



Correlation Analysis on Educational Data for Determining Factors Contributing to Female Students' Academic Performance

Tsehay Admassu Assegie

Department of Computer Science, Faculty of Computing Technology, Aksum University, Axum, Ethiopia

ARTICLE INFO	ABSTRACT
Published Online: 24 March 2021	Educational data mining is a growing research area defined as the application of data mining techniques on raw data from academic institutions to solve educational problems by discovering the information hidden in the educational data. Educational data mining incorporates developing predictive models, visualizing educational data and employing machine learning algorithms and statistical methods to explore hidden relationship using the statistical methods. In this research, we have employed relationship mining to explore the correlation between the features in student dataset. In correlational analysis to investigate the relationship among student dataset features, we have employed Pearson's Correlation Coefficient (R). The dataset consists of senior undergraduate (second year, third year and fourth year) Computer Science, Information Technology and Information System female students' data collected through questionnaire at Aksum University (Axum, Ethiopia). The goal of the correlational analysis is to investigate the feature (s) that have strong relationship to female students' academic performance or cumulative grade point average (CGPA). The investigation on the relationship among the female students' dataset features is important to determine the feature(s) that have strong correlation to the academic performance or the cumulative grade point average (CGPA) of the female student. The findings of this research reveals that class attendance has strong relationship with academic performance of female student.
Corresponding Author: Tsehay Admassu Assegie	
KEYWORDS: educational data mining, correlation analysis, student academic performance, educational data	

I. INTRODUCTION

In recent days, most of female students' attending higher institutions are obtaining Bachelor's Degree and the number of female student attending higher institutions' have increased [1-4]. Despite of the increase in the number female students' admitted to higher institutions of Ethiopia, the academic performance of female students' remained lower as compared to male student. Students' performance can be defined in terms of the success rate or graduation [1] and the performance of student is measured by the learning assessment or CGPA. Most of higher institutions' use final grade of the students' to evaluate students' academic performance. The need for evaluation of academic performance of students' is to ensure effectiveness of the learning process. Delivering a high quality education and improving the academic performance of female students' at higher institutions' is the objective of most of the higher institutions' of Ethiopia. High quality education can be realized only if the students' data is used to discover the hidden knowledge and the determinant factors affecting

female students' academic performance. To identify the determinant factors contributing to female students' academic performance at higher institutions', machine learning is applied to the students' information stored in the higher institutions' data repository and a fact can be found through knowledge discovery process using machine-learning techniques. The fact found from this larger data repository of female students' is significant in the decision making process and drawing conclusions on the determinant factors contributing to the female students' academic performance at higher institutions. Based on the fact discovered from these larger data repository, policies and administrative actions can be taken to adjust and improve female students' academic performance and the productivity of female students' can be improved as well. The objective of this research is to investigate the determinant factors contributing to female undergraduate students' academic performance studying Computer Science, Information Technology and Information System at Aksum University. Moreover, this research

explores the answers to the following research questions: 1) What is the relationship between class attendance and female students’ academic performance (CGPA)? 2) What are the feature(s) strongly related to female students’ academic performance (CGPA)? 3) What is the relationship between female students’ studying habit and their academic performance (CGPA)? 4) What is the relationship between female students’ academic performance (CGPA) and participation in-group works?

II. RELATED WORK

In this section, the pervious works on the determinant factors of female students’ academic performance in different higher institutions of Ethiopia is discussed. A study on female students’ academic performance [2], shows that the number of female students’ that are graduated from higher institutions of Ethiopia is less compared with the number of male students for the same number of admissions of female and male students. The author also found the determinant factors for this and one of the factors for poor academic performance of the female students is socio economic factor such as economic status of the parents. A cross-sectional progress study on factors affecting students’ academic performance [4], showed that the academic performance of students’ or Cumulative Grade Point Average (CGPA) is strongly related to the student’s time management skill. The study showed that a student with better time management skill has better Cumulative Grade Point Average (CGPA) compared to student with less time management skill. In another study [6], the authors highlighted that an important variable significantly affecting female students academic performance is the study habit. The study habit indicates the number of hour the female students are spending on academic activity after classes. The relationship between study habit and academic performance is strongly correlated. The higher the number of hours is spent on study; a student achieves the better academic performance. Literature survey [1-7], shows that student’ different factors such as students learning preferences, attendance and a student parent educational level significantly affects student academic performance is by different factors. A study on factors affecting academic performance of computing technology students shows that, daily-sleep hour, student study habit and English level plays an important role in determining the academic performance of students [8]. Machine learning algorithms are applied to student academic data to identify the performance factor from the dataset by different researchers although the researchers’ focused on both sex (Male and Female). English proficiency has strong relationship with students’ academic performance [9]. The literature shows that other factors like communication barriers and student teacher relationship has an impact on the academic performance of students’. Most of the female students’ does not have a habit of asking for help and questions in the class room [10] and the lack of

