MIJRD Multidisciplinary International Journal of Research and Development

Volume: 01 | Issue: 02 | 2021 - Open Access - Website: www.mijrd.com - ISSN: 2583-0406

Predicting Students Performance in the University System Using Discriminant Analysis Method

P.O. Aye

Dept. of Mathematical Sciences, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria

Email: patrick.aye@aaua.edu.ng

Abstract— This study employed discriminant analysis to predict undergraduate student's performance in the University system. Data on the first year students of the department of Mathematical Sciences were obtained from the Examination and Records unit of the department. The data was analysed using predictive discriminant analysis and it yielded a canonical discriminant function which successfully predicted 87.2% of the graduating student's class of degree. The cross validated classification showed that overall 72.27% were correctly classified. The model performance was validated with statistical tools.

Keywords— Discriminant analysis, Eigen value, Prediction, Canonical discriminant function, Confusion matrix, Hit ratio.

1. INTRODUCTION

The development of model that can adequately predict the performance of an undergraduate student in a university or polytechnic has been of great interest to many researchers or educators, over the years. Several researchers have focused their study area on designing a robust predictive model to predict student's final passing or graduating results. Kasih and Susanto (2012) applied discriminant analysis to predict student's final result. Student performance in some courses in the early stage (the first four semester) of their study in higher education were used as predictor variables. Their study was able to classify correctly into their respective group. Erimafa *et al* (2009) in their study, predicted successfully the graduating performances of undegraduate students in the university system using discriminant analysis method. The work of Usoro (2006) was centred on the application of discriminant function analysis to classify students into their various department based on their individual performance at the first year foundation programme in Polytechnics system. Akaboha and Kwofie (2016) employed discriminant analysis to determine student's performance in their final West African Senior Secodary Certificate Examination (WASSCE). Their study identified six factors such as BECE (Basic Education Certificate Examination) grade in Science, BECE grade in Mathematics, Type of basic education, Duration of the SHS (Senior High School) system, Entry admission age and BECE aggregate score determine the performance of student. Chaubey *et al* (2016) carried out work on method of discriminant function to analyzed factors influencing academic performance of students in the University. Their study revealed that variable attendance is the best predictor followed by graduation marks, study medium, academic gap, and entrance score influencing academic performance. Adebayo and Jolayemi (1998, 1999), have also applied the τ-

Volume: 01 | Issue: 02 | 2021 - Open Access - Website: www.mijrd.com - ISSN: 2583-0406

statistic to investigate how predictable the final year result would be using the first year result or Grade Point Average (GPA) of some selected University graduates.

Most of the early research work and much of the current one are centred on identifying factors influencing/affecting performance of student and causes of student's low academic performance in higher instutution. The focus of this study is to identify students who might be ''at risk'' (AR) and ''not at risk '' (NAR). The first group are the students who are in danger of graduating with Poor Class of Degree, PCD (that is, Third Class, Pass or Fail); and the second group are those that will graduate with Better Class of Degree, BCD (that is, First Class, Second Class Upper Division, Second Class Lower Division), within their first year of study. The method of discriminant analysis was adopted for this research due to the fact that it performed student - identification task better than commonly used educational measures such as regression analysis and correlation because the variable been predicted is categorical.

This study improved on the existing work on the application of discriminant analysis for predicting student's graduation performance by incorporating more predictor variables and validation of the model performance with more statistical tools.

2. MATERIALS AND METHODS

Data used for the study were obtained from the examinatinon and record unit of the department of Mathematical Sciences (Industrial Mathematics Option) in Adekunle Ajasin University, Akungba-Akoko, Ondo state. A first year academic performance of the admitted student for 2009/2010 academic session was studied, which include 75 students out of which one of the students failed to observe a proper registration data, leaving us with total of 74 students for the purpose of analysis of this research. The departmental academic record data were logically studied with special attention being paid to some of the factors (explanatory variable or predictors) influencing the probability of graduating with Better Class of Degree (BCD) and Poor Class of Degree (PCD).

In the first stage of data collection, two groups of students in terms of their graduating class of degree were formed, and four possible predictor variables, including the following, GPA for first semester 100 level, grades in all the faculty, MAT 102 and MAT 106 courses thought to be predictive of performance.

