ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 10(24), DOI: 10.17485/ijst/2017/v10i24/110791, June 2017* ISSN (Online) : 0974-5645

Students' Performance and Employability Prediction through Data Mining: A Survey

Tripti Mishra[1*], Dharminder Kumar[2] and Sangeeta Gupta[3]

[1]MewarUniversity, Chittorgarh – 312901, Rajasthan, India;
mishratripti2007@gmail.com
[2]Department of Computer Science, G. J. University, Hisar – 125001, Haryana, India;
dr_dk_kumar_02@yahoo.com
[3]Guru Nanak Institute of Management, West Punjabi Bagh – 110026, Delhi, India;
sangeet_gju@yahoo.com

Abstract

Objective: To systematically review the work done in the field of academic performance prediction and employability prediction of students in higher education. **Methods:** The survey first explain show higher education has become an exciting field of research and why the prediction of academic performance and employability is beneficial for the institutions. We also explain briefly in how many ways higher education is being provided world-wide. Then we discuss the work done in both the areas of prediction. **Findings:** The survey explores existing research highlights and finds that prediction of academic performance has progressed a lot but employability prediction is yet to mature. **Application:** It further suggests few parameters that have not been considered so far in predicting the performance or employability.

Keywords: Academic Performance, Data Mining, Employability, Higher Education, Prediction, Survey

1. Introduction

In the present knowledge-based epoch, education plays a major role in the progress of a nation's economy and development. It assures to invigorate the country by contributing the reliable and quality workforce to the society. Higher education is the foundation for fostering the talent, the key factor in increasing national human capital quality, and the main way to upgrade a nation's competitive status. Thus, the research on development of higher education is an important work and is actually required. To get an edge over each other, institutions are applying cutting edge technologies like data mining on the huge data generated in class room including academic, behavioral, demographic data of students and faculty data as well. The data generated in educational set up can give deep insight into educational process. Educational Data Mining (EDM) refers mining the data generated in educational set up.

Researches in EDM have benefited the educational setups tremendously.

The review is organized chronologically and categorically to offer insight on how past research efforts laid the groundwork for subsequent studies, including the present research efforts. The detailed review has been carried out, so that; the present research can properly be tailored to add the present body of literature, as well as the scope and direction of the present research effort.

In order to give a comprehensive view of the work done in the field of EDM, many survey papers have been published. The most cited research[1] compiled the work done till 2005 and the other that has discussed vital features of EDM[2]. The work done in EDM till 2010, 2013, 2014 has been complied in the different research papers[3–5] (Table 1).

According to the literature study, EDM research pertains mainly three heads, according to the way data is collected-

- Traditional face to face or the offline education system based on data generated in the classroom.
- E-learning in which the learning is provided through online content based on online activity logs.

* *Author for correspondence*

- Intelligent tutoring system (ITS) and Adaptive Educational Hypermedia System (AEHS) involve online teaching based on students need, his or her progress rather than providing same structured lesson to all the students.

Table 1. Survey Papers in Educational Data Mining

References	Highlights
[1]	Comprehensive survey for traditional educational systems.
[2]	Highlights the work done in EDM from the view point of models, emergence of public data, tools.
[3]	Illustrates problems of education system resolved using data mining and proposed the association of techniques for better predictions.
[4]	Recognized different educational system tasks, discipline, techniques, and algorithms.
[5]	Provides a comprehensive survey, a travelogue from 2002 to 2014 for educational data mining.

E learning and Intelligent Tutoring systems have used EDM mainly to model online behavior of students, track their performance and get feedback from them.

However, this survey explores traditional educational settings as it is still most widely used method of teaching across the globe.

2. Traditional Educational Setup

Traditional educational set up refers to class room teaching which is still the most popular form of teaching by the Institutions across the globe and needs to be explored. Students prefer to take admission in an Institution with high academic performance and where the passing out students have better employability. Prediction of both academic performance and employability can help the management identify students at risk of poor academic performance and low employability. The process of prediction involves application of various data mining algorithms, to predict the dependent variables based on independent factors.

Next two subsections discuss the researcher's inclination towards prediction of both students' performance and their employability. The focus is on the attributes considered and methods/classification algorithms adopted for prediction of performance and employability prediction.

