

# Student's Performance Analysis Using Machine Learning Tools

Atul Prakash Prajapati, Sanjeev Kr. Sharma, Manish Kr. Sharma  
Faculty of Engineering, Dayal Bagh Educational Institute Agra, U.P., India-282005  
National Informatics Center, Delhi, India

Dept of CSE, JB Institute of Technology, Dehradun, Uttarakhand  
[atulprakash21@gmail.com](mailto:atulprakash21@gmail.com), [sanjeevkr.sharma@nic.in](mailto:sanjeevkr.sharma@nic.in), [m\\_sharma17@rediffmail.com](mailto:m_sharma17@rediffmail.com)

## Abstract—

Several tools have been designed till today for the betterment and evaluation of student's performance. The results produced by these tools can help in decision making, that improves student's performance. This paper presents a survey of existing tools and techniques that have been designed in this area. This Paper uses a machine learning tool for analysing and predicting the results based on various factors that can improve the student's performance. This paper also suggests that cognitive modelling is a better way that can improve the decision making capability and it is useful for making quality software and tools for performance analysis.

**Keywords:** Student's performance analysis, Machine learning tools.

## I. INTRODUCTION

As we know in today's environment, there is a lack of quality education, and also the competition is increasing day

by day. So there is a need for quality steps to improve the standard of the students and education also. For this several philosophers provide time to time suggestions and standards for performance improvement. Still, the systems are lacking behind. So researchers had come to a conclusion that the technology can be an important factor for analysing the flaws that are present in the today's system, and why we lack behind. And also the use of technology makes decision-making process easy, as it can generate reports and graphs for analysis purpose. In the year 2016, Poza et al. [1] proposed teaching methodologies based learning tool. In this approach, the teaching/learning process should accomplish both knowledge assimilation and skill development. Previous works demonstrated that a strategy that uses continuous evaluation could meet both objectives. However, those studies did not evaluate and quantify the additional effort required to implement such strategies. This paper evaluates the additional instructor effort required when implementing continuous evaluation in a first-year Computer Fundamentals course in the Computer Engineering degree program at the Technical University of Valencia, Spain. In the year 2016, Elbadrawy et al. [2] proposed a matrix factorization and multi-regression approach based analyser to predict the student's performance. Initially, it was designed for analyzing e-commerce applications. But it can be used to analyse students' performance. It uses a degree planner, which predicts about the students who have very poor performance and may not be able to pass the course. It also forecasts about the future courses by analysing the past

performance. In the year 2015, Ganeshan et al. [3] proposed a web-based analysis system for advising and performance analysis of the students. This system uses techniques that are used in recommendation systems. It divides students into groups having similar features. When a new student comes, this system assigns him a group by

analysing his features and also offers him similar courses. It uses k-means clustering algorithm. In the year 2015, Lopez et al. [5] proposed a data mining approach based model for the academic attrition (loss of academic status) at the University of Colombia. Two data mining models were defined to analyse the academic and non-academic data; the models use two classification techniques, naive Bayes and a decision tree classifier, in order to acquire a better understanding of the attrition during the first enrolments and to assess the quality of the data for the classification task, which can be understood as the prediction of the loss of academic status due to low academic performance. The models aim to predict the attrition in the student's first four enrolments. First, considering any of these periods, and then, at a specific enrolment. Historical academic records and data from the admission process were used to train the models, which were evaluated using cross validation and previously unseen records from a full academic period. In the year 2014, Perikos et al. [6] proposed a data mining approach based performance analysis tool. It analyses the student's learning and produces the semantic rules that can be used further in analysing the overall performance of the student for that particular course. It uses the decision tree approach for the production of semantic rules. This system uses semantic web and ontology techniques for increasing the quality of study material. In the year 2014, Huang et al. [7] proposed a self-help training system for the

