

# Performance of Post Graduate Students Using Multiple Regression Analysis (Case Study)

Fabiha Shaikh<sup>1</sup> M. Anwar Solangi<sup>2</sup> Saifullah Abro<sup>3</sup>  
Department of Basic Sciences & Related Studies,  
Mehran University of Engineering & Technology, Pakistan

## Abstract

The present study is focused to analyze the performance of post graduate students through various statistical parameters. In order to estimate the performance of Post Graduate Students, who are enrolled in different fields of Science, Technology and Humanities at Mehran University of Engineering & Technology (MUET) Jamshoro. The modified model of MLR is used to represent the Number of Degree Holders Students (ND) and would be utilized for those who obtained their master's degree within time duration or well time to be considered as dependent variable, the initially enrolled student denoted as (NS) number of students and with their obtained CGPA to be considered as independent variable. The result of this research would help to estimate the performance of post graduate students. For this study multiple linear regression model can also be generalized as,

$$ND = b_0 + b_1NS + b_2CGPA$$

With Standard Error of Estimate, Co-efficient of Multiple Determination ( $R^2$ ) and Multiple Correlation.

**Keywords:** Student's performance, Multiple Linear Regression (MLR), Hypothesis testing.

**DOI:** 10.7176/MTM/10-8-01

**Publication date:** December 31<sup>st</sup> 2020

## 1. INTRODUCTION

Postgraduate degree is assumed across the world in universities because it is very precious degree least people enrolled themselves for this degree due that's why degrees considers in research and the procedure for getting admission in this discipline is more different than others because research is based on unique ideas in which candidates knows better solution of the problem from previous one to collect data by reliable resources. With authentication they work in research take out the gap and try to resolve according to data by testing, analyzing and from survey for getting good result so that they can complete their degree in time that's why for testing performance of candidates in this degree M.Phil. Applied Mathematics department was checked which belong to Mehran University of Engineering and Technology (MUET) Jamshoro that how many candidates have secured their degree within time. The word regression analysis is a statistical tool for calculating the connection among the variables and it's done to find out the correlation between more than two independent variables which gives the method to be applied for calculating and prognosticate the independent variable which is known as the value of independent variable.

The connection between the expected value of the dependent and independent variable is known as the line of regression .When the dependence more than one independent variable is known as Multiple Regression. In this study, the performance of students by using Multiple Regression Analysis will be discussed. Many researchers have applied multiple regression analysis to predict student academic performance (Stephen J.H.Yang et.al , 2018) , learning performance (Huang, S. and Fang, N. , 2013, Tempelaar. D.T., Rienties. B. and Giesbers. B. , 2015, Zacharis, N.Z. , 2015, Morris, L.V., Finnegan. C. and Wu, S.-S. , 2005, Sorour. S.E., Mine, T., Goda, K. and Hirokawa. S. , 2015, Yoo, J. and Kim, J. , 2012) or to identify at-risk students by predicting the course pass or fail (Marbouti, F., Diefes-Dux, H.A. and Madhavan, K. , 2016, Macfadyen, L.P. and Dawson, S. , 2010, Agudo-Peregrina, A.F. et.al , 2014)

Regression analysis also has supposition of linearity which means that straight line connects between the dependent and independent variable. The linearity between independent and dependent variable for testing by observing scatter plot. Being a linear regression, one desire for testing the significance of the parameters joint for any variables  $X_i$  included in a multiple regression model, the null hypothesis tells the co efficient  $b_i=0$ . The research hypothesis may be one or two sided starting that  $b_i$  is either less than 0 (i.e  $b_i<0$ ), greater than 0 (i.e  $b_i>0$ ), or simply  $b_i\neq 0$ .

## 2. METHODOLOGY

The present study is descriptive in nature. In this study the students of the department of Applied Mathematics of batches from 2014-2017. The data was analyzed by applying the statistical tools like variance, mean, standard deviation, multiple linear regression, multiple correlation, standard error of estimate, adjusted R Squared and the independent t-tests were applied to test the significance between the results.

A Regression which involves two or more independent variables is called MULTIPLE REGRESSION, the linear

multiple regression model formulated as,  
 $Y = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_i + \epsilon$

Where,

Y = Dependent Variable.

$X_i$  = Independent Variables.

b=Parameter.

