IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN: 2278-487X, p-ISSN: 2319-7668
PP 54-58
www.iosrjournals.org

Binary logistic regression modelling in predicting channel behaviour towards instant coffee vending machines

[1]Dr. Ayan Chattopadhyay, [2]Dr. Malini Majumdar and
[3]Mr. G Vishnu Shiva Prasad

[1]Associate Professor (Marketing), Army Institute of Management Kolkata, [2]Associate Professor (Marketing), Army Institute of Management Kolkata, [3]Independent Researcher
Corresponding Author: [1]Dr. Ayan Chattopadhyay

Abstract: *A market oriented predictive decision model can always strengthen a company's endeavour to identify the appropriate channel members for expanding its business. The present study develops a model that predicts such channel decisions, based on binary logistic regression modelling approach. The market for coffee vending machine installation and its scope of expansion has been considered, particularly for the brand Georgia. The objective was to find out the potential channel partners who would be interested in placing coffee vending machine at their retail premise. In the present study healthcare segment comprising of hospitals, nursing homes, doctors' chambers and diagnostic centers from both private and public sectors, have been considered. The research design has been descriptive, wherein data have been collected using a questionnaire as the survey instrument from a respondent group, identified through judgemental sampling. The Model was designed using logistic regression method where three independent variables are footfall, refreshment availability, refreshment availability for tea/coffee and the dependent variable is interest of the potential channel partner. The analysis is done using R. The results showed that footfall has small influence in deciding whether the respondent is interested to become a channel partner or not but other two factors may be considered as deciding factors in predicting the interest.*

Key Words: *Predictive modelling, binary logistic regression, channel partner, instant coffee.*

I. Introduction

Coffee vending machines are an integral part of waiting lounges, particularly in urban India. It is dominantly visible in service sector where the provider of services offers such facility to its customer (also potential customer) as an augmented component to its core service. Utility of such services is gaining paramount importance by the day, more so in healthcare sector comprising of hospitals, nursing homes, clinical laboratories, private clinics/chambers used by doctors, where a lot of potential customers wait for a long time. Vending machines are a highly cost-effective and efficient option of serving quality beverage at affordable price without much involvement of human resource and space. Whether a particular establishment would be interested in installing a vending machine will depend on several factors like footfall (both in terms of quality and quantity), type of refreshments available, availability of other beverages etc. Different brands are vying with each other to get their coffee vending machines installed in these sites. The major players in this sector include Nescafe, Lavazza, Bru, Georgia, Café-Desire, CCD etc to name a few. To ensure objectivity in study, the researchers have zeroed in their attention to brand Georgia (owned by Coca-cola) since it is one of the late entrants in the market. Deployment of such machines involves complex decision making process at the retailer/ establishment end as well on the part of the manufacturer. From the retailer's point of view, estimating ROI of putting a coffee vending machine is difficult as coffee consumption from such machines are a neo Indian phenomenon. Many a times the retailers are risk averse to such unpredictable future while in many cases they are hesitant due to paucity of space in their retail premise. These issues make it even difficult for the coffee brands selling through vending machines to predict the market. The researchers identified the need for a market oriented predictive decision model that could help the company to identify the appropriate channel members interested in installing coffee vending machines. The present study initially focuses on selection of relevant factors that may influence the decision of vending machine installation, followed by developing a model that would help in forecasting the business opportunity for the company in a particular market. Of the different modelling approaches, binary logistic regression modelling has been used in the present case to predict such channel decisions. Two such models were developed and compared basis their suitability from model structure viewpoint and the one with a better fit has been suggested.

