A Survey on approaches for predicting performance of students

Miss. Geetha N[1], Mrs. Swathi .K[2], Mr. Hemanth Y K[3], Dr. Piyush Kumar Pareek[4]

[1]Assistant Professor, Information Science and Engineering, East West College of Technology, Bengaluru, affiliated to Visvesvaraya Technological University Belagavi, India
[2]Assistant Professor, Computer Science and Engineering, K.S.Institute of Technology, Bengaluru, affiliated to Visvesvaraya Technological University Belagavi, India
[3]Assistant Professor, Computer Science and Engineering, East West College of Technology, Bengaluru, affiliated to Visvesvaraya Technological University Belagavi, India
[4]Associate Professor, Computer Science and Engineering, East West College of Technology Bengaluru, affiliated to Visvesvaraya Technological University Belagavi, India

Abstract— *The Paper focuses on predicting performance of students; we have considered different cases and analyzed the methods adopted in various cases by various authors. The main intension behind this survey is to understand latest Research methods adopted along with their results.*

Educational Data Mining (EDM) is an emerging field exploring data in educational context by applying different Data Mining (DM) techniques/tools. EDM inherits properties from areas like Learning Analytics, Psychometrics, Artificial Intelligence, Information Technology, Machine learning, Statics, Database Management System, Computing and Data Mining. It can be considered as interdisciplinary research field which provides intrinsic knowledge of teaching and learning process for effective education.

In this Paper a survey has been carried out in three engineering colleges with an establishment of more than fifteen years, a total of two hundred and forty six students answered the questions and the reliability and validity of questionnaire was found to be good.

Keywords— predicting, survey & Research.

I. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. It sits at the common frontiers of several fields including Data Base Management, Artificial Intelligence, Machine Learning, Pattern Recognition, and Data Visualization. From a statistical perspective it can be viewed as computer automated exploratory data analysis of (usually) large complex data sets. In spite of the somewhat exaggerated hype, this field is having a major impact in education

Severe challenges are being faced by students and alumni in higher education. Institutions would like to know, for example, which students will enroll in particular course programs and which students will need assistance in order to graduate. Are some students more likely to transfer than others? What groups of alumni are most likely to offer pledges? In addition to these challenges, traditional issues such as enrollment management and time-to-degree continue to motivate higher education institutions to search for better solutions.

II. LITERATURE SURVEY

Severe challenges are being faced by students and alumni in higher education. Institutions would like to know, for example, which students will enroll in particular course programs and which students will need assistance in order to graduate. Are some students more likely to transfer than others? What groups of alumni are most likely to offer pledges? In addition to these challenges, traditional issues such as enrollment management and time-to-degree continue to motivate higher education institutions to search for better solutions. One way to effectively address these student and alumni challenges is through the analysis and presentation of data or data mining. Data mining enables organizations to use their current reporting capabilities to uncover and understand hidden patterns in vast databases. These patterns are then built into data mining models and used to predict individual behavior with high accuracy. As a result of this insight, institutions are able to allocate resources and staff more effectively. Data mining may, for example, give an institution the information necessary to take action before a student drops out, or to efficiently allocate resources with an accurate estimate of how many students will take a particular

course. This study addresses the capabilities of data mining and its applications in higher education. Data mining uses a combination of an explicit knowledge base, sophisticated analytical skills and domain knowledge to uncover hidden trends and patterns. These trends and patterns form the basis of predictive models that enable analysts to produce new observations from existing data.

The ability to predict the success of students when they enter a graduate program is critical for educational institutions because it allows them to develop strategic programs that will help improve students' performances during their stay at an institution. Graduate education has become increasingly popular across the spectrum of higher level education. Higher education institutions have always been interested in predicting the paths of students. Thus, they are interested in identifying which students will require assistance as they enter the graduate program. Upon graduation, the students in an educational faculty may either continue in postgraduate programs or become a state or private school teacher. In this way, student performance is critical for ensuring academic success. Student learning in school significantly influences one's future career, particularly for students learning to teach elementary school mathematics. In recent years, prospective teachers have preferred entering postgraduate programs because of having shown more effective teacher performances or having chosen an academic career. A high GPA as an undergraduate is one of the conditions required to be able to enter postgraduate programs. This is important because the ability to predict an undergraduate's success of graduating brings with it the ability to predict their chances of success in being admitted to graduate studies. "To better manage and serve the student population, institutions need better assessment, analysis, and prediction tools to analyze and predict student-related issues.

