

# Efficacy and efficiency of educational data mining through multistrategy machine learning

Sandhya Maitra  
Dept. of Computer science  
Banasthali Vidyapith  
Banasthali, India  
msan324@gmail.com

Sushila Madan  
Dept. of Computer science  
Lady Sri Ram college for women  
New Delhi  
sushila\_yahoo.com

Rekha Kandwal  
Dept. of Computer science  
Mahan Institute of Technologies  
rekha.kandwal@gmail.com

---

*Abstract: The present day higher educational institutions in the country face the tough challenge of surviving in the academia and producing cutting edge professionals for an industry with ever growing demand of meeting the dynamically changing requirements. The absence of a comprehensive data analysis framework and combined analysis of correlated factors affecting the teaching learning process may give rise to false interpretations which are relied upon by the institutions. As a resultant, the system may fail to take effective measures or introduce necessary interventions in time to handle the deviations in the right manner thus defeating the purpose of meeting the program outcomes. In the absence of an accurate comprehensive quality management mechanism individual perceptions bring subjectivity into the quality evaluation and so the results are skewed and give rise to false projections or fabricated interpretations. Major decisions of regulating bodies or institutes themselves are dependent on such analyses which necessitate correct conclusions. Thus quality management of teaching learning process is of utmost importance from the perspective of student, teacher and all other stake holders. It provides a scientific unambiguous basis for quality evaluation and consequently better quality management. A multi strategy machine learning framework goes a long way in supporting evaluation of both qualitative and quantitative aspects by reducing uncertainty in data analysis and facilitating effective quality management.*

**Keywords:** Machine learning, educational data mining, quality management, teaching learning.

---

## I. INTRODUCTION

Teaching learning process is a method for monitoring and judging the overall quality of the teaching and learning based on objective, data and scientific criteria. It is the most powerful instrument of education to bring about desired changes in students. The teaching learning process consists of four basic elements namely assessment, planning, implementation and evaluation. Quality management is the process of ensuring consistency of an organization, product or service by quality planning, quality control, quality assurance and quality improvement [1]. Quality management is focused on product, service quality as well as the means to achieve it. Quality dimensions in teaching learning of higher education include the outputs such as feedback, attendance, results, placements, students level progression as well as the outcomes of the program. Quality Assurance can be achieved by use of statistical process control, prevention mechanisms, audit of quality systems and cause effect analysis. Total Quality management aims for continuous improvements while quality control can be through self evaluation. Juran's quality improvement steps recommend various types of educational data analyses. Software quality dimensions of higher education recommend a detailed analysis of feedback, attendance, deviations, results, outcomes, placements [2]. The data analysis is susceptible to uncertainties of several kinds such as ambiguity, vagueness etc. and there are inadequate measures to handle them. These uncertainties creep in due to non consideration of correlations between various quality parameters during data analysis.

## II. QUALITY MANAGEMENT AND MACHINE LEARNING

The review is performed on computational intelligence powered data mining through usage of machine learning approaches for quality management of teaching learning in Higher Education Institutions. The study is conducted on two aspects namely quality parameters in teaching learning in higher education and educational data mining using machine learning approaches. The Population under consideration is Higher education Institution (student performance) Intervention used is Methods/ techniques for classification and prediction. Outcome is Prediction accuracy on successful prediction techniques. Context is teaching learning, empirical studies such as preliminary Studies and experiments. The research questions proposed in this study are:

- What are the important parameters affecting quality of teaching learning?
- Which of the machine learning methods are used for classification and prediction in quality management of teaching learning in Higher Education Institutions?
- What are the limitations of usage of machine learning in educational data mining so far?

## III. RELATED LITERATURE REVIEW

Reference [3] proposed a multilevel fuzzy evaluation theory based faculty performance evaluation for higher educational institutions. The procedure takes into account three aspects of a teacher namely morality, teaching and scientific research. But student feedback for faculty is not given due consideration which is a very important parameter for assessing a teacher.

