Contextual Application of Learning Analytics: Three Case Snippets from India

Nitin Simha Vihari[1],
M.K.Rao[2],
and Anugamini Shrivastava[3]
Indian Institute of Technology Roorkee, India
[1]nitinvihari@gmail.com
[2]mkraofdm@iitr.ac.in

Abstract - **Learning has always been the basis for the evolution of mankind and subsequently the scientific progress. In the contemporary age of digital space, multitude of data is associated with every formal activity and analytics as the science of data, draws out intelligence predictions from the present observations. In lieu of this, the present study aims to investigate the existing literature of learning analytics among the academic spheres and also presents three case studies, such as Student's performance prediction, Women being a wage earner and Student's Grade Prediction. First case attempts to identify the key predictors of 180 MBA student's performance from a business school of national repute in India, with respect to three parameters such as educational achievement, learning approach and personality type. Second case analyzes a crowd sourced data containing five parameters on 2000 women with 657 as not wage earners. The analysis was conducted using Linear Probability Model (LPM), Logit Model to obtain the marginal effects of variables, Probit Model and finally a comparison of all models was done to pick the best performing model. Third case evaluates a class containing 382 undergraduate students from six different branches of engineering in a university of national repute is considered, to predict their final course grade. Seven predictor variables are identified from the literature and multiple linear regressions was used to identify and rank the predictor variables.**

Keywords - **Learning Analytics, Logit Model, Probit Model, Class Room Analytics, Case Snippets**

I. INTRODUCTION

Learning analytics is an emerging field in which sophisticated analytic tools are used to improve learning and education. Learning Analytics is a research field related to Educational Data Mining (EDM), and is gaining increasing popularity since Horizon Report 2012 described it as a forthcoming trend. Learning analytics (LA) refers to the application of Big Data methodologies and techniques to improvise the learning. LA is based on analyzing the learning behavior of students using a wide data set that takes into consideration- the student enrollment data, the previous academic record of students, student surveys through questionnaires about courses and teaching methods, data from online discussion forums and such. In predicting and analyzing student performance, there are a no. of techniques that could be used such as classification algorithms - Eg: decision tree methods- C4.5, RepTree and J48, k-nearest neighbor classifier, Naive Bayes, Multi-layer perceptron (neural networks), Sequential Minimal Optimization and clustering methods - Eg: Latent Semantic Analysis and K-means clustering methods. Learning analytics draws from, and is closely tied to, a series of other fields of study including business intelligence, web analytics, academic analytics, and educational data mining.

Business Intelligence (BI) deals with the nexus between strategic thinking and information technology (Baker, 2007) to make insightful decision making capabilities. Web Analytics (WA) deals with visitor count of a particular website to understand the efficacy of online initiatives in the digital space. This has been conducted through experimentation, consistent testing and measurement (McFadden, 2005) and enhances the business intelligence capabilities by identifying trends created by millions of online users (Rogers, MacEwan and Pond, 2010). The blend of Web analytics and business intelligence would not only provide the descriptive analytics but only predictive analytic capabilities, by analyzing the demographic data, purchasing records and target advertising and hence influencing the future customer (Mobasher et al., 2000, Cho et al.). Academic Analytics (AA) is an offshoot of business intelligence, which is used purely for academic purposes. The term academic analytics is first coined by Goldstein and Katz (2005) in their study, which captures the importance of technical and managerial factors in the context of an academic institution. The success of a student in his career progression and his retention to the respective institute are key resource areas for an academic institution. Academic analytics shares a thin line of difference with Educational Data Mining (EDM), such as, the posterior attempts to identify patterns in the data and the anterior aims to be more predictive and highlight decision making as its bottom-line. The present study also is based on the academic analytics.

Learning analytics takes over the next generation learning challenges and opportunities, so that the key stakeholders of an academic environment such as instructors, students and other advisors, draws a real time benefit. The contemporary evaluation procedures are mundane and just descriptive in nature. In contrast to the existing scenario, learning analytics offers personalized support by capturing and analyzing data by ongoing basis and also optimizing the time taken. The rationale behind this modeling capability is by combining the past data with present data and extracts useful patterns to the needful (Eckerson, 2006).

