Validation, Testing and Sensitivity Analysis of Binary Logistic Regression Model for Predicting the Performance of Engineering Students

Radheshyam H. Gajghat[1]*, Dr. Chandrahas C. Handa[2], Dr. Rakesh L. Himte[3]
[1]Ph.D. Research Scholar, Department of Mechanical Engineering, K.D.K. Collage of Engg. & Tech., Nagpur, India.
[2]Professor & Head, Department of Mechanical Engineering, K.D.K. Collage of Engg. & Tech., Nagpur, India.
[3]Professor & Head, Department of Mechanical Engineering, Priyadarshini Institute of Engg. & Tech., Nagpur, India.

ABSTRACT: A binary logistic regression (BLR) model was developed to predict the performance of engineering students in university examination. This model shows the mathematical relationship between influencing factors and the performance of engineering students in university examination. Pass/fail result in university examination is taken as performance measure, and personal, pre-admission, institutional and self-learning factors as influencing factors. This BLR model was validated by Artificial Neural Network (ANN) and tested by using newly collected samples. The accuracy of the model was found as 80.95 % which showed a good accuracy for such type of models. The optimum values of each influencing factors were also calculated as 4.01, 3.16, 4.12 and 4.04 for personal, pre-admission, institutional and self-learning factors respectively. In sensitivity analysis, it is observed that personal factors are most sensitive followed by institutional factors whereas pre-admission factors are less sensitive than self-learning factors. This study will help the engineering students to improve their performance in university examination by predicting their probability of passing in advance.

KEYWORDS: Validation, Testing, Sensitivity Analysis, Binary Logistic Regression, Students' Performance.

I. INTRODUCTION

During last decade number of private and government technical institutions has been started to fulfill the increased demand of technical manpower. The country has seen the quantitative growth of engineering institutions at diploma, degree and postgraduate level during this period, mostly a phenomenal growth in number of engineering students. Figure 1 shows the growth of AICTE (All India Council for Technical Education) approved undergraduate technical institutions [1] and figure 2 shows the growth of intake in AICTE approved technical institutions at undergraduate level [2] [3] [1].

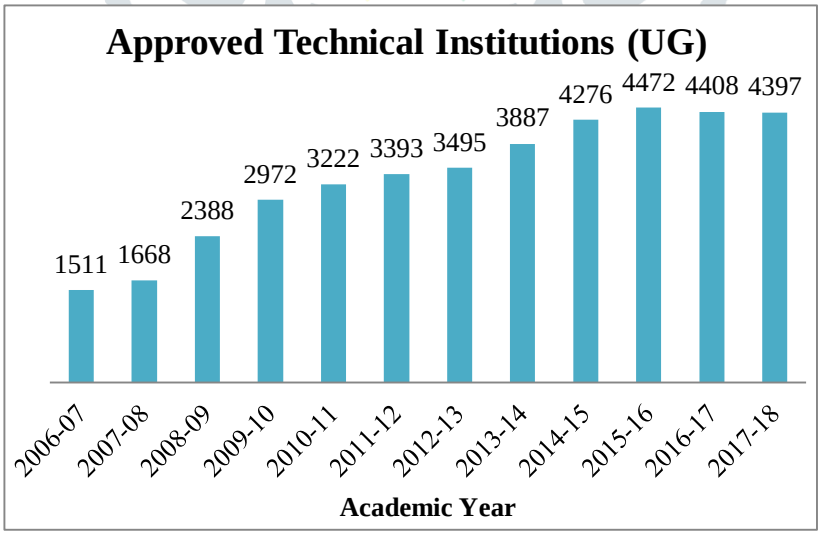


Figure 1: Growth of AICTE Approved Undergraduate Technical Institutions

At the time of independence, there were only 44 and 43 engineering colleges and polytechnics (including pharmacy and architecture institutions) with an intake capacity of 3200 and 3400 respectively. Due to the continuous efforts and initiatives taken by the government during consecutive five-year plans and particularly due to policy changed in the eighties which allowed participation of private and charitable organizations in the setting up of technical institutions on self-financing basis, the growth of technical education has become significant [1].

