Original Research Article

Knowledge Based PerformanceEvaluation and Predictive Model for Undergraduate Students

Abstract:

In educational data mining, the process of analysing and predicting from a pool of acquired data is a big challenge due to the influence of behavioural, environmental, parental, personal and social traits of students. Existing education predictive systems have used patterns generated from mined common factors to predict student performance based on subject, faculty, and grade amongst others while neglecting explicit traits, which defines a student. Such models are general to one whereas predictive features are common but unique to students. Here, a self-academic appraisal and performance predictive (SAAPP) system was developed to analyse and predict the overall performance of students before the expiration of the course duration. The inherent knowledge driven model analyses common available performance predictive internal and external factors, with probabilistic analysis of student academic history and pending courses. The system builds a personal data centric system for individual student through a decision support expert system and a probabilistic optimal grade point analysis for more effective recommendation. The proposed system is more accurate, reliable and precise in student performance classification withtargeted recommendations.

Keywords: Predictive Analytics, Student Performance Evaluation, Educational Data Mining, Recommender System

1. Introduction

Over the years, many researches have been carried out to find out factors affecting students' performances at different institutions of learning. Summarily, the factors could be social, economic, psychological, environment and personal [1]. No doubt, the propensity of these factors varies from one individual or institution of learning to another. While considering the benefits of student performance evaluation to teaching and learning, the internal factors, which involves student and his relationship with the learning environment, alongside external factors, which refers to what happens to the student outside the learning environment [2] has long been regarded as an important research area for student academic performance prediction. Thus, several research questions in education seek to predict student performance centred upon a set of self-regulating variables, these variables may include high school information, background information, parent information, or scores on previous test [3]. Some other approach concentrate on estimating the final grade of a student why using the current available result. Thus, predicting student results is really a process of trying to determine the eventual academic success or risk of a student in an institution [4]. Predominantly, the development of predictive system for students' performance evaluation has been through the data mining techniques. Data mining techniques are used to extract useful information and patterns from educational database in predicting student's performance [5]. This process plays a key role in learning analytics or educational data mining. Educational data mining (EDM) as an emerging discipline focuses on applying data mining tools and techniques to educationally related data. [6]. The EDM Classification is used to categorize the students in order to shape their learning styles and inclination. This process seek to find ways to make advantageous use of the enlarging amount of data about learners to understand the process of learning and the social and motivational factors surrounding learning. While several approaches through data mining have focused onpattern analysis to define and predict student's performance for effective instructional interventions by the instructor to the student [7], to the best of our knowledge, no system exist for student's self-appraisal and evaluation towards performance prediction. Hence, in the scope of this research, by trying to understand distinct learners, a computational approach that combines data and knowledge inherent therein is used to evaluate a student performance towards a more accurate prediction. Thus, in section 2, related works in educational data mining is presented, in section 3; the developed self-appraisal model is present with summarized experiment and evaluation in section 4. In section 5, the research is concluded.