II. LITERATURE REVIEW

In a review of learning analytics, Ferguson Rebecca (2012) examined the factors like Big data, online learning and economic concerns that cause learning analytics. Their analysis showed the process through which data driven analytics are created and how the learning analytics are materialized. Rebecca (2012) also established the degree to which LA is related to education data mining and academic analytics. Further, the authors recognised the challenges that can be resolved by the use of learning analytics like setting up unambiguous ethical strategies along with acknowledging upcoming learning technologies, understanding the perspectives of learners and working on extensive learning data sets. For a clearer understanding, Wolfgang Greller et al., (2012) provided a generic outline for learning analytics. The authors provided six critical sub aspects of learning analytics namely Objectives, Data, Instruments, Internal Limitations, External Constraints and Stakeholders and supported the ethical outlook of LA with a motive to defend the learners. Discussing about the attention data in the learning milieu, Erik Duval (2011) showed how attention metadata captured via posts, comments and messages can be stored and utilised. This study considered two approaches- first the learning dashboards and learning recommenders. In learning dashboards, a visual overview of user activities is provided, both individually and in connection to their peers. Learning recommenders is utilised to gauge and suggest resources grounded on the data collected on user behavior pattern. Putting forward a specific scaffold for course management system, Jie Zhang (2010) presented a framework which enables an analysis of student's usage and access pattern of eLearning system within a specific period of time. This framework included Data infrastructure and data analytical modules. On one hand, Data infrastructure module used Hadoop framework for distributed computation, distributed data storage and data broker service, Data analytical modules enables collection of data from cloud, adaption, refinement and optimization of data analytics flows and patterns of mining usage. Implicating for course manager, the study guided course development with due consideration and customisation of the material for individual students; identification of effective methods of learning and the preferred learning gadgets. Evaluating the learning pattern of students, Alyssa Friend Wise et al., (2013) investigated the way in which students contribute and respond to messages in online discussion in learning environment. The authors represented that the learning analytics is entrenched in and created from learning environment, thus causing an integrated model. In their key findings, this study revealed that the reflective journal, a way to record student goals, periodical evaluation score and student feedbacks to reflective questions, contributes towards productive class environment and effective participation of students. Moreover, it sorted invisible activity validation (like ability to capture listening data) as a valuable outcome of the study.

Tim Roger et al., (2014) examined the student at the risk of failure by conducting a comparative analysis on index method versus linear multiple regression method on a sample of student. The findings of this study showed that correlation estimates for both the methods were significantly related and fell in the same range of line. Further, it was shown that Index method, has slighter variability in prediction, although with moderately lower correlation, which is very important reflection in the prediction analysis. Sharon Slade et al., (2013) discussed the way in which learning analytics enables educational institution to collect information regarding their students learning behavior, apply it to upraise the student retention rate and execute in time intervention to cause student success. However, collection of student data required as input to predictive models face a number of ethical issues and challenges that has been the focus of this paper. Conducting a study on fifteen year period graduate student set of data, Mohammed M. Abu Tair et al., (2012) explored a case study on knowledge mining from educational data. This study showed that data mining is applicable for classification and clustering, identification of relationships, and execution of analysis for outliers. It highlighted how Rule Induction and Naïve Bayesian classifier can be utilized for classification and K-means as clustering method. The outlier analysis demonstrated that the outliers were product of rare events and not due to errors.

Kabakchieva D [2013] also emphasised on application of data mining technologies on student data which comprises student's personal, pre-university and university specific characteristics. The author utilised Rule learners, decision tree technique, Bayes classifier and Nearest neighbor techniques to analyse the data and provide the results. In this line, Osmanbegovic et. al. [2015] employed and compared three data mining strategies (Bayesian classifier, neural networks and the decision tree method - J48) through the use of WEKA software package. In this study, data was collected from student surveys, students past success and present success while the influence of input variables was examined through Chisquare, One-R, Info gain and Gain ratio tests. The outcome explained that GPA attribute affected the output most followed by the attributes - entrance exam, study material and average weekly hours devoted to study. Further it was also observed that Naïve Bayes envisaged better than the others and Multilayer Perceptron algorithm (neural networks) confirmed lowest prediction accuracy.

III. CASE SNIPPETS

A. Case 1

The present case attempts to identify the key predictors of 180 MBA student's performance from a business school of national repute in India, with respect to three parameters such as educational achievement, learning approach and personality type. The relationship among the three predictor variables and response variable was conducted using structural equation modeling. The responses were taken from the 180 students using a survey based questionnaire, which was constructed using adapted scales such as two-factor version of the study process questionnaire (R-SPQ-2F), consisting of 20 items, was given by Biggs et al (2001), International Personality Item Pool (IPIP) Five-Factor Personality Inventory, consisting of 20 items, was developed by Buchanan, (2001) and other demographic information of the students. The observed variable performance was measured using scores obtained in Common Admission Test. Personality traits have been finding relevance especially in psychology studies from the early twentieth century (Zhang, 2003). Among many sets of personality factors, Big five personality such as conscientiousness, openness to experience, extraversion, agreeableness and neuroticism has special treatment in the organizational studies and psychology literature (Costa and McCrae, 1995). The present case adopts the same five factors to measure the student's academic performance (McKenzie et al., 2004). Learning is a predominant factor for successful career progression. Approaches to learning are the key strategies for success and also encompasses the intention of the respective learner in which the information is being processed to knowledge (Garrison et al., 1995). There are two broad types of learning processes as suggested by Marton and Saljo (1976) are deep approach and surface approach. Deep approach underlines the synthetic outcome of learning, when it is compared with other experiences at more of a critical level. In contrast to it, surface approach defines learning in more of a memorizing the facts in a completely isolated manner (Fourie, 2003; Biggs et al. (2001).

The mean of the respondents was found out to be 27.82 years, with a standard deviation of 5.44. The gender composition was 62.73% and 37.27% as male and female employees respectively. Table I represents the summary of demographic information with respect to gender, age, level of education, job position and years of work experience.

