Sentiment Analysis on the Performance of Engineering Students in University Examination: A Non-parametric Approach Using Two-way Analysis of Variance Model

Vijay Bharath*
Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur Tamil Nadu, India. E-mail: vijaybha@srmist.edu.in

Shanthini
Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. E-mail: shanthia@srmist.edu.in

Subbarayan
Directorate of Research, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. E-mail: subbaraa@srmist.edu.in

Received November 12, 2020; Accepted December 10, 2020
ISSN: 1735-188X
DOI: 10.14704/WEB/V18SI01/WEB18061

Abstract

Higher Education System in India has undergone a rapid change after the year 2000 in India. The Government of India has initiated to start more number of higher learning institutes for the growth of the education sector. In the early years of 2000 a number of engineering colleges / universities were opened for strengthening engineering education by the government and by private management. The engineering educational authorities in particular have formulated a number of guidelines for improving quality of engineering education. The improvements were made in respect of integration of different courses and different streams of engineering disciplines. In this article an attempt is made to study the effects of such integrated efforts based on Sentiment Analysis. The aspects of diversified course structures and the engineering streams are identified with respect to pass percentage of students in university examinations. Friedman Two-way analysis of variance model is used for the analysis of the sampled data. Inferences based on the hypotheses tested are presented.

Keywords

Sentiment Analysis, Education, Engineering Streams, Two-way ANOVA Model, Ranking.

Introduction

Higher education system in India has a distinct place among the educational systems in the globe. It has largest number of institutions compared to other countries. The system is witnessing rapid changes in the post-independence era. Higher education in general and engineering education in particular requires a total transformation to fulfill the aspirations and needs of the society. The engineering education system is very much diversified in terms of Courses, Management and geographical coverage. It is important to note that new technologies offer vast opportunities for quality improvement in the higher education setup. The education policy makers are constantly developing strategies and mechanism for improvement of quality education in the engineering educational institutions viz., colleges and universities. The quality aspects consists of formulating effective teaching and academic programmes and establishing better infrastructure facilities in the campuses of the colleges and universities. In the context of quality improvement of the engineering education sector academicians and management of the engineering institutions play an important role. The quality improvement consists of the following aspects.

i. Developing deep knowledge in the core disciplines.
ii. Basic understanding of the Engineering subjects.
iii. Analytical thinking based on principles of mathematics and its application for Engineering.
iv. Inclusion of skill development subjects.

The main objective of the present study is to analyze the above aspects based on the sentiment analysis. The study is structured as given below. Section 2 deals with sentiment analysis and its associated aspects for improving engineering education. In section 3 we have presented the methodology in identifying / studying the role and importance of the specified subjects and its characteristics. Section 4 contains details relating to Data Structure. Empirical analysis of the data and results are given in Section 5. A conclusion based on the study is presented in Section 6.

Sentiment Analysis and its Characteristics

Meaning of Sentiment Analysis

Sentiment Analysis (SA) primarily involves identification and classification of opinions of users based on a text. The sentiments expressed by the users may be of the type viz., positive, negative and neutral or emotions of the type viz., happy, sad, angry or disgusted towards a particular subject. The importance of SA in different field is felt by researchers. SA plays a major role in the field of education.

Basic Components of Sentiment Analysis in the Education are:

a. Assignments

b. Quizzes
c. Student Seminars
d. