International Journal of Education and Development using Information and Communication Technology (IJEDICT), 2018, Vol. 14, Issue 1, pp. 33-51

Building and evaluating logistic regression models for explaining the choice to adopt MOOCs in India

Sangeeta Trehan and Rakesh Mohan Joshi
Indian Institute of Foreign Trade, India

ABSTRACT

Logistic regression is a popular tool used to build and evaluate binary choice models. It has been applied in a variety of situations and contexts involving dichotomous choice. In the current paper, we apply it to explain and predict the individual choice of adopting online learning through a Massive Online Open Course (MOOC), a specific artefact in the domain of online, open learning. The MOOC holds promise for a developing country like India to scale quality education to keep up with an overwhelming demand of its large population. Hence the importance of the choice to enter MOOCs in the Indian context cannot be overemphasized. Factors possibly driving MOOC adoption were selected after an extensive literature study and a binary logistic regression model was applied to analyze their influence using data from a primary survey. A variety of different subsets of possible explanatory variables under consideration were experimented with. Due diligence was exercised with respect to model selection and evaluation and a 'best' model was identified and discussed. Among other things, our proposed model suggested that a respondent's online communicative efficacy was the strongest predictor of MOOC usage and similarly their preference for self-directed learning. In terms of practice and policy, the results of this study underline the need for strategies for enhancing digital literacy, online efficacy, self-directed learning and e-readiness of the prospective higher education aspirants in order to help with wider MOOC outreach and adoption in a developing country like India.

Keywords: *Massive Open Online Course; MOOC; model building; model evaluation; logistic regression; Akaike Information Criterion; Schwartz Criterion; Receiving Operating Characteristic curve; ROC curve; cross validation*

INTRODUCTION

Massive Open Online Courses (MOOCs) have been growing every year since 2012 globally (Shah 2016). Trehan et al. (2017) report that the concept of the MOOC is perceived as valuable for a developing country like India in several ways. But for MOOCs to become a part and parcel of the future of higher education in India, the learner choice to adopt the MOOC is critical. Also, it is imperative to discover and promote the drivers of MOOC adoption and fulfill the potentially new demands of a MOOC community. This motivated the current authors to consider the learner-level characteristics and traits facilitating MOOC adoption and to build and evaluate logistic regression models for the binary choice of MOOC adoption (i.e., the learner choice to enroll in a MOOC).

India mainly remains a consumer of the branded MOOC without substantially jumping on the MOOC bandwagon itself (Trehan et al. 2017) - learners from India have subscribed to international MOOCs in a massive way. India has figured among the top three nations with respect to overall enrolments since the inception of the popular x-MOOC form in 2012 (Bhattacharyya 2013; Ho et al. 2014). In 2016 the population of India's MOOC learners was the third largest internationally after the U.S.A and Brazil and ahead of China at the fourth place (Shi & Yu 2016). Similar MOOC activity statistics and statistics about certificate earners found in other international reports and papers (Ho et al. 2014; Jordan 2014; Ho et al. 2015 and Chuang & Ho 2016) evidence the fact that MOOCs have been a popular choice amongst learners from India.

Despite this, apart from some broad generalizations such as that an Indian MOOC-participant, by and large, is from metropolitan and urban areas, is well-educated, having a college degree and, in most cases, employed too (Christenson & Alcorn 2013), not enough is known regarding what factors characterize and drive a potential MOOC-user to make the behavioural choice of 'taking the plunge' into MOOCs.

The authors conducted a survey to study the profiles of MOOC users and non-users from India as well as the characteristics facilitating MOOC adoption. In the current paper we report about our attempt to build and evaluate a useful logistic regression model for the binary choice of MOOC adoption. The paper is organized as follows. In the next section we discuss the study background and design. The third section introduces the logistic regression model. In the next two sections we report and discuss the results from the model construction and evaluation phases. Next results for the 'best' model are presented and discussed. The final section unpicks the wider discourse and lacunae around MOOCs and discusses the implications of our study and related research for practice and policy. The limitations of our study and some suggestions for future research are also noted in this last section.

BACKGROUND AND METHODOLOGY

Literature study of the international and India-centric MOOC literature was conducted for background study on MOOC. The corpus of 60 journal articles on MOOCs published from January 2008 to May 2014 used by Raffaghelli et al. (2015) was the starting point. To this were added a set of 30 more papers, books and reports, besides blogs and other online resources for background study on MOOC. Scholarly literature arising out of user-focused, educational theory-grounded research on MOOC, it was noted, was limited and emergent (Lewis 2014). On the other hand, MOOCs are acknowledged to be a web-based information system and a specific e-learning artifact for distance learning. Hence, in order to gain perspective on learners' choice to adopt MOOCs, we also sifted through the extant literature on factors facilitating distance and e-learning. We do not have the opportunity to report about our comprehensive review of this literature here. But this allowed identification of literature gaps and helped select parameters for our study.

Several learner-level attributes related to their demography, personality and capability, preferred learning-style and learning motivation have been discussed in the literature, mostly as determinants of learner engagement and performance or course outcome in distance and/or online learning (Kim & Schniederjans 2004; Baruch, Bezalel & Barth 2007; Keller & Karau 2013; Scardilli 2013 and Pope 2014). Four broad classes of such attributes discussed in the past literature are noteworthy here, namely, Internet self-efficacy (Peng, Tsai, & Wu 2006), personality attributes (Keller & Karau 2013), learning motivation (Lim & Kim 2003) and learning styles (Grasha & Yangarber-Hicks 2000). Our research objective is to utilize these learner attributes to characterize MOOC users and non-users and as determinants of the choice to adopt MOOCs. So for the purpose of our survey, after due diligence we chose a set of ten independent variables that consisted of two Internet self-efficacy variables (one each for general and communicative self-efficacy), three learner personality-related variables (perseverance, creativity and inclination for learning new and different things), two proxies for measuring learning motivation (life goal and scheduled and planned approach to learning), two learning style-related variables (preference for self-directed learning and collaborative learning style) and one variable related to preference for educational videos as a medium for learning since video content is extremely important in MOOCs (Guo, Kim, & Rubin 2014). Age, gender, education and employed were the four control variables. Thus we had fifteen study variables in all including the dependent variable (MOOC usage).