TABLE I
SUMMARY OF RESPONDENT'S DEMOGRAPHICS

Variable	Type	%
Gender	Male	62.73
	Female	37.27
Age	20-26	27.15
	27-33	56.72
	34-40	17.21
	41-47	5.9
	48 &Above	None
Educational level	Graduate	75.5
	Post Graduate	NA
Job Position	Middle level manager	21.5
	High Level manager	4.89
Years of Work Experience	0 to 1	42.5
	1 to 2	29.1
	3 to 4	16.5
	5 to 6	7.5
	6 and Above	4.5

Due to the very nature of multiple source design of the study and arrangement of items as questions at different parts of the questionnaire, there exists a possibility of the presence of Common Method Bias (CMB), which in turn might impend the validity of the study results. CMB is the indication of single latent factor resilient a better fit, when compared to the remaining factor of the research model (Podsakoff et al. 2003). In order to check the presence of CMB, the study used Harman's Single factor approach (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) using varimax rotation. The authors have found that more than one factor showcases the eigen values, which were greater than 1 and the first factor accounted for 22.54 % out of the total variance explained, 74.32 %. Therefore the presence of common method bias is not a potential threat in the present study. Table II represents the descriptive statistics of the response set, which includes the control variables and the four central constructs of the present study along with their inter correlations and reliability coefficients.

TABLE II
SUMMARY OF DESCRIPTIVE STATISTICS

S.no	Variable	Mean	S.D	1	2	3	4	5	6
1	Age	27.82	5.44	-					
2	Gender	0.41	.67	-.16	**.91**				
3	Personality	5.26	.72	.32**	.41**	**.82**			
4	Performance	4.82	.63	-.27*	.38*	.17*	**.86**		
5	Deep Learning	5.14	.72	.42**	-.12**	.39**	-.25**	**.89**	
6	Surface Learning	5.06	.81	.36**	.19*	.31**	.12*	.14*	**.77**

Note. Bold values indicate the corresponding reliabilities (Cronbach's alpha). Age is measured in years. Gender is indicated as 1 = male, 0 = female. (*p < .05. **p < .01.)

1) Measurement Model

Measurement model encompasses three significant aspects such as Reliability, Discriminant Validity and Convergent validity (Hair et al., 2006). The estimates of this model represent the latent constructs as a weighted sum of all the observed constructs, which are as part of the research model. Confirmatory Factor Analysis (CFA) was used to access the sufficiency of convergent and discriminant validity of the four latent constructs. Fornell and Larcker (1981) suggested the procedures to test both validities of the respective scales. Discriminant validity was tested using the construct correlations (Kling, 2001). Convergent Validity is defined as the degree to which the items of any two given latent constructs of the same research model, which are theoretically related and in fact related (Campbell & Fiske, 1959). The results indicate that the construct correlations of all the latent constructs are more than or equal to 0.7 and the Average.

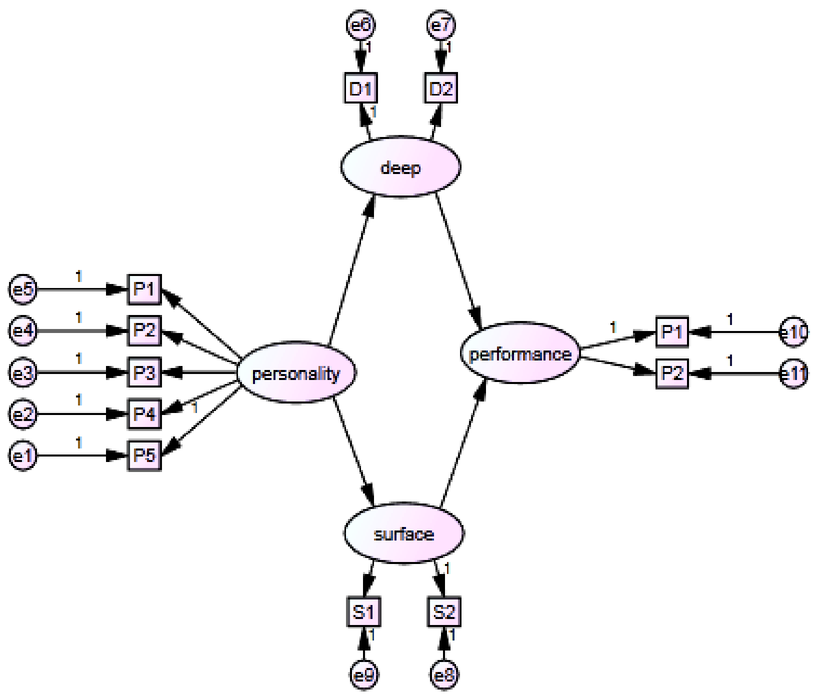

Fig 1. Proposed Research Model

Variance Extracted (AVE) values are more than 0.5. The AVE values of each respective construct were found out to be greater than the squared correlations, with respect to the remaining factors in the model (Gefen et al. 2000). This confirms the sufficiency of the discriminant andconvergent validity of the data and the same is represented in table III. The goodness of fit of the model was tested by using $\chi2/df$ = 3.92, Comparative Fit Index (CFI) = 0.93, Root Mean Square Error of Approximation (RMSEA) = 0.068 and Standardized Root Mean Square Residual (SRMR) = 0.073 respectively. All the path coefficients were significant at 1% and 5% level of significance and fit indices were within the acceptable limits as suggested by Hu & Bentler (1999). Fig. 1 represents the proposed research model and table III indicates the summary results of the measurement model.

