

# Application of Logistic Regression model to determine the Academic Performance of MBA Students of Department of Management Studies, NIT Tiruchirappalli

<sup>1</sup>Surendheran R, <sup>2</sup>Monisha Ravi

<sup>1,2</sup>Dept. of Management Studies, National Institute of Technology Tiruchirappalli, Tamil Nadu, India

## Abstract

This study is conducted to determine the students' capability in Academics based on their past performance in studies and certain other related factors among students pursuing their Master of Business Administration in National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. The data is collected by sending a questionnaire to a sample of 64 students, in order to collect the details regarding their Age, Gender, Under-Graduation Percentage, 12th Standard Percentage, 10th Standard Percentage, Work Experience and CAT Examination Percentile. A Binary Logistic Regression model is used to determine the probability of the student faring rate based on the above stated factors. The model found is  $\text{Log} = -0.3041 - 0.1238 (\text{Age}) - 0.5799 (\text{Gender}) - 0.2024 (\text{Under Graduation Percentage}) + 0.3966 (12\text{th Standard Percentage}) + 0.0795 (10\text{th Standard Percentage}) + 1.0061 (\text{Work Experience}) - 0.3063 (\text{CAT Percentile})$ . This would, ultimately, be used to predict the nature of perseverance a student has with regard to performing in Academic domain, under the tag of being a "Good" or a "Bad" student. These factors play a significant role at 5% level of significance. Thus, a Logistic Regression model to predict the Academic Performance will be an effective tool for decision making process.

## Keywords

Binary Logistic Regression, Academic Performance, Good or Bad Performance, Decision making, Counselling and guidance, pre-requisites to MBA.

## I. Introduction

Under the guidance of trained and experienced faculty members and mentors, education is a process of acquiring skills, knowledge, methods and values. In India, huge importance is given to acquire knowledge right from the Vedic period. Now that the student population has increased dramatically over the past few decades, it is vital to understand if the quality of the students is at par with the requirements of the growing economic scenario, as there is a huge possibility of unemployment, in the future. It is known that a student's performance can be judged based on past records in academics, due to the intensive focus on assessment results and test preparation, along with a few other dependent factors like Gender, Age and Work Experience. Hence, this is a study that is conducted to predict the performance of Post Graduate students, pursuing Master of Business Administration from National Institute of Technology, Tiruchirappalli, Tamil Nadu, India. The prediction of variables that have an effect on the academic performance of students is essential as assistive programs can be implemented to avoid dropouts or failing scores among the student community, for a better success rate in the course, which will also help with respect to choosing the right field during the course of time.

## II. Literature Review

The main objective of Ashine et al. (2016) is to assess and identify major factors which influence the academic achievement of students

at College of Natural and Computational Science of Wolaita Sodo University (WSU). Out of 1,497 students, a sample of 316 students was taken from nine departments of College of Natural and Computational Science using stratified random sampling with proportional allocation to size of department. It was concluded that to increase and improve students' academic achievement, some steps for securing student first choice of department and counselling as well as guiding about peer influence in university should be taken into consideration.

The work done by Aromolaran et al. (2013) is an empirical survey study conducted to determine the socio-economic factors influencing student academic performance in Yaba College of Technology, Yaba, Lagos. The data used for this work was collected from 600 students across different departments using a 28-item structured questionnaire, administered by quota sampling method, a non-probability sampling technique. It was concluded that the students should be orientated on the possible factors (both demographic and socio-economic) that may impede their success in academic pursuit.

The study by Mabula et al. (2015) investigated the performance of secondary school students in Mathematics at Selected Secondary Schools in Mtwara Municipality and Ilemela District by keeping Absenteeism, Conduct, Type of School and Gender as explanatory Factors. Findings show that two out of four explanatory factors used in the study (absenteeism and misconduct) significantly predict student performance in mathematics based on binary logistic regression fitted in the study.

## III. Methodology

A structured questionnaire was given to 64 students of an MBA 2<sup>nd</sup> year class and this was used in the data gathering process for the study. MATLAB was used for statistical data analysis conducted (correlation test, chi-square and Logistics Regression).