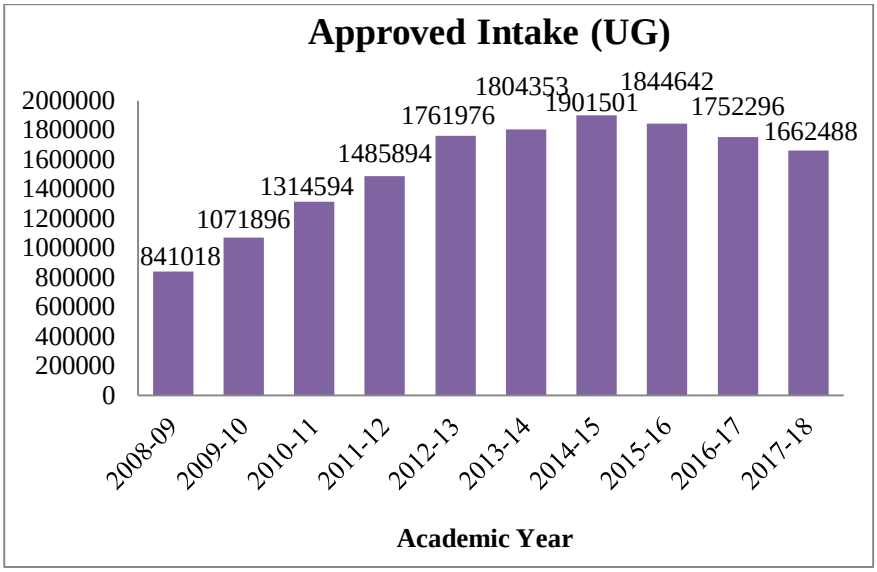


Figure 2: Growth of Intake in AICTE Approved Technical Institutions at UG Level

Due to the availability of large number of engineering seats in the country, even 12th class students with lower scores get chance to take admission in engineering courses, there by affecting the results. The poor result has also adversely affected the placement. As such now result of students has become the highest concern of engineering education system. If it is possible to know in advance which students are likely to fail, the corrective action such as arranging extra and personal improvement classes, use of advanced tools for teaching etc. can be taken by the college management or the teachers to improve the results. This will certainly help in improving the placements. Good placement is the most important factor that will help the college to attract the students [4].

There are certain parameters like family background, personal characteristics, high school academic background, college environment etc. which have significant impact on the performance and results of students. The previous studies proved that student's university results can be improved by predicting and controlling the influencing factors which affect their academic performance. Most of the studies are focused on students' performance in the foreign universities, which may not be suitable for Indian universities due to the differences in their academic, social and cultural environment. So there was need to develop such a model which will be more suitable for the students of Indian universities [5].

II. DEVELOPMENT OF BINARY LOGISTIC REGRESSION MODEL

In this study binary logistic regression (BLR) method was used to develop the mathematical model. Here, pass/fail in the university examination was taken as the dependent variable and personal, pre-admission, institutional, self-learning factors were taken as independent factors for predicting the students' performance at university level examination.

The data of 419 engineering students of technical institutions of Chhattisgarh collected through the questionnaire [6] was used to formulate the mathematical model by using binary logistic regression method. This binary logistic regression model (R^2 = 0.842) was formulated [7] as

$$Y = \frac{e^{(10.484X_1+1.929X_2+2.948X_3+2.225X_4-63.947)}}{1+e^{(10.484X_1+1.929X_2+2.948X_3+2.225X_4-63.947)}}$$

Where Y is the probability of passing on fail i.e.1 on 0

X_1, X_2, X_3 & X_4 are Personal, Pre-admission, Institutional and Self-learning factors respectively

e is the base of the natural logarithm,

The above designed binary logistic model can predict the probability of passing of engineering students in university examination. The optimum predicted accuracy for pass/fail result dataset was found at cut-off value of 0.65. Therefore, the sample categorised as pass if predicted probability is equal or more than 0.65. All influencing factors contributed significantly at significance level 0.05 ($p < 0.05$).