'MOOC usage' was conceived as a dichotomous (0, 1) variable. 'Age' was measured as a continuous ratio scale variable and 'Education' as a categorical variable. Model variables related to Internet self-efficacy, personality, learning style and learning motivation were measured as scale variables using a single-item 5-point or 7-point Likert scale. Use of single-item measures was made instead of formal multi-item scales. For instance, with respect to measuring Internet self-efficacy we took a cue from the exploratory factor analyses conducted by Peng, Tsai, & Wu (2006) which resulted in two main factors of general and communicative self-efficacy. 'General Internet Self-Efficacy' (labeled Info_Process) measures one's confidence in their use of the Internet in general during information searching, processing and way-finding through the Internet whereas 'Communicative Internet Self-Efficacy' (Int_Comm) measures one's confidence in and facility for Internet-based communication or interaction. We devised simple single-item measures on a 5-point scale to measure these respectively, taking values 1 for those who self-report having 'very poor' efficacy, through 5 for those who self-report having 'very good' such efficacy. The remaining six variables were conceived as binary variables. The questionnaire items were pre-tested for reliability and validity and refined. The final survey instrument was used to collect data from among the current and past students of three premier management and technical institutes in Delhi, Chennai and Udaipur during 2016. There were no pre-set criteria for subjects, except having an email account and being 18 years or older. Responses gathered from both online and offline respondents were clubbed for analysis. The survey garnered a total of 441 unique responses out of which 28 were incomplete and/or inconsistent and hence were dropped. The following analysis is based on 413 complete responses. We presented a descriptive analysis of the self-reported profiles of MOOC users and non-users based on the results of this survey elsewhere. Here we use the same survey data to further investigate the choice to enter a MOOC using a discrete choice model. R and Tanagra are primarily used for data preparation and statistical analysis as needed.

A LOGISTIC MODEL

A Brief Description

Table 1: *Description of a Logistic Model*

Response (MOOC usage)	Independent or Predictive Variables ($\boldsymbol{x}$ *and* $\boldsymbol{z}$)	Odds ratio for MOOC usage	Log of the odds ratio for MOOC usage	Unknowns/ parameters
A discrete, dichotomous variable with value 1 for those who have enrolled for at least one MOOC and 0 otherwise.	The ten chosen independent variables , $X_1 \dots X_{10}$ and the four selected demographic variables $Z_1, \dots, Z_4$	$p/(1\text{-}p)$ where p is the probability for MOOC usage to be 1	$Y = \log(p/(1\text{-}p)) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \gamma_1 Z_1 + \cdots + \gamma_k Z_k + \epsilon$	p, the probability; $\beta_0, \beta_1 \dots \beta_n$ and $\gamma_1, \dots \gamma_k$ the coefficients estimated from the data and variance of ϵ, the random error term.

A logistic model is useful whenever the response has a two-level outcome or event and is thought to be influenced by one or more independent or predictive variables. In order to use regression, the dependent variable is transformed into a continuous value that is a function of the probability of the event occurring (Rud 2001). Similar to linear regression, logistic regression is based on a statistical distribution. Therefore, it enjoys the same benefits as linear regression as a robust tool for developing analytical models (Rud 2001). Logistic regression was considered suitable here since the dependent variable (MOOC usage) is not continuous; rather it is a discrete, dichotomous variable with value 1 for those who have enrolled for at least one MOOC and 0 otherwise. Table 1 describes the model as applied in our case.

Some Testable Hypotheses

Past research has explored relationships about the effect of the proposed explanatory variables on the learners' online/ distance learning choices, preferences and outcomes (Battalio 2009; Bonk et al. 2013; Kaveri et al. 2015; Milligan, Littlejohn, & Margaryan 2013; Sadera 2014; Scardilli 2013). On similar lines, we devised ten hypotheses (H1 through H10) to test the impact of the chosen independent variables on the user adoption of MOOCs.

H1: Individuals who are identified by personality trait of 'perseverance' are more likely to be MOOC users than non-users.
H2: Entry into MOOCs is positively related to 'creativity'.
H3: Individuals who are inclined 'to learn new and different things' are more likely to be MOOC users than non-users.
H4: Individuals with higher order General Internet Efficacy are more likely to be MOOC users than non-users.
H5: Individuals with higher order Communicative Internet Efficacy are more likely to be MOOC users than non-users.
H6: Individuals with higher preference for self-directed learning as a learning style are more likely to be MOOC users than non-users.
H7: Individuals having higher learning motivation manifested by scheduling and planning carefully are more likely to be MOOC users than non-users.
H8: Individuals who indicate 'videos as their preferred medium for learning online' are more likely to be MOOC users than non-users.
H9: Learners with collaborative learning style preference are more likely to be MOOC users than non-users.
H10: Individuals who give extremely high importance to feeling a sense of accomplishment as their life goal are more likely to be MOOC users than non-users.

RESULTS AND DISCUSSION - MODEL CONSTRUCTION

This section describes the results of our process of constructing a model for the decision to enter MOOCs - both without variable selection and with variable selection through strategies of forward and backward logit. Both these are automatic variable selection methods which allow one to specify how independent variables are entered into the analysis. In forward logit we start with a null model (with no predictor variables, only an intercept) and at each step add the variable that gives the biggest improvement to the current model. In backward logit, on the other hand, we begin with a full model with all the predictors and then iteratively remove the least useful predictors (the variables with the largest p-value) one by one until a stopping rule is reached.

Construction of the Regression Models

A series of models were built with use of all 413 data values choosing various subsets of the independent variables under consideration starting from model # 1 (0 Descriptors/ Predictors –

only intercept) through the model with all 14 Descriptors/ Predictors and the model outputs for alternative models were examined. As for model selection, Akaike Information Criterion (AIC) (Bozdogan 1987) and Schwarz Criterion (SC) (Pauler 1998) are two widely used information criteria. They reflect a trade-off between the goodness of fit and the complexity of the model. The smaller they are the better the fit of the model is (from a statistical perspective) as they compensate between the lack of fit and the number of parameters in the model. For example, the Akaike criterion reads $-2\log(\ell)+2k$, where k is the number of parameters. Different models were identified to be the 'best' using each of these criteria separately.

The Odds Ratios and Tests of Hypotheses for the 'Full' Model

Using AIC, the 'full' model with 14 predictors is the model selected (AIC = 388.481 is the least). Table 2 provides odds ratios and 95% confidence intervals for them for testing the hypothesized relationships. Odds ratios have interesting interpretations. For instance, among the respondents, men were more than twice (2.15 times) as likely to enroll into MOOCs as women. For every year of increase in age, the likelihood to enrol in MOOC increased by about 1.2 times. Individuals who reported higher order online communication capability (Int_Comm) by one point on the scale of measurement were about 8 times more likely to be MOOC users than non-users (H4 is upheld). Similarly, those with self-reported higher order online way-finding and information processing capability (Info_Process) by one point were about twice as likely to be MOOC users than non-users (H5 is upheld). Individuals who reported high preference for self-directed learning as a learning style were 2.78 times more likely to be MOOC users than non-users (H6 is upheld). Similar hypotheses about Video (H8) and New_Learn (H3) were also upheld while hypotheses H1, H2, H7, H9 and H10 about Perseverance, Creative, Planning, Collab and Life_Goal did not seem to be upheld by our data.