questioning have negative effect on the students’ performance..

III. RESEARCH METHODOLOGY

The methodology for data collection, interpretation and analysis of the collected dataset is described in this section. To find out the factors contributing to the female students’ academic performance or cumulative grade point average (CGPA), a relationship mining technique is employed. Correlational analysis using Pearson’s correlation coefficient is employed to find out strongly related features(s) to academic performance or cumulative grade point average (CGPA) of female students in the data repository for different observations. This correlation helps us to determine the features strongly related with female students academic performance or cumulative grade point average (CGPA) which is a target feature, since performance is determined based on the cumulative grade point average (CGPA) of female students. The data repository used in this research is collected using questionnaire. The number of data used in the research is 58 (undergraduate female students’ studying Computer Science, Information Technology and Information Systems at Aksum University). The type of data collected includes both qualitative and quantitative data. The qualitative data includes student academic status or performance such as Excellent, Very Good, Good and Satisfactory, participation in group works and class attendance, study habit and the quantitative data like CGPA and class year. A Python programming language is used to develop a program that computes the Pearson’s correlation coefficient of the features of female students’ and their academic performance or CGPA.

A. Correlation Model

The Pearson correlation coefficient (named after Karl Pearson) can be used to summarize the strength of linear relationship between two data observations or data samples. The Pearson correlation coefficient is calculated using the formula shown in equation (1).

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2 \sqrt{n(\sum y)^2 - (\sum y)^2}} \quad (1)$$

Where: x and y are two continuous variables and r denotes the correlation coefficient between variables x and y., n denotes the number of observations. For example to find the correlation coefficient, r between class year and performance or CGPA let us consider the class year and CGPA values given in table 1.

Table 1: Academic performance vs class year

<i>CGPA(Performance)</i>	<i>Class year</i>	<i>CGPA(Performane)</i>
3.3	4	3.3
2.9	2	2.9
3	3	3

“Correlation Analysis on Educational Data for Determining Factors Contributing to Female Students’ Academic Performance”

In this case, the class year is independent variable or x and the CGPA is considered as y or dependent variable. Solving for the enumerator we found the value of 9.9 and substituting the x and y values to the denominator we found value of 3.3. The correlation r for -0.3, which shows negative correlation between class year and student performance.

B. Dataset Description

The female students’ academic performance data is collected through a questionnaire survey conducted during the first semester at the Department of Computing Technology, Aksum University, academic year 2019-2020, among the second, third and fourth year undergraduate female students. After eliminating incomplete data, the sample comprised of 65 female students’ data. The features of the student’s data repository used in this research are summarized in table 2.

Table 2: Academic performance ranking classifications

CGPA	Performance Category
≥ 3.75	Excellent
≥ 2.75 and ≤ 3.5	Very Good
≥ 2.5 and ≤ 2.75	Good
≥ 2.0 and ≤ 2.5	Satisfactory

Table 3 features of the dataset

Table 3: Dataset feature

Feature Name	Description
Student ID	A unique variable used to identify each student
CGPA	Commutative Grade Point Average
Department/Program	Program under which the student is studying, CS-Computer Science, IS-Information System, IT-Information Technology
Class Year	2-2 nd year 3-3 rd year 4-4 th year
Class Attendance	1-Attends regularly 0-Attends sometimes 2-Misses Sometimes
Students’ studying habit	1-Study’s regularly 0-Study’s sometimes 2-Does not study at all (Never study)
Academic Performance	1-Excellent 0-Satisfactory 2-Very Good 3-Good

IV. RESULT AND DISCUSSION

The Pearson’s correlation coefficient in figure 1 demonstrates the relationship between female students’ dataset features. As

demonstrated in figure 1, the academic performances or cumulative grade point average (CGPA) is strongly correlated to class attendance, group work. This implies that the higher the number of class a student attended the higher CGPA. Therefore, by implementing some polices on student class attendance, better performance or cumulative grade point average (CGPA) can be achieved.