$\epsilon$ =Error.

The generalized multiple linear regression model for this study formulated as,

$$ND = b_0 + b_1NS + b_2CGPA$$

### 2.1. TO COMPUTE STANDARD ERROR OF ESTIMATE:

$$s_{r,12} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 3}}$$

### 2.2. TO COMPUTE THE MULTIPLE CORRELATION CO-EFFICIENT:

$$R = \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1} \cdot r_{YX_2} \cdot r_{X_1X_2}}{1 - r_{X_1X_2}^2}}$$

Where,

$$r_{YX_1} = \frac{\sum X_1Y}{\sqrt{\sum X_1^2} \sqrt{\sum Y^2}}$$

$$r_{YX_2} = \frac{\sum X_2Y}{\sqrt{\sum X_2^2} \sqrt{\sum Y^2}}$$

$$r_{X_1X_2} = \frac{\sum X_2X_1}{\sqrt{\sum X_2^2} \sqrt{\sum X_1^2}}$$

### 2.3. TO COMPUTE THE ADJUSTED MULTIPLE CO-EFFICIENT OF DETERMINATION:

Adjusted R-SQUARED is used to compare the model with different number of predictors. The formula of ADJUSTED R-SQUARED is given below,

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k}$$

Where

$$R^2 = \frac{\sum \hat{Y}^2}{\sum Y^2} = 1 - \frac{\sum e_i^2}{\sum Y^2} = \frac{b_1 \sum YX_1 + b_2 \sum YX_2}{\sum Y^2}$$

k=no: of independent variables.

n= no: of observation.

### 3. RESULTS AND DISCUSSION

(Table 1) representing the mean, standard deviation and variance of the CGPA of students for the three disciplines from Batches 2014-2017.

(Table 2) showing the equations of Regression line (Model) in batches from 2014-2017 in three disciplines.

(Table 3) showing the test hypothesis values at 95% confidence interval for the difference between means for the results of three disciplines from batches 2014-2017. The table shows that the hypothesis is accepted for all batches.

Graph and Bar Chart (1-5), showing the CGPA of students from batches 2014-2017.

Discipline	Batches	No: of Students	MEAN	SD	VAR
M.Phil Applied Mathematics	Batch 14	08	3.67375	0.298613	0.08917
	Batch 15S	08	3.72125	0.386095	0.14907
	Batch 15F	10	3.678	0.213323	0.045507
	Batch 16	20	3.5835	0.270735353	0.073297632
	Batch 17	07	3.448571	0.279787	0.078281

(Table 1)

Discipline	Batches	Trend Line
M.Phil Applied Mathematics	Batch 14	$y = 0.0132x + 3.6143$
	Batch 15S	$y = 0.0263x + 3.6029$
	Batch 15F	$y = -0.0166x + 3.7693$
	Batch 16	$y = -0.0141x + 3.7318$
	Batch 17	$y = -0.0093x + 3.4857$

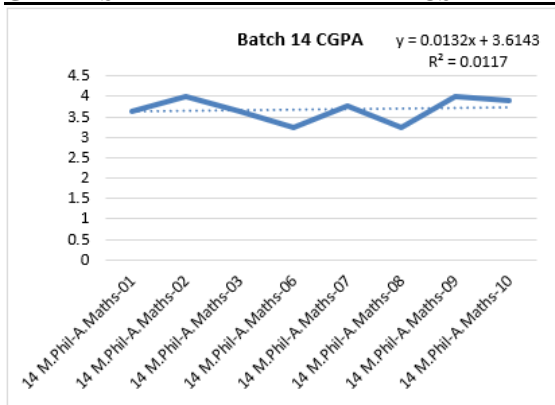
(Table 02)