***Table 2**: Odds Ratios and 95% Confidence Intervals - the 'Full' Model*

Attribute	Coef.	Low	High
Gender	2.1524	1.1806	3.9242
Education	2.1062	0.9922	4.4709
Employed	0.1533	0.0624	0.3764
Age	1.1952	1.0344	1.3811
Info_Process	1.9845	1.1367	3.4646
Int_Comm	7.9675	4.5262	14.0252
Perseverance	0.3975	0.1895	0.8342
Creative	0.5654	0.2899	1.1028
New_Learn	1.5891	0.9468	2.6674
Self-directed	2.7825	1.7851	4.3370
Planning	1.0588	0.6078	1.8445
Video	1.6827	0.7981	3.5477
Collab.	0.3123	0.1481	0.6585
Life_Goal	0.3952	0.2447	0.6382

Variable Selection through Forward-Logit and Backward-Logit

It is widely held that a good model should explain the data 'well enough' while being simple. We noted that the full model with lowest AIC had all 14 independent variables, some of which were insignificant. So in order to find the best subset of predictors, we looked at methods for variable selection. We applied both forward-Logit and backward-Logit to be able to select the most relevant attributes for explaining MOOC usage. Some statistics for the resulting models (Models

#8 & #9 in Table 9) are displayed in the last two rows of Table 9. Between the two, the model resulting from applying backward-Logit (henceforth also called the 'Reduced Model' or Model #9) is preferable as per both the criteria – AIC and SC. Indeed, having the least value of SC = 424.118 among all possible subsets, it is the preferred model even in comparison to the full model on the basis of the Principle of Parsimony (Seasholtz & Kowalski 1993) or Occam's razor (Heylighen 1997; Braithwaite 2007). It has the six predictors - Gender, Employed, Age, Int-Comm, Self-directed and Life Goal. The odds ratios and tests of hypotheses for this reduced model are presented subsequently in the section titled Results and Discussion - the Best Model. Further, we carried out detailed diagnostics and evaluation of the various models. In the next section we describe the results of these diagnostics applied on the two (full and reduced) shortlisted models above.

RESULTS AND DISCUSSION - MODEL EVALUATION AND DIAGNOSTICS

In the previous section we presented some results from the model construction exercise. It was noted that the 'full' model was the 'best' model as per AIC while the reduced model, Model #9 in Table 9, found by variable selection through backward-Logit was the 'best' one as per SC. In this section we perform model evaluation and diagnostics of these two models. We use three kinds of tools and techniques including train-test process and k-fold cross-validation for knowing how good the proposed model is, how well it fits the data, which predictors are most important and if the predictions are accurate.

1) Goodness of Fit (Likelihood Ratio (LR) test, Pseudo R2's and Hosmer-Lemeshow tests)
2) Statistical Tests for Individual Predictors (Wald tests and Variable Importance), and
3) Validation of Predicted Values (Classifier Performances, ROC Curve, The Train-Test Process and K-Fold Cross Validation)

Evaluation and Diagnostics for the 'Full' Model, the Best Model as per the AIC Criterion

Goodness of Fit

A logistic regression is said to provide a better fit to the data if it demonstrates an improvement over a model with fewer predictors. The Likelihood Ratio (LR) test of the null hypothesis H_0 (that the reduced model with no predictors is true) led us to reject H_0. We found strong evidence in favour of the current model as the observed difference in model fit was highly statistically significant (p-value =0.0000). The model seems to have low to moderate predictive power as McFadden's R^2 value is on the lower side (0.37) and all the other Pseudo R^2's too are moderate (Cox and Snell's R^2 = 0.4042 and Nagelkerke's R^2 = 0.5390). However, based on the Hosmer-Lemeshow criterion, it seemed our model did not fit the data well since the p-value for the statistic value of 64.0614 with 8 degrees of freedom for testing the null hypothesis of a 'good' fit was 0.000.

Statistical Tests for Individual Predictors

The relative importance of individual predictors in the model was assessed using Wald tests and the absolute value of the t-statistic for each predictor. The null hypothesis for the Wald test postulates that an independent variable may be dispensed with. The five attributes of Planning, Video, Creative, New_Learn and Education with large p-values of Wald tests did not seem to contribute much to the fit of the model and could therefore be dispensed with. Similar to the result for the Wald tests above, utilizing the absolute value of the t-statistic for each model parameter also we identified the attributes of Planning, Video, Creative, New_Learn and Education as variables with relatively lesser importance ($|t|$ statistic less than 2).

Validation of Predicted Values

Classifier Performances

Our model serves as a binary classifier – with people classified as a MOOC user or not (positive value being 'user'). We used several statistics using the true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) data from the so-called 'confusion matrix' (Table 3) to describe the model's classification performance on our dataset for which the true values are known (Table 4).

Table 3: *Confusion Matrix – 'Full' Model*

	User	Non-user	Sum
User	TP=166	FN=36	actual yes =202
Non-user	FP=37	TN=174	actual no =211
Sum	Predicted yes=203	Predicted no=210	Total=413

Table 4: *Classifier Performances – 'Full' Model*

Sr #	Parameter		Question	Estimator	Estimate
1	Accuracy		"Overall, how often is the classifier correct?"	(TP+TN)/total	(166+174)/413 = 0.823245
2	Misclassification Rate/ Error Rate (= 1- Accuracy)		"Overall, how often is it wrong?"	(FP+FN)/total	(36+37)/413 = 0.176755
3	Recall	True Positive Rate/ Sensitivity	"When it's actually yes, how often does it predict yes?"	TP/actual yes	166/202 = 0.821782
		Specificity	"When it's actually no, how often does it predict no?"	TN/actual no	174/ 211 = 0.824645
4	Precision		"When it predicts yes/no, how often is it correct?"	TP/predicted yes; TN/predicted no	166/203 = 0.817734 174/210 = 0.828571
5	1- Precision		"When it predicts yes/no, how often is it wrong?"	FP/ predicted yes; FN/predicted no	37/203 = 0.182266 36/210 = 0.171429
6	Prevalence		"How often does the yes condition actually occur in our sample?"	actual yes/total	202/413 = 0.489104

Our model's classification performance on our dataset seems moderately good - all desirable statistics like Accuracy, Specificity, Sensitivity and Precision are close to 82% while the negative statistics like misclassification/ error rate and 1-Precision are between 17-18%.

ROC Curve

The Receiving Operating Characteristic (ROC) is a measure of classifier/ prediction performance. The area under the ROC curve (or the AUROC) that ranges from 0.50 to 1.00 is the metric of interest ultimately. Values of AUROC above 0.80 indicate that the model does a good job of discriminating between the two categories comprising our target variable. We carried out a 70%-30% random learning-test splitting of the dataset, built the prediction model on the learning set and used the test set (i.e., unselected cases) to build the ROC curve. Thus 289 cases comprised the learning set for computing “scores” and the remaining 124 unused cases comprised the test set for assessing the model. Out of these 60 were positive (i.e., users) and 64 were negative (i.e., non-users). The ROC table and the corresponding graphic were generated (Figure 1). The model seems to discriminate well between MOOC users and non-users as the AUROC is 0.8626.

