

# Data analytics on performance of computing students

Oluwafemi Samson Balogun<sup>†</sup>  
School of Computing  
University of Eastern Finland  
Kuopio, Finland  
samson.balogun@uef.fi

Solomon Sunday Oyelere  
School of Computing  
University of Eastern Finland  
Joensuu, Finland  
solomon.oyelere@uef.fi

Donald Douglas Atsa'am  
Department of Maths, Sta and CS  
University of Agriculture  
Makurdi, Nigeria  
donatsaam@alumni.edu.tr

## ABSTRACT

This study was conducted to determine the relationship between computing students' initial and final academic performance to support decision making in higher education institutions. The data used in this research contains the initial GPA and CGPA of 710 computing students for a period of 5 years. Test of normality for the final graduating results and linear regression models were fitted to the data to predict the overall performance based on the initial result. This study revealed that there are strong linear relationships between the GPAs and CGPAs of computing students over the period of this study. The contribution of this work is that the result enables the student and the teacher to understand the trend of students' academic performance for the purpose of decision-making. Keeping track of the performance of the students help provide support whenever is needed. This results aimed to decrease student dropout by means of facilitating student to predict their probability of success in computing courses after enrollment. In addition, teachers will be able to boost student performance in their courses as a result of enhanced determination of student's abilities to learn the course and fine-tuning teaching approaches and techniques.

## CCS CONCEPTS

Applied computing → Education

## KEYWORDS

Computing education, Data analytics, Academic performance, Linear regression analysis, Correlation analysis

## ACM Reference format:

Oluwafemi Samson Balogun, Solomon Sunday Oyelere, Donald Douglas Atsa'am. Data analytics on performance of computing students. In *Koli Calling '19*, November 21–24, 2019, Koli, Finland. © 2019 Copyright is held by the owner/author(s). ACM ISBN 978-1-4503-7715-7/19/11. <https://doi.org/10.1145/3364510.3366152>

## 1 Background

Numerous factors are regularly considered as key in the decision-making of university education. A portion of the

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*Koli Calling '19*, November 21–24, 2019, Koli, Finland

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-7715-7/19/11.

<https://doi.org/10.1145/3364510.3366152>

variables that determine student overall performance, success, failure or dropout are favorable condition for effective learning, student-teacher interaction, teacher satisfaction, student engagement and satisfaction with the educational setting. Educational data mining methods are employed to understand patterns and trends of students' learning. Large-scale data emerging from learning settings are collected and used to on the one hand, identify the weaknesses within the educational settings, and on the other hand, improving the performance of teachers and students [1]. In fact, one particular goal of educational data mining is to predict the performance of students using different prediction algorithms. Prediction algorithms analyze the entire dataset to generate accurate models for each potential outcome. Prediction algorithms usually provides signs about trends of students' academic endeavor such as student performance, pedagogical activities, learning style and behaviors, teaching quality, effectiveness and efficiency, potential for success or dropout and so on [2], [3], [4].

This paper depicts the apparent association between the final graduating grade of computing students and their initial results in the university. This study examine the impact of the initial Grade Point Average (GPA) of computing students in their first year on the final graduation CGPA. This study answers the research question, to what extent can the final year CGPA of a computing student be predicted using the GPA of the first year? Correlation analysis was used to determine if there is a linear relationship between the initial GPA and the final CGPA of computing students. While linear regression analysis was used to test the normality of the final graduating results and linear regression models were fitted to the data to predict the overall performance based on the initial result [5].

## 2. Data source, results and interpretation

The data used in this research contains the initial GPA of computing students for a period of 5 years and their graduating CGPA from the departments of Computer Science (CS), Information and Communication Technology (ICT), and Electrical Engineering (EE). The data used for this work is a secondary data collected from Modibbo Adama University of Technology, Yola, Adamawa State, Nigeria.

### 2.1 Results of correlation analysis

The correlation coefficients between the GPAs of the students from the first year till the point of graduation for all the three departments are presented in Table 2.

Table 1: Correlation Analysis

ICT Final CGPA		CS Final CGPA		EE Final CGPA	
First GPA	P-value	First GPA	P-value	First GPA	P-value
Pearson	0.000	Pearson	0.000	Pearson	0.000
0.773		0.797		0.795	

The result in Table 1 show that there are positive linear relationships among the GPAs across all levels. The linear relationships are also strong except for few occasions where there exist fair relationships.