However, using the method of stepwise discriminant analysis (forward stepwise analysis), we found that only two of this variables made significant independent and combined contributions. Below is the analysis of the data as processed by a modern computer application called Statistical Package for Social Sciences (SPSS).

3. DATA ANALYSIS

The analysis of data was carried out using Linear Discriminant Function method introduced by Fisher (Fisher, 1936). The method was adopted because of it ability to discriminate correctly between two groups (Anderson, 1958).

Multidisciplinary International Journal of Research and Development

Volume: 01 | Issue: 02 | 2021 - Open Access - Website: www.mijrd.com - ISSN: 2583-0406

Using an arbitrary linear discriminant function given by

$$Z = U_1X_1 + U_2X_2 + \cdots + U_nX_n \quad [1]$$

where Z is the discriminant function, U is the weight of the variable.

The group statistics and variance- covariance matrix for the data was computed. The outputs obtained using SPSS statistical package are shown in Tables 1.

Table 3.1: Group Statistics for Grades obtained in the Proficiency Test

Groups		Mean	Standard Deviation	Unweighted	Weighted
1	Second Semester	56.0405	13.54995	74	74.000
	First Semester	64.2973	14.58910	74	74.000
2	Second Semester	58.9459	9.72906	74	74.000
	First Semester	46.4459	13.11816	74	74.000
Total	Second Semester	57.4932	11.84509	148	148.000
	First Semester	55.3716	16.47314	148	148.000

$$S^1 = \begin{bmatrix} 92.41 & 17.40 \\ 17.40 & 57.49 \end{bmatrix} \quad \text{and} \quad S^2 = \begin{bmatrix} 74.12 & 11.24 \\ 11.24 & 67.45 \end{bmatrix}$$

S^1, S^2 are the variance-covariance matrices for group 1 and 2 as obtained in Table 1.

To determine the vector of discriminant weight in equation [1], we compute:

(a) Pooled Sum of Squares and Cross Product Matrix, W

$$W = (N_1 - 1)S^1 + (N_2 - 1)S^2 = \begin{bmatrix} 132.414 & -17.40 \\ -17.40 & 138.852 \end{bmatrix} \quad [2]$$

Where N_1 is the number of cases in group 1, N_2 is the number of cases in group 2.

(b) Inverse of Matrix, W

$$W^{-1} = \frac{1}{|W|}C = \begin{bmatrix} 0.00802400 & 0.00151027 \\ 0.00151027 & 0.00782003 \end{bmatrix} \quad [3]$$

Where C is the transpose of the co-factors of the pooled sum-of-squares and cross- products matrix, W

(c) Mean Vector, D

The deviation of mean vectors of Group 2 from Group1 gives:

$$d = \begin{pmatrix} X_{1G1} - X_{1G2} \\ X_{1G2} - X_{1G2} \end{pmatrix}$$

$$= \begin{bmatrix} 3.20 \\ 4.13 \end{bmatrix} \quad [4]$$

So that:

$$U = W^{-1}d = \begin{bmatrix} 0.03871156 \\ 0.05976899 \end{bmatrix} \quad [5]$$

Thus, substituting the above value of the discriminant weights, UW in equation [1], we obtain the Fisher's

linear discriminant function:

Z = 0.03871156 (First Semester) + 0.05976899 (Second Semester) [6]

Multivariate test of significance

Problems arising in multivariate population are direct generalization from the univariate case. Thus, test such as test for quality of group means, equality of variance - covariance matrices, significance of discrimination between two groups and strength of relationship between predictive variables and outcome groups are needed to validate the performance of the model.

Wilks' Lambda and F Statistics for Test of Equality

Equality of group means test for the predictive variables are presented in Table 2. The Table shows that there is a significant difference between the two groups.

This mean that the two variables are able to discriminate the students' performance significantly in two groups.

Table 2: Tests of Equality of Groups Matrices

	Wilk's Lambda	**F**	**Sig.**
Second Semester	.985	2.245	.136
First Semester	.704	61.262	.000

Box's M Test for Equality of Covariance Matrices

To examine whether there is an equal variance between the group. The result of Box's M test are presented in Table 3. Results reject to the general assumption and null hypothesis that " there is no equal variance between group" and shows that there is an equal variance between the groups with Box's M value of 8.142 and P value 0.046.

