

# Correlation Between Text Book Usage and Academic Performance of Student in Higher Education Using ‘R’

Shanti Verma and Jignesh Doshi

**Abstract** In this paper authors try to focus one of the key factors to improve student performance i.e. text book usage. In this study authors explore association between students final Semester Grades and text books usage. Scope of study is limited to only MCA program of Gujarat Technological University students and the text book of subjects that are not issued for longer period by institute library. Authors also take assumption that books are not purchased by students as they are too costly. The aim of this paper is to use correlation and regression methods to analyze the dataset containing 60 students of MCA semester III students and try to find out that does the text book usage by student affects the academic performance. The primary result of experimental analysis shows that if the text book usage increases, performance of students also increases. The results are important for the teachers to motivate students towards library as well as institutions to focus on library utilization.

**Keywords** Correlation · Linear regression · Academic performance · Higher education

## 1 Introduction

As per All India Council for Technical Education (AICTE), number of education institutes and intake have increased exponentially in past decade [1]. Poor results and quality are key challenges for universities nowadays. Each student can be classified using into various groups using learning styles [2, 3].

---

S. Verma (✉)  
LJ Institute of Computer Applications, Ahmedabad, India  
e-mail: verma.shanti@gmail.com

J. Doshi  
LJ Institute of Management Studies, Ahmedabad, India  
e-mail: doshijig@gmail.com

Data mining or Knowledge Discovery in Database (KDD) is a collection of advanced analytical techniques to discover knowledge from large amount of data [3, 4, 5]. Data mining techniques are applied successfully in various areas like manufacturing, production, customer relationship management, tele-communication, education for improvements [6, 7].

Many researchers have performed various types of analysis in education area [8]. In this paper, authors have applied correlation and regression methods to find relation between reading textbooks and academic performance.

This paper is organized as below: Introduction to topic is provided in Sect. 1, Objectives are defined in Sect. 2, Literature review is discussed in Sects. 3, 4 and 5 discusses experiment and results performed using correlation and regression using R programming. Conclusion is provided in Sect. 6 and Future work in Sect. 7.

## 2 Objectives of Study

In current scenario there is a trend of using different online available data that are not reliable. It is very important to find that how text book usage affects academic Performance of student [9]. Thus, this study aims to investigate the association between text book usage and academic performance in higher education. The objectives are:

1. To find the degree of association between text book usage and academic performance of student.
2. To determine the linear regression equation between text book usage and academic performance of student.

## 3 Literature Review

### 3.1 *Related Work: Education Mining*

Data Mining can be used in academic to enhance and evaluate the learning process students [10, 11, 12]. Various data mining techniques are used by various researchers in various areas for analysis like:

- (a) Bhardwaj and Pal have performed research using Bays Classification to found whether new comer students will be performer or not [13].
- (b) Using data analysis, Galit tried to predict the results and to warn students at risk before their final exams [14].
- (c) Decision tree model was used by Al-Radiate, et al to predict the final grade of students. Three different classification methods namely ID3, C4.5, and the Naive ayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models [15].

- (d) Pandey and Pal applied association rule and try to find the interestingness of student in opting class teaching language [16].
- (e) As per Ayesha, Mustafa, Sitar and Khan, we may use k-means clustering algorithm to predict student's learning activities [5].
- (f) Jignesh explored Association rule, chi-square and lift techniques to predict failures [17].
- (g) Hijazi and Naqvi performed research using simple linear regression analysis to discover that the factors like mother's education and student's family income are highly correlated with the student academic performance or not [18].

## ***3.2 Related Work: Correlation and Regression***

Correlation technique is used by researchers to find degree association between two variables. Regression coefficients are useful to predict dependent variable value for some independent variable. Both techniques are used by many researchers to predict academic performance of student [18].

- (a) Alan Chea et al. Used Data mining technique: correlation to associate various variables that influence academic performance. These association results were used to predict student's academic success [8].
- (b) Norman Poh, Ian Smythe used Regression techniques to predict student performance. They predict student performance on basis of student performance based on self-efficacy, socio-economic background, learning difficulties, and related academic test results [7].