M.Phil Applied Mathematics						
Batches	Batch 14	Batch 15S	Batch 15F	Batch 16	Batch 17	
MEAN	3.67375	3.72125	3.678	3.5835	3.44857	
SD	0.298613	0.386095	0.213323	0.270735	0.27979	
N	8	8	10	20	7	
Formulated Value		Tabulated Value				Remarks
t(14, 15S)	0.28	t(0.025, 14)	2.145			ACCEPT
t(14, 15F)	0.04	t(0.025, 16)	2.12			ACCEPT
t(14, 16)	0.77	t(0.025, 26)	2.056			ACCEPT
t(14, 17)	1.50	t(0.025, 13)	2.16			ACCEPT
t(15S, 15F)	0.30	t(0.025, 16)	2.12			ACCEPT
t(15S, 16)	1.08	t(0.025, 26)	2.056			ACCEPT
t(15S, 17)	1.54	t(0.025, 13)	2.16			ACCEPT
t(15F, 16)	0.96	t(0.025, 28)	2.048			ACCEPT
t(15F, 17)	1.92	t(0.025, 15)	2.131			ACCEPT
t(16, 17)	1.13	t(0.025, 25)	2.06			ACCEPT

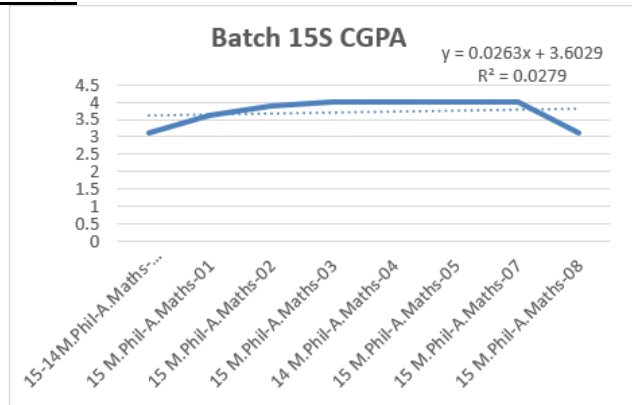
(Table 03)

ACCEPT shows that there is no difference between the performance of the students or their CGPA.

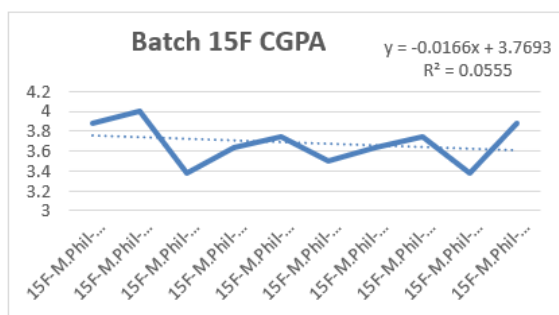
**GRAPHS APPLIED MATHEMATICS DEPARTMENT**



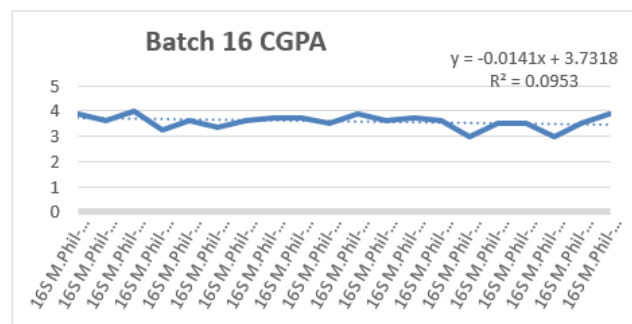
(Graph 01)



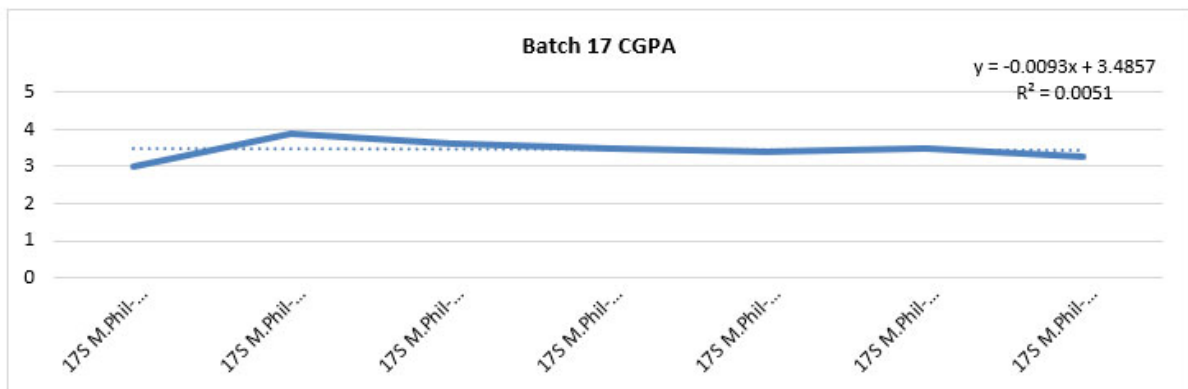
(Graph 02)