A neural network is a well-developed modeling technology, and during the past decades it has been widely used in technical applications that involve predictions and classifications. The neural network model is especially attractive for modeling complex systems because of its favorable properties: its abilities to approximate universal functions, accommodate multiple non-linear variables with unknown interactions, and generalize well [1]."

A negative binomial model was used to develop Utah-specific crash prediction models based on both the three-year and five-year sample periods. A backward stepwise regression technique was used to isolate the variables that would significantly affect highway safety. The independent variables used for negative binomial regression included the same set of variables used in the HSM predictive model along with other variables such as speed limit and truck traffic that were considered to have a significant effect on potential crash occurrence. The significant variables at the 95 percent confidence level were found to be average annual daily traffic (AADT), segment length, total truck percentage, and curve radius [2][3].

Students commencing tertiary education enter through a number of traditional and alternative academic pathways. As a result, tertiary institutions encounter a broad range of students, varying in demographic, previous education, characteristics and academic achievement. In recent years, the relatively constant increase in tertiary applications in Australia has not translated to an increase in student retention or graduate numbers. The Health Sciences discipline typically falls within this paradigm, prompting various approaches to promote academic success and overall student retention. In this study, the demographic and previous education of health science students at an Australian University, were analyzed along with first year science grades from a core first year anatomy and physiology unit. A generalized linear model (GLM) demonstrated statistically significant relationships between performance in the unit (measured by grade point average) and year 12 Australian Tertiary Admissions Rank (ATAR) subjects (human biology and chemistry; $p<0.001$) and gender ($p<0.001$). No significant performance correlation was observed with household socioeconomic status, as measured by socio-economic indexes for areas. Taken together, the results from this study facilitate estimation of academic success by some parameters prior to their commencement at University [4].

The LTPP program includes two experiment categories: General Pavement Study (GPS) experiments and Specific Pavement Study (SPS) experiments. Within each experiment category, several types of studies have been evaluated by the LTPP program. The GPS experiments consist of in service pavement test sections throughout the United States and Canada. They embrace an array of site selection factors that provide information for a national database to meet the objectives of the LTPP program. The SPS experiments have their own set of unique goals, construction needs, and experimental approaches that cannot be achieved by the GPS experiments. The SPS experiments are intensive studies with specific evaluation variables [5].

Mindfulness enhances emotion regulation and cognitive performance. A mindful approach may be especially beneficial in high-stakes academic testing environments, in which anxious thoughts disrupt cognitive control. The current studies examined whether mindfulness improves the emotional response to anxiety-producing testing situations, freeing working

memory resources, and improving performance. In Study 1, Authors have examined performance in a high-pressure laboratory setting. Mindfulness indirectly benefited math performance by reducing the experience of state anxiety. This benefit occurred selectively for problems that required greater working memory resources. Study 2 extended these findings to a calculus course taken by undergraduate engineering majors. Mindfulness indirectly benefited students' performance on high-stakes quizzes and exams by reducing their cognitive test anxiety. Mindfulness did not impact performance on lower-stakes homework assignments. These findings reveal an important mechanism by which mindfulness benefits academic performance, and suggest that mindfulness may help attenuate the negative effects of test anxiety [6]

Using detailed data on recipients of bachelor's degrees from MIT between 2006 and 2012, I examine the selection of students into finance or science and engineering (S&E). I find that academic achievement in college is negatively correlated with a propensity to take a job in finance and positively correlated with a propensity to pursue a graduate degree or taking a job in S&E. This pattern is primarily driven by differences in skill development during college, not by differences in academic qualifications at college entry. In both high school and college, the two groups participate in different activities: students who ultimately choose finance are substantially more likely to be varsity-sports leaders in high school; they are also more likely to join fraternities and sororities, a decision typically made at college entry. Sizable differences in academic performance begin in freshman year and persist throughout college. The 2008 financial crisis, which substantially reduced the availability of entry-level positions in finance, prompted some students with relatively low college-entry qualifications to major in S&E instead of management or economics and/or to improve their academic performance. But there is no evidence that those with top qualifications changed their skill development in response to the crisis. Taken together, the results demonstrate that the preferences and skills of graduates who pursue finance are not comparable to those of graduates who choose a career in S&E [7]