## **4 Experiment**

### ***4.1 Data Extraction and Transformation***

In this paper authors have selected data sets from one of the Largest University "Gujarat Technological University" of Gujarat and Library log records of "L. J. institute of computer Applications", one of the. largest intake college of Master of computer Applications Program in Gujarat. The Data sample is taken from post graduate course "Master of Computer Applications" semester III of 60 students. Steps for Data Extraction, Collection and Transformation

1. Data Selection
  - a. Data of 2 institutes out of 40
  - b. Total 60 data collected out of 240

**Table 1** Academic grade coding

Grade	AA	AB	BB	BC	CC	FF
New value	5	4	3	2	1	0

- c. Semester 3 students data taken only for two subjects Statistical Methods (SM), SOOADM
2. Data Collection
  - a. University result data was collected from GTU web portal ([www.gtu.ac.in](http://www.gtu.ac.in))
  - b. Library log collected from LJMCA Library Software
3. Data Transformation
  - a. Convert student result grade as quantitative data (Table 1).
  - b. From Library log records find the frequency of book issued of each student in subjects SOOADM and SM.
  - c. Initially data was entered in Notepad as tab delimited file.
4. Tool Used
  - a. Data Mining Tool- 'R'

## 4.2 Technique: Correlation

Ho: Grades and Frequency of book used are positively correlated

Ha: Grades and Frequency of book used are not positively correlated

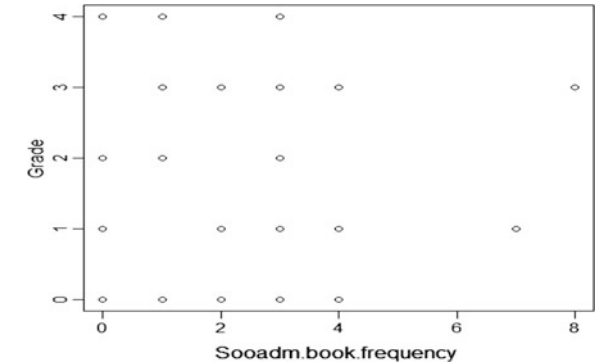
Steps to find the correlation value in 'R'

1. Make a Two column tab delimited file. First Column name is Frequency of Book used as independent variable and second is Grade as dependent variable.
2. Draw scatter plot.
3. Find correlation coefficient (Fig. 1 and Table 2).

In the above scatter plot author find out that in most of the cases as frequency of book used is increased, grades also increased. In another words there is a positive correlation between SOOADM text book frequency and grades (Fig. 2 and Table 3).

In the above scatter plot author find out that in most of the cases as frequency of book used is increased, grades also increased. In another words there is a positive correlation between SM book frequency and grades.

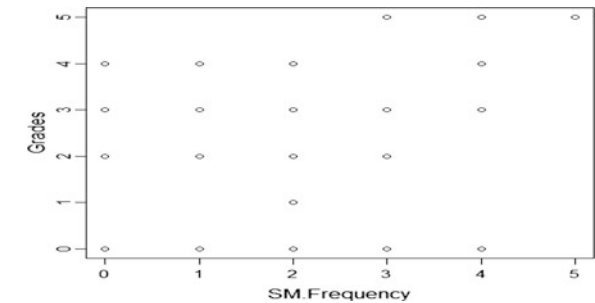
**Fig. 1** Scatter plot between SOOADM text book usage and grade



**Table 2** Correlation coefficient for subject SOOADM

	Sooadm.book.frequency	Grade
Sooadm.book.frequency	1.00	0.0640
Grade	0.0640	1.00

**Fig. 2** Scatter plot between SM text book usage and grade



**Table 3** Correlation coefficient for subject SM

	SM.Frequency	Grade
SM.Frequency	1.00	0.5270
Grade	0.5270	1.00

4.3 Technique: Linear Regression

Regression coefficients for Subject SOOADM

(Intercept)	Grade
1.55323	0.07381

Regression line for Subject SOOADM is  
**Grade(Y)= 1.5532 + 0.0738 \* Sooadm.book.frequency (x)**  
Regression coefficients for Subject SM

(Intercept)	Grades
0.6372	0.4588

Regression line for Subject SM is  
**Grades (Y) = 0.6372 + 0.4588 \* SM.Frequency (x)**  
Here authors find that if the value of independent variable is increased, dependent variable value also increased but not in the same rate as subject SOOADM.

5 Results and Discussion

We summarize experimental results as in Table 4.  
Authors discover:

+ve Correlation

Text book usage and academic performance are positively correlated means if text book usage increase/decreases, academic performance of student also increase/ decreases.