III. RESULTS AND DISCUSSION

3.1 Validation of BLR Model for Pass/Fail Result by ANN

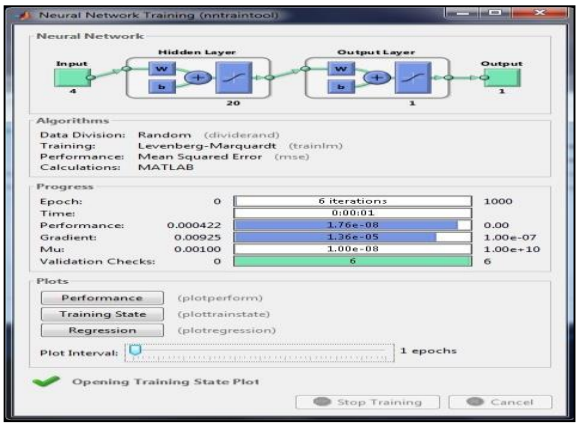
The validation of binary logistic regression model was done by using Artificial Neural Network and confirmed by stakeholders [8]. The separate sample of 46 students as shown in table 1 was used for model validation. The predicted output was calculated by using BRL model [9] which was further used in ANN model for validation purpose [10]. ANN model has given the value of overall R^2 as 0.9988 which indicates that the BLR model has predicted accurate result and thus validated. Screen shots of validation of BLR Model by using ANN in MATLAB are shown in figure 3(a-d). The BLR model validation details are given in table 2.

Table 1: Data for Validation of BLR Model for Pass/Fail Result by using Separate Samples

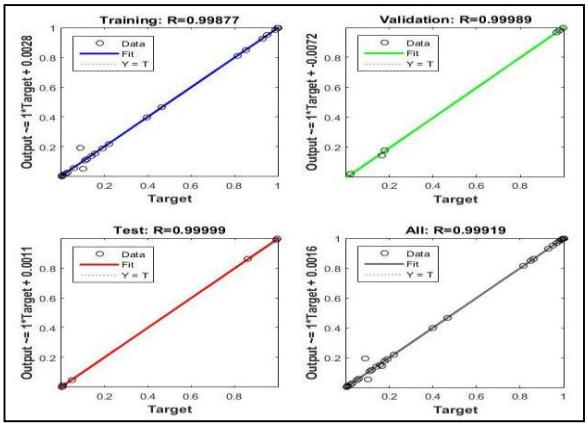
Sr. No.	X_1	X_2	X_3	X_4	Y	Y_{BLR}
1	3.54	2.00	3.80	3.50	0	0.0189
2	3.54	2.75	4.20	3.90	1	0.3933
3	4.28	3.00	4.20	3.70	1	0.9994
4	3.17	2.50	3.07	2.90	0	0.0000
5	3.80	3.00	3.60	4.10	1	0.8115
6	3.54	3.00	4.20	3.30	0	0.2165
7	3.52	1.50	3.73	3.70	0	0.0074
8	3.72	3.50	3.13	3.30	0	0.1621
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
39	4.28	3.25	4.53	4.60	1	1.0000
40	3.72	2.25	3.67	3.50	0	0.1154
41	3.52	2.00	4.00	3.30	0	0.0174
42	3.93	2.50	4.13	4.10	1	0.9688
43	3.59	2.50	3.07	3.10	0	0.0038
44	3.48	3.50	2.93	3.80	0	0.0259
45	4.13	2.50	4.00	3.90	1	0.9905
46	3.87	2.00	3.53	3.10	0	0.0991