As institutions seek to leverage analytics capabilities and improve student retention and graduation, they must recognize an important reality: it's not enough to simply have the technology. Schools must be prepared to ask the hard questions and challenge "campus legends," the sort of longstanding assumptions that aren't actually grounded in data. Schools must be willing to say, "Our old approach hasn't achieved optimal results, and we need to improve it." In short, schools must be willing and able to collect and analyze the data – and ready to make changes based on what they find. Making decisions based on detailed, actionable, and empirical data allows institutions to spend their budgets in the most effective ways possible. As institutions implement programs to help students, the retention and graduation rates will go up. As retention and graduation rates rise, more funding and more tuition dollars will be available to reinvest in programming and infrastructure oriented toward student success. In this way, data-driven decision-making becomes the key to a virtuous cycle of continuously improving student retention programming and positive student outcomes [8].

Data is a key source of intelligence and has competitive advantage for every higher educational organization. With the explosion of electronic data available to educational organizations and the demand for better and faster decisions, the role of data driven intelligence is becoming central in educational organizations. Contrary Data mining or Knowledge Discovery in Database (KDD) is the process of converting the raw data into useful knowledge required supporting decision making. It automates the process of knowledge discovery, making us more productive in our search for useful information than we would be otherwise. It also increases the confidence with which we can make business decisions [9].

III. INDENTATIONS AND EQUATIONS

3.1 Data Collection & Analysis

A) Selection of organization

- Three Engineering colleges were selected based on convenient sampling.
- These colleges are also very well established and into existence since fifteen years

B) Sampling population

- As many as 246 samples were included as part of data for the study. These samples were collected from final year engineering students

C) Data collection

- An exhaustive questionnaire was prepared and data was collected with regard to predicting performance

D) Stages of Data collection

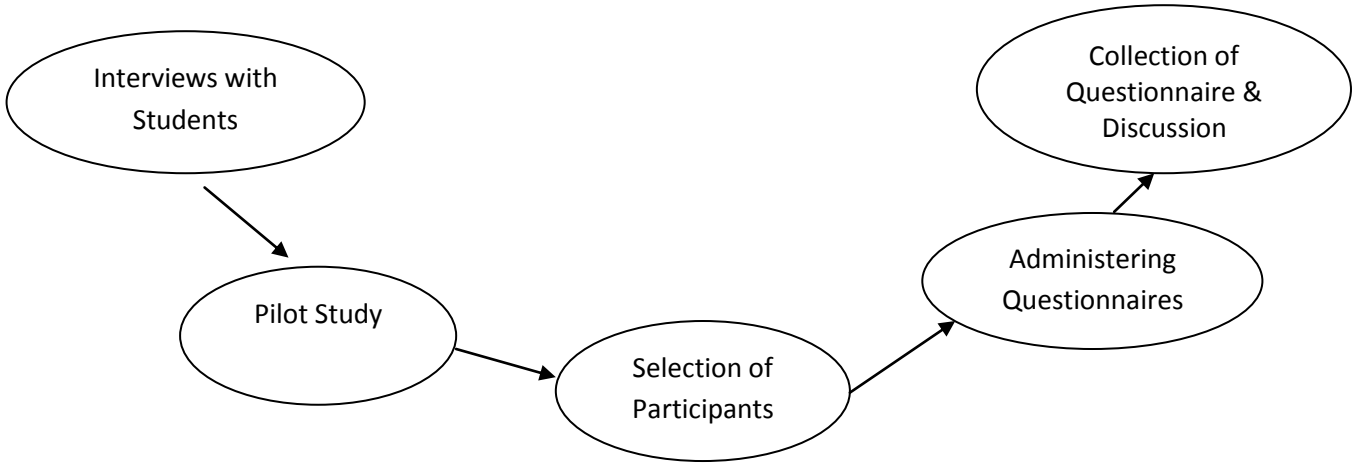

FIGURE 1: STAGES OF DATA COLLECTION FOLLOWED BY AUTHOR

3.2 Hypothesis:

H1 - A positive relationship exists between Teaching and students.

H2 – Innovative methods will help in enhancing results

IV. RESULTS

The following Tables & Figures show the opinion of the participants. SPSS software package is used for statistical analysis. There are no missing values in the data collected.