TABLE III
RESULTS OF MEASUREMENT MODEL

Variables	Reliability	AVE	AVE Sq. Root
Personality	0.82	0.62	0.78
Performance	0.86	0.71	0.84
Deep Learning	0.89	0.68	0.82
Surface Learning	0.77	0.54	0.72

2) Structural Model

Structural Equation Modeling (SEM) was adopted to test the linkage between personality factors and performance via deep learning and surface learning. A step by step process was followed to validate this two mediator model. Initially, the direct effect of five personality factors on performance was tested, without involving either of the mediators. The standardized path coefficient for this relationship was found out to be (β =0.14, $p<$ 0.01) positively significant. Soon after this, the two mediators were invoked into the model along with the existing direct path. The results in the table IV showed adequate goodness of fit. Deep learning partially mediates the relationship between personality factors and performance (β =0.27, $p<$ 0.05) and surface learning displayed no mediation between personality factors and performance (β = -0.32, $p<$ 0.01).

TABLE IV
SUMMARY RESULTS OF STRUCTURAL MODEL

Structure Model	$\chi2$	d*f*	*P*	$\chi2/$ d*f*	NNFI	CFI	GFI	AGFI	RMR
Hypothesised Model	416.82	134	<0.01	3.11	0.84	0.92	0.88	0.79	0.084

Furthermore, the path coefficients of the three factors of the big five personality model, extraversion, openness and conscientiousness are high towards the deep learning approach. The findings of the present study are in line with the study conducted by Duff et al., (2004) and hence it can be concluded that personality traits influences the student's academic performance via learning approaches.

B. Case-2

A crowd sourced data containing five parameters on 2000 women (marital status, working, age, number of children & education) was obtained and it were recorded that 657 were recorded as not wage earners. The analysis was conducted using Linear Probability Model (LPM), Logit Model to obtain the marginal effects of variables, Probit Model and finally a comparison of all models was done to pick the best performing model.

A survey is conducted on 2000 women, to understand the factors that enable women to be the wage earners. The variables considered in this study which can influence a women's working chances are age, marital status, number of children and education. Totally four independent variables and one dependent variable (work) is taken for the analysis. Since the dependent variable in the study is categorical in nature, i.e it can take values of 0 and 1 only, Logit model and Probit Model can be applied. The results of both the model are analyzed and concluded that logit model is better than Probit Model. The coding of the variables is allocated as follows: a.) 1/0 – women works or not b) 1/0 – married /unmarried c) years of schooling is considered for education.

1) Linear Probability Model (LPM)

LPM is a special case of a binomial regression model. Here the observed variable for each observation takes values which are either 0 or 1. The probability of observing a 0 or 1 in any one case is treated as depending on one or more explanatory variables. For the linear probability mode, this relationship is a particularly simple one, and allows the model to be fitted by simple linear regression. Hence the parameters can be estimated using least squares. A coefficient is the change in the probability that Y = 1 for a one-unit change of the independent variable of interest, holding everything else constant. The disadvantage of this model is that it assumes linearity and predicts probabilities outside the boundaries of 0 and 1. Sometimes the probability values of LPM will be -1 and 2 also. This is the reason we go for Logit and Probit models. Estimating the simple OLS equation in the E-Views software, the following results are generated.

Work C Age Married Children Education

Dependent Variable: WORK
Method: Least Squares
Date: 11/17/13 Time: 09:30
Sample: 1 2000
Included observations: 2000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.207323	0.054111	-3.831436	0.0001
AGE	0.010255	0.001227	8.358393	0.0000
MARRIED	0.111112	0.021948	5.062567	0.0000
CHILDREN	0.115308	0.006772	17.02848	0.0000
EDUCATION	0.018601	0.003250	5.723597	0.0000

R-squared	0.202623	Mean dependent var	0.671500
Adjusted R-squared	0.201024	S.D. dependent var	0.469785
S.E. of regression	0.419920	Akaike info criterion	1.104990
Sum squared resid	351.7833	Schwarz criterion	1.118992
Log likelihood	-1099.990	Hannan-Quinn criter.	1.110131
F-statistic	126.7381	Durbin-Watson stat	1.957237
Prob(F-statistic)	0.000000		

Fig 2. LPM Results

All the variables are significant at 5% level of significance. The overall model is also significant. The Durbin Watson Value is close to 2, that indicates no presence of autocorrelation. Holding everything else constant one unit change in Age, will result in 0.01 or 1% increase in probability of women becoming a wage earner. Holding everything else constant one unit change in getting Marriage, will result in 0.111 or 11.1% increase in probability of women becoming a wage earner. Holding everything else constant one unit change in having Children, will result in 0.115 or 11.5% increase in probability of women becoming a wage earner. Holding everything else constant one unit change in Education, will result in 0.0186 or 1.86% increase in probability of women becoming a wage earner. Having compared all the four independent variables, the order of influence of women becoming a wage earner would be, Excluding the minor difference between children and marriage, we can conclude that both these factors deicide weather a women be a wage earner or not.