Table 3: Box's M Test of Equality of Covariance Matrices

Box's M		8.142
F	Approx.	2.674
	df1	3
	df2	3836880.000
	Sig.	.046

Tests null hypothesis of equal population covariance matrices

Strenght of Relationship Between Predictive Variables and Outcome Groups:

To measure the strenght of relationship between predictive variables and outcome groups. Canonical discriminant functions were used in the analysis and results presented in Table 4. Result shows that there is a high degree of positive canonical correlation (0.716) exists between predictive variables and outcome groups.

Table 4: Canonical Discriminant Function and Canonical Correlation

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Corrreation
1	1.053	100.0	100.0	.716

Statistical Significance and Discrimination Power of the Model

Table 5 shows the results of Chi-square test with " Successive Root Removed ". Table 5 shows a Wilks' Lambda value 0.487 which is closed to 0.5. Though it ranges from 0 - 1 but, value close to 0 is considered well which indicates better discriminating power of the model. The Chi-square test shows that there is highly significant discrimination between the two groups at 5% level of significance with Chi-square value = 104.307 and p = 0.000.

Table 5: Wilks' Lambda and Chi-square test

Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1	.487	104.307	2	.000

Table 6: Classification Function Coefficients

	GROUPS	
	1	2
Second Semester	.251	.437
First Semester	.186	-.016
(Constant)	-13.708	-13.194

Fisher's linear discriminant functions

The classification discriminant functions given by SPSS are as follows:

1 (score) = 0.251SecondSemester + 0.186FirstSemester - 13.708

2 (score) = 0.437SecondSemester – 0.016FirstSemester - 13.194

Table 7: Adjusted Standardized Canonical Discriminant Function Coefficients

	Function
	1
Second Semester	-.091
First Semester	.099
(Constant)	-.252

Unstandardized coefficients

Table 7 above shows each of the semesters with their corresponding adjusted linear discriminant function coefficients. The magnitudes of these coefficients indicate the partial contribution (the unique, controlled

association) of each of the two semesters to the discriminant function. A cursory look at Table 7 clearly reveals that each of the two semester's unique contribution to the linear discriminant function is different and significant, in terms of their unique contribution to the linear discriminant function.

First semester has the highest contribution with a coefficient of 0.099 followed by second semester with a coefficient of -0.091 respectively.

Table 8: Confusion matrix for actual and predicted categories of class of degree

Class of Degree	Predicted Class of Degree		Total
	BCD	PCD	
2^1	52	0	52
2^2	50	12	62
3rd	14	12	26
Pass	0	8	8
Total	116	32	148

Table 9: Confusion matrix for actual and predicted class of degree with percentage

		GROUPS	**Predicted Group Membership**		**Total**
			1	**2**	**1**
Original	Count	1	64	10	74
		2	9	65	74
	%	1	86.5	13.5	100.0
		2	12.2	87.8	100.0

a 87.2% of original grouped cases correctly classified.

4. DISCUSSION OF RESULTS

The data were processed using computer package called statistical package for social sciences (SPSS) and confusion matrix for the analysis sample and hold-out sample are shown in Table 8 - 9 above.

In Table 8, the rows totals are the observed categories for the class of degree and the columns totals are the predicted categories for the class of degree.

It was observed that 102 out of 116 students predicted to graduate with Second Class Upper (2^1) or Second Class Lower division (2^2) did so.

This represents Hit Ratio of 87.93%. Also, of 24 students predicted to graduate with Third Class or Pass, some 33 did so. This also represents a Hit Ratio of 72.72%.

In Table 9, success in identifying students that will graduate with Better Classes of Degree (BCD) was 86.5%. Also, success in identifying students that will graduate with Poor Class of Degree (PCD) was 87.8%.

5. CONCLUSION

The importance of the derived discriminant function for the study was assessed using canonical discriminant function coefficients, Wilks' Lambda test, Box's M test and Chi-square test. In testing the classification performances of the discriminant function, overall hit- ratio, which is the same thing as percentage of the original group cases correctly classified was used.

The overall percentage of correct classifications which is 86.5% for Better Class of Degree (BCD) and 87.8% for Poor Class of Degree (PCD) is a measure of predictive ability which shows that discriminant analysis can be used to predict student's graduating class of degree from knowledge of variable(s) that have relationship with performance.

The use of discriminant analysis in this manner that is, conducting discriminant analysis for predictive purpose enables us to identify the students who might be termed at risk; these are students that will graduate with Poor Class of degree (PCD).