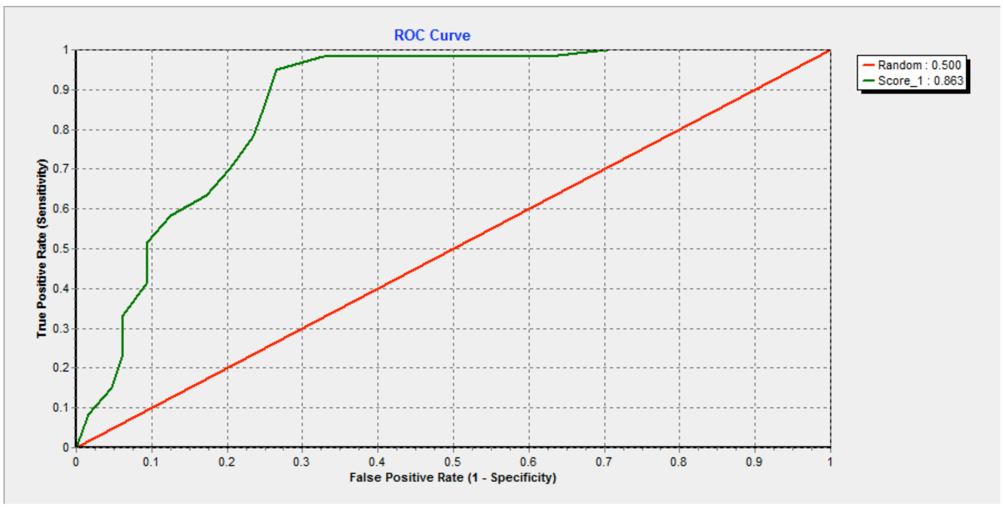

***Figure 1**: The ROC Curve Using the Full Model with 70-30% Learning-Test Split*

The Train-Test Process

The train-test process is a model-building and evaluation technique in which the data is divided into two subsets, called training (or learning) set and a test set. The model is built on the training set and tested on the test set. This process may be repeated several times and performance of each model in predicting the hold-out set assessed by tracking the test error rate. We attempted to carry out the train-test process with 90%-10% and 80%-20% random splits respectively. Each time we performed 10 repetitions of the process and tracked the model performance by calculating the performance metric of Error Rate (i.e., 1- Accuracy). The two ways of splitting produced similar results with the overall test error rates of nearly 0.2. For instance, with 90%-10% random split the train size was 371 while test size was 42. 10 trails were repeated in which the Error Rate ranged from 0.1429 to 0.2619 with an Overall Test Error Rate of 0.1976.

K-Fold Cross Validation

A k-fold cross-validation is a specialized variation of the train-test process that uses the data efficiently for training as well as evaluation of the model. It assesses how well the model performs in predicting the target variable on different subsets of the data (Mic 2015). We performed a 10-fold cross-validation with 10 trials, the most common variation of cross validation. The data was partitioned into 10 equally sized segments (called 'folds'). One fold was held out as testing data for validation while the other nine folds were used to train the model and then used to predict the target variable in the testing data. This process was repeated 10 times, while tracking the performance of the model in predicting the hold-out set using the performance metric of Error Rate (i.e., 1- Accuracy). The Error Rate ranged from 0.2024 to 0.2171 with an overall rate in ten trials equal to nearly 21%. Table 5 presents the confusion matrix along with the overall CV error rate.

Table 5: *Overall CV Error Rate and Confusion Matrix for a 10-fold Cross Validation of the Full Model*

Error rate			0.2088			
Values prediction			Confusion matrix			
Value	Recall	1-Precision		user	non-user	Sum
User	0.7769	0.2080	user	1557	447	2004
non-user	0.8049	0.2095	non-user	409	1687	2096
			Sum	1966	2134	4100

Evaluation and Diagnostics for the Reduced Model, the Best Model as per the SC Criterion

In this subsection we briefly report some salient results of evaluation and diagnostics for the reduced model (#9), analogous to those for the 'full' model in the last subsection. The SC value for the reduced model is superior to the 'full' model (424.118 as against 448.833). However, other model fit statistics (like Pseudo R^2s) are slightly worse. Table 6 displays the results of statistical tests for individual predictor. We note both through small p-values of Wald tests as well as the absolute t values that all variables contribute to the explanatory and predictive power of the model and hence cannot be dispensed with.

Table 6: *Regression Coefficients, │t│Statistics and Results of Wald Test – Reduced Model*

Attribute	Coef.	Std-dev	Absolute t	Wald	Signif
constant	-10.653151	1.5504	6.8711	47.2115	0.0000
Gender	0.886539	0.2787	3.1814	10.1213	0.0015
Employed	-1.832352	0.4005	4.5751	20.9318	0.0000
Age	0.251538	0.0598	4.2032	17.6665	0.0000
Int_Comm	2.071365	0.2422	8.5539	73.1697	0.0000
Self-directed	0.810094	0.1841	4.4001	19.3605	0.0000
Life_Goal	-0.725407	0.1910	3.7979	14.4242	0.0001

Error rate for this model (0.184) is comparable to that (0.177) for the 'full' model. Table 7 provides the confusion matrix and some other performance metrics for model #9.

Table 7: *Model Classifier Performances – Reduced Model*

Error rate			0.1840			
Values prediction			Confusion matrix			
Value	Recall	1-Precision		user	non-user	Sum
User	0.8267	0.1971	user	167	35	202
non-user	0.8057	0.1707	non-user	41	170	211
Value	Recall	1-Precision	Sum	208	205	413

As before, the ROC curve (Figure 2) was built and used as a measure of classifier/ prediction performance. Size of the learning set was 70% of the observations or 289 cases while that of the test set used for building the ROC curve was 30% or 124 cases. The AUROC metric was only slightly lower (0.858 as compared with 0.863 earlier).