2. Related works

Since Educational Data Mining (EDM) has emerged as a research area in recent years for researchers from different and cognate research areas all over the world, its processing application does not differ much from other areas of Data Mining, like business, genetics, medicine, etc. [8]. The data used in the mining exercise may be personal, academic or both. The essence is to understand student's behaviour, assist instructors, improve teaching, social infrastructures and program, evaluate and improve curriculum and learning processes. The application of data mining in education as discussed [9] is used to extract meaningful information from huge data set for decision making processing through analytical tool views. The dataset always contain different influencing factors that determines the performance class of a student after mining. [10] presented a study that evaluates the impact of students' attendance in class, family income, parent's level of education, availability of trained teachers in school, sex of the student and the distance of the student home from school as factor to determine student performance. [11] also identified interruption of electricity supply, overcrowded lecture rooms, unfavourable learning environment, incessant strike, and closure of school among others as institutional factors that affect performance of students in some Nigerian Universities. While staffing, teaching and learning materials, motivation, attitude of teachers were identified as institutional factors that determine students performances in mathematics [12], learning preference, age, gender and entry qualification was also seen as a factor that affects students' performance [13]. With researchers having these factors in mind among others, a number of models have been developed to predict student performance. A data mining approach has been adopted and applied by [14] to discover students' performance models in supervised and unsupervised assessment tools of a course in an engineering degree program. In addition, a validated set of mathematical models to predict student academic performance in engineering dynamics was developed by [15]. The study also showed that radial basis function (RBF) network models and support vector machine models have better generalizability than multiple linear regression (MLR) models and multilayer perceptron (MLP) network models. By using association rule mining, [16] suggested a method of evaluation of student's performance. [17] proposed a framework for predicting students' academic performance of first year bachelor students in Computer Science course. The students' demographics, previous academic records, and family background information were the factors considered for the study. Decision Tree, Naïve Bayes, and Rule Based classification techniques were applied to the students' data in order to produce the best students' academic performance prediction model. [18] applied different data mining algorithms on preoperative assessment data to predict students' success in a course. From the research, it was discovered that Naïve Bayes classifier outperforms prediction tree and neural network model. Another comparative study on the precision of Decision Tree and Bayesian Network algorithms for presaging the academic performance by [19]. In this analysis, it was discovered that the Decision Tree is 3-12% more precise than the Bayesian Network. In addition, a Multi Agent Data Mining wasproposed by {20] to predict the performance of the students based on their data with high precision of prognostication and provide an aid to the weaker students by optimization rules. However, the existing methods have been applied on a collective dataset of all students wherein the unique or distinct behaviour of a student alongside his academic history among others are relatively not considered. Consequently, in section 3, the self-appraisal and performance predictive model is presented.

3. The Developed SAAPP Model

No doubt, a modern day higher institutionmust have systems to store student's information.These databases always contain useful knowledge that can be extracted for effective predictions and decision-making. However, for students' performance evaluation and prediction, most of the used dataset hides detailed information about key performances indicators of students with available dataset mostly relying on previous academic performance. Since, certain predictive factors like family size, gender, productive time, food habit, academic strength and weaknesses are confidential and sensitive attributes, they are not often allowed access for performance prediction processing. Here, the developed SAAPP model for student's self-appraisal and performance prediction is a dual but parallel level prediction and performance evaluation algorithm. Initially, the predictive metrics from the student are fed into a rule based decision support expert system to determine the performance class of a student. The developed SAAPP is domain-specific, which depends on knowledge base and reasoning algorithm. In other to enhance decision-making, its knowledge base consisting of IF-THEN production rules, which is used on the predictive metrics supplied by students to define their performance class. Answers to each of these corresponding metrics, whose samples are presented in table 2 are weighted on a scale of one (1) or five (5) in order to obtain the likelihood performance class of a student.In the process, the identification number of the student is used to retrieve the academic history of a student before a probabilistic calculation is done on the remaining courses to graduation while using the present cumulative grade point average (CGPA) as bases.Thus, a predictive CGPA is obtained through Grade Point (GP) randomization for different Course Unit (CU) using the equation 1;

$$cgpa = \frac{CTQP + RQP}{CTCU + RCU} \quad (1)$$

Where, $CTQP$ is the current Total Quality Point, RQP the Remaining Quality Point for the student, $CTCU$ is the current Total Credit Unit while RCU is the Remaining Credit Unit. Here, Quality points (QP) are derived by multiplying the Credit Units (CU) for a course by the Grade Point (GP) earned by the student in that course. In table1, marks and grade point are represented.

Table 1: Marks and Grade Point Analysis

Mark/Score	Grade Point (GP)
70 – 100	4
60 – 69	3
50 – 59	2
44 – 49	1
0 –44	0

E.g. If a student earns a $\frac{CTQP + RQP}{CTCU + RCU}$ 12 $\frac{RQP}{RCU}$ the most optimal $CTQP$ is the Cu 13 RQP is the R

CGPA = (40 + 45) / (12 + 13) = 3.4

At the end, the most optimal grades that will earn a student the most excellent graduation point is presented as output alongside its recommendation. In figure 1, theearlier processes are further illustrated.