TABLE 1
MEAN & STD. DEVIATION FOR FEW QUESTIONS

	N	Mean	Std. Deviation
How well do professors teach in college	246	19.15	.815
How effective is the teaching outside your major at this college?	246	18.85	.962
How helpful is your mentor	246	17.91	1.237
How easy is it to register for courses at this college?	246	18.92	1.068
Valid N (list wise)	246		

TABLE 2
RESULTS FOR Q: HOW WELL DO PROFESSORS TEACH IN COLLEGE

	Frequency	Percent	Valid Percent	Cumulative Percent
Not at all well	1	.4	.4	.4
Extremely well	30	12.2	12.2	12.6
Somewhat well	38	15.4	15.4	28.0
Very well	112	45.5	45.5	73.6
Not so well	65	26.4	26.4	100.0
Total	246	100.0	100.0	

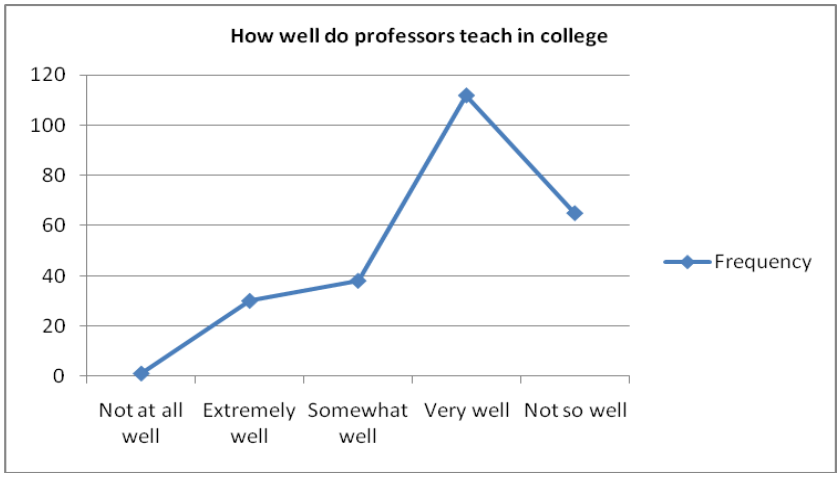

FIGURE 2: RESULTS FOR UNDERSTANDING LEARNING LEVEL OF STUDENTS

TABLE 3

HOW EFFECTIVE IS THE TEACHING OUTSIDE YOUR MAJOR AT THIS COLLEGE

	Frequency	Percent	Valid Percent	Cumulative Percent
Not at all well	4	1.6	1.6	1.6
Extremely well	16	6.5	6.5	8.1
Somewhat well	44	17.9	17.9	26.0
Very well	126	51.2	51.2	77.2
Not so well	56	22.8	22.8	100.0
Total	246	100.0	100.0	

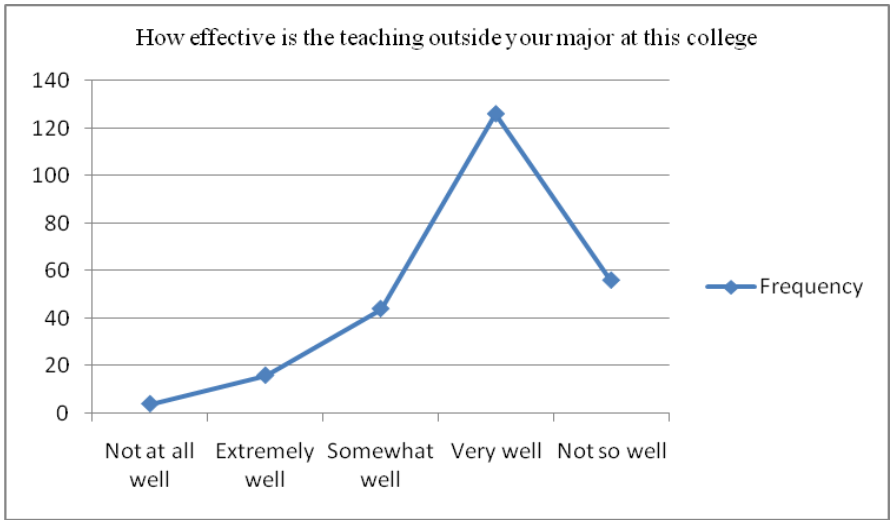

FIGURE 3: UNDERSTANDING OTHER DEPARTMENTS

TABLE 4

HOW HELPFUL IS YOUR MENTOR

	Frequency	Percent	Valid Percent	Cumulative Percent
Not at all helpful	4	1.6	1.6	1.6
Not so helpful	16	6.5	6.5	8.1
Somewhat helpful	56	22.8	22.8	30.9
Very helpful	96	39.0	39.0	69.9
Extremely helpful	74	30.1	30.1	100.0
Total	246	100.0	100.0	