students of nursing course. This training system helps the student to learn the techniques of transferring a patient from bed to wheelchair. This system uses video and checklist method for demonstrating the skills. This system uses two Kinect sensors one for measuring posture of the trainee, and the other one is for the patient. In the year 2014, S. Bai et al. [8] proposed a performance evaluation system for analysing the performance of the faculty members. According to this paper, faculty performance directly affects the performance of the students. So for this, they used an ontology-based system that uses semantic web rule language for designing the semantic rules. For testing, the integrity of this system a sample dataset is used for the public sector university of Pakistan. In the year 2014, Cheng et al. [9] proposed a multi-touch puzzle game for the primary class students to teach them the basic geographical concepts. It has two scaffolding tools having different levels of difficulty, which can develop the understanding levels of students. In the year 2014, Kaur et al. [10] proposed a rule based expert system for performance analysis. This paper focuses on the essence of an analysis tool that can evaluate students' performance. Because individual interaction to the students is not possible in the degree institutes due to the large strength of students. So this paper focuses on the key factors that can affect the performance of a student. This analysis tool uses fuzzy rules for performance analysis. This analysis is performed based on five key factors, family issues, university environment, teaching methodology, university system and personal reasons. In the year 2014, Mei et al. [11] proposed a reference aid system. Several lists of formulaic sequences have been proposed, mainly for developing teaching and testing materials. However, their limited numbers and insufficient usage information seem unable to benefit formulaic language use. To address these issues we have developed GRASP, a reference aid for formulaic expressions, to promote learners' productive competence. Users are allowed multiword inputs to target their desired phrases or collocations. Utilizing natural language processing techniques, our system categorizes and displays the structures and sequences in a hierarchical way. The corresponding example sentences are also provided. The formulaic structures serve as a quick access index. The formulaic sequences and corpus examples illustrate the real-world language use. Importantly, automatic summarization from language data lends support to the idea of data-driven learning. A single-group pre-posttest design was adopted to assess the effectiveness of GRASP on 150 Chinese-speaking college freshmen. In the year 2014, Mustafa et al. [13] proposed a new methodological approach for the field integrating learning and education, with other research areas, such as neurobiological, cognitive, and computational sciences. Specifically, presented work is an interdisciplinary piece of research aiming to simulate appropriately a challenging and critical issue concerned with academic

performance in e-learning systems. Namely, considering face to face tutoring phenomenon observed while an interactive e-learning process is performed. Referring to strong interest announced by educationalists to know how neurons' synapses inside the brain are interconnected. Together to perform communication processing among brain regions. Herein, special attention has been developed towards the dynamical academic evaluation of timely based brain learning via face to face (FTF) interactive tutoring. In the year 2014, Chew Li et al. [12] proposed their management system to manage the students' records. Currently, even though there is a student management system that manages the students' records in University Malaysia Sarawak (UNIMAS), no permission is provided for lecturers to access the system. This is because the access permission is only to top management such as Deans and Deputy Deans of Undergraduate and Student Development due to its privacy setting. Thus, this project proposes a system named Student performance analysis system (SPAS) keep track of students' result in the Faculty of Computer Science and Information Technology (FCSIT). The proposed system offers a predictive system that can predict the students' performance in course TMC1013 System Analysis and Design, which in turns assists the lecturers from Information System department to identify students that are predicted to have bad performance in course TMC1013 System Analysis and Design. In the year 2014, Simpson et al. [14] proposed an evolution tool for performance analysis. Mathematics and physics courses are recognized as a crucial foundation for the study of engineering and often are prerequisite courses for the basic engineering curriculum. But how does a performance in these prerequisite courses affect student performance in engineering courses? This study evaluated the relationship between grades in prerequisite math and physics courses and grades in subsequent electrical engineering courses. Where significant relationships were found, additional analysis was conducted to determine minimum grade goals for the prerequisite courses. In the year 2013, Azzi et al. [15] focused on the role of experimental work in the field of education. This paper points out the basic need and essence of a laboratory in any engineering institute. But due to monetary problems not all institutes can afford the laboratory with the regular teaching course that affects severely the performance of the students. So as a solution they proposed the concept of Virtual Electronic Laboratories. This virtual laboratory concept uses a Bayesian Network tool for the performance assessment of the students. In the year 2013, Chen-Hsuan et al. [16] explained the importance of the scaffolding approach. This approach takes the previous knowledge of the students as input and produces the suggestions to improve the reading and concept building skills of the students. The basic aim of this approach is to improve the performance of each and every student of the class. For testing, this system 54 students was selected from an

undergraduate course. Further, they found that the learning skills of the students are improved by the use of this tool. In the year 2012, Doctor et al. [17] proposed a Fuzzy rule-based technique for improving students' performance by providing proper feedback. It monitors group as well as individual's performance during study related activity and provides the feedback. It also analyses the teaching strategy of the teachers. In the year 2012, Barney Khurum et al. [18] explained that how these two approaches 'oral feedback' and 'Rubrics'

can improve the performance of students. The statistical output produced by these approaches helps in decision making and in solving the grading issue.