2) Logit Model

As with LMP and Probit, a predicted value is the predicted probability that Y = 1 given X. However, here the predicted values are calculated using the cumulative probability distribution function of the standard logistic.

$$\hat{Y} = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 X)}} = \frac{1}{1+e^{-z}}$$

Thus, as with Probit, the predicted values are bounded between 0 and 1. A coefficient is the change in the z-value for a unit change in the independent variable of interest, holding everything else constant (where 1/1+e−z is the estimated probability of Y=1). The estimated equation for Logit model in E-views is given below:

Work C Age Married Children Education

Dependent Variable: WORK
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 11/17/13 Time: 09:31
Sample: 1 2000
Included observations: 2000
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-4.159247	0.332040	-12.52635	0.0000
AGE	0.057930	0.007221	8.022434	0.0000
MARRIED	0.741777	0.126471	5.865221	0.0000
CHILDREN	0.764488	0.051529	14.83612	0.0000
EDUCATION	0.098251	0.018652	5.267537	0.0000

McFadden R-squared	0.188204	Mean dependent var	0.671500
S.D. dependent var	0.469785	S.E. of regression	0.416267
Akaike info criterion	1.032914	Sum squared resid	345.6898
Schwarz criterion	1.046917	Log likelihood	-1027.914
Hannan-Quinn criter.	1.038056	Deviance	2055.829
Restr. deviance	2532.445	Restr. log likelihood	-1266.223
LR statistic	476.6162	Avg. log likelihood	-0.513957
Prob(LR statistic)	0.000000		
Obs with Dep=0	657	Total obs	2000
Obs with Dep=1	1343		

Fig 3. Results of Logit Model

Results indicate that all the Variables are significant. Over Model is also significant as evident from the F- Statistic. In Logit model, R-Squared will not be of much use. Interpretation in odds will be more meaningful, which can be obtained by taking anti log of the slope coefficients. A one year increase in age corresponds to a 1.05 increase in the *odds* of a woman being a wage earner and the corresponding probability (Pi value) is 51.2%. A one year increase in age corresponds to a 2.09 increase in the *odds* of a woman being a wage earner and the corresponding probability (Pi value) is 67.63%. A one year increase in age corresponds to a 2.14 increase in the *odds* of a woman being a wage earner and the corresponding probability (Pi value) is 68.15%. A one year increase in age corresponds to a 1.1 increase in the *odds* of a woman being a wage earner and the corresponding probability (Pi value) is 52.38%.

3) Probit Model

A Probit model is a type of regression where the dependent variable can only take two values, for example married or not married. The name is from *probability* + *unit*. The purpose of the model is to estimate the probability that an observation with particular characteristics will fall into a specific one of the categories; moreover, if estimated probabilities greater than 1/2 are treated as classifying an observation into a predicted category, the Probit model is a type of binary classification model. A Probit model is a popular specification for an ordinal or a binary response model. As such it treats the same set of problems as does logistic regression using similar techniques. The Probit model, which employs a Probit link function, is most often estimated using the standard maximum likelihood procedure, such an estimation being called a Probit regression. The predicted values are calculated using the cumulative probability distribution function of the standard normal.

$$\hat{Y} = \Phi(\hat{\beta}_0 + \hat{\beta}_1 X) = \Phi(z)$$

Thus, the predicted values are bounded between 0 and 1. A coefficient is the change in the z-value for a unit change in the independent variable of interest, holding everything else constant (where Φ(z) is the estimated probability of Y=1). The estimated equation for Probit model in E-views is given below:

Work C Age Married Children Education

Dependent Variable: WORK
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 11/17/13 Time: 09:33
Sample: 1 2000
Included observations: 2000
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.467365	0.192563	-12.81326	0.0000
AGE	0.034721	0.004229	8.209604	0.0000
MARRIED	0.430857	0.074208	5.806078	0.0000
CHILDREN	0.447325	0.028742	15.56362	0.0000
EDUCATION	0.058365	0.010974	5.318351	0.0000

McFadden R-squared	0.188878	Mean dependent var	0.671500
S.D. dependent var	0.469785	S.E. of regression	0.416219
Akaike info criterion	1.032062	Sum squared resid	345.6103
Schwarz criterion	1.046064	Log likelihood	-1027.062
Hannan-Quinn criter.	1.037203	Deviance	2054.123
Restr. deviance	2532.445	Restr. log likelihood	-1266.223
LR statistic	478.3219	Avg. log likelihood	-0.513531
Prob(LR statistic)	0.000000		

Obs with Dep=0	657	Total obs	2000
Obs with Dep=1	1343		

Fig 4. Results of Probit Model

Results indicate that all variables are significant at 5% level of significance. The overall model is also significant. One year increase in age of women, the Probit will increase by 0.034. The higher the probability of the probit, more is the chance to say yes.