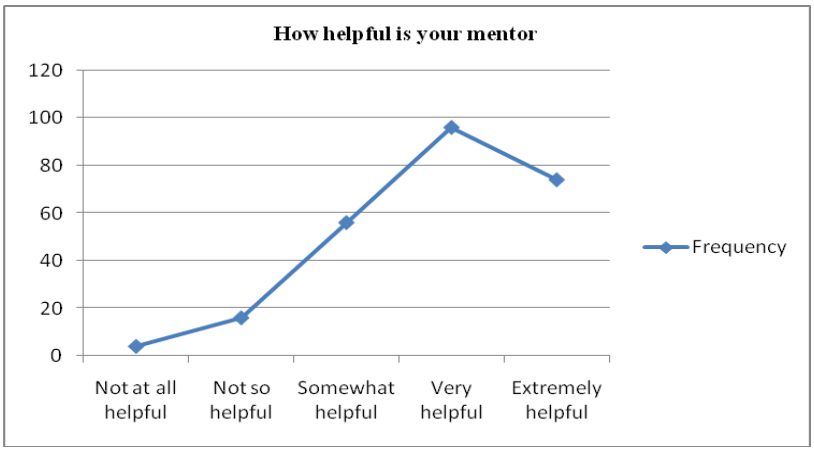

FIGURE 4: UNDERSTANDING MENTORS BEHAVIOR WITH STUDENTS

TABLE 5
HOW EASY IS IT TO REGISTER FOR COURSES AT THIS COLLEGE?

	Frequency	Percent	Valid Percent	Cumulative Percent
Extremely easy	12	4.9	4.9	4.9
Very easy	14	5.7	5.7	10.6
Not so easy	36	14.6	14.6	25.2
Not at all easy	103	41.9	41.9	67.1
Somewhat easy	81	32.9	32.9	100.0
Total	246	100.0	100.0	

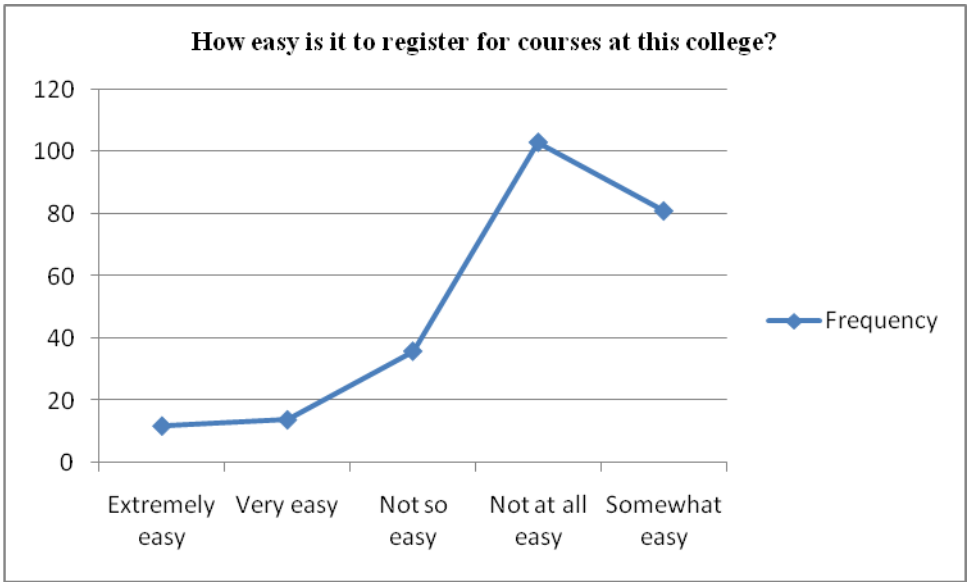

FIGURE 5: UNDERSTANDING ADMISSION PROCESS OF COLLEGE

4.1 Analysis & Interpretation

- The Data collected has been primarily tabulated & Master table was prepared
- Sample was tested for reliability using Cronbach's alpha
- Percentage analysis is the basic tool for analysis
- Regression analysis a statistical process for estimating the relationships among variables is used

4.2 Regression Analysis

MODEL SUMMARY-1

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.743[a]	.552	.550	.847

Dependent Variable(X): A positive relationship exists between Teaching and students

Independent Variable(Y): How well do professors teach in college

In Model Summary 1 – 0.74 means that 74% of the variation of y-values around the mean is explained by the x-values. In other words, 74% of the values fit the model.