3. Prediction of Students' Performance

Students' academic performance is a mature field now with many researchers contributing to it. Moving in chronological increasing order from 2007 onwards the researchers have considered various parameters for prediction and using different algorithm they have tried to predict the result of specific course of a particular university.

Authors[6] have considered students from two different institutions, one International and one small institution, to check the effect of research separately on both. In both the cases decision tree provided better accuracy than Bayesian Network.

A simple student performance assessment and monitoring system based on various data mining techniques was developed with the predictor attributes including students' demographic details, course average score in 1st to 5th semester overall gain performance, etc[7]. Decision Tree C5 showed highest accuracy, followed by Classification and Regression, Trees (CART), Artificial Neural Network (ANN), Chi-Squared Automatic Interaction Detection (CHAID).

Taking more factors like university matriculation exam, GCE (General Certificate of Education) Score, Senior Secondary Certified Examination score SSCE), grades in O level subject, location of University from home, gender and age, Cumulative Grade point was to classified as Good, Average and Poor. Artificial Neural Network (ANN) was used for prediction and 74.5 %accuracy was attained in performance prediction[8].

Prediction of school students' performance has been considered with a total of 33 parameters including socio -demographic details like (parental marital status, father's job, mother's job, quality of family relationship, attitude towards study (No. of hour, past failure) Internet facility, family support, free time after school, health, alcohol consumption etc[9]. To predict the performance, Decision was used. Attendance, parents job, previous year performance was found to be the key factors that affect the current achievement.

The study[10] uses feature selection process of Waikato Environment for Knowledge Analysis (WEKA) tool which has inbuilt set of methods and considers Student's gender, Eyesight, Community, Physical handicap, food habit, family details, mode of transport, medium of

instruction, sports activity etc. as predictive factors.It was also observed that classification methods like Naïve Bayes, one R voted perception performed much better with feature selected subset than where all variables were considered.

Voting technique is a method of applying more than one algorithm in succession and then taking the best result. The researchers[11] have concluded that Decision Stump along with Hidden Naïve Bayes is most suited for academic performance prediction in case of a New Zealand Polytechnic, where, apart from academic performance, demography, disability was also considered as predictive factors. In a similar research[12] CART algorithm has been found to give highest accuracy. Further, the study[13] added few more predictive parameters like ethnicity, and student's current job condition to predict performance. In this case tree was found less accurate than regression and analysis.

Chi Squared Automatic Interaction Detector (CHAID) was used make high school result prediction with predicting factors being taken as health of the student, tuition availability, facilities to study at home etc. Prediction accuracy was not very good 44.69% and key influences were found to be mother's education, location from school etc[14].

The study[15] aimed to discover individual characteristics that decide their success using Microsoft Decision Tree.11 attributes that includes registered information high school information; Turkish, University entrance exams degree and University placement info family living conditions and financial status etc. were considered. Microsoft Decision Tree (MDT) has been used to predict GPA with just two categories successful and unsuccessful.

Four classification methods Artificial Neural Network, Decision trees, Support Vector Machine and logistic regression along with ensemble techniques (i.e. bagging, busting and information fusion) to analyze 16,000 students records[16]. 39 variables including demographic data, TOEFL, SAT Score, Loan etc. were used to predict attrition/ retention. Support Vector machine provided Highest Accuracy followed by decision tree, Artificial Neural network and regression.

In another research paper attribute importance was emphasized by ranking them, using correlation based feature subset selection and consistency subset selection (COE) and using them further find accuracy of various classifiers[17]. Unlike other researchers, authors do not consider demographical details of the students, but concentrate only on grades of various courses as attributes to predict categories of students as first class, the other consisting second class upper lower and third class. In this case Naive Bayes was found to show better accuracy than Decision Tree. In a contrast authors[18] have used previous academic achievements in exams and present academic assessment in Lab and class assignments to predict end semester marks, whereas another research paper[19] has made academic performance prediction of Engineering Drawing course. SVM is well suited for individual students' performance prediction and regression is to be used for prediction of entire class.