+ve Regression Coefficients

It states that text book usage and academic performance are positively dependent to each other.

Predicted Value

Here authors take text book usage value 7 and check academic performance. For SOODAM subject value is 2.069863 means that if students use text book effectively at least 7 times in semester then he/she will get at least BC grade in final semester Exam. For subject SM predicted value is 3.8488 means if students use SM text book at lease 7 times in semester then he/she will get at least BB grade in final semester exam.

Table 4 Experimental result summary

	Correlation coefficient	Regression coefficient	Predicted value when text book usage is 7
SOOADM	0.06403358	Intercept 1.55323	2.069863
		Grade 0.07381	
SM	0.5279365	Intercept 0.6372	3.8488
		Grades 0.4588	

## 6 Conclusion

The objective of identifying student academic performance on data mining the text book usage is achieved. We believe that these results could be important to predict academic performance of student on the basis of text book usage.

Here authors validate the null hypothesis that text book usage and academic performance are positively correlated. It is expected that results of the research are useful for teachers to motivate students towards text book usage.

## 7 Future Work

We can take more variables that are dependent with academic performance and use Multivariate analysis to improve results.

## References

1. All India Council for technical Education (AICTE) approval process handbook (2013–14), page 4–7.
2. Richard M. Felder, Rebecca Brent: “Understanding Student Differences”: <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers>, Journal of Engineering Education, 94 (1), 57–72 (2005).
3. Richard M. Felder, G. N. Felder, E.J. Dietz: “The Effects Of Personality Type On Engineering Student Performance And Attitudes”, <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers>: Journal of Engineering Education, 91(1), 3–17 (2002).
4. Galit.et.al, “Examining online learning processes based on log files analysis: a case study”. Research, Reflection and Innovations in Integrating ICT in Education 2007.
5. Han, J. &Kamber, M. (2006), Data Mining: Concepts and Techniques, 2nd edition, Morgan Kauf-mann.
6. G K Gupta, Introduction to data mining with case studies, Second Edition, PHI Learning.
7. Norman Poh, Ian Smythe “To what Extent we predict Student’s performance? A Case Study in colleges in South Africa”, 2014 IEEE.
8. S. T. Hijazi, and R. S. M. M. Naqvi, “Factors affecting student’s performance: A Case of Private Colleges”, Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.
9. Alan Cheah Kah Hoe, Mohd Sharifuddin Ahmad, Tan Chin Hooi, Mohana Shanmugam, Saraswathy Shamini Gunasekaran, Zaihisma Che Cob, Ammuthavali Ramasamy, “Analyzing Students Records to Identify Patterns of Students’ Performance”, Research and Innovation in Information Systems (ICRIIS), 2013 International Conference, IEEE PP. 554–557.
10. Alaa el-Halees, “Mining students data to analyze e-Learning behavior: A Case Study”, 2009.
11. Brijesh Kumar Baradwaj, Surabh Pal, Mining Educational Data to Analyze Students Performance, International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011, Page 633–69.
12. Verma, Shanti. “Deciding Admission Criteria For Master of Computer Applications Program in India using Chi-Square Test.” Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. ACM, 2016.

13. U. K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp. 686–690, ISSN:0975-9646, 2011.
14. U. K. Pandey, and S. Pal, "A Data mining view on class room teaching language", (IJCSI) International Journal of Computer Science Issue, Vol. 8, Issue 2, pp. 277–282, ISSN:1694-0814, 2011.
15. Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, "Data mining model for higher education system", European Journal of Scientific Research, Vol. 43, No. 1, pp. 24–29, 2010.
16. Q. A. Al-Radaideh, E. W. Al-Shawakfa, and M. I. Al-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.
17. Jignesh Doshi, Result Mining: Analysis Of Data Mining Techniques In Education, International Journal of Technical Research and Applications e- ISSN: 2320-8163, Volume 2, Issue 3 (May-June 2014), PP. 25–28.
18. Norman Poh, Ian Smythe "To what Extent we predict Student's performance? A Case Study in colleges in South Africa", 2014 IEEE.



Proceedings of International Conference on  
Communication and Networks

ComNet 2016

Modi, N.K.; Verma, P.; Trivedi, B. (Eds.)

2017, XX, 757 p. 345 illus., Softcover

ISBN: 978-981-10-2749-9