H0 – There is no relationship exists between Teaching and students

H1 - A positive relationship exists between Teaching and students

MODEL SUMMARY - 2

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
2	.810[a]	.657	.655	.525

Dependent Variable(X): Innovative methods will help in enhancing results

Independent Variable(Y): How helpful is your mentor

In Model Summary 2 – 0.81 means that 81% of the variation of y-values around the mean is explained by the x-values. In other words, 81% of the values fit the model.

H0 - Innovative methods has no effect in improving results

H2.1 - Innovative methods will help in enhancing results

Alternate Hypothesis is accepted.

V. CONCLUSION

Our result shall determine a correlation between the GPA at the undergraduate level and the one at the graduate level. This outcome shall emphasize the relevance of indicators of undergraduate achievements for graduate admissions decision-making. We shall attribute this improvement primarily to the completeness of our data, to the strong consecutive nature of the Computer Science curriculum, and to the fact that data were collected within one institution. Our outcomes shall be identification of most significant explanatory variable. Innovative methods will help in enhancing results and A positive relationship exists between Teaching and students.

ACKNOWLEDGEMENTS

The satisfaction and euphoria that accompany the successful completion of any task will be incomplete without the mention of the individuals, we are greatly indebted to, who through guidance and providing facilities have served as a beacon of light and crowned our efforts with success. We are thankful to Management of East West Institute of Technology Bengaluru, Principal & Director Dr. K Channakeshavalu ,VTU Research Centre of Department of Computer Science & Engineering at EWIT.

REFERENCES

[1] "Using Neural Network and Logistic Regression Analysis to Predict Prospective Mathematics Teachers' Academic Success upon Entering Graduate Education", by Elif Bahadur published in academic success upon entering graduate education. Educational Sciences: Theory & Practice, 16, 943-964.

[2] " CRASH PREDICTION MODELING FOR CURVED SEGMENTS OF RURAL TWO-LANE TWO-WAY HIGHWAYS IN Utah "Prepared For: Utah Department of Transportation Traffic & Safety, Research Divisions Submitted By: Mitsuru Saito,October 2015.

[3] REPORT ON ANNUAL INDICATORS THE UNIVERSITY OF MASSACHUSETTS ,University Performance Measurement System, July 2015

[4] Predicting Academic Success of Health Science Students for First Year Anatomy and Physiology. Ryan S Anderton, Tess Evans & Paola T Chivers, www.sciedupress.com/ijhe International Journal of Higher Education Vol. 5, No. 1; 2016

[5] TRANSPORTATION & DEVELOPMENT INSTITUTE (T&DI) Of THE AMERICAN SOCIETY OF CIVIL ENGINEERS (ASCE) And LONG-TERM PAVEMENT PERFORMANCE (LTPP) INTERNATIONAL DATA ANALYSIS CONTEST

[6] Mindfulness, anxiety, and high-stakes mathematics performance in the laboratory and classroom David B. Bellinger , Marci S. DeCaro , Patricia A.S. Ralston , http://dx.doi.org/10.1016/j.concog.2015.09.001 1053-8100/_ 2015 Elsevier Inc.

[7] Are the "Best and Brightest" Going into Finance? Skill Development and Career Choice of MIT Graduates by Pian Shu,Harward Business School,Working Paper 16-067

[8] Report on Analytics for Higher Education: Solving Six Critical Issues with Blackboard Analytics, Blackboard 2015

[9] Prognostication of Student's Performance: Factor Analysis Strategy for Educational Dataset, International Journal of Emerging Engineering Research and Technology Volume 4, Issue 1, January 2016, PP 12 -21 ISSN 2349-4395 (Print) & ISSN 2349-4409, Aniket Muley, Parag Bhalchandra