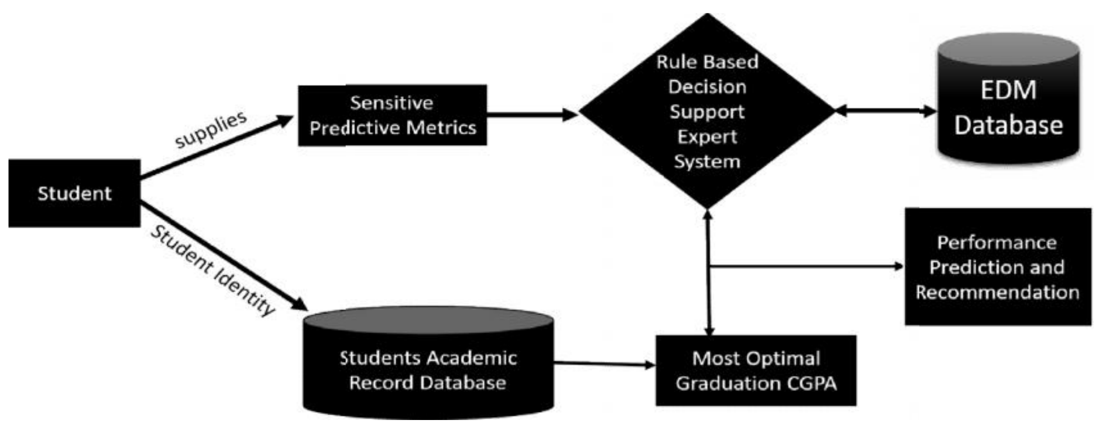

Figure 1: SAAPP model

From Figure 1, each student supplies the required confidential academic predictive metrics through the mobile application. These predictive metrics are categorised into emotional, environmental and learning factors. Here, sample of some of the metrics and possible values that a student can supply is presented in Table 2.

Table 2: Sample predictive Metrics and Possible Values

VARIABLES	DESCRIPTION	POSSIBLE VALUES
Sex	Student gender	Male or female
Fh	Food habit of the student	Picky, not picky; Once, twice, thrice or more
Fs	Family size of the student	1, 2, 3, 4, >4
Ps	Parent status of the student	Divorced, separated, single parent, married
Pai	Parent annual income	Very poor, poor, medium, high
Meq	Mothers education qualification	No form of education, primary education, secondary education, higher education
Feq	Fathers education qualification	No form of education, primary education, secondary education, higher education
Conc	Level of concentration	20%, 40%, 60%, 80%, 100%
Retn	Level of Retention	20%, 40%, 60%, 80%, 100%
Comp	Level of Comprehension	20%, 40%, 60%, 80%, 100%

The scores obtained from key confidential performance factors from student alongside the academic performance predictive processing are further subjected to If then rules in order to make concise recommendation for student. Each rule takes the form of: IF <requirement> THEN <outcome>. Where requirement describes the predictive characteristics of student's performance class, the outcome represents the most suitable recommendation for each class of student based on the total score obtained after the analysis via the educational data mining (EDM) database. The EDM databases contains patterns, which have been trained over a period via machine learning. It also has the capacity to adapt to changing parameters.Overall, the developed system provides mobility for student's self-appraisal, evaluation, academic result prediction and possible recommendation. Unlike existing systems, which are not available to student but mainly used for general performance evaluation, predictions and decision-making. With the accurate estimation of students' grades, being depended on the present result and its respective future courses, in the selection of next term courses, here, the method relies on the performance that the students achieved in previously taken courses. Although, for the purpose of recommendation, a random calculation for the most optimal possible grade at the end of course duration is also performed. A unique aspect of the model is that the obtained results are specific to each student-course tuple, which creates a personalized degree pathway to facilitate successful and timely graduation.

4. Experiment and Evaluation

By considering the predictive metric alongside academic record of randomly selected students, we experimented to discover the accuracy rate for the developedpredictive model, here; accuracy is defined in terms of the most optimal potential obtainable final CGPA. In addition, the role of predictive factors on the academic performance of students is considered. However, initially, the usability degree for which the developed mobile application can be used to achieve quantified objectives is determined. The outcome is presented in Figure 2