(Graph 03)

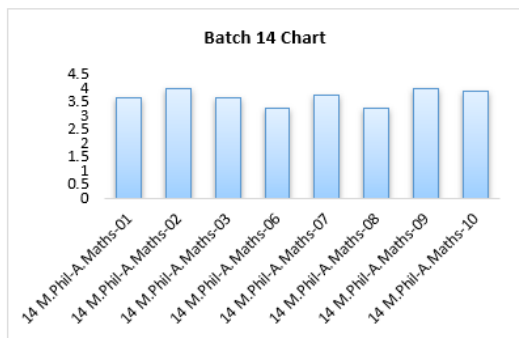


(Graph 04)

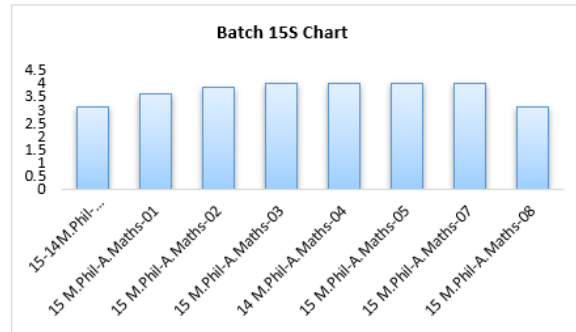


(Graph 05)

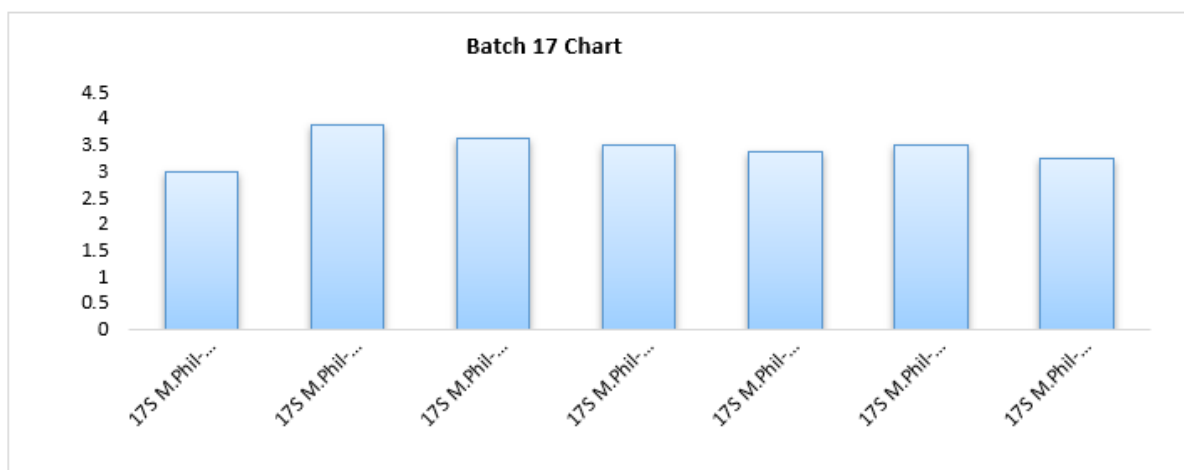
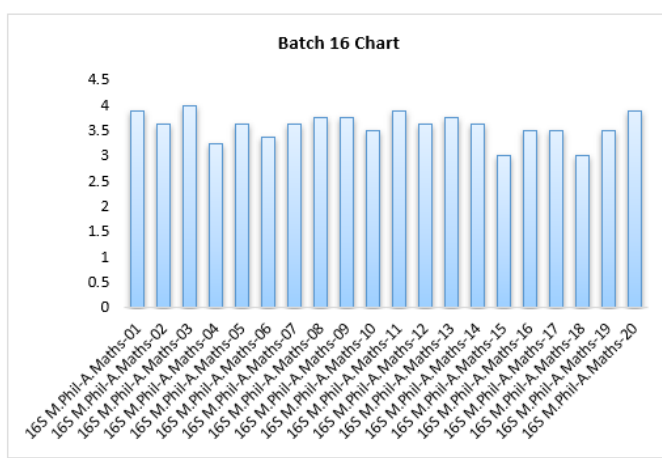
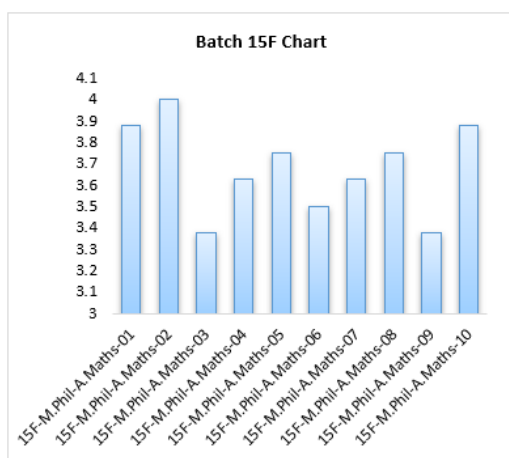
**BAR CHART APPLIED MATHEMATICS DEPARTMENT**