### A. Logistic Regression

The logistic regression model aims to describe the relationship between one or more independent variables that may be continuous, categorical, or binary. Due to its contribution, the logistic regression model has become a useful tool in business, medicine, epidemiology, sociology and marketing and so on, as it can predict the ailment disease, the granting of a credit to a specific individual, the success or failure of a business, among many others [1].

A binary logistic regression model for the academic performance of the student is formulated by

$$L_i = \text{Log} \frac{P_i}{1-P_i} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + U_i$$

From this,

$$P_i(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + U_i)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + U_i)}$$

Where,

$L_i = 1$  if the CGPA is greater than or equal to 8.5

$P_i$  = the probability of the student having good or bad academic performance

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$  = partial regression coefficients

$X_1$  = Age

$X_2$  = Gender

$X_3$  = Under-Graduation Percentage

$X_4$  = 12th Standard Percentage

$X_5$  = 10th Standard Percentage

$X_6$  = Work Experience

$X_7$  = CAT Percentile

$U_i$  = Error term

## B. Variables Identification

### 1. Dependent Variable

The dependent variable of this study is “academic performance”. For the study, the response variable “academic performance” has two binary outcomes coded 0 if a student might become a bad performer (CGPA<8.5) and 1 if student has the possibility of becoming a good performer (CGPA>=8.5).

### 2. The Explanatory Variables

The predictor variables considered in this analysis include Age, Gender, Under-Graduation Percentage, 12th Standard Percentage, 10th Standard Percentage, Work Experience and CAT Examination Percentile. The Common Admission Test (CAT) is a computer based test held in India. It tests Quantitative Ability (QA), Verbal Ability (VA) and Reading Comprehension (RC), Data Interpretation (DI) and Logical Reasoning (LR) of a candidate for selecting students for Business Administration programs provided by various B-Schools all over India.

### 3. Data Collection Methods

Primary data was collected through a structured questionnaire for

a quantitative method.

### (i). Data Analysis

Feature scaling was used to normalize the data due to the variations in the data collected and MATLAB was used for the analysis.

### (ii). Study Design

The research design is quantitative in nature and a conclusive research design can be employed.

### (iii). Source of Population

The students of the Post Graduate Course of Master of Business Administration admitted in the National Institute of Technology, Tiruchirappalli during the academic year 2015 were considered as the population.

### (vi). Sample Size

The Study Subjects are 64 students which comprises the whole population, as the sample is feasible.

### C. Purpose of the Study

Using binary logistic regression to examine the presence and magnitude of associations between the student’s profile and proficiency for scoring at least 8.5 or above, the model classifies the student likely to be “Good” or “Bad” in the Post Graduate course.

### IV. Data Analysis Using Matlab

The data collected is presented in the Table 1 below. Gender has been marked as “1” for Male and “0” for Female. The data for Age, 10<sup>th</sup> Standard Percentage, 12<sup>th</sup> Standard Percentage, Under-Graduation Percentage, CAT Examination Percentile and Work Experience have undergone a feature scaling for normalization. Also, “Good” performance is indicated by “1” and “Bad” performance is indicated by “0”.

Table 1: Data Summary

Age	Gender	Under Graduation %	12th %	10th %	Work Experience (in Months)	Good/Bad
24.29041	1	66	67	68	0	0
25.10959	1	84	64.6	81.6	0	0
23.97534	1	65	77.4	87.28	0	0
25.37808	1	70	78	67	0	0
24.79178	0	61	74	76	0	1
23.94795	1	81.5	81.33	68.6	0	0
23.39178	0	73.4	83	93.2	0	1
25.18356	1	70	88.4	89.16	0	0
22.63288	1	82.4	82	87	0	1
24.38082	1	85.3	92.4	91	0	1
25.46027	1	60.8	94.1	86.83	0	0
23.89315	1	74.6	83.1	82	0	0
23.22466	1	80	83.4	91	0	1
23.19726	0	83	87.91	85	0	1
23.65205	0	73	77.6	77.9	0	0
24.69315	0	67.8	56	64.14	0	0
22.48767	0	80.1	70	78	0	0
26.01644	1	67.2	91.8	91.2	0	1