Table 2: Validation Details of BLR Model for Pass/Fail Result by ANN

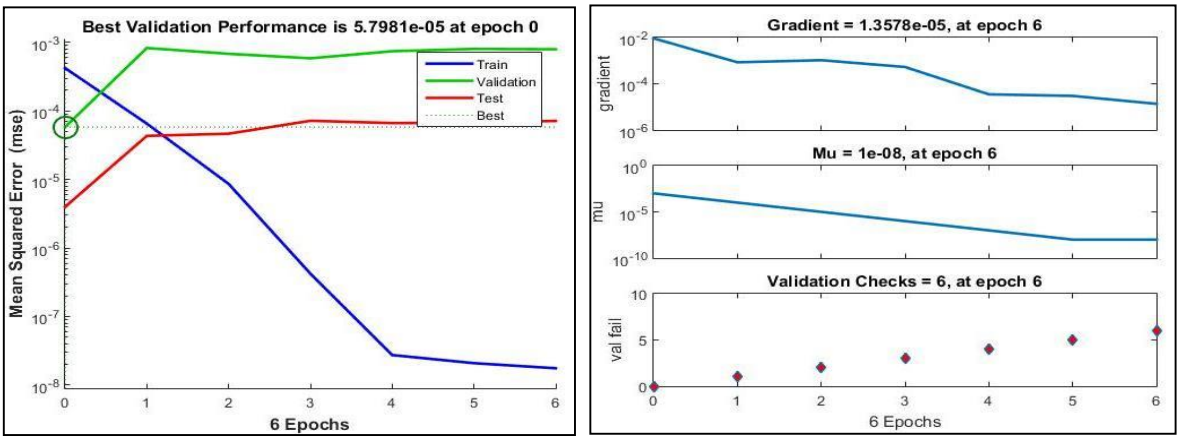
Overall R^2	0.9988
Training R^2	0.9975
Validation R^2	0.9998
Testing R^2	0.9999
Mean Square Error	0.000057981
No. of Hidden Layers	2
No. of Neurons	20
Input Variables	4 (X_1, X_2, X_3, X_4)
Output Variables	1 (Y - Pass/Fail)



a. Neural Network



b. Regression



c. Performance **d. Training State**

Figure 3: Screen Shots of Validation of BLR Model for Pass/Fail Result

3.2 Testing of BLR Model by using Newly Collected Samples

Accuracy of Binary logistic regression model was calculated by using 42 newly collected samples, from fourth semester engineering students. Their result of third semester was taken as performing factor. The predicted probability for each student was calculated by using BLR Model as shown in table 3. The cut-off value of this model was 0.65. By using this cut-off value, the predicted probability was converted into pass/fail results. As per BLR Model out of 42 students, 23 students were passed and 19 students were failed. Then this result of BLR Model was cross checked by comparing actual results. The prediction was correct for 34 samples and wrong for 8 samples. Thereby, the accuracy of this model was found as 80.95 %, showed a very good and satisfactory prediction for such type of models.

Table 3: Testing of BLR Model for Pass/Fail Result by using Newly Collected Samples