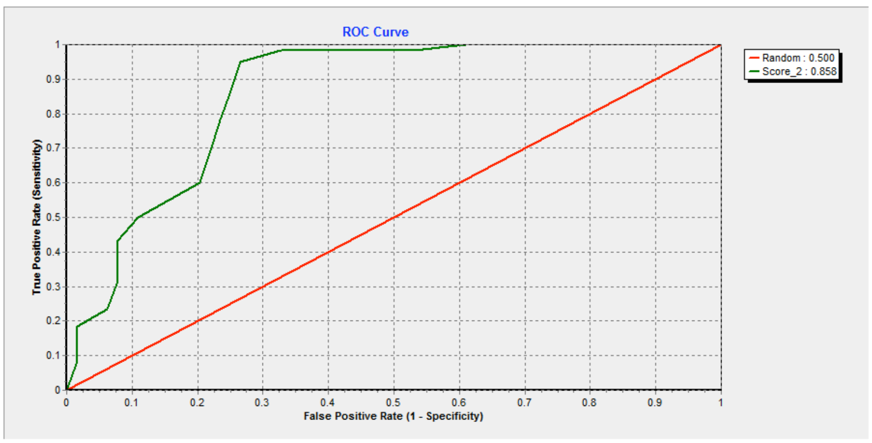

Figure 2: *The ROC Curve Using Model # 9 with 70-30% Random Learning-Test Split*

As with other models, we evaluated the prediction error rate of Model #9 also with 90-10% train-test process and ten-fold cross-validation. The error rate in ten trials ranged from 0.1190 to 0.2619 with the overall test Error Rate being 0.2024. Further, the Cross Validation (CV) Error Rates in the ten-fold cross-validation ranged from 0.1927 to 0.2122 with an overall CV Error Rate of 20.49%. Table 8 presents the overall CV error rate and the confusion matrix.

Table 8: *Overall CV Error Rate and Confusion Matrix for a 10-fold Cross Validation of Model #9*

Error rate			0.2049			
Values prediction			Confusion matrix			
Value	Recall	1-Precision		user	non-user	Sum
user	0.7844	0.2061	user	1572	432	2004
non-user	0.8053	0.2038	non-user	408	1688	2096
			Sum	1980	2120	4100

RESULTS AND DISCUSSION - THE “BEST” MODEL

We built a variety of logistic regression models for MOOC usage during model-building and subsequently assessed them. Table 9 summarizes the results of model evaluation and diagnostics – the model AIC, SC and overall CV error rate in 10-fold cross validation – for a few of the several models explored by us in our search for the most useful one.

Table 9: *Summary of Model Evaluation of a Few Models for MOOC Usage Explored*

Model #	# Descriptor/ Predictive Variables	Descriptor/ Predictive Variables	Model AIC	Model SC	Overall Cross Validation Error Rate
1	0	Intercept Only	574.343	578.367	-
2	4	Gender, Education, Employed, Age	517.195	537.313	0.3327
3	6	Gender, Education, Employed, Age, Self-directed, Planning	498.048	526.212	0.2954
4	8	Gender, Education , Employed, Age, Self-directed, Planning, Video, Collab.	494.199	530.410	0.3068
5	10	Gender, Education , Employed, Age, Self-directed, Planning, Video, Collab., Info_Process, Int_Comm	400.784	445.042	0.2117
6	13	Gender, Education , Employed, Age, Self-directed, Planning, Video, Collab., Info_Process, Int_Comm, Perseverance, Creative, New-Learn	402.116	458.444	0.2085
7* Best as per AIC	14 (the ‘Full’ Model)	Gender, Education , Employed, Age, Self-directed, Planning, Video, Collab., Info_Process, Int_Comm, Perseverance, Creative, New-Learn, Life Goal	388.481	448.833	0.2088
8	6	Based on independent variables from forward Logit :- Int_Comm, Self-directed, Perseverance, Life Goal, Gender, Collab	405.470	433.634	0.2541
9* Best as per SC	6	Based on independent variables from backward Logit :– Gender, Employed, Age, Int_Comm, Self-directed , Life_Goal	395.954	424.118	0.2049

Models #7 and #9 were identified previously as the most promising models as per the AIC and SC criteria respectively. Table 10 makes model comparison between these two models at a glance.

Table 10: *Model Comparison between Models #7 and #9 Using Information Criteria, Pseudo-R^2, AUROC, Train-Test Process & Cross Validation*

Criterion	Model # 7 with 14 Variables	Model # 9 with 6 Variables
AIC	388.481	395.954
SC	448.833	424.118
McFadden's R^2	0.3737	0.3326
Cox and Snell's R^2	0.4042	0.3693
Nagelkerke's R^2	0.5390	0.4925
AUROC	0.863	0.858
Overall Prediction Error Rate with 90-10% Train Test Process with ten trials	0.1976	0.2024
Range of 10-fold CV Error Rates	0.0147	0.0195
Overall 10-fold CV Error Rate	0.2088	0.2049

The choice of the best model depends on the criterion used to compare and evaluate models and the purpose of modelling. According to the pseudo-R^2s, Model #9 resulting from variable selection through backward logit seems less powerful. But when we consider the criteria which take into account the model complexity such as SC, it is in reality preferable. The ROC curve from model #9 is very similar to the previous one. The area under the ROC curve or the AUROC metric is above 0.80 in both cases indicating that both models do a good job of discriminating between the two target variable categories of MOOC users and non-users. The accuracy is slightly less for Model #9 as compared to the previous one (the overall prediction error rate with 90-10% train-test process with 10 trials being slightly higher 0.2024 for Model #9 as against 0.1976 for Model #7), but the new model #9 comprises only 6 variables. The interpretation of coefficients is easier. Also the overall 10-fold Cross Validation (CV) error rate is lower with model #9, though the range of CV error rates is slightly larger. Hence, we propose model #9 as a parsimonious, simplest possible model to explain the adoption of MOOCs as per our data.

The Odds Ratios and Tests of Hypotheses for the 'Best' Model

As mentioned previously, the backward logit procedure was used to arrive at the predictors in the reduced model (Model #9), which is our 'best' model. A probability cut-off value of 0.01 was used for the backward elimination of predictors at each step. It led to the selection of the aforesaid six predictors out of the set of 14. The odds ratios and 95% confidence intervals for them for this model are presented below (Table 11).

Table 11: *Odds Ratios and 95% Confidence Intervals – the 'Best' Model*

Attribute	*Coef.*	*Low*	*High*
Gender	2.4267	1.4055	4.1900
Employed	0.1600	0.0730	0.3509
Age	1.2860	1.1437	1.4460
Int_Comm	7.9356	4.9370	12.7557
Self-directed	2.2481	1.5671	3.2250
Life_Goal	0.4841	0.3329	0.7039

It may be noted from Table 11 that among the respondents, men were more than twice (2.4 times) as likely to enroll into MOOCs as women. For every year of increase in age, the likelihood to enrol in MOOC increased by nearly 1.3 times. Individuals who reported higher order Communicative Internet Efficacy ('Int_Comm') by one point on the scale of measurement were about 8 times more likely to be MOOC users than those with lower reported 'Communicative Internet Efficacy' (H5 is upheld). Similar hypothesis about 'Self-directed' is upheld too. For every one point increase in the preference for self-directed learning, the likelihood to enrol in MOOC increased by more than twice (nearly 2.25 times) (H6 is upheld). However, hypothesis H10 about Life_Goal did not seem to be upheld by our data.