II. Literature Review

The purpose of literature review was to establish a sound theoretical background on the concept of logistic regression and its application in predicting a business outcome. In doing so the market for vending machine installation and its scope of expansion was kept in mind. The past research works on the related field helped to develop a concrete methodology in developing a predictive model for the real life case of coffee vending machine installation in select channels. Some of the important research works conducted have been captured in this section. "*Consumer perception on HUL vending machine & its use in house hold - a study in bhubaneswar*" ***(Sahu, 2015)*** captured the consumer perception, market growth rate and market profitability of vending machine usage with respect to HUL's Lipton Vending machine in Bhubaneswar. The study concluded with the potential and prospective business in organizations. *"Viability of Coffee Vending Machine: An Assessment"****(Carino, 2014)*** aimed to determine the viability of coffee vending machine in Batangas City; specifically, in terms of capitalization, location, number of years in operation, hours of operation, average monthly income, average number of customers per day. The study determines the factors that contribute to the viability of coffee vending machine in terms of sales performance, operational performance, and financial performance, and also identifies the problems encountered in the operation of the coffee vending machine. **"***Perceived Quality and Attitude Toward Tea & Coffee by Consumers"***(Han, 2012)** determine the consumers' perception and attitude toward tea and coffee. Fishbien's multi-attribute attitude model and t-test were used to measure hypothesis and compare attitude toward tea and coffee. The study concluded that consumers had an overall more positive attitude towards coffee compared with tea with regards to availability, different flavour, and environment of shop attributes. Emerging challenges and prospects of FMCG product development in India ***(Nagarajan G, 2013)*** provides inputs for a clear understanding of the consumer mind-set towards FMCG products. It focuses on some of the fundamental issues pertaining to the emerging challenges and prospects of marketing FMCG products in India. *"FMCGs sector in India: a strategical view"****(Singh, 2011)*** examines Marico Ltd and Nestle India Ltd, Colgate Palmolive India Ltd, Britannia Industries Ltd, Godrej Consumer Products Ltd and Dabur with respect to their distribution network, penetration levels, operating cost, per capita consumption and intense competition between the organized and unorganized segments. *"A logistic regression model to predict freshmen enrollments"* ***(Sampath, 2002)*** presents the steps involved in developing a logistic regression model based on student test scores, performance at high schools, and other demographics to predict whether or not a student will eventually enrol if admitted. "*Breast cancer analysis using logistic regression"* ***(H. Yusuff, 2012)*** studies the diagnosis of breast cancer from mammograms is complemented by using logistic regression. *"Logistic Regression Analysis of Graduate Student Retention"****(Sheridan, 1993)*** utilized to predict the retention of 477 master's and 124 doctoral candidates at a large Canadian university. Selected demographic, academic and financial support variables were used as independent variables. The dichotomous dependent variable was whether the student successfully completed the degree. *"Predicting social trust with binary logistic regression"****(Boamah, 2015)*** predicts social trust with five demographic variables from a national sample of adult individuals who participated in The General Social Survey (GSS) in 2012. The five predictor variables were respondents' highest degree earned, race, sex, general happiness and the importance of personally assisting people in trouble. The study assesses the impact of the predictors on the likelihood that respondents would report that they have low social trust. Many other studies including that of *"Employee Attrition Risk Assessment using Logistic Regression Analysis"****(Khare, 2011)***, "*Predictive Models for Equipment Fault Detection in the Semiconductor Manufacturing Process"* ***(Ramadoss, 2016)*** and *"Predictive modelling to forecast student outcomes and drive effective interventions in online community college courses"****(Smith, 2011)*** have used logistic regression to predict certain outcomes. Though many research works have been done across the globe to estimate the potential of tea/ coffee vending machines, most of which revolved around the perception of customers towards tea/ coffee vending machine, the researchers did not find predictive modeling being used as an approach to identify the potential of vending machine placement in select channels. The same has thus been considered as the gap area where investigation may be carried out.

III. Objective

To develop model that may be used to predict potential of business development in select channel using binary logistic regression. The aim is to find out the potential customers who would be interested in placing coffee vending machine.

IV. Esearch Framework

1. Research Design

Out of different study designs, cross sectional and longitudinal study, the former is preferred to the latter as it gives the population characteristics at a particular point in time. The present work, being descriptive in nature, aims to find out certain characteristics of population as on a particular time frame without getting into the estimation of causal relationship.