Sr. No.	X_1	X_2	X_3	X_4	Y	Predicted Prob. Yp	Binary Yp (Cut-off = 0.65)	Correct Prediction?
1	4.07	3.25	4.33	3.90	1	0.9984	1	Y
2	4.02	2.50	4.13	3.80	1	0.9757	1	Y
3	4.04	3.25	4.33	4.20	1	0.9989	1	Y
4	4.02	3.00	3.67	4.60	1	0.9937	1	Y
5	4.39	3.00	4.27	4.60	1	1.0000	1	Y
6	3.98	3.25	4.13	3.90	1	0.9927	1	Y
7	3.76	2.00	3.73	3.30	0	0.0912	0	Y
8	4.04	3.00	4.13	4.00	1	0.9952	1	Y
9	3.57	2.25	3.53	3.10	0	0.0074	0	Y
10	3.85	2.75	4.00	4.00	1	0.9171	1	Y
11	4.50	2.50	4.53	4.70	0	1.0000	1	N
12	4.17	3.25	3.87	3.90	1	0.9979	1	Y
13	4.24	2.50	4.47	4.20	1	0.9996	1	Y
14	4.24	3.25	4.13	4.10	1	0.9997	1	Y
15	3.43	3.25	3.47	3.60	0	0.0315	0	Y
16	3.50	3.25	3.13	3.30	1	0.0122	0	N
17	3.89	3.75	4.00	4.00	1	0.9917	1	Y
18	3.89	3.00	3.60	3.30	1	0.6467	1	Y
19	3.02	2.25	3.13	2.30	0	0.0000	0	Y
20	4.15	3.00	4.13	4.00	0	0.9985	1	N
21	3.91	3.25	3.73	3.70	0	0.9307	1	N
22	3.87	3.25	4.13	3.80	1	0.9719	1	Y
23	3.54	3.50	3.60	2.90	0	0.0489	0	Y
24	3.57	2.50	3.27	2.90	0	0.0035	0	Y
25	3.07	2.75	3.93	2.70	0	0.0001	0	Y
26	3.61	2.75	3.87	3.60	1	0.1999	0	N
27	4.00	2.50	3.47	4.30	1	0.9316	1	Y
28	4.04	2.75	4.07	3.60	1	0.9773	1	Y
29	3.59	2.25	3.13	3.70	0	0.0108	0	Y
30	3.50	2.50	3.13	2.10	0	0.0002	0	Y

Sr. No.	X_1	X_2	X_3	X_4	Y	Predicted Prob. Yp	Binary Yp (Cut-off = 0.65)	Correct Prediction?
31	3.96	3.25	3.87	3.50	1	0.9527	1	Y
32	3.70	2.75	3.53	3.50	1	0.1571	0	N
33	3.11	2.25	2.73	2.30	0	0.0000	0	Y
34	4.07	2.50	3.33	3.40	1	0.7109	1	Y
35	3.61	2.75	3.40	3.20	0	0.0253	0	Y
36	3.33	2.50	3.20	3.30	0	0.0006	0	Y
37	3.54	3.00	3.40	3.30	0	0.0258	0	Y
38	4.09	3.00	3.73	4.60	1	0.9974	1	Y
39	3.91	3.00	3.80	4.00	0	0.9517	1	N
40	3.43	2.75	3.60	3.50	0	0.0145	0	Y
41	2.87	2.25	3.27	2.30	0	0.0000	0	Y
42	4.11	2.75	2.60	3.60	1	0.5302	0	N

Correct Predictions = 34, Wrong Predictions = 8

Accuracy of the BLR Model = (34/42) x 100 = 80.95 %

3.3 Optimum Level of Influencing Factors

Through this exercise an attempt was made to identify the influence of various influencing factors (X) on the examination result. In this research study, 419 samples were used for model formulation out of which 263 samples of pass students and 156 of fail students. Average weightage given by pass and fail students for all the influencing factors were calculated as shown in table 4.

Table 4: Optimum Values of Influencing Factors and Gap between Pass and Fail Students

Influencing Factors ➡	X_1	X_2	X_3	X_4	Prob. of Passing (Y)
Average Score of Fail Students	3.53	2.56	3.54	3.41	0.0186
Average Score of Pass Students (Optimum)	4.01	3.16	4.12	4.04	0.9951
Gap between Pass & Fail Students	0.48	0.6	0.58	0.63	
Gap in %	11.97%	18.99%	14.08%	15.59%	
Rank	4	1	3	2	

The probability of passing for fail and pass students was calculated by using Binary logistic regression model for pass/fail result. As shown in above table 4, the probability of passing of fail students was 0.0186, which is very close to 0 and probability of passing of pass students was 0.9951, which is very close to 1. The average values of influencing factors calculated for pass students give almost 100 % probability of passing. So these values were taken as optimum values of influencing factors.

