IRJSE *Int. Res. J. of Science & Engineering*, 2020; Special Issue A7: 800-804 SJIF Impact Factor 6.68
ISSN: 2322-0015
RESEARCH ARTICLE **OPEN ACCESS**

The Science for Prognostication of Student's Performance through Educational Data Analytics.

Bhalchandra Parag[1], Muley Aniket[2], Joshi Mahesh[3], Khamitkar Santosh[1], Kulkarni Govind[4], Khurpe Prashilkumar[1], Lokhande Sakharam[1]

[1] School of Computational Sciecnes, [2] School of Mathematical Sciences,
[3] School of Educational Sciences, S.R.T.M.University ,Nanded, MS, 431606, India
[4]Department of Computer Sceince, LLDM College, Parli Vaijanath, Dist , Beed,431515
Email: srtmun.parag@gmail.com, aniket.muley@gmail.com, maheshmj25@gmail.com, s.khamitkar@gmail.com, govindcoolkarni@gmail.com,prashilkhurpe@gmail.com, sana_lokhande@rediff.com,

Manuscript Details
Available online on http://www.irjse.in
ISSN: 2322-0015

Cite this article as:
Bhalchandra Parag et al. The Science for Prognostication of Student's Performance through Educational Data Analytics., *Int. Res. Journal of Science & Engineering*, February 2020 | Special Issue A7: 800-804.

ABSTRACT

Data mining via academic analytics is speedily increasing and lot of educational institutes has started synthesis of their data to understand hidden patterns, predict possibilities, and for self learning. In these lines, we have coined a word, prognostication, where we aim to forecast regarding performance of students via analysis of social, economical, personal and academic attributes. The definitive goal is to find attributes in terms of set of crucial aspects those severely affect the student's performance. The results facilitate us to use discovered variables for creating more conducive environment for academic learning process.

Keywords: Data mining, KDD,educational data analysis.

INTRODUCTION

The role of data driven intelligence is becoming central in educational organizations. It is also known as Data mining or Knowledge Discovery in Database (KDD) which is considered as the process of finding hidden and useful knowledge from large amount of data [1, 2,7]. Now days, all educational organizations, institutions or Universities have been computerized and they have database systems capturing all essential data from all vital parts.

Even though, higher educational institutes face very hard for predicting accurate knowledge from databases. Thus Data mining finds applications in educational industry [7] mainly to explore hidden knowledge in crucial aspects including student support, course registration processes, alumina associations, designing new courses, etc. We can apply Data Mining to traditional processes also. Such applications are very challenging [8]. Despite the Government of India, Government of Maharashtra, UGC, New Delhi, etc are pouring lot of funds for improving education standards, academic performance of students in India is not improving [1,3]. Hence it is need of time to escalate other hidden variable, where performance of students can have some sort of correlations. Over the years, University like Swami Ramanand Teerth Marathwada University, Nanded have taken attempts to create large amount of databases from the data pertaining to academic activities, educational activities and administration contexts [1]. If we apply educational data mining algorithms then it will be very interesting to see hidden patterns in these databases. Hence a research project, titled, "Assistive prognostications implemented to Students through Data Analytics implemented over Campus Educational Data" was submitted as a self enquiry. The project explores opportunities to implement academic analytics over campus educational data to find social, economical and personal variables that affects performance of the student. These results are then used for assistive measures called prognostications. The performance analysis and prognostication share boundaries with Computational, Statistical and Educational Sciences [1]. This work is an example of a joint interdisciplinary work to be undertaken with the help of two more schools, viz, School of Mathematical Sciences and School of Educational Sciences. In routine sense, our work relies on implementation of data mining algorithms over university's database in order to discover social, habitual and economical aspects associated with performance of students. This research study needs to build a database. Data in this database comes from sources called questionnaire, which consisted of questions related to the social, personal and economical information of students [1]. The original questionnaire has 98 questions. But as a part to test feasibility study, a trial questionnaire was circulated first to a limited group of students. This questionnaire had 43 questions. A research fellow was appointed who was trained to counsel students and encourages them to fill the questionnaire. A trail student's dataset was created with 360 records and 46 fields by closed questionnaire method[1]. This includes student's personal details like social, intellectual, demographic, habitual, health and economical data[1]. Experimentation was carried out in SPSS Ver. 22 and R Miner software [4].