Most of these findings seem to be borne out by some past findings about MOOC participants' profiles and pre-requisites. For instance, most reported Western studies about demographics of MOOC learners have shown how majority of MOOC learners were male, relatively older and having good prior knowledge and Internet communication skills, among other things (Gasevic et al. 2014; Ebben & Murphy 2014). Writing in the Indian and Chinese contexts, Trehan et al. (2017) bring out the importance of information and social media literacies of the learner population for MOOC adoption and success. Miller (2010) investigated the relation between prior social media literacies and engagement and the value experienced by participants in Massive Open Online Courses in the Canadian context. They found that the means of the low-engaged and high-engaged groups differed in the expected direction (although were not found to be statistically significant by them). Su, Huang, & Ding (2016) examined the effects of MOOC learners' social searching results on learning behaviours and outcomes. Reporting results from a July 2013 survey by the University of Pennsylvania which included responses from about two thousand students Christenson & Alcorn (2013) note that among Indians taking MOOCs more than three-fourths were male and with a college degree. In their investigation of user adoption of MOOCs in India, Kaveri et al. (2015) showed those with better internet skills and an existing preference for learning through videos were seen to be significantly more likely to adopt MOOCs.

IMPLICATIONS FOR PRACTICE AND POLICY & SUGGESTIONS FOR FUTURE RESEARCH

The notion of MOOC has been acknowledged as having value and potential for a developing country like India - "While staying with the MOOC technology and a minor shift in pedagogy, Higher Education (HE) institutions in India ... may explore MOOCs/ blended MOOCs as a way to complement efforts to improve quality and scale in their respective systems. Beyond formal HE, MOOCs have a larger potential role in the non-formal and informal education and indeed in general development too" (Trehan et al. 2017, pp.158). However, in practice, several gaps exist that detract from the value of the MOOC, among them, the digital divide, poor online efficacy and, in general, poor 'e-readiness' of the potential MOOC-user population and lack of MOOC engagement and completion among those who choose to enroll in a MOOC. Unlike traditional courses and programmes at a physical institution of higher learning, MOOC courses have been available freely, at a fraction of the regular cost of enrolment in the corresponding traditional course or programme. So adoption of and participation in MOOC happens at the user's free will. Low student motivation and low completion rates in MOOCs have been a cause for concern (Balsh 2013; Jordan 2013; Khalil & Ebner 2014; Jacoby 2014) and identified as the core MOOC issues in the literature (Ebben & Murphy 2014; Hew & Cheung 2014). The extent to which one participates in a MOOC after enrollment is guided by several factors including the individual's inclination or interest in a particular topic of study, their personality traits and the wish to form a social identity (Cormier & Siemens 2010; Bruff 2013; Belanger & Thornton 2013). If MOOCs are to be used increasingly as a substitute for or even a complement to traditional classroom-based learning, learners' participation and completion levels need to improve in practice. Improving MOOC participation and completion rates, although obviously important in practical terms, however, was not the subject of current study where we limited ourselves to explaining one's initial choice to adopt (x-)MOOCs. The former is an area which begs to be investigated in a future study.

In practical terms, the findings of this work highlight some of the learner-level characteristics and capabilities that facilitate adoption of online learning through MOOCs. We found, for instance, that individuals reportedly more efficacious with respect to online communication were more likely to be MOOC users than those who were less efficacious. Similarly, preference for self-directed learning is a strong predictor of enrolling in MOOC. In order to benefit from online learning (through MOOCs or other modes) learners have to be ready and have to possess some basic skills. Scholars have formalized the construct of e-readiness as a measure of the degree to which a community may be eager and prepared to make benefit of using information and communication technologies (ICT) (Dada 2006). Past researchers (Guglielmino & Guglielmino 2003; Stephen, Mutula & Brakel 2006; Dada 2006; Hanafizadeh, Hanafizadeh & Khodabakhshi 2009) have noted that ICT competencies and access to technology and resources are among the factors that should be considered for e-readiness. Guglielmino & Guglielmino (2003) emphasized the role of self-directed learning and discussed how to identify learners who are ready for e-learning and facilitate them. Hung, Chou, Chen & Own (2010) developed an online learning readiness scale having five dimensions, namely, self-directed learning, motivation for learning, computer/internet self-efficacy, learner control and online communication self-efficacy. Reviewing the past literature, one finds a paucity of studies about the effects of e-readiness and possible factors that affect the outcomes of e-Learning, especially in the context of higher education (Darab & Montazer 2011). In one such study analyzing the role of e-readiness factors in e-learning outcomes, Keramati, Afshari-Mofrad and Kamrani (2011) identified technical, organizational and social readiness factors and also found that organizational readiness factor had the most important effect on outcomes. Gulbahar, among others, emphasized the importance of the notion of e-readiness for producing new approaches and strategies for increase in effectiveness and efficiency of e-learning processes. Amalgamating key factors for e-readiness across past research studies, Gulbahar (2012) proposed five factors that should be considered for measuring students' level of e-readiness – namely, individual properties, ICT competencies, access to technology, motivation and attitude and factors that affect success – and a 5-point Likert-type "e-Readiness scale" involving 26 items based on these five dimensions. A few empirical studies have been conducted, mainly in Turkey, utilizing this scale to examine the readiness and other related constructs with respect to e-learners (Ilgaz & Gulbahar 2015; Kalelioglu 2017). There are no reported empirical studies yet about MOOC-readiness of Indian population. In future, it would be interesting to measure various dimensions of MOOC-readiness among samples of subpopulations of different educational, economic and cultural background. One would like to know answers to questions like 'How do learners from one subpopulation fare with respect to prior self-directed learning experience/ online communicative efficacy/ e-readiness as compared to learners from other subpopulations?'

Online efficacy of the population (i.e., potential MOOC-users) is also hampered by the phenomenon of 'digital divide'. If the MOOC has to bridge the gap in distribution of 'quality' education between rural areas and big cities in India, then the digital divide must be bridged too through appropriate policy and market intervention. In 2009-10, only 3.5 households in 1000 rural households in rural India had internet connectivity at home as per the National Sample Survey Organization (NSSO) Level and Pattern of Consumer Expenditure Report 2011 and even in 2014 the NSS found that 94% people in rural India did not own a computer. Although latest national digital literacy data for India were not available at the time of writing this article, we know India has made some big strides in this direction in recent times in terms of laying the basic infrastructure for wider digital literacy through initiatives like National Optical Fibre Network (NOFN) under the umbrella programme of Digital India (PIB 2014) and the National Digital Literacy Mission (NDLM) (Ghosh 2014); (IANS Jun 26, 2016). A recent report by Deloitte and

Associated Chambers of Commerce and Industry of India (ASSOCHAM) admitted that “Although the use of digital technology is on the rise in India, there still exists a wide ‘digital divide’ between urban and rural India which needs to be bridged urgently”(Deloitte, Nov 2016, pp. 5). And further about rural mobile connectivity, “currently, over 55,000 villages remain deprived of mobile connectivity. This is largely due to the fact that providing mobile connectivity in such locations is not commercially viable for service providers” (Deloitte, Nov 2016, pp. 13). The National Knowledge Network (NKN) needs to be pushed to the rural and semi-urban areas of India for addressing last mile connectivity there. Offline versions of MOOCs on popular platforms must be offered allowing users to view and interact with the core content without needing an Internet connection, much as what the open-source project Khan Academy Lite (KA Lite) did in 2014 to make the MOOC more affordable for all: Thus, strategies for enhancing digital literacy, online efficacy and self-directed learning of the prospective higher education aspirants would serve to help with wider MOOC outreach and adoption in a developing country like India.