2. Research Methodology

The ensuing study uses Logistic Regression, also called logit regression or logic model is a regression model where the dependent variable is categorical. Binomial or binary logistic regression ***(David Anderson, 2014)*** includes those models where dependent variables can assume only two discreet values '0' and '1' which may represents outcomes such as interested or not interested, pass or fail, accepted or not accepted, win or lose etc. Logistic regression can be multinomial as well where the dependent variable might have more than two outcome categories. The binomial or binary logistic model is primarily used to evaluate or estimate or determine the probability of a binary response or outcome based on one or more independent or predicated variables. Logistic regression in many ways is similar to ordinary regression or multiple regressions where a set of independent variables are used to predict the outcome of the dependent variables. However in logistic regression the relationship is non-linear and may be expressed as:

$$\mathrm{E}\ (\mathrm{y}) = \frac{\mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots +\beta_n x_n}}{1+\mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots +\beta_n x_n}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$\beta_0$is the intercept. β_1,β_2,………………………………β_n are regression coefficients.

Here E(y) can assume two discreet values '0' and '1' in case of binary or binomial logistic regression which may be written as (for two and three independent variables):

$$\hat{y} = Estimate\ of\ p\left(y = 1 \middle| x_{1,}x_{2,}x_3\right) = \frac{\mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3}}{1+\mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3}} \text{ for three independent variables}$$

$$\hat{y} = Estimate\ of\ p\left(y = 1 \middle| x_{1,}x_2\right) = \frac{\mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2}}{1+\mathrm{e}^{\beta_0+\beta_1 x_1+\beta_2 x_2}} \text{ for two independent variables}$$

The independent variable may either be continuous or categorical but the above form of logistic regression violates the linear assumption of regression. Also the linear regression does not make sensical predication for a binary dependent variable. Thus it is important to convert a binary variable into a continuous one so that it may take any real value (+ve or –ve). In order to achieve that binomial logistic regression first takes the odds of an event happening for different levels of each independent variable and then takes the ratio of those odds. The logarithm of this ratio is referred to as logit (also called log-odds) which is used to create continuous criteria of the dependent variables. The logit transformation is referred to as the link function in logistic regression though the dependent variable of logistic regression is binomial. The logit is a continuous criterion upon which linear regression is performed. The logit function for equation (1) may be represented or calculated as:

$$logit[\,E(y)] = \ln\left[\frac{E(y)}{1-E(y)}\right] = Ratio\ of\ odds$$

The odds ratio for independent variables represents the change in odds for 1 unit change in the independent variables holding all other independent variable constant. The odds ratio also enables us to compare the odds for different events.

$$logit[E(y)] = \ln\left[\frac{\dfrac{e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}}}{1-\dfrac{e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}}}\right]$$

$$logit[E(y)] = \ln\left[\frac{\dfrac{e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}}}{\dfrac{1}{1+e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}}}\right]$$

$$logit[E(y)] = \ln\left[e^{\beta_0+\beta_1x_1+\beta_2x_2 \ldots\ldots\ldots\ldots\ldots+\beta_nx_n}\right]$$

$$logit[E(y)] = \beta_0 + \beta_1x_1 + \beta_2x_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. + \beta_nx_n$$

Thus, logit of the probability of an event is a simple linear equation. A unique relationship exists between the odds ratio for a variable and its corresponding regression coefficient which is

$\boldsymbol{Odds\ ratio = e^{c\beta_i}}$ where 'c' represents the change in independent variables.

3. Sampling

Out of the two different broad sampling methods, namely, probabilistic and non-probabilistic sampling methods, the researchers used judgmental sampling, a non-probability sampling technique, in the present study. This sampling method uses researcher's judgement in drawing elements within the sample. Such representation is expected to serve the purpose of addressing the research objectives set. Hence, this type of sampling technique is also known as purposive sampling or authoritative sampling. The process involves nothing but purposely handpicking elements from the population based on the researcher's knowledge and judgment. It is one of the viable sampling techniques in obtaining information from a very specific group of people. In the present study healthcare segment comprising of hospitals, nursing homes and diagnostic centers from both private and public sectors, have been considered.