METHODOLOGY

Our Main and Trail questionnaires have questions with predefined options [1,9] related to student's attributes as defined in Pritchard and Wilson [5, 9]. The attributes in questionnaire have relations with the performance of students. The questionnaire was reviewed by experts from School of Educational Sciences[1]. There was some trial testing of questionnaire on some small groups of students. These trials helped us to understand over all feedback as such. Two revisions with trial testing have been made to devise out the final version of questionnaire. The final questionnaire consisted of 43 closed type questions [1,9]. The questionnaire were distributed to students and demonstrated for feedback. After this, a dataset was integrated with 360 student records and each record consists of 46 fields. Originally, there were 43 fields, equal to the total number of questions in questionnaire. Three additional fields were added for seeking information of students. Microsoft Excel 2007 software is used to record the dataset. Data set values like Yes / No were converted in to numeric values like 1 0r 0. Other numerical codes in the range 0 ,1,2,3,4 …6 were also given depending upon the number of possible answers a question can have. Likewise other answers are also converted into numeric values. During the pre-processing of data, we found some ambiguities and false information in student's data [1]. A snapshot of questionnaire is as given in Figure 1. Since we aim for discovery of attributes which affects performance of students, we primarily investigated literature across the globe to see what other people have done. We have referred to the renowned sources, including the Psychology of Asian Learners [3]. For proper understanding the performance terminology, we primarily relied on the

work of Shoukat Ali at al.[1,6]. A similar works of Graetz et al. [1,7] suggested that the social status of parents and performance of students are reliant with each other. Another literature, Considine and Zappala [1, 8] have noticed that parent's economical conditions have positive effects on the performance. The work of Staffolani and Bratti [1, 9] highlights, that the future achievements of students have correlation with previous year's scores. The combined finding that affects performance as per [1,6] highlights many social, economical and self related factors that have effect on student's academic performance. We felt that, once we discover these hidden truths, we can undertake appropriate, corrective actions to improve overall performance. This is the motto behind our prognostication work[10]. The overall research methodology is shown in Fig 2.

RESULTS AND DISCUSSION

The hierarchical clustering process is initiated using Ward's method [1, 7]. It is appeared that, the UG percentage is the highest cluster which contains number of small sub clusters which affect student's performance. While analyzing the performance of students, we found that, at the first level of clustering, student's family environment including family income, family support and family relations matters a lot[1]. These attributes have direct affect on the performance. Further, the use of internet, the available study material, own notes and additional tutorials also have corresponding positive effect on the performance. The personal computer and personal book library also shows self help seeking attitude of the student and have the closure association with performance [1]. These highlight prognostications avenues for increasing performance of the students. The SPSS22.0v is used to analyze the data set.

Secondly, general Academic Analytics were carried out by the examination of cross-classified category data which is common in evaluation and research, with Karl Pearson's family of chi-square tests representing one of the most utilized statistical analyses for answering questions about the association or difference between categorical variables[4,5]. Hence we have made an attempt to make the use and interpretations of the family of chi-square tests developed by Pearson, focusing primarily on the chi-square tests of independence and homogeneity of variance. The SPSS22.0v is used to analyze the data set. We have observed that, there is a significant difference between gender wise distribution of students and parents earning sources. Similarly, there is significance difference between earning sources and students parenting[4,5]. If earning is more, more proper will be the parenting. It is also observed that students coming from well educated families have preferred to enter computer science courses. This is because of alertness of the parents. Educated parents are more concerned to computer courses because of knowledge regarding IT professional's earning and salary packages [4,5]. Similarly, the children of parents belonging to Agriculture and other fields also prefer computer science courses as they think that they may get job quickly which obliviously would help their family earning to grow[1]. So, if parenting is proper, there will be good motivation to students which ultimately accelerates performance of students. Thus our study has made visible the hidden correlation between the non academic variables like parenting, earning of family with the performance of students [4,5].Thirdly, Discriminate Analysis was done. The Discriminant function analysis, a.k.a. discriminant analysis or DA, is used to classify cases into the values of a categorical dependent, usually a dichotomy [2,6]. If discriminant function analysis is effective for a set of data, the classification table of correct and incorrect estimates will yield a high percentage correct. Multiple discriminant function analysis (MDA) is used when the dependent has three or more categories. Discriminant function analysis is found in SPSS under Analyze, Classify and Discriminant. One gets DA or MDA from this same menu selection, depending on whether the specified grouping variable has two or more categories. Discriminant analysis is performed to test based on certain parameters viz. whether they scholarship, self library, self PC, where they lived, does the use internet, does they get free time to study, free time spends with friends, does they have failure background, fathers income and mothers income[6]. It has been observed that 67.4% of data was correctly classified as Yes and No of the system by the discriminant function. Mother's income and free time