In another study[20] language proficiency in English, credit selection and whether a student is unmarried, married, divorced have shown their effect on academic performance prediction and decision tree showed better accuracy over neural network.

The research[21] used various factors that were ranked according to their effectiveness in predicting the placement result of Turkish secondary school. The predictors included marks Spirituality and morals, Turkish language and entry level exam. Decision tree worked well with the problem.

Demographic variable, score of high school entrance exam and attributes related to their attitude towards studying etc. were considered for predicting the result of 1st year students of economics course[22]. A set of four algorithms were applied. It was found that past academic records, entrance exam marks, hours put into studies are having maximum impact on prediction accuracy.

Students' academic performance in five categories were predicted[23] using Decisions Tree, Functions, Rules, Bayes Net etc. and Random tree provided the best result. In a more recent study[24], factors considered were demographic profile, previous academic scores, entrance exam result and among the various algorithms applied j48 gave highest accuracy of prediction.

In yet another research paper[25] at school level performance was predicted using Parental status, Mode of transport, Groups of subjects, type of school, previous marks and applying algorithms decision tree, KNN, SVM etc.

Academic performance is always one of the primary predictors of employability. Thus all the factors affecting academic performance also affect the employability. Next we explore the work done on employability.

4. Prediction of Students' Employability

Today, the reputation of an Institution is judged by its academic success, its ability to retain students and to provide employment for its students. The term "Employability" still has no precise definition. Employability has been described in many ways, like, the ability to secure a job, getting a job within a specified time period after graduating, the ability to skill map oneself according to the job need, or the willingness of the student to extend the graduate learning at work [26]. Alternatively, employability is defined as the ability of students to secure a job during on campus placements[27]. Research in employability prediction is in nascent stage. It mostly involves identification of skills or attributes required from the perspective of employers and is obtained from employer through questionnaire and interviews. Mostly statistical methods have been applied and research is more of descriptive than predictive.

The importance of psychological factors along with personal and organizational awareness has been emphasized in a report by Higher Education Academy with the Council for Industry and Higher Education (CIHE) in United Kingdom[28].another study considered the effect of working environment on performance of the employee[29].

Whether an employee will be able to meet the expectation and should be hired is predicted in research paper[30] which is based on few attributes extracted from candidate's curriculum vital key words in application and interview.

It has been concluded that as the paradigm is shifting from product based to service based industry especially in Information Technology, and hence the curriculum and method of delivering lecture must evolve in order to enhance employability[31].

A descriptive study in the research[32] indicates that employers look forward to employees with Personal attributes that include loyalty, commitment, honesty, integrity, enthusiasm, reliability, personal presentation, common sense, positive self-esteem, and a sense of humor, motivation, adaptability, a balanced attitude towards work and home life and ability to deal with pressure.

Correlation and ordinal regression has been used to conclude that a students' non-technical education consisting of reasoning, logical ability, and soft skills were stronger predictor of their employability than their technical education consisting of their academic performance[33]. Poor English language competency is found to be a major reason for the low employability[34]. This has been supported by one more study[35] which concludes that knowledge of GPA and English Language competency are required for the students in software industry to continue with their employment and also the female candidates were better performer than male candidates in campus placement drives.

Psychology[36] seeks an answer to the kind of skill requirement which is essential for enhancing job prospects. A sound effect of overall personality of the employee along with career competence, self-efficacy is seen on employability. This has been supported by the study[37] where employers give priority to soft skills, team work, etc.

Another researcher[38] concludes that J48 is most suited algorithm to predict the employability based on demographic profile, employees job satisfaction, academics etc.

It is comprehensible that not much work has been done in the direction of employability prediction mainly due to lack of authentic data, hence this study of employability prediction and model development will contribute significantly in educational data mining.