24.70411	1	82.1	87.1	88.8	12	0
25.07671	0	70	84.67	93.8	0	0
24.68493	1	79.6	62.5	55	0	0
24.59178	1	70	69.6	74.6	0	0
26.40548	1	64.4	94.25	88.54	0	1
23.71507	1	64	93.2	96.2	0	0
24.8	0	63.6	94.75	91.4	0	0
23.83836	1	72.9	82.2	79	0	0
22.6274	1	61.3	87	79.2	0	0
23.19726	0	81.3	77	85	0	0
22.77534	1	71	83.5	86.85	14	0
25.59726	1	59	70	82.5	0	0
25.59178	1	63	76	88	0	0
22.75068	0	65	66.6	78.6	0	0
26.06849	1	72	64	83.8	0	0
24.47397	1	73.5	90.75	87.2	0	0
23.47945	1	76.75	85.4	84.5	0	0
24.82192	1	85.4	87.8	89	0	0
23.37534	0	7.35	78.4	84.6	0	1
22.34247	0	77.2	61.2	83	0	0
23.30685	0	85.4	79.4	86.2	0	0
23.45205	1	76.75	81.4	70.4	6	0
25.89315	1	74.9	69.4	85.2	7	0
25.15342	0	70.01	75.8	87.4	8	0
25.24932	1	68.9	79.4	86.2	8	0
24.52329	1	73.5	90.3	90.4	10	0
24.33151	0	80.75	64.7	71.8	12	0
24.48493	1	86.9	77.4	80	12	0
24.53425	1	71	84.71	93.75	14	0
26.73973	1	61	92.3	85	14	0
24.3726	0	91	63.6	79.2	15	1
25.18904	1	70.11	70	69.8	16	1
25.92877	1	83.54	83	85.23	21	1
27.06849	0	70.9	86	94.2	24	1
25.01096	1	70	81.33	72.8	21	1
24.84384	0	66.4	95.9	92.3	23	1
25.46575	0	67	83.2	87.8	29	1
25.17534	0	68.1	93.8	94.8	23	1
25.36164	1	77.2	81.4	70.4	30	1
26.02466	1	72.9	61.2	83	24	1
25.78082	0	8.27	80	80	25	1
25.89863	1	77.9	79	80	28	1
32.22192	1	67	89	85	32	1
26.49863	0	67.5	87.6	88.2	40	1
28.16712	0	72.9	87.8	89	40	0
28.16438	1	68.9	78.4	84.6	42	0

### A. Statement of Hypothesis

1.  $H_0$ : There is no association between student Academic Performance (CGPA) and student Age.
2.  $H_1$ : There is significant association between student Academic Performance (CGPA) and student Age.
3.  $H_0$ : There is no association between student Academic Performance (CGPA) and student Gender.
4.  $H_1$ : There is significant association between student Academic Performance (CGPA) and student Gender.
5.  $H_0$ : There is no association between student Academic Performance (CGPA) and Under-Graduation Percentage.

6.  $H_1$ : There is significant association between student Academic Performance (CGPA) and Under-Graduation Percentage.
7.  $H_0$ : There is no association between student Academic Performance (CGPA) and 12th Standard Percentage.
8.  $H_1$ : There is significant association between student Academic Performance (CGPA) and 12th Standard Percentage.
9.  $H_0$ : There is no association between the student Academic Performance (CGPA) and 10th Standard Percentage.
10.  $H_1$ : There is significant association between student Academic Performance (CGPA) and 10th Standard Percentage.
11.  $H_0$ : There is no association between student Academic Performance (CGPA) and Work Experience.
12.  $H_1$ : There is significant association between student academic performance (CGPA) and Work Experience.
13.  $H_0$ : There is no association between student academic performance (CGPA) and CAT Examination Percentile.
14.  $H_1$ : There is significant association between student academic performance (CGPA) and CAT Examination Percentile.