to study these to parameters are playing important role in discrimination or classification of student's satisfactory performance [6].

1	Course code	MSc (5) , MCA (6)
2	Your name	
3	Gender (sex)	Male (1) Female (0)
4	Marital status	Married (2) unmarried(3)
5	Age	
6	Home address	Urban(1) rural (2) foreign(3)
7	Mobile no.	
8	Personal email id	
9	Degree passer and percentage	General B.Sc. / (1) B.Sc.(computer CS) / (2) BCA / BCS/ (3) Other / (4) (5)
	Percentage	
10	Degree collage name	
11	Father's Education	Below or SSC/ (1) HSC/ (2) Graduate/ (3) Post Graduate/ (4) other (5)
12	Fathers job and annual income	Service / (1) Business/ (2) Agriculture/ (3) In house/ (4) Other/ (5)
	Income	0-1 lakh (1) , 1.1-2 lakh(2), 2.1-5 lakh(3) , 5lakh - above (4)
13	Mothers education	Below or SSC/ (1) HSC/ (2) Graduate/ (3) Post Graduate/ (4) other (5)
14	Mothers job and annual income	Service / (1) Business/ (2) Agriculture/ (3) In house/ (4) Other/ (5)
	Income	0-1 lakh (1) , 1.1-2 lakh(2), 2.1-5 lakh(3) , 5lakh - above (4)
15	Family size	
16	Family relationship	Excellent / (1) Good/ (2) Satisfactory/ (3) Bad/ (4) Very Bad (5)
17	Family support to your education	Excellent / (1) Good/ (2) Satisfactory/ (3) Bad/ (4) Very Bad (5)
18	Reason to choose this course	Career in IT/ (1) Near to Home/ (2) Reputation of course / (3) Blind Decision/ (4) Parents wish: (5)
19	Travel mode and time needed	Bus/ (1) Railway/ (2) City Bus/ (3 but taken as 1) Rickshaw/ (4) Self Vehicle / (5) walking (6)

