable for predicting teachers' performance in relation to other algorithms in this work. Our results also show the performance of the teachers in this study. The outcome of this project demonstrated that data mining techniques can be applied in the prediction of teachers' performance. This project can be improved upon by using other classification algorithms and other data mining techniques such as, Naïve Bayes classifier, genetic algorithm as well as data from other universities. The data set can also be expanded with more distinctive attributes to get improve upon the classification accuracy and make the work more robust.

References

[1]C. Romero, S. Ventura, 2007 "Educational data mining: A Survey from 1995 to 2005", Expert Systems with Applications, (33), pp. 135-146.

[2]Baker, R. S. J. D., and K. Yacef. 2009 "The State of Educational Data Mining in 2009: A Review and Future vision", Journal of Educational Data Mining 1 (1): 3–17.

[3]Rajni J. and Malaya D.B , 2013 "A Survey on Educational Data Mining and Research Trends", International Journal of Database Management Systems (IJDMS).

[4]VarunKumar and Anupamachada, 2011 "An Empirical Study of the Applications of Data Mining Techniques in Higher Education", International Journal of Advanced Computer Science and Applications, Vol. 2, No.3.

[5]Chin Chia Hsu and Tao Huang, 2006 "The use of Data Mining Technology to Evaluate Student's Academic Achievement via multiple Channels of Enrolment", An empirical analysis of St. John's University of Technology.

[6]Osofisan A.O. and Olamiti A.O, 2009 "Academic Background of Students and Performance Computer Science Programme", Nigerian University.

[7]Asanbe M.O, Osofisan A.O, William W.F, 2016 "Teachers' Performance Evaluation in Higher Educational Institution using Data Mining Technique", International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868.

[8]Ankita A Nichat, Dr.Anjali B Raut. 2017 "Predicting and Analysis of Student Performance Using Decision Tree Technique", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 4.

[9]Ahmed Mohamed Ahmed, Ahmet Rizaner, Ali Hakan Ulusoy. 2016 "Using data mining to predict instructor performance", 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS.

[10]Aranuwa Felix Ola, Prof. Sellapan Pallaniappan. 2013 "A data mining model for evaluation of instructors' performance in higher institutions of learning using machine learning algorithms", International Journal of Conceptions on Computing and Information Technology Vol. 1, Issue 2, ISSN: 2345 – 9808.

[11]Randa Kh. Hemaid, Alaa M. El-Halees. 2015 "Improving Teacher Performance using DataMining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 2.

[12]Mardikyan S, and Badur B. 2011 "Analyzing teaching Performance of Instructors Using Data Mining techniques", Informatics in Education.

[13] Musthafa Agaoglu. 2016 "Predicting Instructor Performance Using Data Mining Techniques In Higher Education", IEEE Access, Volume 4.

[14]A. M. Abaidullah, N. Ahmed, E. 2015 "Identifying hidden patterns in students' feedback through cluster analysis", Int. J. Comput. Theory Eng.,vol. 7, no. 1, pp. 1620.

[15] Randa Kh. Hemaid, Alaa M. El-Halees. 2015 "Improving Teacher Performance using Data Mining", International Journal of Advanced Research in Computer and Communication Engineering,Vol. 4, Issue 2.

[16] A. Pena-Ayala. 2014 "Review: Educational data mining: A survey and a data mining-based analysis of recent works", Expert Syst. Appl., vol. 41.

 [17] N. Delavari, S. Phon-Amnuaisuk, and M. R. Beikzadeh. 2007 "Data mining application in higher learning institutions", Inform. Edu.-Int. J., vol. 7

[18] M. Goyal and R. Vohra. 2012 "Applications of data mining in higher education", Int. J. Comput. Sci. Issue, vol. 9.