Table 2: Correlation Test

	Good/Bad	Age	Gender	UG %	12th %	10th %	Work Ex	CAT
Good/Bad	1							
Age	0.228533758	1						
Gender	-0.185577943	0.150938711	1					
UG %	-0.161670882	-0.131869682	0.187847037	1				
12th %	0.20188592	0.171445549	0.108025292	-0.095482638	1			
10th %	0.1547634	0.119471823	-0.120479257	-0.054345057	0.622439941	1		
Work Ex	0.413858592	0.627562983	-0.111582416	-0.069855921	0.148850468	0.103324091	1	
CAT	-0.098595167	0.146692038	0.199886493	0.086988584	0.027625562	0.024898735	0.115584815	1

### B. Correlation Test

The student academic performance is categorized into two categories- "Good" and "Poor". Those students whose CGPA is less than 8.5 are categorized as "Poor", while those students whose CGPA is greater than or equal to 8.5 are categorized as "Good" on a scale of 10.

According to the sample, students with CGPA less than 8.5 are 84.375% of the total respondents and those in the category of 8.5 and above are 15.625% of the total respondents. This implies that the bulk of the respondents need to improve to get into the 8.5 and above CGPA bracket.

### C. Log Likelihood

The most common assessment of overall model fit in logistic regression is the likelihood ratio test, which is simply the chi-square difference between the null model (i.e., with the constant only) and the model containing the predictors [2].

Under Model Summary, the -2 Log Likelihood statistics is 21.3162. This statistic's main motto is to measure how poorly the model predicts the student academic performance in "Okay" status. Smaller the statistic, better is the model.

Table 3: Model Summary

-2 Log Likelihood Value	21.3162
Status of the Model	Okay

### D. Goodness of Fit (Chi Square Test)

P stands for probability, and the P-value is the probability that the test outcome would take a value as extreme or more extreme than that observed.

As it is observed from the table, P-value is 0.0289 at 5% level of significance, which is insignificant and therefore the fitted logistic regression model is a good fit.

Table 4: Goodness of fit Test

Chi Square	Degrees of freedom	P value
14.0671	6	0.0289

### E. Chi Square Analyses of Hypothesis

A small p-value (typically  $\leq 0.05$ ) indicates strong evidence against the null hypotheses, and hence, the null hypotheses can be rejected.

### F. Variables in the Equation

Table 5: Variables in the Equation

Variables	Beta Values	Standard Error	Z Values
Constant	-0.3041	0.5453	-0.5576
Age	-0.1238	0.4017	-0.3083
Gender	-0.5799	0.7008	-0.8275
UG %	-0.2024	0.3134	-0.6458
12th %	0.3966	0.4197	0.9449
10th %	0.0795	0.4315	0.1842
Work Exp	1.0061	0.4217	2.386
CAT	-0.3063	0.3084	-0.9932

Numerical problems such as multicollinearity among the independent variables (constant excluded) are checked by examining the value of standard errors for the B-coefficients. Standard error larger than 2 indicates numerical problems. None of the independent variables in this analysis has a standard error larger than 2, indicating that there is no numerical problem of the model in the current study concerned regarding the academic performance [4].

Looking at the regression coefficients, 10th Standard Percentage has the least association with the model for Academic Performance with just 0.0795 and Work Experience has the maximum association with the model with 1.0061 as the value. Further, positive regression coefficient means that a change in the value of the independent variable will have a positive impact on the academic performance, i.e. for an example, with an increase in 12th Standard Percentage, there will be a 0.3966 increase in the Academic Performance. And with a negative coefficient, there will be a negative impact on the Academic Performance, i.e. for



an instance, with an increase in CAT Percentile, there will be a 0.3063 decrease in the Academic Performance.

When Z values are considered, tabular Z value needs to be lesser than the individual Z value found during the analysis, for the variable to be significant. As per the study, Work Experience has significant importance in model with the value being 2.386, which is much higher than the tabular Z value of 1.6449.