Fig.1. Sample Questionnaire

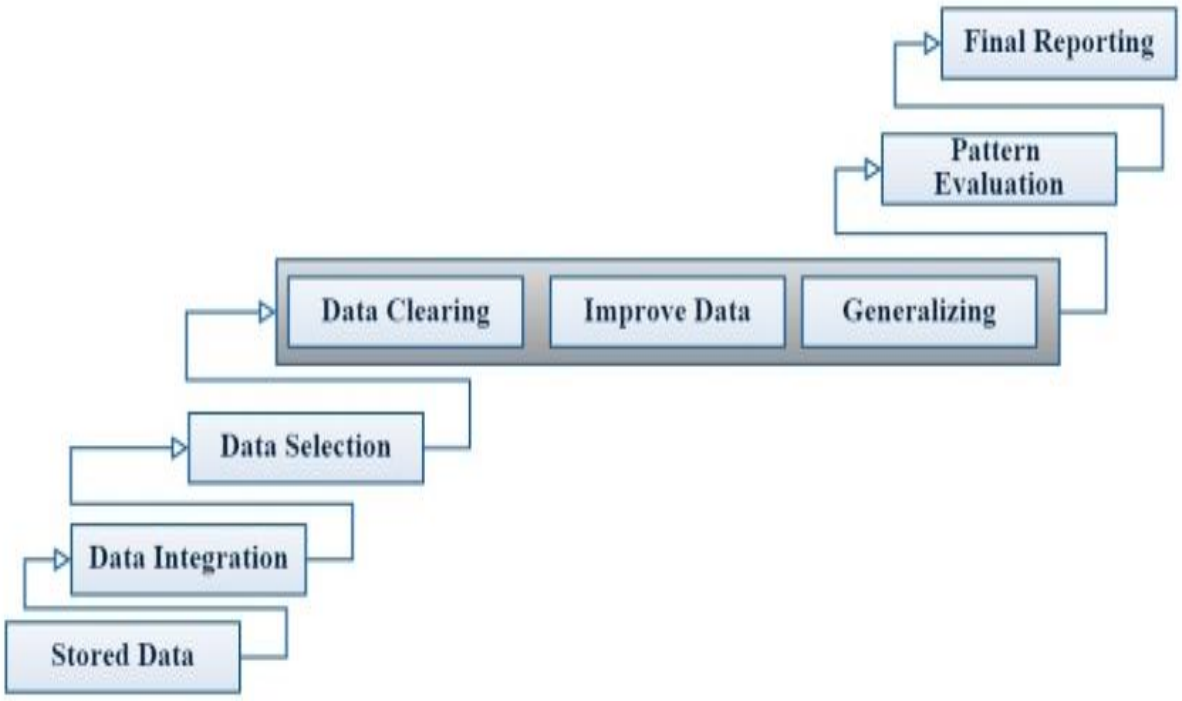

Fig.2. Research Methodology

CONCLUSION

The study narrates three experiments made over self devised out dataset containing 43 fields and 360 records to highlight implementation of data mining over educational data. The study aimed to discovery of knowledge for the sake of prognostications related with performance of students. The exercise was carried out to find role of other hidden factors affecting performance o students, which included

social, economical and personal. The interdisciplinary approach was adopted to get such insight. The study has scientifically made visible the hidden aspects which affects the performance which shall definitely improve students in their weaker aspects. Limitation of this study comes from the reality that this research was only carried out on the students of computer science and that maybe other data sets will show different characteristic.

Conflicts of interest: The authors stated that no conflicts of interest.

REFERENCES

1. Prognostication of Student's Performance: An Hierarchical Clustering Strategy for Educational Dataset, Computational Intelligence in Data Mining—Volume 1: Proceedings of the International Conference on CIDM, 5-6 December 2015 (pp.149-157)
2. Margaret Dunham, Data Mining: Introductory and Advanced Topics, by Margaret H. Dunham, Pearson publications, 2002.
3. Psychology of Asian Learners [3], King, Ronnel B, Springer Publications,2016
4. Analyzing students performance using Academic Analytics, Mahesh Joshi, et al, 2016 International Conference on ICT in Business Industry & Government (ICTBIG), IEEE Publications
5. Academic Analytics Implemented for Students Performance in Terms of Canonical Correlation Analysis and Chi-Square Analysis, Aniket Muley et al, Advances in Intelligent Systems and Computing book series (AISC, volume 625), 2017
6. Classification Through Discriminant Analysis Over Educational Dataset, Parag Bhalchandra, et al, nformation and Decision Sciences, pp.99-106, 2018
7. Bratti, M. and Staffolani, S. 2002, 'Student Time Allocation and Educational Production Functions', University of Ancona Department of Economics Working Paper No. 170.
8. Ma, Y., Liu, B., Wong, C. K., Yu, P. S., & Lee, S. M. (2000), Targeting the right Students using data mining, Sixth ACM SIGKDD International Conference, Boston, MA (Conference Proceedings; p. 457-464.
9. B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer and, W. F. Punch, Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA, in Proceedings of ASEE/IEEE Frontiers in Education Conference, Boulder, CO: IEEE, 2003.
10. Kotsiantis S. 2009. Educational Data Mining: A Case Study for Predicting Dropout – Prone Students. Int. J.Knowledge Engineering and Soft Data Paradigms, 1(2), 101–111.