4. Data Collection

Primary data forms the basis of the present research work and the same was collected using questionnaire as the survey instrument. Before initiating the final survey, a pilot survey was done to identify and eliminate the flaws in the questionnaire. The full scale survey was implemented after making minor changes in the questionnaire. The study used a structured questionnaire with a mix of open and close ended questions. Undisguised face to face interview was conducted before which appointments are were taken from the shortlisted respondents. The sample size required was calculated from the expression: N = [{ t^2 x p (1 – p) } / m^2] where N: Sample size required, t: confidence level at 95% (standard value of 1.96), m: margin of error at 5% (standard value of 0.05) and p: estimated prevalence of consumer knowledge about Georgia Coffee vending machines (15%). N was found to be 196. In the full-scale survey, 230 health care establishments in Kolkata city were included in the study. A total of 184 filled in questionnaires were received. Responses of 150 questionnaires were finally considered for analysis owing to their completeness.. Internal consistency estimates of reliability of primary data were found out and Cronbach's α was found to be in acceptable range (0.83). The researchers in their ensuing study have used R 3.4.0 version for conducting logistic regression.

V. Analysis

The logistic regression equation with three independent variables; footfall, availability of refreshments and current source of tea/ coffee is expressed in (1). The β values obtained are: $\beta_0 = 0.94,\ \beta_1 = 0.74,\ \beta_2 = 2.66,\ \beta_3 = -5.02$; where β_0 is constant, $\beta_1$corresponds to footfall, β_2 corresponds to availability of refreshment and $\beta_3$corresponds to current source of tea/ coffee. y is the dependent variable and x_1, x_2 and x_3 are the independent variables corresponding to footfall, availability of refreshments and current source of tea/ coffee respectively.

$$\mathrm{E}\ (\mathrm{y}) = \frac{e^{-0.94+0.74x_1+2.66x_2-5.02x_3}}{1+e^{-0.94+0.74x_1+2.66x_2-5.02x_3}}$$

$$logit[\,E(y)] = \ln\left[\frac{E(y)}{1-E(y)}\right] = Ratio\ of\ odds$$

$$\boldsymbol{logit[E(y)] = -0.94 + 0.74x_1 + 2.66x_2 - 5.02x_3}\text{--------------- } \mathbf{(1)}$$

Equation (1) thus represents a predictive model with three predictor variables. The R output with all the three independent variables x_1, x_2 and x_3 is shown below.

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-0.9360	0.6902	-1.356	0.175056
Footfall	0.7377	0.7886	0.935	0.349531
Availability of Refreshments	2.6619	0.7739	3.440	0.000582 ***
Current Source of Tea/Coffee	-5.0160	1.2132	-4.135	3.56e-05 ***
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for binomial family taken to be 1)			

Fig – 1; Coefficients of Logistic Regression with 3 variables; ***Source:*** *R Output of primary data*

Null deviance was found to be 122.469 with a Residual deviance of 50.502. Akike Information Criteria (AIC) was found to be 58.502 with Chi-Square value of 71.96 at 3 degrees of freedom and a p-value of 1.618127e-15. Both availability of refreshments and current source of tea and coffee have been found to statistically significant as seen from the p value in Fig. 1; however the p-value of footfall is found to be statistically insignificant. Thus, the same analysis was conducted by dropping footfall from the list of independent variable. AIC value, which is an estimator of relative quality of the statistical model relative to other models, is found to be 58.502. Also the difference in deviance between the residual deviance for the model with predicators and the null model was tested using chi-square test. The test statistic is distributed chi-squared with degrees of freedom equal to the degrees of freedom between the current and the null model that is the number of predicator variables in the model. The chi-square of 71.96687 with 3 degrees of freedom and associated p-value (1.618127e-15) of less than 0.001 indicates that the model represented by equation 1 as a whole fits significantly better than an empty model.