It also identifies MAT 102 and MAT 106 as having a booster effect on final graduating Cumulative Grade Point Average (CGPA), as well as brought to light the difficult in understanding its concept.

In conclusion, this study shows that discriminant analysis provides results that are both more interpretable and statistically sound.

Table 10: Historical Data

VALUES OF GPA, MAT 102 AND MAT 106 FOR TWO GROUPS							
GROUP 1 (N_1 = 38)				**GROUP 2 (N_2 = 37)**			
NO	**GPA**	**MAT 102**	**MAT 106**	**NO**	**GPA**	**MAT 102**	**MAT 106**
1	2.83	40	60	1	2.75	44	50
2	3.54	46	70	2	3.37	36	50
3	2.91	54	70	3	2.12	26	40
4	2.5	38	50	4	3.45	46	60
5	2.59	46	62	5	1		
6	3.25	72	73	6	2.83	38	60
7	3.7	66	74	7	3.37	36	50
8	3.39	46	62	8	1.87	50	60
9	2.66	44	60	9	2.41	52	60
10	3.54	42	60	10	2.83	58	72
11	2.25	46	60	11	1.91	48	43
12	2.79	50	60	12	1.91	38	40

13	3.41	78	70	13	2.08	32	60
14	2.95	50	60	14	1.16	28	46
15	2.73	52	60	15	1.62	28	50
16	2.58	36	60	16	3.04	54	76
17	3.08	48	66	17	1.79	44	48
18	2.7	62	70	18	2.5	56	70
19	2.7	52	60	19	1.5	30	45
20	2.95	38	45	20	1.91	46	40
21	3.14	46	70	21	2.2	36	60
22	2.45	62	70	22	2.66	50	60
23	4.08	72	81	23	1.2	22	50
24	2.83	40	51	24	1.16	56	53
25	3.91	62	57	25	1.75	42	60
26	3.33	43	50	26	1.87	50	40
27	2.83	38	60	27	3.45	44	50
28	3.62	56	57	28	1.58	50	45
29	3.54	50	60	29	1.58	34	40
30	2.2	50	50	30	1.83	30	50
31	3.5	36	60	31	1.54	26	29
32	3.91	50	70	32	2.25	36	61
33	3.58	53	50	33	2.75	36	50
34	2.69	50	60	34	3.09	56	60
35	3.54	68	73	35	1.75	51	40
36	2.7	32	60	36	1.87	42	40
37	3.83	56	70	37	2.29	34	60
38	2.83	42	60				

REFERENCES

[1] J. Kasih, S. Susanto. ‘‘Predicting Students’ Final Results through Discriminant Analysis’’. World Transactions on Engineering and Technology Education, Vol. 10, No. 2, 2012.

[2] J. T. Erimafa, A. Iduseri, and I.W. Edokpa. ‘‘Application of Discriminant Analysis to predict the class of degree for graduating students in a University system, ’’ International Journal of physical sciences vol. 4(1), pp. 016-021, January, 2009.

[3] A. E. Usoro. ‘‘Discriminant Analysis and Its Application to the Classification of Students on the Basis of their Academic Performances’’. J. Res. Phy. Sci. 2(3): 53-55, 2006.

[4] A. A. Akabona, S. Kwofie. ''Students' Performance in Ghana: A Discriminant Analysis''. International Journal of Business and Social Research, Vol. 06, Iss. 09, 2016.

[5] D. S. Chaubey, H. C. Kothari, and S. Kapoor. ''Analysis of the Factors Influencing Academic Performance of Student Using Discriminant Function''. International Journal of Engineering and Management Research, Vol.6. Iss. 3, pp 111-116.

[6] S. B. Adebayo, E. T. Jolayemi. ''On the Effect of Rare Outcome on the Agreement/Concordance Indices'' Nig. J. of Pure and Appl. Sci. 13: 718-723, 1998.

[7] S. B. Adebayo, E. T. Jolayemi. ''On the Effect of Rare Outcome on the Measure of Agreement Index'' J. Nig. Stat. Assoc. 13: 1-10, 1999.

[8] R. A. Fisher. ''The use of multiple Measurements in Taxonomic problems''. Ann. Eug. 7: 178-188, 1936.

[9] T.W. Anderson. ''An introduction to Multivariate Statistical Analysis'' New york: John Wiley, 1958.