The gap between pass and fail students was also calculated by subtracting the average score of fail students from average score of pass students which is 11.97 %, 18.99 %, 14.08 % and 15.59 % for personal, pre-admission, institutional and self-learning factors respectively. Pre-admission factors are showing the maximum gap i.e. 18.99 % but the values of most of these factors can't be changed because these are based on pre-academic background. The second highest gap i.e. 15.59 % is showing for self-learning factors where there is a lot of scope for improvement by changing the self-learning habits. There is also a moderate gap i.e. 14.08 % for institutional factors which can be filled by improving the institutional academic environment. Personal factors have the lowest gap i.e. 11.97 % which can be filled by improving the personal characteristics.

3.4 Sensitivity Analysis

The sensitivity analysis is carried out to identify as to which parameter is sensitive to change. The sensitivity analysis assesses the effect on correct and incorrect predictions if the prediction probability threshold is altered [11]. The optimum values of influencing factors were calculated for approximately 100 % probability of passing are 4.01, 3.16, 4.12 and 4.04 as personal, pre-admission, institutional and self-learning factors respectively. The effect of these factors on the probability of passing was checked by varying the values of one factor by 10 % in each step while the values of other factors were kept unchanged as shown in table 5.

A graph was plotted between influencing factors (X_1, X_2, X_3, X_4) on X axis and probability of passing (Y) on Y axis as shown in figure 4. The average rate of change of probability was calculated and shown in table 6.

Table 5: Impact of Influencing Factors on Probability of Passing

X_1	X_2	X_3	X_4	Y	Rate of Change
4.01	3.16	4.12	4.04	0.99521	0.23879
3.61	3.16	4.12	4.04	0.75642	0.71210
3.21	3.16	4.12	4.04	0.04432	0.04363
2.81	3.16	4.12	4.04	0.00069	0.00068
2.41	3.16	4.12	4.04	0.00001	0.00001
2.01	3.16	4.12	4.04	0.00000	0.00000
1.60	3.16	4.12	4.04	0.00000	0.00000
1.20	3.16	4.12	4.04	0.00000	0.00000
0.80	3.16	4.12	4.04	0.00000	0.00000
0.40	3.16	4.12	4.04	0.00000	----------
4.01	3.16	4.12	4.04	0.99521	0.00398
4.01	2.84	4.12	4.04	0.99123	0.00725
4.01	2.53	4.12	4.04	0.98399	0.01306
4.01	2.21	4.12	4.04	0.97093	0.02313
4.01	1.90	4.12	4.04	0.94780	0.03980
4.01	1.58	4.12	4.04	0.90800	0.06511
4.01	1.26	4.12	4.04	0.84289	0.09823
4.01	0.95	4.12	4.04	0.74465	0.13147
4.01	0.63	4.12	4.04	0.61319	0.15033
4.01	0.32	4.12	4.04	0.46286	----------
4.01	3.16	4.12	4.04	0.99521	0.01116
4.01	3.16	3.71	4.04	0.98406	0.03581
4.01	3.16	3.30	4.04	0.94824	0.10356
4.01	3.16	2.88	4.04	0.84469	0.22719
4.01	3.16	2.47	4.04	0.61750	0.29354
4.01	3.16	2.06	4.04	0.32396	0.19943
4.01	3.16	1.65	4.04	0.12453	0.08402
4.01	3.16	1.24	4.04	0.04051	0.02813
4.01	3.16	0.82	4.04	0.01238	0.00867
4.01	3.16	0.41	4.04	0.00371	----------
4.01	3.16	4.12	4.04	0.99521	0.00689
4.01	3.16	4.12	3.64	0.98832	0.01653
4.01	3.16	4.12	3.23	0.97179	0.03836
4.01	3.16	4.12	2.83	0.93343	0.08253
4.01	3.16	4.12	2.42	0.85090	0.15185
4.01	3.16	4.12	2.02	0.69904	0.21308
4.01	3.16	4.12	1.62	0.48596	0.20810
4.01	3.16	4.12	1.21	0.27787	0.14246
4.01	3.16	4.12	0.81	0.13541	0.07548
4.01	3.16	4.12	0.40	0.05993	----------