Logistic regression equation by dropping one statistically insignificant predictor variable, i.e. footfall has been shown in (2). The β values obtained are: $\beta_0 = -0.48$, $\beta_1 = 2.89$, $\beta_2 = -4.85$; where β_0 is constant, $\beta_1$corresponds to availability of refreshment, $\beta_2$corresponds to the current source of tea/ coffee. y is the dependent variable and x_1 and x_2 are the independent variables corresponding to availability of refreshments and current source of tea/ coffee respectively.

$$\mathrm{E}\ (\mathrm{y}) = \frac{\mathrm{e}^{-0.48+2.89x_1-4.85x_2}}{1+\mathrm{e}^{-0.48+2.89x_1-4.85x_2}}$$

$$logit[\,E(y)] = \ln\left[\frac{E(y)}{1-E(y)}\right] = Ratio\ of\ odds$$

$$\boldsymbol{logit[E(y)] = -0.48 + 2.89x_1 - 4.85x_2}\text{-------------------------- } \mathbf{(2)}$$

Equation (2) represents an alternative predictive model with two predictor variables. The R output with two independent variables x_1 and x_2 is shown below.

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-0.4801	0.4794	-1.001	0.31661
Availability of Refreshments	2.8918	0.7595	3.808	0.00014 ***
Current Source of Tea/ Coffee	-4.8541	1.1656	-4.164	12e-05 ***
	Signif.codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for binomial family taken to be 1)			

Fig – 2; Coefficients of Logistic Regression with 2 variables; ***Source:*** *R Output of primary data*

Null deviance was found to be 122.469 with a Residual deviance of 51.501. Akike Information Criteria (AIC) was found to be 57.501 with Chi-Square value of 70.96 at 2 degrees of freedom and a p-value of 3.886856e-16. Even after dropping footfall from the list of predictor variables, both availability of refreshments and current source of tea and coffee have been found to statistically significant as seen from the p-value in Fig. 2. AIC value for model 2 is found to be 57.501. Also the difference in deviance between the residual deviance for the model with predicators and the null model was tested using chi-square test. The chi-square of 70.96752 with 2 degrees of freedom and associated p-value (3.886856e-16) of less than 0.001 indicates that the model 2 as a whole fits significantly better than an empty model. Comparison of both the models has been done using AIC criteria (an estimator of relative quality of statistical). Model represented by equation 2 shows improvement in AIC value over model 1. Since all test of significance for model 2 (which has a better AIC value compare to model 1) have been found to be favourable, the researcher proposes (2),

($logit[E(y)] = -0.48 + 2.89x_1 - 4.85x_2$), as the predictor model for business development of instant coffee through vending machine in select channels. The probability of various events, as obtained from equation (2) is shown in Fig 3.

$\square_0$	$\square_1$	$\square_2$	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{E\ (y) = \dfrac{e^{-0.48+2.89x_1-4.85x_2}}{1+e^{-0.48+2.89x_1-4.85x_2}}}$	**Probability, E(y)**
-0.48	2.89	-4.85	1	1	$e^{-2.44}/(1+e^{-2.44})$	0.080
			1	0	$e^{2.41}/(1+e^{2.41})$	0.918
			0	1	$e^{-5.33}/(1+e^{-5.33})$	0.005
			0	0	$e^{-0.48}/(1+e^{-0.48})$	0.382

Fig – 3; Probability Calculation for Modeled Equation; ***Source:*** *R Output of primary data*

It is seen that probability of a channel partner opting for vending machine installation is very low (0.08) when there is availability of both nearby refreshments as well as current source of tea/ coffee. It is also seen that probability that a channel partner would opt for vending machine installation is highest (0.91) when there is availability of nearby refreshments but no source of tea/ coffee. Probability that a channel partner would opt for vending machine installation with availability of nearby refreshments but without current source of tea/ coffee (0) is the lowest at 0.005 while the same is found to be 0.38 when there is neither any availability of nearby refreshments nor any current source of tea/ coffee.