## II. METHODOLOGY

This paper selects a data set. This dataset has a set of attributes having number's on (0-9) point scale, awarded based

Fig. 1. Schematic diagram of a data analysis and prediction model

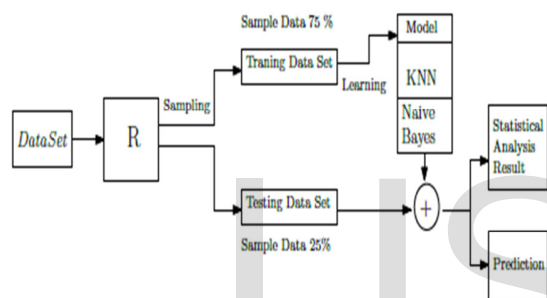


Fig. 1. Schematic diagram of a data analysis and prediction model.

TABLE I  
 TEST DATASET

Attendance	Assignment	Project	Lab-Performance	Seminar	Performance
0.6	0.8	0.4	0.0	0.0	OK
0.8	0.8	0.6	0.5	0.5	Good
0.3	0.0	0.6	0.8	0.6	Good

on the student’s performance in their respective subjects. After that, a group of experts has assigned a performance rating by considering the marks obtained by the students in the respective field. Experts have categorised students into two categories, ‘OK’ for students having poor performance ‘Good’ for the students who have a good performance. After that, a statistical analysis tool ‘R’ is used for analyzing this data set that will help in predicting future data sets. R divides the data set into two parts training data set and testing data set respectively. Finally, all the constraints have been taken into consideration. This thesis presents the logical step by step method to develop the project by using proper validation and testing, to meet the requirement of the project.

### III. EXPERIMENTAL DESIGN AND RESULT

In this paper, a sample dataset has taken from the records of Anand Engineering College, Agra. This dataset five factors (Attendance, Assignment, Project, Lab-Performance, Seminar) have been chosen that affects a student’s overall performance. Further a statistical analysis tool R has been chosen, for building a model data set. This data set is further divided into two parts (Test Data-Set, Training Data-Set). Following schematic table (Table-1, Table-2) represents the sample test data set and training data set respectively. Training and test dataset is divided in the ratio of (75 : 25) respectively. Finally the (Table-3) shows the result of Naive Bayes approach. It concludes that, error ration is (20 : 5). So the performance of the model is 75%:

TABLE II  
 TRAINING DATASET

Attendance	Assignment	Project	Lab-Performance	Seminar	Performance
0.3	0.7	0.0	0.9	0.9	OK
0.1	0.3	0.8	0.5	0.5	OK
0.0	0.2	0.1	0.2	0.7	Ok

### IV. CONCLUSION AND FUTURE WORK

The work can be concluded as follows. This paper covers all the objectives discussed above and full all the constraints.

TABLE III  
 NAIVE BAYES PREDICTION TABLE

Predicted	Good	Ok
Good	3	2
OK	3	17

The main concern of the teachers regarding the performance issue of the students is covered by introducing the analysis tool R. R is a statistical analysis tool that takes a data set as an input and builds a model that is used for predicting the future data set. This analysis tool R uses machine learning approaches like Naive Bayes, K-nearest neighbour first for building a model. Introduction section covers the basic review part, that discusses the tools and approaches that have been designed in this area. Section methodology describes the machine learning tool and the approach for choosing a data set. Section experimental design and result shows the schematic diagrams of tables that are used for performance analysis. Finally, this paper suggests that one should use the cognitive modelling for designing the knowledge-base. So that they can produce better results.