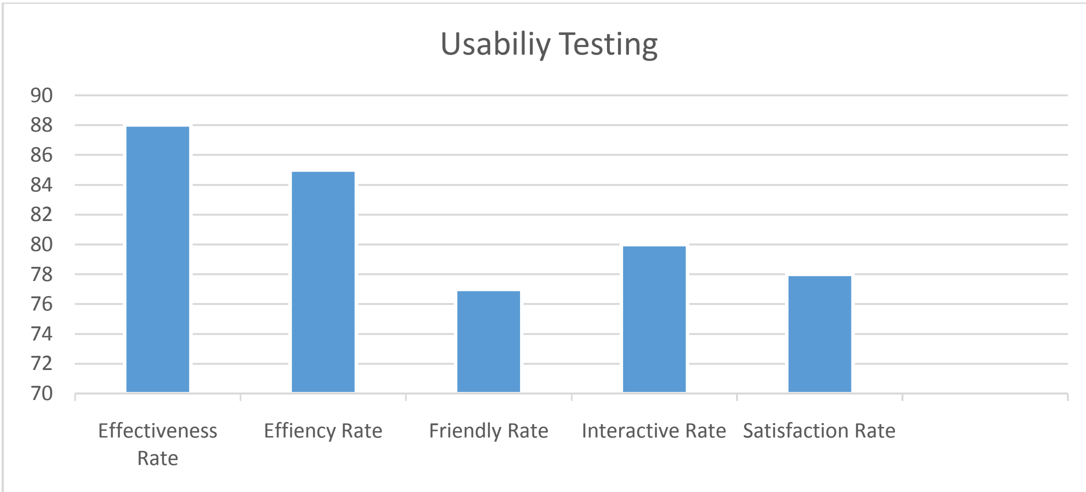

Figure 2: Usability Ratings

Based on randomly selected number of students across levels, in table 3, we presented the current average CGPA, average percentage increase and drop in CGPA for our respondents throughdata analysis.

Table 3: Average student performance rating

Level	Percentage Number of Student	Average CGPA	Average % CGPA Increase	Average % CGPA drop
100 L	55	2.89	86	14
200 L	43	3.0	68	32
300 L	62	2.58	43	57
400 L	50	2. 35	29	61

Based on the result illustrated in Table 1, 210 students were crossed examined. The accuracy rate for the proposed algorithm in relevance to the predictive metrics as presented in table 2is 84.5%.

Later, we evaluated the influence of Emotional Factors, Learning Factors, and Environmental Factors on student academic performance based on the selected predictive metrics. The result obtained as illustrated in figure 3 shows that emotional factors, with 64%, hasmuch influence on the majority of our respondent while learning factors follows with 23% while the environmental Factor (13%) is the least.