Thus, the logistic regression model for the study is

$$\text{Log} \frac{P}{1-P} = -0.3041 - 0.1238 (X_1) - 0.5799 (X_2) - 0.2024 (X_3) + 0.3966 (X_4) + 0.0795 (X_5) + 1.0061 (X_6) - 0.3063 (X_7).$$

Where,

- X<sub>2</sub>= Gender
- X<sub>3</sub>= Under-Graduation Percentage
- X<sub>4</sub>= 12th Standard Percentage
- X<sub>5</sub>= 10th Standard Percentage
- X<sub>6</sub>= Work Experience
- X<sub>7</sub>= CAT Percentile

And hence, we get the probability equation as

$$P_i(x) = \frac{\exp(-0.3041-0.1238(X1)-0.5799(X2)-0.2024(X3)+0.3966(X4)+0.0795(X5)+ 1.0061(X6)-0.3063(X7))}{1+ \exp(-0.3041-0.1238(X1)-0.5799(X2)-0.2024(X3)+0.3966(X4)+0.0795(X5)+ 1.0061(X6)-0.3063(X7))}$$

**G. Classification Table**

Table 6: Accuracy and Error

Correctly predicted (%)	82.8125
Apparent Error Rate (%)	17.1875

Thus, the model is good enough for prediction purposes in the future, since it has an error rate of just 17.1875% and an accuracy of 82.8125%, when the considered cut off is 0.5.

**V. Conclusion and Recommendations**

The most powerful variable in determining academic performance of a student would be Work Experience and the least important variable would be 10th Standard Percentage. Some of the recommendations to improve the performance can be as follows:

- Counselling and guiding about course structure by university should be taken into consideration, so that the students are briefed upon the pre-requisites which would help them decide to opt or not for the course, for a better career path.
- Stalk holders should set programs to strengthen self-concept to make the students confident about their potential. The motivational factor is understood as having an orientation towards achievement that sets high goals and strive to reach them and as a skill, the interest to achieve a good academic performance. This will influence the choice in tasks, proposed goals, effort, and persistence of actions towards reaching such goals. Therefore, professors must consider the importance of reassuring the student’s self-esteem when learning new concepts. [1]
- Other factors like intelligence, study habits and attitude of a student towards a B-School, different aspects of their personality, socio economic status, home environment, curriculum intentions and school and classroom environments

also affect student performance. Hence, further study can be done by including these qualitative factors.

**References**

- [1] Urrutia-Aguilar, M.E.; Fuentes-Garcia, R.; Martinez, V. D. M.; Beck, E.; León, S. O.; Guevara-Guzman, R., “Logistic Regression Model for the Academic Performance of First-Year Medical Students in the Biomedical Area”*Creative Education*, 7, 2202-2211, 2016. [Online] Available: <http://dx.doi.org/10.4236/ce.2016.715217>
- [2] Ashine, K.M.; Zewude, B.T., “Binary Logistic Regression Analysis in Assessment and Identifying Factors That Influence Students’ Academic Achievement: The Case of College of Natural and Computational Science, Wolaita Sodo University, Ethiopia.”, 2016. [Online] Available: <http://www.isicenter.org>
- [3] Aromolaran, Adeyemi, D.; Benjamin, E.; Isaiah, K.; Olukotun, Okon, O.; Oyeyinka, “Binary Logistic Regression of Students Academic Performance in Tertiary Institution in Nigeria by Socio-Demographic and Economic Factors”, 2013.
- [4] Mabula, S, “Modeling Student Performance in Mathematics Using Binary Logistic Regression at Selected Secondary Schools- A Case Study of Mtwara Municipality and Ilemela District.”, 2015. [Online] Available: <http://www.iiste.org>



Surendheran R, pursuing MBA from National Institute of Technology (NIT) Tiruchirappalli, Tamil Nadu, India, and a founder of two starts-ups, is into Operations and Big Data Analytics. An Industrial Engineer from heart, Operations Excellence is one of the major areas of his career. His aim in life is to become a successful Entrepreneur and a mentor in the future.



Monisha Ravi, a student of National Institute of Technology Tiruchirappalli (NIT), Tamil Nadu, India, is pursuing her MBA and has taken interest in Marketing and Data Analytics. Trying to further her venture into these streams, along with her Engineering background in Electronics and Communication, she wishes to become a notable Entrepreneur.