As for the demographic characteristics of MOOC participants, our model finds that older men are more likely to engage with this kind of learning. However, our data does not suggest anything about why younger people and females are less likely to participate. Perhaps this could also be an area for further research. Besides, our study had other limitations which restricted the robustness and generalizability of the results. The first limitation emanated from the measurement model used. As noted in the background and methodology section, we refrained from using formal multi-item scales for measuring the core model variables and used single-item 5-point or 7-point Likert scale instead for simplicity. It would be interesting to replicate the study with more comprehensive and reliable measure of the constructs in question by use of the formal multi-item scales like ‘general self-efficacy’ and ‘communicative self-efficacy’ scales by Peng, Tsai & Wu (2006), the Learning Motivation Questionnaire (LMQ) developed by Lim & Kim (2003) or the Grasha-Riechmann Student Learning Styles Scale (Grasha 1996), for instance and compare the results obtained therefrom. Further the four independent variables (Perseverance, Creative, Collab. and Video) that were conceived as binary variables in this study could perhaps be measured as scale variables and the results could be compared with the present results. The second limitation arose due to the method of data collection. There were no pre-set criteria for subjects, except having an email account and being 18 years or older. But accessing the present and past students of premier management and technical institutes introduced a certain bias in our sample. Hence, the results may not be claimed to be representative of Indian youth in general. For the future, it would be appropriate to adopt a more nuanced approach like proportionate stratified sampling so as to include learners from many more geographies and social strata of Indian society for better insight into the diverse drivers and experiences of MOOC adoption across India.

REFERENCES

Balsh, T. 2013, “Why the “Low MOOC Completion Rate” Statistic is a Bogus Argument”, Retrieved 27th July, 2017, available at http://augmentedtrader.wordpress.com/2013/07/24/why-the-low mooccompletion-rate-statistic-is-a-bogus-argument/

Baruch, O., Bezalel, R. & Barth, I. 2007, “Introverts, extroverts, and achievement in a distance learning environment”, *The American Journal of Distance Education*, *21*(1), pp.3-19.

Battalio, J. 2009, “Success in distance education: Do learning styles and multiple formats matter?”, *The Amer. Jrnl. of Distance Education, 23*(2), pp.71-87.

Belanger, Y. & Thornton, J. 2013, “Bioelectricity: A Quantitative Approach”, Duke University’s First MOOC, Duke Center for Instructional Technology.

Bhattacharyya, R. 2013, “US-based MOOCs consortia in tie-ups with local universities for taking content online”, *The Economic Times*, 19 July.

Bonk, C.J., Lee, M.M. & Sheu, F.R. 2013, “Self-directed online learning: MOOCs, open education, and beyond”.

Bozdogan, H. 1987, “Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions”, *Psychometrika, 52*(3), 345-370.

Braithwaite, J. J. 2007, “Occam’s razor:The principle of parsimony”, *Academia. edu*.

Bruff, D., 2013, “Lessons learned from Vanderbilt’s first MOOCs”, *2013-10-17, http://cft, vanderbilt, edu/2013/08/lessons-learned-from-vanderbilts-first-moocs*.

Christenson, G. & Alcorn, B. 2013, “Can MOOCs Help Expand Access to Higher Education in India?”, Centre for the Advanced Study of India, University of Pennsylvania. [online] Available at https://casi.sas.upenn.edu/iit/christensenalcorn. (Accessed February 5, 2017)

Chuang, I. & Ho, A. D. December 23, 2016, “HarvardX and MITx: Four Years of Open Online Courses -- Fall 2012-Summer 2016”, Available at SSRN: https://ssrn.com/abstract=2889436 or http://dx.doi.org/10.2139/ssrn.2889436

Cormier, D. & Siemens, G. 2010, “Through the Open Door: Open Courses as Research, Learning and Engagement”, *Educause Review*, *45*(4), pp. 30-39.

Dada, D., 2006, “E-Readiness for developing countries: Moving the focus from the environment to the users”. *The Electronic Journal on Information Systems in Developing Countries, 27*(6), pp.1-14.

Darab, B. & Montazer, G. A., 2011, “An eclectic model for assessing e-learning readiness in the Iranian universities”. *Computers & Education, 56*(3), pp.900-910.

Deloitte, November 2016, “Digital India. Unlocking the trillion Dollar Opportunity”, Retrieved 3[3d] February, 2018, available at https://www2.deloitte.com/in/

Ebben, M. & Murphy, J. S. 2014, "Unpacking MOOC scholarly discourse: a review of nascent MOOC scholarship". *Learning, Media and Technology,* vol. 39, no. 3, pp. 328-345.

Gasevic, D., Kovanovic, V., Joksimovic, S. & Siemens, G. 2014, "Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative". *The International Review of Research in Open and Distributed Learning,* vol.15, no. 5.

Ghosh, D. 2014, “Digital Literacy: the Way Forward”. Retrieved from http://www.nasscom.in/digital-literacy-way-forward?fg=206554, 2014.(http://pib.nic.in/newsite/PrintRelease.aspx?relid=108926)

Grasha, A.F. 1996, *Teaching with style: A practical guide to enhancing learning by understanding teaching & learning styles* (128). Pittsburgh: Alliance Publishers.

Grasha, A.F. & Yangarber-Hicks, N. 2000, "Integrating teaching styles and learning styles with instructional technology", *College Teaching, 48*(1), pp. 2-10.

Guglielmino, P.J. and Guglielmino, L., 2003. "Are your learners ready for e-learning". *The AMA handbook of e-learning*. New York: American Management Association.

Gülbahar, Y., 2012, "Study of developing scales for assessment of the levels of readiness and satisfaction of participants in e-learning environments". *Ankara University Journal of Faculty of Educational Sciences*, *45*(2), pp.119-137.

Guo, P.J., Kim, J. & Rubin, R. 2014, "How video production affects student engagement: An empirical study of MOOC videos", In *Proceedings of the first ACM conference on Learning@ scale conference*, pp.41-50. ACM.

Hanafizadeh, P., Hanafizadeh, M. R. & Khodabakhshi, M., 2009, "Taxonomy of e-readiness assessment measures". *International Journal of Information Management, 29*(3), pp.189-195.

Hew, K.F. & Cheung, W.S. 2014, "Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges". *Educational Research Review*, vol. 12, pp. 45-58.