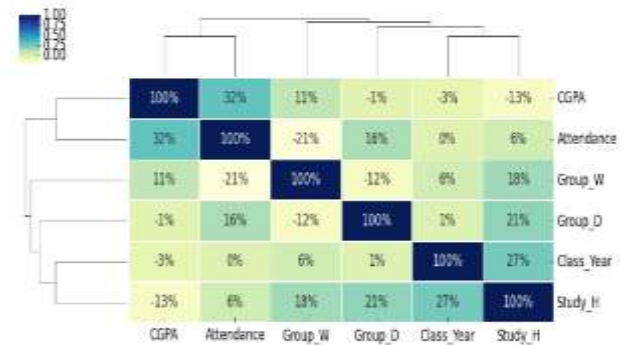


Fig. 1. Correlation of the student dataset features

V. CONCLUSION

In this research, we have applied Pearson correlation to educational data and we have investigated features with positive or negative correlation to female student academic performance or cumulative grade point average (CGPA). The outcome of the research reveals that class attendance has strong relationship with performance. In addition to the class attendance, group work has role in achieving better academic performance. This study is significant for academic institutions to improve the female students’ academic performance or cumulative grade point average (CGPA) by implementing polices and administrative actions that encourage students’ group working behaviour and class attendance monitoring mechanisms.

REFERENCES

1. Amirah Mohamed Shahiria, Wahidah Husaina , Nur’aini Abdul Rashid, A Review on Predicting Student’s Performance using Data Mining Techniques, The Third Information Systems International Conference, Procedia Computer Science, 2015.
2. Wudu Melese, Getahun Fenta, Trend and cause of female students’ dropout from teacher education institutions of Ethiopia, The case study of Jimma university, 2009.
3. Ogbogu Christiana Osaikhiuwu, Institutional Factors Affecting the Academic Performance of Public Administration Students in a Nigerian University, Public Administration Research; Vol. 3, No. 2, 2014.
4. Gebrehiwot, D. B., Teklay, A. & Kebede, T, Factors affecting academic performance of female students

- at Mekelle University, Ethiopia, Fifth RUFORUM Biennial Regional Conference, 17-21 October 2016.
5. H. Esin Erdem, A cross-sectional survey in progress on factors affecting s academic performance at a Turkish university, Akdeniz Language Studies Conference 2012.
 6. Fatemeh Mashayekhi, Shideh Rafati, Mahdie Mashayekhi, Foozieh Rafati, Mohamad Reza Mohamadisardoo, Emad Yahaghi, International Journal of Current Research and Academic Review, The relationship between the study habits and the academic achievement of students in Islamic Azad University of Jiroft Branch, Volume 2 Number 6, June-2014.
 7. S. Valli Jayanthi, Santhi Balakrishnan, Angela Lim Siok Ching, Noor Aaqilah Abdul Latiff, A.M.A. Nasirudeen, Factors Contributing to Academic Performance of Students in a Tertiary Institution in Singapore, American Journal of Educational Research, 2014.
 8. Yahya M. AlMurtadha, Abdelrahman Osman Elfaki, Osman A. Abdalla, Khaled M. Alhawit, Factors Influencing Academic Achievement of Undergraduate Computing Students, International Journal of Computer Applications, July 2016.
 9. Pattaramon Vuttipittayamongkol, Predicting Factors of Academic Performance, Second Asian Conference on Defense Technology, 2016.
 10. Farhan Alshammari, Reynita Saguban, Eddieson Pasay-an, Ahmed Altheban, Layla AlShammari, Factors affecting the academic performance of student nurses: A cross-sectional study, Journal of Nursing Education and Practice 2018.