Graduate employability has been the subject of little empirical research. There are a number of difficulties in defining and measuring graduate employability, which means that there is a paucity of research that looks at its predictors and outcomes. Previous work has proposed that emotional competence improves graduate employability. Researchers in psychology have shown that the emotional skills of a student are also important factors for performance prediction. However not much work has been done in the field of EDM to validate the existing knowledge or construct new knowledge about emotional skill being predictor of performance. One reason for not considering emotional skill is lack of authentic data. The need of hour is to construct authentic primary data that has factors of academic integration and social integration and emotional skills and study the prediction of academic performance and employability in tandem. The authors of this paper have identified this gap and published two papers in this regard where apart from social and academic integration, emotional skills like leadership, self-esteem, empathy; decision making

capability, time management and stress management are also considered as predictor variables. A model was derived for performance prediction[39] based on academic, social and above emotional skills. Similarly, Employability prediction model has been developed by the authors[40] using authentic primary data that involves social, academic as well as emotional skills. Future work requires developing tools for prediction of performance and employability.

5. Conclusion

This paper discussed the work done in educational data mining categorically in traditional education. Within each category the works are again discussed chronologically. We consider research areas of traditional education, as our study involves traditional education.

In traditional education, performance prediction is in matured state with contribution from many researchers. However, there is paucity of research in the field of employability prediction. As both performance and employability of students graduating from an institution decide the market value of the institution, research is required to develop comprehensive models for performance and employability tool and develop a system that will be able to predict both performance and employability. From the literature review, it is clear that most commonly used predictors are socio economic / demographic profile and past academic record of the students. Apart from this, number of hours dedicated to studies, distance of the institution from home, loan, internet facility etc. has been considered by the individual researchers in their studies. Thus in general researchers in the field of EDM have focused on academic and social integration of students for performance and employability prediction. The effect of emotional skills on academic performance and employability needs to be explored further. The future work includes survey of tools, available for prediction of academic performance and employability.

6. References

1. Romero C, Ventura S. Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2010 Nov; 40(6):601–18. Crossref
2. Baker RS, Yacef K. The state of educational data mining in 2009: A review and future visions. JEDM-Journal of Educational Data Mining. 2009 Oct; 1(1):3–17
3. Romero C, Ventura S. Educational data mining: A survey from 1995 to 2005. Expert systems with applications. 2007 Jul; 33(1):135–46. Crossref
4. Pe-a-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications. 2014 Mar; 41(4):1432–62. Crossref
5. Thakar P. Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. International Journal of Computer Applications. 2015 Jan; 110(15):60–8.
6. Nghe NT, Janecek P, Haddawy P. A comparative analysis of techniques for predicting academic performance. 37th Annual Frontiers in Education Conference-Global Engineering: Knowledge without Borders, Opportunities without Passports 2007 Oct, T2G-7, 2007.
7. Ogor EN. Student academic performance monitoring and evaluation using data mining techniques. Electronics, Robotics and Automotive Mechanics Conference, 2007 Sep. p. 354–9. Crossref
8. Oladokun VO, Adebanjo AT, Charles-Owaba OE. Predicting students' academic performance using artificial neural network: A case study of an engineering course. The Pacific Journal of Science and Technology. 2008 May; 9(1):72–9.
9. Cortez P, Silva AM. Using data mining to predict secondary school student performance. 2008; 1–8.
10. Ramaswami M, Bhaskaran R. A study on feature selection techniques in educational data mining. ArXiv preprint arXiv. 2009 Dec; 1(1):7–11.
11. Paris IH, Affendey LS, Mustapha N. Improving academic performance prediction using voting technique in data mining. World Academy of Science, Engineering and Technology. 2010 Feb; 62:820–3.
12. Kovacic Z. Early prediction of student success: Mining students' enrolment data. 2010; 647–65.
13. Kovačić ZJ, Green JS. Predictive working tool for early identification of 'at risk'students. 2010; 1–73.
14. Ramaswami M, Bhaskaran R. A CHAID based performance prediction model in educational data mining. ArXiv preprint arXiv. 2010 Feb; 7(1):1–9.
15. Guruler H, Istanbullu A, Karahasan M. A new student performance analysing system using knowledge discovery in higher educational databases. Computers & Education. 2010 Aug; 55(1):247–54. Crossref
16. Delen D. A comparative analysis of machine learning techniques for student retention management. Decision Support Systems. 2010 Nov; 49(4):498–506. Crossref
17. Affendey LS, Paris IH, Mustapha N, Suleiman MN, Muda Z. Ranking of influencing factors in predicting students' academic performance. Information Technology Journal. 2010; 9(4):832–7. Crossref
18. Baradwaj BK, Pal S. Mining educational data to analyze students' performance. ArXiv preprint arXiv: 1201.3417. 2012 Jan; 2(6):63–9.
19. Huang S. Predictive modeling and analysis of student academic performance in an engineering dynamics course. 2011; 1–136.