Table 6: Effectiveness Ranking of Influencing Factors

Influencing Factors	*Xmax*	*Xmin*	*(Xmax – Xmin)*	*Ymax*	*Ymin*	*(Ymax – Ymin)*	Ave Rate of Change (%)	Rank
X_1	4.01	0.401	3.6090	0.99521	7.68E-15	0.99521	27.58	1
X_2	3.16	0.316	2.8440	0.99521	0.46286	0.53235	18.72	4
X_3	4.12	0.412	3.7080	0.99521	0.003707	0.99151	26.74	2
X_4	4.04	0.404	3.6360	0.99521	0.059925	0.93529	25.72	3

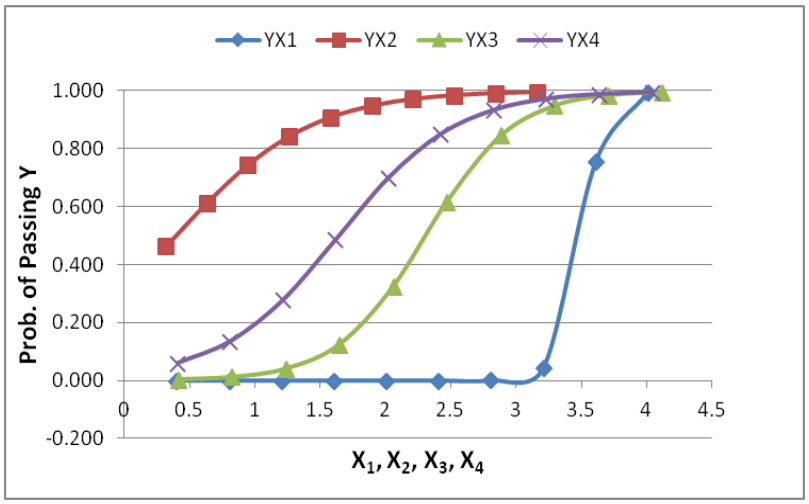


Figure 4: Effect of All Factors on Probability of Passing

The above figure 4 shows that the probability of passing can't be achieved only by focusing on any one influencing factor. It means the students have to focus on all factors simultaneously to achieve 100 % passing probability. It is also observed that rate of change of probability up to the value of 3.21 as personal factors (X_1) is negligible and probability of passing is equal to 0. At the value of 3.21, there is sudden rise in the rate of change of probability up to value of 3.61. Between these two values the rate of change of passing probability is highest and more than the rate of change of personal factors score. Then this rate decreases slightly up to 4.01. This indicates that students have to enhance their personal factors' score about 3.61 for increasing the good chances of passing. For the pre-admission factors (X_2), it is observed that the rate of change of probability is less than the rate of change of pre-admission factors score. The rate of change of probability is increasing rapidly up to the score of 1.26, then slightly decreases the rate till score 1.9. After this score, there is very little improvement in the probability. So it is suggested that pre-admission factors score up to 1.9 is very significant and play an important role in passing the university examination.

For institutional factors (X_3), it is observed that initially up to score of 1.24; the rate of change of probability is very small and then increases rapidly up to the score of 3.3. Between these scores both the rates, rate of change of probability and rate of change of institutional factor score, are almost same. From the score of 3.3 to 4.12, probability increases very slowly and gradually reaches to 100 %. This analysis suggests that students should score the institutional factors up to minimum 3.3 for increasing the chances of passing. For self-learning factors (X_4), it is observed that the rate of probability is almost linear and increases gradually from the score of 0.40 to 2.83. Between these score, the rate of change of self-learning factor score is slightly higher than the rate of change of probability. From score 2.83 and onwards, the probability is increasing very slowly and finally reaches to 100 %. This means the students need to score minimum 2.83 to achieve a good probability of passing. Considering combined effect of all factors on the performance of engineering students, it can be concluded that personal factors show more sensitivity followed by institutional factors whereas pre-admission factors are less sensitive than self-learning factors. Initially and at the end, the rate of change of probability is found less than the rate of change of influencing factors. At the middle portion, the rate of change of probability is more. The ranking of the influencing factors was done on the basis of average change of rate of probability which also shows the sensitivity of the factors.