(Chart 01)

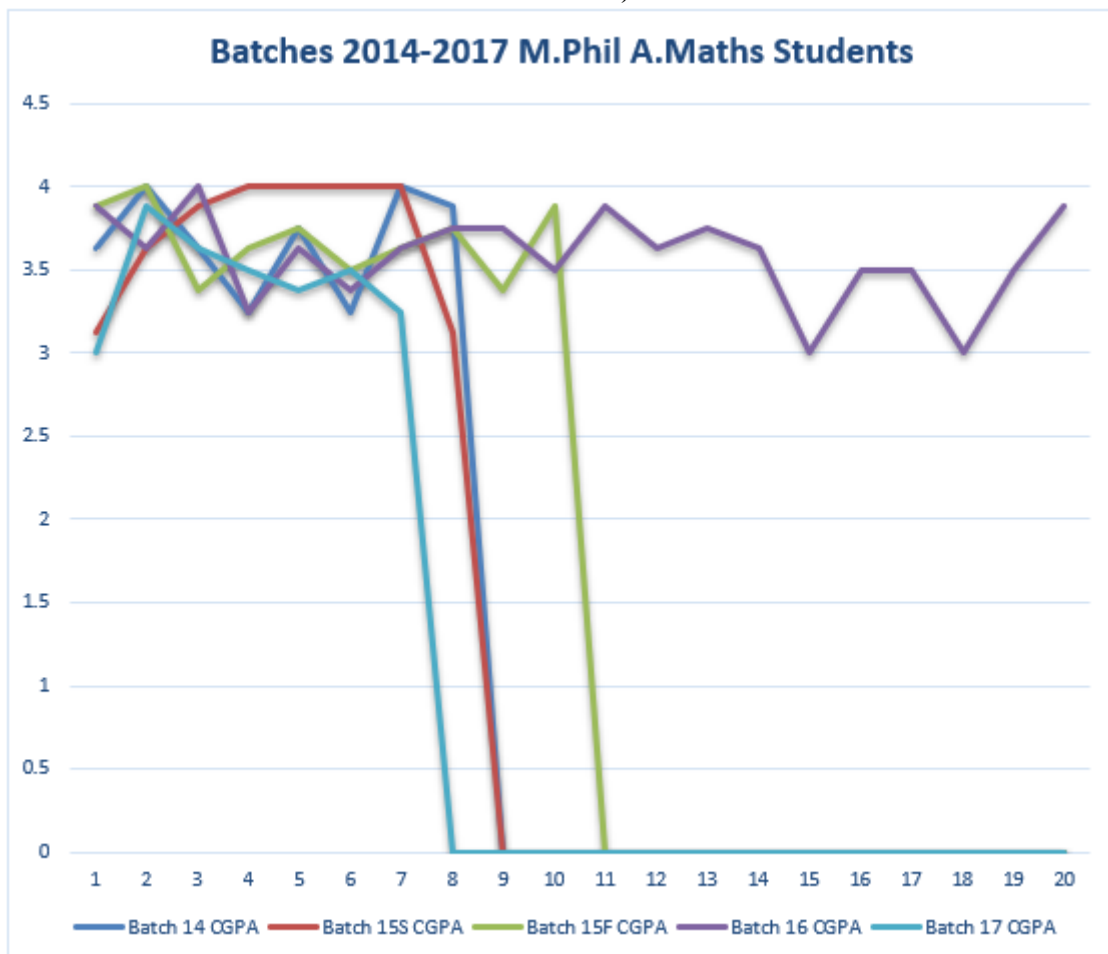


(Chart 02)