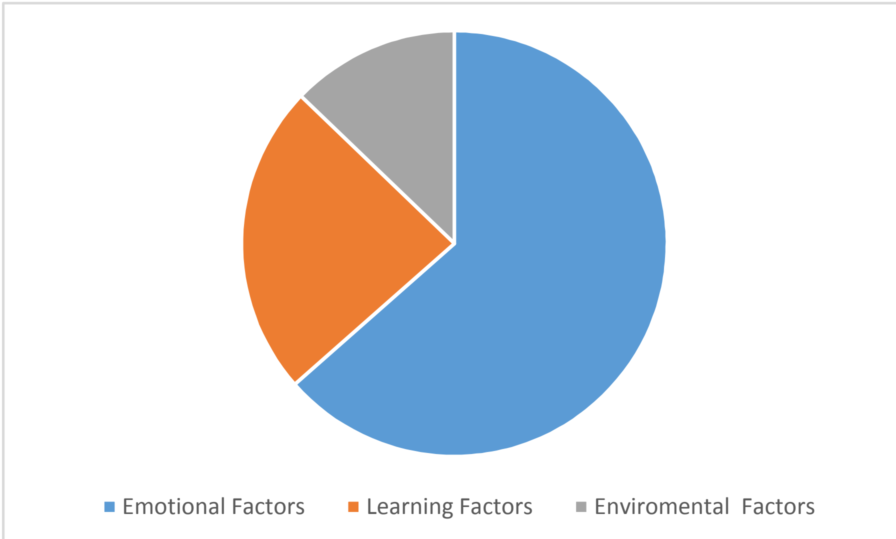

Figure 3: Predictive Metric Influence Rate on Students

5. Conclusion

With the rapid increase of data in educational environment, educational data mining as emerged as a developing tool for analysing the unique types of data that come from educational settings. Here, the educational data mining students' performance is predicted based on the confidential criteria and academic record, using a rule based decision support expert system and mathematical distribution. The focus is to go beyond descriptive statistics and reporting on what has happened through collective data mining to building a personal data centric system for individual student through a decision support expert system and a probabilistic optimal grade point analysis. Overall, by combining predictive metrics with student academic records of accomplishment, an effective recommendation can be provided to students across levels. This will serve has guide to students on how to achieve excellence within the course duration. In future, we hope to apply ensemble machine learning models on acquired dataset.

Reference

1. Mushtaq, S. N. K. (2012). Factors affecting students' academic performance. Global Journal of Management and Business Research, 12(9).
2. Ogunsola, O. K., Osuolale, K. A., & Ojo, A. O. (2015). Parental and Related Factors Affecting Students' Academic Achievement in Oyo State, Nigeria. World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 8(9), 3129-3136.
3. Shahiri, A. M., & Husain, W. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, *72*, 414-422.

4. Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. Frontline Learning Research, 1(1), 42-71.
5. Khan, I. A., & Choi, J. T. (2014). An Application of Educational Data Mining (EDM) Technique for Scholarship Prediction. *International Journal of Software Engineering and Its Applications*, *8*(12), 31-42.
6. Sonali Agarwal, G. N. Pandey, and M. D. Tiwari, "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012
7. Kannammal, J. Macklin Abraham Navamani (2014), Predicting Performance of Schools by Applying Data Mining Techniques on Public. Research Journal of Applied Sciences, Engineering and Technology , pp. 262-271.
8. Mohammed M. Abu Tair, Alaa M. El-Halees. (2012), Mining Educational Data to Improve Students' Performance:. International Journal of Information and Communication Technology Research, pp. 140-146.
9. v. Sarala, Dr. v.v. Jaya Rama krishnaiah, (2015) "Empirical study of data mining techniques in education system", international journal of advances in computer science and technology (ijacst), vol. 4 no.1, pages: 15 –21 2015
10. Raychaudhuri, A., Debnath, M., Sen, S., & Majumder, B. G. (2010). Factors affecting students' academic performance: A case study in Agartala municipal council area. Bangladesh. E-Journal of Sociology, 7(2), 34-41.
11. Christiana, O. O. (2014). Institutional Factors Affecting the Academic Performance of Public Administration Students in a Nigerian University. Public Administration Research, 3(2), 171.
12. Mbugua, Z. K., Kibet, K., Muthaa, G. M., & Nkonke, G. R. (2012). Factors contributing to students' poor performance in mathematics at Kenya certificate of secondary education in Kenya: A case of Baringo county, Kenya.
13. Mlambo, V. (2012). An analysis of some factors affecting student academic performance in an introductory biochemistry course at the University of the West Indies. The Caribbean Teaching Scholar, 1(2).
14. Anwar M. A., Naseer Ahmed, "Information Mining in Assessment Data of Students' Performance", ISSN: 2319-5967 ISO 9001:2008 Certified International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 1, Issue 2, November 2012
15. Huang, S. (2011). Predictive modeling and analysis of student academic performance in an engineering dynamics course.
16. Suchita Borkar, K. Rajeswari. 2013, Predicting Students Academic Performance Using Education Data Mining. International Journal of Computer Science and Mobile Computing, pp. 273-279
17. Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The Prediction of Students' Academic Performance Using Classification Data Mining Techniques. Applied Mathematical Sciences, 9(129), 6415-6426.

18. Osmanbegović, E., & Suljić, M. (2012). Data mining approach for predicting student performance. Economic Review, 10(1).
19. Nguyen Thai Nghe Paul Janecek and Peter Haddawy, "A Comparative Analysis of Techniques for Predicting Academic Performance ",37th ASEE/IEEE Frontiers in Education Conference 2007
20. Abdullah AL-Malaise, Areej Malibari, and Mona Alkhozae, "Students' Performanceprediction System Usingmultiagent Datamining Technique ", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.4, No.5, September 2014