IV. CONCLUSIONS

On the basis of this study and analysis of results, some important conclusions can be drawn which will be helpful to solve the issues that are responsible for students' poor academic performance in university examination. This work will help the management, academicians and engineering faculties to frame a better and effective educational policy and modify the existing teaching learning process which will improve the overall performance of engineering students and hence enhance their employability. Based on this research work, the following conclusions are drawn and presented as below:

1. Table 2 shows the value of overall R^2 as 0.9988 given by ANN Model which indicates that the BLR model has predicted accurate result and thus validated.
2. Table 3 shows testing of binary logistic regression model for pass/fail result. The accuracy of this model was checked by using newly collected samples from the students of fourth semester. Out of 42 samples, 32 samples were predicted correctly. Therefore, calculated accuracy of this model is 80.95 % which is a good accuracy for such type of model.
3. Table 4 shows the optimum level of influencing factors (average score of pass students) as 4.01, 3.16, 4.12 and 4.04 for personal (X_1), pre-admission (X_2), institutional (X_3) and self-learning (X_4) factors respectively which gives almost 100 % of probability of passing for engineering students. It also shows a gap of 11.97 %, 18.99 %, 14.08 % and 15.59 % for the similar factors between pass and fail students. So there is necessary to make a proper

policy accordingly by the engineering colleges to reduce this gap to increase the probability of passing of engineering students.

4. Figure 4 and table 5 shows that 100 % of probability of passing can't be achieved by focussing only on one influencing factor. It means the students should have to focus on all factors simultaneously to achieve 100 % of probability of passing. From table 6, it is observed that personal factors are most sensitive followed by institutional factors whereas pre-admission factors are less sensitive than self-learning factors.

In nutshell, the developed binary logistic regression model is good enough to predict the performance of engineering students in their university examination with high accuracy.

References

[1] *All India Council for Technical Education Approval Process Handbook 2018-19.*

[2] *All India Council for Technical Education Approval Process Handbook 2013-14.*

[3] *All India Council for Technical Education Approval Process Handbook 2015-16.*

[4] R. R. Kabra and R. S. Bichkar, "Performance prediction of engineering students using decision trees," *International Journal of Computer Applications,* vol. 36, no. 11, 2011.

[5] N. Harb and A. El-Shaarawi, "Factors Affecting Students' Performance," *Munich Personal Archive (MPRA),* 2006.

[6] R. H. Gajghat, C. C. Handa and R. L. Himte, "Design and Development of a Questionnaire to Study the Effect of Various Influencing Factors on Performance of Engineering Students," *i-manager's Journal on Educational Psychology*, vol. 12, no. 2, pp. 126-135, 2018.

[7] R. H. Gajghat, C. C. Handa and R. L. Himte, "Development of a Model for Predicting the Performance of Engineering Students in University Examination," Communicated.

[8] C. C. Handa, R. D. Askhedkar, J. P. Modak and M. P. Singh, "Development and Validation of Computer Based Entrepreneurial Aptitude Test," *International J.of Multidispl.Research \& Advcs. in Engg.*, vol. 2, no. 1, pp. 245-257, 2010.

[9] V. N. Kalbande and C. C. Handa, "Developing a Model to Predict Employability of Engineering Students in Campus Placement for IT Sector," *International Journal of Advance Research in Engineering, Science \& Technology*, vol. 2, no. 6, pp. 201-206, 2015.

[10] V. N. Kalbande and C. C. Handa, "Developing a Model to Correlate Employability of Engineering Graduate with Employability Skill Sets in Campus Placement for IT Sector," *Online International Interdisciplinary Research Journal*, vol. 6, pp. 128-138, 2016.

[11] Q. Li, H. Swaminathan and J. Tang, "Development of a classification system for engineering student characteristics affecting college enrollment and retention," *Journal of Engineering Education,* vol. 98, no. 4, pp. 361-376, 2009.