(Chart 05)

**(GRAPH OF M.Phil APPLIED MATHEMATICS STUDENTS PERFORMANCE FROM BATCHES 2014-2017)**



**(SUMMARY OUTPUT OF M.Phil APPLIED MATHEMATICS STUDENTS FROM BATCHES 2014-2017)**

SUMMARY OUTPUT

*Regression Statistics*

Multiple R	0.97
R Square	0.95
Adjusted R Square	0.90
Standard Error	1.36
Observations	5.00

$$ND = -59.12 + 0.73 NS + 16.92 CGPA$$

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2.00	69.51	34.76	18.86	0.05
Residual	2.00	3.69	1.84		
Total	4.00	73.20			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-59.12	23.00	-2.57	0.12	-158.07	39.83	-158.07	39.83
NS	0.73	0.13	5.77	0.03	0.19	1.28	0.19	1.28
CGPA	16.29	6.31	2.58	0.12	-10.86	43.44	-10.86	43.44

Based on Regression Analysis Result, the Regression equation, with the coefficient value for each

independent variable and the response variable ('Y' represents ND) is,

$$\text{ND} = -59.12 + 0.73 \text{NS} + 16.92 \text{CGPA}$$

Standard Error = 1.36,  $R^2 = 95\%$ ,  $R^2_{(\text{adj})} = 90\%$

From the ANOVA in the above table it is observed that the P-value is 0.03-0.12 which is implied that the model estimation is significant level of 0.05. The p-value estimated coefficients of NS, CGPA are 0.03 and 0.12 respectively. The coefficient of the multiple determination ( $R^2$ ) describes that 95% only of the total variation of the Y (ND) values about their mean are explainable by predictor variable, indicating goodness of fit of the multiple regression model.

## CONCLUSION

The above research is based on performance of the post graduate student's for various batches from 2014-2017 at MUET, Jamshoro. From the investigated study and testing hypothesis it's concluded that there is a strong significance relationship between the results of the parameters (No: of students enrolled, Average CGPA) for the Master's Students in the Department of BSRS (M.phil Applied Mathematics) MUET Jamshoro.

## REFERENCES

- [1] Yang, S.J., Lu, O.H., Huang, A.Y., Huang, J.C., Ogata, H. and Lin, A.J., 2018. Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis. *Journal of Information Processing*, 26, pp.170-176.
- [2] ABRO, S., SOLANGI, M. and SHAIKH, A., 2016. An Investigation on the Performance of Students in Mathematics: A Case Study. *Sindh University Research Journal-SURJ (Science Series)*, 48(2).
- [3] Marbouti, F., Diefes-Dux, H.A. and Madhavan, K., 2016. Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, pp.1-15.
- [4] Zacharis, N.Z., 2015. A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, 27, pp.44-53.
- [5] Tempelaar, D.T., Rienties, B. and Giesbers, B., 2015. In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior*, 47, pp.157-167.
- [6] Sorour, S.E., Mine, T., Goda, K. and Hirokawa, S., 2015. A predictive model to evaluate student performance. *Journal of Information Processing*, 23(2), pp.192-201.
- [7] Agudo-Peregrina, Á.F., Iglesias-Pradas, S., Conde-González, M.Á. and Hernández-García, Á., 2014. Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in human behavior*, 31, pp.542-550.
- [8] Huang, S. and Fang, N., 2013. Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education*, 61, pp.133-145.
- [9] Yoo, J. and Kim, J., 2012, June. Predicting learner's project performance with dialogue features in online q&a discussions. In *International Conference on Intelligent Tutoring Systems*(pp. 570-575). Springer, Berlin, Heidelberg.
- [10] Macfadyen, L.P. and Dawson, S., 2010. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & education*, 54(2), pp.588-599.
- [11] Morris, L.V., Finnegan, C. and Wu, S.S., 2005. Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, 8(3), pp.221-231.