### 2.2 Results of regression Analysis

The final CGPA is made to be the dependent variable while the GPA is made to be independent variable. The regression line for ICT, CS and EE scores are:

ICT: final CGPA = 0.648+0.7920 first GPA

CS: final CGPA = 0.131+0.8697 first GPA

EE: final CGPA = 0.402+0.8048 first GPA

Interpretation: For every unit increment in first year GPA of ICT, CS and EE students, there will be an increase of 0.7920, 0.8697 and 0.8048 respectively in the final CGPA.

Tables 2, 3 and 4 represent the model summary of ICT, CS and EE is used to measure the closeness of the data to the fitted regression line.

Table 2: Model summary for ICT

S	R <sup>2</sup>	R <sup>2</sup> adjusted	R <sup>2</sup> (Predicted)
0.401796	65.21%	65.11%	64.84%

Table 3: Model summary for CS

S	R <sup>2</sup>	R <sup>2</sup> adjusted	R <sup>2</sup> (Predicted)
0.395086	63.27%	63.16%	63.89%

Table 4: Model summary for EE

S	R <sup>2</sup>	R <sup>2</sup> adjusted	R <sup>2</sup> (Predicted)
0.368264	61.53%	61.41%	60.94%

Interpretation: This implies that about 65.21%, 63.27% and 61.53% of the variability in the final CGPA is being explained by the other GPAs for ICT, CS and EE students respectively.

Table 5 represent the test of equality of means for ICT, CS and EE, which is used to measure the relationship between the initial GPA and the final CGPA of computing students.

Table 5: Test of equality of means for the three department

ICT	Term	Coefficient	Std. Error Coefficient	t	p – value
	Constant	0.648	0.113	5.75	0.000
First GPA	0.402	0.402	25.50	0.000	
CS	Term	Coefficient	Std. Error Coefficient	t	p – value
	Constant	0.131	0.135	0.93	0.353
	First GPA	0.8697	0.0356	23.45	0.000
EE	Term	Coefficient	Std. Error Coefficient	t	p – value
	Constant	0.402	0.135	2.98	0.003
	First GPA	0.8048	0.0342	23.56	0.000

Interpretation: All the variables contribute significantly to the rejection of the null hypothesis for the three departments.

Table 6 represent the analysis of variance for the fitted regression model, which is used to show if the regression model fits for ICT, CS and EE students.

Table 6: Analysis of variance for the fitted regression model for the three department

Department	Source	Degree of freedom	Sum of Squares	Mean Square	F	P-value
ICT	Regression	1	105	105	652.17	0.000
	Error	347	56.02	0.161		
	Total	348	161.02			
	Source	Degree of freedom	Sum of Squares	Mean Square	F	P-value
CS	Regression	1	93.29	93.29	597.63	0.000
	Error	347	54.16	0.1561		
	Total	348	147.45			
	Source	Degree of freedom	Sum of Squares	Mean Square	F	P-value
EE	Regression	1	75.25	75.25	554.94	0.000
	Error	347	47.06	0.1356		
	Total	348	122.31			

Interpretation: Since the p-value in Table 6 is 0.000 which is less than 0.05, then, we conclude that the regression model significantly fits the data for the three departments.

### 3. Implication of the study and conclusion

Inferential statistics has been implemented on the academic performance of computing students in a Nigerian university. This study revealed that there are strong linear relationships between the GPAs and CGPAs of computing students over the period of this study. All the results of the regression models significantly fit the data used and each of the regression parameters contributes to the significance of the model. Teacher could use the result of this study to understand the trends in the academic performance of the computing student to provide the needed support at each stage of the students' academic progression. Besides, the results of this study will serve as motivation for the teachers and look out for signs of trouble in each student. It will also help in keeping a track record of each students by predicting their academic performance so as to mentor, advice and nurture the students. The result of this study will encourage competition among the students and motivates students to work harder to improve their performance in order to succeed.

### REFERENCES

- [1] T. Daradoumis, J.M. Marqués Puig, M. Arguedas, L. Calvet Li-nan, 2019. Analyzing students' perceptions to improve the design of an automated assessment tool in online distributed programming. *Comput. Educ.* 128, 159-170.
- [2] S. Roy, A. Garg, 2018. Predicting academic performance of student using classification techniques. In: Paper Presented at the 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics. UPCON 2017.

- [3] D. Kim, M. Yoon, I.H., Jo, Branch, R.M., 2018. Learning analytics to support self-regulated learning in asynchronous online courses: a case study at a women's university in South Korea. *Comput. Educ.* 127, 233-251.
- [4] F. Yang, F.W.B. Li, 2018. Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Comput. Educ.* 123, 97-108.
- [5] P. M. Arsad and N. Buniyamin 2014. Neural network and linear regression methods for prediction of students' academic achievement. In the IEEE Global Engineering Education Conference, IEEE, Istanbul, Turkey, 916-921.