20. Cheewaprakobkit P. Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program. Proceedings of the International Multi Conference of Engineers and Computer Scientists. 2013; 1:13–5.
21. Şen B, Uçar E, Delen D. Predicting and analyzing secondary education placement-test scores: A data mining approach. Expert Systems with Applications. 2012 Aug; 39(10):9468–76. Crossref
22. Osmanbegović E, Suljić M. Data mining approach for predicting student performance. Economic Review. 2012 May; 10(1):3–12
23. Shah NS. Predicting Factors That Affect Students' academic Performance by Using Data Mining Techniques. Pakistan business review. 2012 Jan; 631(68):631–810.
24. Kabakchieva D. Predicting student performance by using data mining methods for classification. Cybernetics and information technologies. 2013 Mar; 13(1):61–72. Crossref
25. Ramesh VA, Parkavi P, Ramar K. Predicting student performance: a statistical and data mining approach. International journal of computer applications. 2013 Jan; 63(8):35–9. Crossref
26. Harvey L. Defining and measuring employability. Quality in higher education. 2001 Jul; 7(2):97–109. Crossref
27. Gokuladas VK. Technical and nontechnical education and the employability of engineering graduates: an Indian case study. International Journal of Training and Development. 2010 Jun; 14(2):130–43. Crossref
28. Rees C, Forbes P, Kubler B. Student employability profiles. New York: The Higher Education Academy; 2006. p. 1–83.
29. Kahya E. The effects of job performance on effectiveness. International Journal of Industrial Ergonomics. 2009 Jan; 39(1):96–104. Crossref
30. Chien CF, Chen LF. Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. Expert Systems with applications. 2008 Jan; 34(1):280–90. Crossref
31. Mukhtar M, Yahya Y, Abdullah S, Hamdan AR, Jailani N, Abdullah Z. Employability and service science: Facing the challenges via curriculum design and restructuring. International Conference on Electrical Engineering and Informatics. 2009 Aug; 2:357–61. Crossref
32. Shafie LA, Nayan S. Employability awareness among Malaysian undergraduates. International Journal of Business and Management. 2010 Aug; 5(8):119–23.
33. Gokuladas VK. Predictors of employability of engineering graduates in campus recruitment drives of Indian software services companies. International Journal of Selection and Assessment. 2011 Sep; 19(3):313–9. Crossref
34. Othman Z, Musa F, Mokhtar NH, Ya'acob A, Latiff RA, Hussin H. Investigating University Graduates' English Language Competency towards Employability: A Proposed Research Method. International Journal of Learning. 2010 Aug; 17(7):429–40.
35. Potgieter I, Coetzee M. Employability attributes and personality preferences of postgraduate business management students. SA Journal of Industrial Psychology. 2013 Jan; 39(1):1–10. Crossref
36. Yusoff YM, Omar MZ, Zaharim A, Mohamed A, Muhamad N. Employability skills performance score for fresh engineering graduates in Malaysian industry. Asian Social Science. 2012 Dec; 8(16):140–5. Crossref
37. Jantawan B, Tsai CF. The Application of Data Mining to Build Classification Model for Predicting Graduate Employment. ArXiv preprint arXiv: 1312.7123. 2013 Dec; 11(10):1–7.
38. Pool LD, Qualter P. Emotional self-efficacy, graduate employability, and career satisfaction: Testing the associations. Australian Journal of Psychology. 2013 Dec; 65(4):214–23. Crossref
39. Mishra T, Kumar D, Gupta S. Mining Students' Data for Prediction Performance. Fourth International Conference on Advanced Computing & Communication Technologies. 2014 Feb; 255–62. Crossref
40. Mishra T, Kumar D, Gupta S. Students' Employability Prediction Model through Data Mining. International Journal of Applied Engineering Research. 2016; 11(4):2275–82.