### REFERENCES

- [1]Poza-Lujan, Jose-Luis and Calafate, Carlos T. and Posadas-Yague. "Assessing the Impact of Continuous Evaluation Strategies: Tradeoff Between Student Performance and Instructor Effort", IEEE Transactions on Education, vol.59, pp.17-23, Feb 2016.
- [2]Elbadrawy, Asmaa and Polyzou, Agoritsa and Ren, Zhiyun and Sweeney. "Predicting Student Performance Using Personalized Analytics", IEEE, vol. 49, pp. 61-69, Apr.2016.
- [3]Ganeshan, Kathiravelu and Li, Xiaosong. "An intelligent student advising system using collaborative filtering", 2015 IEEE Frontiers in Education Conference (FIE), pp. 1-8, Oct. 2015.
- [4]Barney, Sebastian and Khurum, Mahvish and Petersen, Kai and Unterkalmsteiner, Michael and Jabangwe, Ronald. "Improving Students With Rubric-Based Self-Assessment and Oral Feedback." IEEE Transactions on Education, vol. 55, pp.319-325, Aug 2016.
- [5]Lopez Guarin, Camilo Ernesto. "A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining", IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 10, pp. 119-125, Aug 2015.

- [6]Grivokostopoulou, Foteini. "Utilizing semantic web technologies and data mining techniques to analyze students learning and predict final performance" 2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), pp. 488-494, Dec 2014.
- [7]Huang, Zhifeng and Nagata, Ayanori and Kanai-Pak, Masako and Maeda, Jukai and Kitajima."Self-Help Training System for Nursing Students to Learn Patient Transfer Skills" IEEE, vol. 7, pp. 319-332, Oct 2014.
- [8]Bai, Samita and Rajput, Quratulain and Hussain, Sharaf and Khoja, Shakeel A. "Faculty performance evaluation system: An ontological approach", 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), pp. 117-124, Nov 2014.
- [9]Cheng-Yu Hung, Cheng-Yu and Kuo, Fang-O and Sun, Jerry Chih-Yuan and Pao-Ta Yu, Pao-Ta. "An Interactive Game Approach for Improving Students Learning Performance in Multi-Touch Game-Based Learning" IEEE Transactions on Learning Technologies, vol. 7, pp.31-37, Jan 2014.
- [10]Kaur, Parwinder and Agrawal, Prateek. "Fuzzy rule based students' performance analysis expert system", 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 100-105, Feb. 2014.
- [11]Mei-Hua Chen, Mei-Hua. "An Automatic Reference Aid for Improving EFL Learners' Formulaic Expressions in Productive Language Use" IEEE Transactions on Learning Technologies, vol. 7, pp. 57-68, Jan 2014.
- [12]Sa, Chew Li and bt. Abang Ibrahim, Dayang Hanani. "Student performance analysis system (SPAS)", The 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M), pp. 1-6, Nov. 2014.
- [13]Mustafa, Hassan M. H. "Dynamical evaluation Of academic performance in e-learning systems using neural networks modeling (time response approach)", 2014 IEEE Global Engineering Education Conference (EDUCON), pp. 574-580, Apr 2014.
- [14]Simpson, Jane and Fernandez, Eugenia. "Student performance in first year, mathematics, and physics courses: Implications for success in the study of electrical and computer engineering", 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, pp. 1-4, Oct 2014.
- [15]Achumba, I. E. and Azzi, D. and Dunn, V. L. and Chukwudebe, G. A. "Intelligent Performance Assessment of Students' Laboratory Work in a Virtual Electronic Laboratory Environment." IEEE Transactions on Learning Technologies, vol. 6, pp. 103-116, Apr 2013.
- [16]Chen, Hsuan-Hung and Chen, Yau-Jane and Chen, Kim-Joan. "The Design and Effect of a Scaffolded Concept Mapping Strategy on Learning Performance in an Undergraduate Database Course" IEEE Transactions on Education, vol. 56, pp. 300-307, Aug 2013.
- [17]Doctor, Faiyaz and Iqbal, Rahat. "An intelligent framework for monitoring student performance using fuzzy rule-based Linguistic Summarization." 2012 IEEE International Conference on Fuzzy Systems, pp. 1-8, Jun 2012.
- [18]Barney, Sebastian and Khurum. "Improving Students With Rubric-Based Self-Assessment and Oral Feedback", IEEE Transactions on Education, vol. 55, pp. 319-325, Aug 2012.