VI. Conclusions

A predictive model was designed using binomial logistic regression approach where three independent variables considered include footfall, availability of nearby refreshments, refreshment availability for tea/coffee and customer behaviour (channel partner's willingness to install vending machine) as the dependent variable. The results showed that availability of nearby refreshments and refreshment availability for tea/coffee are significant in predicting customer behaviour. Positive sign of regression co-efficient for variable x_1 (x_1: availability of nearby refreshments) suggests a direct proportionality between x_1 and the dependent variable. Thus, more the options of nearby refreshments, more is the possibility of channel partner opting for coffee vending machine. One might intuitively say that refreshment availability for tea/coffee is indirectly proportional to channel partner's willingness to install vending machine and the same was found from the negative sign of regression co-efficient for variable x_2 (x_2: refreshment availability for tea/coffee). Finally, it is concluded that probability that a channel partner would opt for vending machine installation is highest when there is availability of nearby refreshments but there is no source of tea/ coffee. Though footfall is imperative to any business, the same has emerged to be non-significant and the same may be attributed to the fact that availability of nearby refreshments ensures footfall and separate consideration of the same is not required.

VII. Limitations & Scope For Further Studies

Like any other research work, the present paper too has some limitations. Since a select channel was included in the study, the outcome cannot be generalized for all business channels. The sample size considered in the study was restricted owing to selection of a specific channel. Future studies may consider wider channels. Also, one may also consider more variables compared to those considered in the present study. The present study was conducted in urban area and the same be extended to semi-urban or rural areas. Furthermore, the present study is based on the concept of binomial logistic regression and multinomial logistic regression may form the basis of future studies on similar area.

References

[1]. Andren A. Carino, E. R. (2014). Viability of Coffee Vending Machine: An Assessment. Research Academy of Social Sciences, 119-129.

[2]. David Anderson, D. Sweeney and Thomas Williams. (2014). *Statistics for Business and Economics* (11ed.). New Delhi, India: Cenage Learning.

[3]. H. Yusuff, N. M. (2012). Breast cancer analysis using logistic regression. IJRRAS, 14-22.

[4]. Han, I. M. (2012). Perceived Quality and Attitude Toward Tea & Coffee by Consumers. International Journal of Business Research and Management (IJBRM, 100-112.

[5]. Joseph Adwere Boamah, S. H. (2015). Predicting social trust with binary logistic regression. Research in Higher Education Journal.

[6]. Nagarajan G, K. S. (2013). Emerging challenges and prospects of fmcg product development in India. International Journal of Marketing, Financial Services & Management Research, 41-52.

[7]. Ramadoss, S. M. (2016). Predictive Models for Equipment Fault Detection in the Semiconductor Manufacturing Proces. IACSIT International Journal of Engineering and Technology, 273-285.

[8]. Rupesh Khare, D. K. (2011). Employee Attrition Risk Assessment using Logistic Regression Analysis. 2nd IIMA International Conference on Advanced Data Analysis, Business Analytics and Intelligence. Ahmedabad.

[9]. Sahu, S. C. (2015). Consumer perception on HUL vending machine & its use in house hold- a study in Bhubaneswar. Intercontinental journal of marketing research review, 2321-0346, 2347-1670.

[10]. Sheridan, S. W. (1993). Logistic Regression Analysis of Graduate Student Retention. The Canadian Journal of Higher Education, 45-64.

[11]. Singh, A. et al. (2011). FMCGs sector in India: a strategical view. Asian Journal of Management Research, 2(1), 612-621.

[12]. Vernon C. Smith, E. (2011). Predictive modelling to forecast student outcomes and drive effective interventions in online community college courses. Journal of Asynchronous Learning Networks, 16(3), 51-61.

[13]. Sampath, V. A. F. (2002). A logistic regression model to predict freshmen enrollments. Annandale.