# Efficacy and efficiency of educational data mining through multistrategy machine learning

Reference [4] propose a fuzzy comprehensive evaluation based on an index system taking into account all elements of teaching quality, attitude, method, content and effectiveness. To make the process transparent and objective, logic mathematical methods and index weights are determined using analytic hierarchy process providing a good fusion of qualitative and quantitative analysis. It requires filtering of student feedback for faculty from negative factors which influence the feedback. Reference [5] proposed a teaching quality evaluation model based on theory of fuzzy mathematics with class room teaching quality evaluation system which facilitates qualitative as well as quantitative evaluation. There was no consideration of student student feedback for faculty. Reference [6] propose a method for obtaining priority vector for intuitionistic fuzzy complementary judgement matrix. It is further applied to analytic hierarchy process for class room teaching quality evaluation. There was no procedure to validate feedback obtained from students on classroom delivery. Reference [7] elaborate on innovation education and propose an evaluation index system to evaluate the quality from four perspectives i.e., government, university, society and student. The evaluation method is based on fuzzy analytical hierarchy process. The problem with this approach is that the index system requires constant revision. Reference [8] performed a survey on usage of data mining techniques in higher education. They observed that data mining is an indispensable tool in education industry as it helps learn many trends prevalent in education data pertaining to various aspects such as teaching learning process. The research gap is that no single tool suffices to cater to all kinds of educational systems. Further the tools are too complex for usage by educators and require integration into e-learning environment as well. Reference [9] study the quality of predictive models offered by machine learning algorithms for student retention management. It is observed that decision trees produce easy to interpret classification rules compared to other models. The experimental results show that predictive models produce short and accurate prediction list for student retention besides identifying students requiring special attention thus reducing dropout rate. The research gap is that weaknesses of the predicted dropout students are not evident to help facilitate timely corrective measures. Reference [10] reviews TQM and emphasizes on how higher education institutions can improve the quality by adopting principles of TQM and proposes a conceptual TQM model for excellence in Higher Education Institutes. The model is based on the five variables of Commitment of top management, Course delivery, Campus facilities, Customer feedback and improvement which lead to student satisfaction. The implementation aspect needs to be specified. Reference [11] studied and analyzed the effectiveness of different classification algorithms in prediction of student academic performance which is based on various factors. It is observed that Neural network based classification showed maximum precision followed by Naïve bayes algorithm and ID3 algorithm in comparison with Reptree, SimpleCart, Decision table and J48 machine learning algorithm. But a multi strategy machine learning approach can overcome weaknesses of each of these models. Two classification methods namely decision tree and fuzzy

genetic algorithm are used by [12] for predicting students' academic performance in both bachelors and masters degree which serves as feed forward mechanism for teaching faculty to pay more attention to these students before it is too late. It also helps in placements of expert students to reputed firms. The decision tree showed more students at risk class while the genetic algorithm gave more passed students due to classification of students between risk and safe state as safe. The research gap is that the decision tree adopts a pessimistic approach while genetic algorithm follows a totally optimistic approach both of which may result in uncertainty creeping into results. Faculty are not equipped with specific factors leading to poor performance in case of each student. Reference [13] applied ID3 and C4.5 separately to analyse the course evaluation questionnaires and classify satisfactory and unsatisfactory student performances and their weaknesses in the subjects. The instructors are at an advantage to identify students at risk of poor performance and take timely measures to improve their performance. It is further observed that for small data sets both algorithms exhibited same accuracy while for large data sets C4.5 was marginally more accurate than ID3. The experiment identifies potential poor performers but not the specific reasons for those weaknesses. Reference [14] performed a survey of decision tree, C4.5, Naïve Bayesian, RIPPER and SVM prediction algorithms and presented a comparison of their best performances on different aspects. The Naïve Bayes algorithm is found to be superior than others in terms of FP rate, Precision, F-M, Recall and MCC. The applicability of the algorithms in quality management of teaching learning in Higher Education Institutions is yet to be explored. Reference [15] used decision tree Classification technique on students' assessment outcome to identify students at risk of poor performance to support the quality management of teaching learning process. The research gap is that the specific weaknesses of students are not identified by this approach. Reference [16] create a model that successfully classifies students into one of two categories, depending on their success at the end of their first academic year, and finding meaningful variables affecting their success. This model is based on information regarding student success in high school and their courses after completing their first year of study, as well as the rank of preferences assigned to the observed faculty, and attempts to classify students into one of the two categories in line with their academic success.

## IV. DISCUSSION

The areas of teaching learning to be addressed for data analysis according to TQM in Higher Education Institutions are feedback analysis, result analysis, attendance analysis and student progression. Student Feedback for class room delivery of faculty needs to be devoid of influences which may cloud the actual measure. It should be validated. We can use indexes to quantify qualitative aspects. Further correlations should be explored among various student parameters such as results, attendance, feedback and skills acquired. Correlated combined data analysis shall be performed on interdependent quality parameters. The Quality of teaching learning should be

# Efficacy and efficiency of educational data mining through multistrategy machine learning

managed through rigorous feed forward and feedback mechanism based on aforementioned approaches. Machine learning approaches to data mining show promising results. Nevertheless single approaches still suffer from inefficiencies especially to deal with uncertainties. Multi strategy machine learning approaches are the need of the hour. Among the popular machine learning approaches a Decision tree, Naïve Bayes, Neural Network give good performance for classification and prediction.