Heylighen, F. 1997, "Occam's razor", *Principia cybernetica web*.

Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. 2014, "HarvardX and MITx: The first year of open online courses, fall 2012-summer 2013", *HarvardX and MITx Working Paper No. 1*.

Ho, A. D., Chuang, I., Reich, J., Coleman, C. A., Whitehill, J., Northcutt, C. G., Williams, J. J., Hansen, J. D., Lopez, G., & Petersen, R. March 30, 2015, "HarvardX and MITx: Two Years of Open Online Courses Fall 2012-Summer 2014", Available at SSRN: https://ssrn.com/abstract=2586847 or http://dx.doi.org/10.2139/ssrn.2586847

Hung , M., Chou, C., Chen, C. & Own, Z., 2010, "Learner readiness for online learning: Scale development and student perceptions". *Computers & Education, 55*, pp.1080–1090.

IANS, Jun 26, 2016, Retrieved from http://timesofindia.indiatimes.com/tech/tech-news/New-Digital-Literacy-Mission-soon-for-60mn-in-rural-India/articleshow/52925005.cms

Ilgaz, H. and Gülbahar, Y., 2015. "A snapshot of online learners: e-Readiness, e-Satisfaction and expectations". *The International Review of Research in Open and Distributed Learning, 16*(2).

Jacoby, J. 2014, "The disruptive potential of the Massive Open Online Course: A literature review". *Journal of Open, Flexible and Distance Learning,* vol. 18, no.1, p. 73.

Jordan, K., 2013. MOOC completion rates: The data.

Jordan, K. 2014, "Initial trends in enrolment and completion of massive open online courses", *The International Review of Research in Open and Distributed Learning, 15*(1).

Kalelioglu, Filiz, 2017, "Online learning experience of the learners: their readiness, expectations and level of satisfaction". *International Journal of Management and Applied Science*, 3(3).

Available online http://www.iraj.in/journal/journal_file/journal_pdf/14-349-149691246912-16.pdf

Kaveri, A., Gunasekar, S., Gupta, D. & Pratap, M. 2015, "Decoding the Indian MOOC learner" In *MOOCs, Innovation and Technology in Education (MITE), 2015 IEEE 3rd International Conference*, pp.182-187. IEEE.

Keller, H. & Karau, S. 2013, "The importance of personality in students' perceptions of the online learning experience", *Computers in Human Behavior*, *29*(6), pp. 2494-2500.

Keramati, A., Afshari-Mofrad, M. & Kamrani, A., 2011, "The role of readiness factors in e-learning outcomes: An empirical study". *Computers & Education, 57*(3), pp.1919-1929.

Khalil, H. & Ebner, M. 2014, "MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review", In J. Viteli & M. Leikomaa (Eds.), *Proceedings of EdMedia 2014--World Conference on Educational Media and Technology*, vol. 1, pp. 1305-1313, Tampere, Finland: Association for the Advancement of Computing in Education (AACE).

Kim, E. & Schniederjans, M. 2004, "The role of personality in Web-based distance education courses", *Communications of the ACM - Homeland security*, 47, pp. 95-98.

Lewis, M. 2014, "Learning styles, motivations and resource needs of students enrolled in a massive open online class", *North Carolina, Chapel Hill. Recuperado de https://cdr. lib. unc. edu/indexablecontent/uuid: 6caa790d-8235-4dd6-bd24-54ddf9bbf33e [Links].*

Lim, D. H. & Kim, H. J. 2003, "Motivation and learner characteristics affecting online learning and learning application", *Journal of Educational Technology Systems, 31*(4), pp. 423–439.

Mic. September 15, 2015, "Predicting creditability using logistic regression in R: cross validating the classifier (part 2)", R-bloggers, [online] https://www.r-bloggers.com/predicting-creditability-using-logistic-regression-in-r-cross-validating-the-classifier-part-2-2/

Miller, T. 2010, Social Media Literacies and Perceptions of Value in Open Online Courses. *Bonnie Stewart ED*, pp. 612-701.

Milligan, C., Littlejohn, A. & Margaryan, A. 2013, "Patterns of engagement in connectivist MOOCs", Journal of Online Learning and Teaching, 9(2), pp. 149-159.

Pauler, D. K. 1998, "The Schwarz criterion and related methods for normal linear models", *Biometrika, 85*(1), pp. 13-27.

Peng, H., Tsai, C.C. & Wu, Y.T. 2006, "University students' self-efficacy and their attitudes toward the Internet: the role of students' perceptions of the Internet", *Educational studies, 32*(1), pp.73-86.

PIB, 2014, "Digital India – A programme to transform India into digital empowered society and knowledge economy". *Press Information Bureau Government of India*, Aug 2014

Pope, J. 2014, "What Are MOOCs Good For?", *MIT Technology Review*, *118*(1), pp. 69-71.

Raffaghelli, J.E., Cucchiara, S. and Persico, D., 2015, “Methodological approaches in MOOC research: Retracing the myth of Proteus”. *British Journal of Educational Technology, 46*(3), pp.488-509.

Rud, O. P. 2001, “*Data mining cookbook: modeling data for marketing, risk, and customer relationship management”,* John Wiley & Sons.

Sadera, E. 2014, “Dawn of the MOOC: the challenge of online learning”, *University of Auckland Business Review*, *17*(1), pp. 6.

Scardilli, B. 2013, “MOOCs: Classes for the Masses”, *Information Today*, 308.

Seasholtz, M. B. & Kowalski, B. 1993, “The parsimony principle applied to multivariate calibration”, *Analytica Chimica Acta, 277*(2), pp. 165-177.

Shah, D. December, 2016, “Monetization Over Massiveness: Breaking Down MOOCs by the Numbers in 2016: A Review of MOOC Stats and Trends in 2016”, edSurge.

Shi, X. & Yu, S. 2016, “THE RISING OF CHINA'S MOOC: OPPORTUNITIES AND CHALLENGES TO THE HEIs”, *Current Politics and Economics of Northern and Western Asia*, *25*(1), pp. 61.

Stephen, M., Mutula, S. M. & Brakel, P., 2006, “An evaluation of e-readiness assessment tools with respect to information access: Towards an integrated information rich tool”. *International Journal of Information Management, 26*, pp.212–223.

Su, Y. S., Huang, C. & Ding, T. 2016, "Examining the Effects of MOOCs Learners’ Social Searching Results on Learning Behaviors and Learning Outcomes". *Eurasia Journal of Mathematics, Science & Technology Education, 12*(9), pp. 2231-2243.

Trehan, S., Sanzgiri, J., Li, C., Wang, R. & Joshi, R. M. 2017, “Critical discussions on the Massive Open Online Course (MOOC) in India and China”, *International Journal of Education & Development using Information & Communication Technology, 13*(2).

Original article at: http://ijedict.dec.uwi.edu/viewarticle.php?id=2376