Similarly, all the remaining coefficients can be explained. The estimated probability for one individual, can be calculated as follows, Taking a women of age= 33, education= 10, Marriage=1, children=1.

$$\text{Probit} = -2.46 + 0.034(33) + 10(0.05) + 1(0.43) + 1(0.44)$$
$$\text{Probit} = 0.032$$

We can convert this figure into probability sense, we use Z-Table. $P(Z>0.032) = 0.50 - 0.0120 = 0.488$ and Probability (Probit) = 0.488. So this particular woman has a chance of 48.8 % to be a wage earner. In this way, the estimated probabilities are calculated for all the other 1999 women.

4) Which model is better?

As is evident from the above results, although the coefficients are slightly different between the logit and Probit models, the z-statistics and significance levels are almost identical. Since it is easier to interpret the Logit results, it might be preferable to use those.

TABLE V
MODEL COMPARISON

Factor	Logit	Probit
McFadden R-squared	0.188204	0.188878
S.E. of regression	0.416267	0.416219
LR statistic (4 df)	476.6162	478.3219

By comparing the results, we can see that all the three factors are almost the same. Since the complexity of calculating and interpreting the coefficients of Probit is cumbersome, we prefer to go with the Logit Model.

C. Case 3

A Management science (institutes global elective credit course) class containing 382 undergraduate students from six different branches (Electronics & Communications, Computer Science, Electrical and Electronics, Mechanical, Civil and Bio technology) of engineering in a university of national repute is considered, to predict their final course grade. Seven predictor variables (Class Standing, Grade Point Average, Anticipated Grade, requirement of the course for a student's major, Approaches to Study Inventory and Course Valuing Inventory) are identified from the literature and multiple linear regressions was used to identify and rank the predictor variables.

In order to collect the data from the 382 undergraduate students, the authors have adopted two scales which are highly prominent in the literature (ETL Project, 2002; Nehari & Bender, 1978): Course Valuing Inventory (CVI) and Approaches to Study Inventory (ASI). The authors have taken the liberty to contextualize the scales with respect to the verbatim and the Indian under graduate students. CVI helps to ascertain the affective component of learning and ASI limits itself to understand the study related academic behaviors. CVI contains a total of 36 items spread across four dimensions such as cognitive content, cognitive valuing, affective personal and behavioral (Nehari & Bender, 1978; Sturges et al., 2012). High mean values on the second order latent construct indicates that the students' level of valuing. ASI, which is more inclined towards the learning approaches of the students', was developed by Entwistle, Hanley, & Hounsell (1979) with 60 indicators. A contextualized and more shortened version was used in this study and finally contains 18 items to measure the latent construct of ASI (ETL Project, 2002). The combined effect of these two significant components would influence the student's academic performance.

A total of 382 under graduate students who are enrolled in the institute's global elective course titled 'Management Science', are asked to complete the survey questionnaire which consists of 62 items (18 of ASI, 36 of CVI and 8 of demographics). The students are from the six branches of engineering sciences such as electronics & communications, electrical, computer science, mechanical, civil and bio technology departments. Responses are collected for about ten days and the authors received 356 completely filled questionnaires, leaving a response rate of 93.1%. The higher response rate is due to the stringent academic environment.

TABLE VI
DESCRIPTIVE STATISTICS

Constructs	Mean	Std. deviation	1	2	3	4	5
Cognitive content	3.56	0.082	1.00				
Cognitive valuing	3.13	0.129	.342**	1.00			
Affective personal	4.13	0.136	.217**	.592**	1.00		
Behavioral	3.92	0.303	.519*	.423*	.451*	1.00	

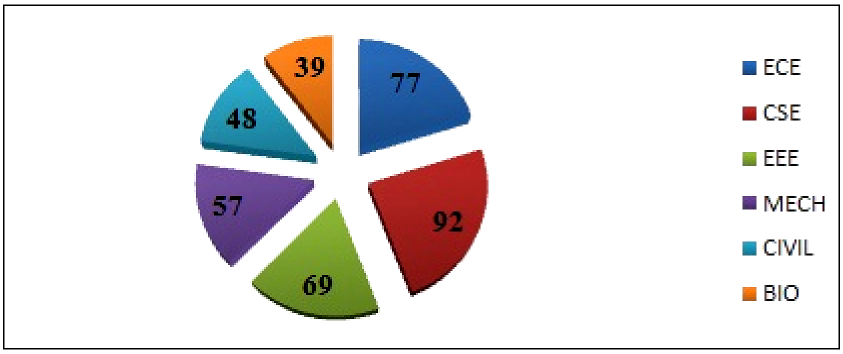

Fig 5. Class Demographics

The descriptive statistics represented as part of table VI, indicated that all the variables have significant relationships with each other. Fig 5, indicated the class demographics with respect to various branches of engineering. A multiple regression model was conducted to predict the final course grade with seven independent variables. The results (Table VII) indicated that the model was significant and all the independent variables are also significant at 5% level of significance and the coefficient of determination (R^2) was found 0.53. Table VII represents the standardized regression coefficients along with their respective standard errors and significance values. Grade point average followed by anticipated grades is the strongest predictors of the students final grade and course value inventory remained insignificant.