## V. CONCLUSION AND FUTURE WORK

A Model based on Multistrategy machine learning techniques is proposed for filtering, classification and analysis of correlated quality parameters namely feedback, attendance, results to provide a more realistic and comprehensive basis for their interpretation. A comprehensive evaluation index system will be designed which shall capture the quantitative as well as the qualitative data. Further it is proposed to build an effective framework based on a combination of machine learning techniques to evaluate the index, to forecast the direction of teaching learning performance and allow for adjustments to handle uncertainties such as ambiguities prevalent in the analysis of data for betterment of the teaching learning process. Subsequently the model shall support effective quality management of teaching learning process in higher educational institutions of the country.

## VI. REFERENCES

- [1] Rose, K. Rose, *Project Quality Management: Why, What and How*. Boca Raton, Fla.: J. Ross, 2005, pp. 41. ISBN 1-932159-48-7.
- [2] M. Sanjaya, "Quality Assurance in Higher Education: An introduction", NAAC, Bangalore, 2018.
- [3] J. Zhao and X. Li, "Investigation on Comprehensive Evaluation of University Teachers' Achievements", in *International Conference on Machine Learning and Cybernetics*, Dalian, China, 2006, pp. 1720 - 1723. DOI: 10.1109/ICMLC.2006.258969
- [4] S. Sanyuan, W. Changlong and L. Dongping, "Comprehensive Evaluation of CDIO Model Teachers' Classroom Teaching Quality Based on Fuzzy Analytic Hierarchy Process", in *International Conference on Innovation Management*, Wuhan, China, 2009, pp. 88 - 91. DOI:10.1109/ICIM.2009.38
- [5] H. Liu, X. Huang and Y. Zhang, "Fuzzy mathematics based teaching quality evaluation system", in *International Conference on Computer Application and System Modeling*, Taiyuan, China, 2010, pp. V11-146 - V11-149. DOI : 0.1109/ICCSM.2010.5623241
- [6] W. Li, C. Zou and C. Zhang, "One priority method for intuitionistic fuzzy complementary judgement matrix and its application", in *Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Shanghai, China, 2011, pp. 849 - 853. DOI: 10.1109/FSKD.2011.6019749
- [7] Y. Hezhan, "Application of Fuzzy Analytical Hierarchy Process in innovation education quality evaluation of higher education institution", in *International Conference on Information Management, Innovation Management and Industrial Engineering*, Sanya, China, 2012, pp. 438 - 441. DOI: 10.1109/ICIIM.2012.6339697
- [8] M. Goyal and R. Vohra, "Applications of Data Mining in Higher Education", *International Journal Of Computer Science Issues*, vol. 9, no. 2, pp. 113-119, 2012.
- [9] S. Yadav, B. Bharadwaj and S. Pal, "Mining Education Data to Predict Student's Retention: A comparative Study", *International Journal of Computer Science and Information Security*, vol. 10, no. 2, pp. 113-117, 2012. Retrieved from <http://sites.google.com/site/ijcsis/>
- [10] T. Sudha, "Total Quality Management In Higher Education Institutions", *International Journal of Social Science & Interdisciplinary Research*, vol. 2, no. 6, pp. 121-131, 2013.
- [11] J. Ruby and K. David, "Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms - A Case Study", *International Journal for Research in Applied Science & Engineering Technology*, vol. 2, no. 11, pp. , 173-179, 2014. Retrieved from <http://www.ijraset.com>
- [12] H. Hamsa, S. Indiradevi and J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm", *Procedia Technology*, vol. 25, pp. 326-332, 2016. DOI: 10.1016/j.protcy.2016.08.114
- [13] A. Nichat and A. Raut, "Predicting and Analysis of Student Performance Using Decision Tree Technique", *International Journal Of Innovative Research In Computer And Communication Engineering*, vol. 5, no. 4, pp. 7319-7327, 2017. DOI: 10.15680/IJIRCCCE.2017.0504144
- [14] P. Eswara Rao and S. Sankar, "Survey on Educational Data Mining Techniques", *International Journal Of Engineering And Computer Science*, vol. 6, no. 4, pp. 21034-21041, 2017. DOI: 10.18535/ijecs/v6i4.41

# Efficacy and efficiency of educational data mining through multistrategy machine learning

- [15] G. Kavitha and L. Raj, "Educational Data Mining and Learning Analytics - Educational Assistance for Teaching and Learning", *International Journal Of Computer & Organization Trends*, vol. 41, no. 1, pp. 21-25, 2017.
- [16] J. Mesarić and D. Šebalj, "Decision trees for predicting the academic success of students", *Croatian Operational Research Review*, vol. 7, no. 2, pp. 367-388, 2016.  
DOI: 10.17535/crorr.2016.0025