TABLE VII
MULTIPLE REGRESSION RESULTS

Variable	B	SE	β	t	p
CS	0.21	0.16	0.15	3.26	0.007
GPA	0.16	0.18	0.36	2.17	0.042
Required	1.02	0.28	0.29	1.35	0.013
CTM	1.04	0.13	0.19	3.14	0.011
AG	0.18	0.19	0.31	4.52	0.001
ASI	0.12	0.05	0.27	3.07	0.027
CVI	0.13	0.14	0.06	1.06	0.213

Note. (N = 382). CS- Class Standing, GPA- Grade Point Average, AG- Anticipated Grade, ASI- Approaches to Study Inventory and CVI - Course Valuing Inventory.

IV. DISCUSSION AND CONCLUSION

The present study attempts to empirically put forth two different case studies as an offshoot of learning analytics. Decision making in twenty first century has become so handy with the exponential advancement of the computing power. Traditional way of analyzing issues are very much mundane and time consuming. The domain of analytics as the science of data, is the next big thing in the decades to come. The first case as part of the study deals with the performance of the MBA students from a business school of national repute in India. Due to the increasing number of under graduates opting for MBA as a bright opportunity, the business schools feel a pressing need to align them towards the cut throat competitive business world. The performance with right attitude would make him/her go places. Hence, an empirical investigation was conducted to identify the hall mark of predictors which have the power to instigate an individual's true capabilities. The results indicated that a good understating and practice of the big five personality factors and a deep leaning approach, would leave no stone unturned to clinch his/her success path. The second case aims to test the probability of women being a wage earner from a crowd sourced data on five given parameters. Linear probability model, logit model and probit model were used as techniques to understand the innate relationships. By comparing the results, the authors concluded that all the three factors are almost the same. Since the complexity of calculating and interpreting the coefficients of Probit is cumbersome, the authors prefer to go with the Logit Model. Finally, the third case deals with the factors that predict the students' final grade. Seven factors are identified from the past literature and undergraduate students from a university of national repute who opted for management science elective, are considered for this study. Results indicated that students' grade point average is the strongest predictor of students final grade.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Alyssa, F.W., Yuting, Z., and Simone, N.H. (2013). "Learning analytics for online discussions: a pedagogical model for intervention with embedded and extracted analytics". Third International Conference on Learning Analytics and Knowledge, pp. 48-56.

[2] Biggs, J., Kember, D., and Leung, D. (2001). "The revised two-factor study process questionnaire: R-SPQ-2F". British Journal of Educational Psychology, Vol. 71, Pt 1, pp. 133-49.

[3] Buchanan, T. (2001). "Online implementation of an IPIP five factor personality inventory". <http://users.wmin.ac.uk/-buchant/wwwffi/introduction.html>.

[4] Baker, B. (2007). "A conceptual framework for making knowledge actionable through capital formation". D.Mgt. dissertation, University of Maryland University College, United States Maryland. ABI/INFORM Global. (Publication No. AAT 3254328), Accessed 19 October 2010.

[5] Campbell, D.T. and Fiske, D.W. (1995). "Convergent and discriminant validation by the multitrait-multimethod matrix". Psychological bulletin, 56(2), pp. 81.

[6] Costa, P.T. and McCrae, R.R. (1995). "Domains and facets: hierarchical personality assessment using the revised NEO personality inventory". Journal of Personality Assessment, Vol. 64, pp. 21-50.

[7] Duff, A., Boyle, E., and Dunleavy, J.F. (2004). "The relationship between personalities, approach to learning and academic performance". Personality and Individual Differences, Vol. 36, pp. 1907-20.

[8] Eckerson, W.W. (2006). "Performance dashboards: Measuring, monitoring, and managing your business". Hoboken, New Jersey: John Wiley & Sons.

[9] Erik, D. "Attention please! Learning analytics for visualization and recommendation". 1st International Conference on Learning Analytics and Knowledge, 2011, pp. 9-17.

[10] Edin, O. and Mirza, S. "Data Mining Approach for predicting student performance". Economic Review - Journal of Economics and Business, Vol. 10, Issue 1, May 2012, pp. 3-12.

[11] Enwistle, N., Hanley, M., and Hounsell,

D. (1979). "Identifying distinctive approaches to studying". Higher Education 8: 365-80

[12] ETL project reports. (2002). <http://www.etl.tla.ed.ac.uk//publications.html>. Accessed 25 October 2013.

[13] Ferguson, R. "Learning analytics: drivers, developments and challenges". International Journal of Technology Enhanced Learning, 4(5/6), pp. 304-317.

[14] Fornell, C. and Larcker, D.F. (1981). "Evaluating structural equation models with unobservable variables and measurement error". Journal of marketing research, pp. 39-50.

[15] Garrison, R., Andrews, J., and Magnusson, K. (1995). "Approaches to teaching and learning in high education". <www.ucalgary.ca/pubs/Newsletters/Currents/Vol2.1/ approaches.html>.

[16] Goldstein, P.J. and Katz, R.N. (2005). "Academic Analytics: The Uses of Management Information and Technology in Higher Education". ECAR Research Study Vol. 8.

[17] Gefen, D., Straub, D., and Boudreau, M. C. (2000). "Structural equation modeling and regression: Guidelines for research practice". Communications of the association for information systems, 4(1), pp. 7.

[18] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., and Tatham, R.L. (2006). "Multivariate data analysis". (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.

[19] Hu, L.T. and Bentler, P.M. (1999). "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives". Structural Equation Modeling, 6 (1), pp. 1-55.

[20] Jie, Z., William, C.T., Bu Sung Lee, K. K.L., Julita, V., and Chee, K.L. "A Framework of User-Driven Data Analytics in the Cloud for Course Management". 18th International Conference on Computers in Education, Putrajaya, Malaysia, Nov 29 - Dec 3 2010, Asia-Pacific Society for Computers in Education.

[21] Kabakchieva, D. "Predicting Student Performance by Using Data Mining Methods for Classification". Cybernetics and Information Technologies, Vol. 13, No. 1, pp. 61-72.

[22] Kling, K.C. (2001). "The role of personality, academic ability, and gender in predicting academic achievement". Poster presented at the 2001 meeting of the International Society for the Study of Individual Differences. Edinburgh: July.

[23] McFadden, C. (2005). "Optimizing the Online Business Channel with Web Analytics". <http://www.webanalyticsassociation.org/members/blog_view.asp?id=533997&post=89328&hhSearchTerms=definition+and+of+and+web+and+analytics>.

[24] Mobasher, B., Colley, R., and Srivastava, J. (2000). "Automatic personalization based on web usage mining". Communications of ACM 43(8), pp. 142-151.

[25] Mohammed, M., Abu Tair, A.M., and El, H. "Mining Educational Data to Improve Students Performance: A Case Study". International Journal of Information and Communication Technology Research, Vol. 2, No. 2, February 2012, ISSN 2223-4985, pp. 140-146.

[26] McKenzie, K., Gow, K., and Schweitzer, R. (2004). "Exploring first-year academic achievement through structural equation modeling". Higher Education Research and Development, Vol. 23 No. 1, pp. 95-112.

[27] Marton, F. and Saljo, R. (1984). "Approaches to learning". in Marton, F., Hounsell, D. and Entwistle, N. (Eds), The Experience of Learning, Scottish Press, Edinburgh, pp. 36-55.

[28] Nehari, M. and Bender, H. (1978). "Meaningfulness of a Learning Experience: A Measure for Educational Outcomes in Higher Education". Higher Education, 7 (1), pp. 1-11.

[29] Osmanbegović, E., Suljić, M., and Agić, H. (2015). "Determining Dominant Factor for Students Performance Prediction by Using Data Mining

Classification Algorithms". Tranzicija, 16(34), pp. 147-158.

[30] Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., and Podsakoff, N.P. (2003). "Common method biases in behavioral research: a critical review of the literature and recommended remedies". Journal of applied psychology, 88(5), pp. 879.

[31] Rogers, P.C., McEwen, M.R., and Pond, S. (2010). "The design and evaluation of distance education". In Veletsianos, G. (Ed.) Emerging technologies in distance education, pp. 231-247.

[32] Sharon, S. and Prinsloo, P. "Learning analytics: ethical issues and dilemmas". American Behavioral Scientist, Oct 2013, Vol. 57, No. 10, pp. 1510-1529.

[33] Sturges, D., Maurer, T., and Dobson, J. (2012). "Allied Health Students' perceptions of learning in undergraduate Human Anatomy and Physiology classes". International Conference for Academic Disciplines, Toronto, Canada.

[34] Slade, S. and Prinsloo, P. (2013). "Learning analytics: ethical issues and dilemmas". American Behavioral Scientist, 57(10) pp. 1509-1528.

[35] Tim, R. and Cassandra, C. "Modest analytics: using the index method to identify students at risk of failure". Fourth International Conference on Learning Analytics and Knowledge, 2014, pp. 118-122.

[36] Wang, T. and Ren, Y. (2009). "Research on Personalized Recommendation Based on Web Usage Mining Using Collaborative Filtering Technique". WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS 1(6).

[37] Wolfgang, G. and Hendrik, D. "Translating Learning into Numbers: A Generic Framework for Learning Analytics". Educational Technology & Society, Vol. 3, Issue 5, ISSN: 1436-4522, pp. 42-57.

[38] Zhang, H. and Almeroth, K. (2010). "Moodog: Tracking Student Activity in Online Course Management Systems". Journal of Interactive Learning Research, 21(3), Chesapeake, VA: AACE. pp. 407-429.

[39] Zhang, L. (2003). "Does the big five predict learning approaches?". Personality and Individual Differences, Vol. 34, pp. 1431-46.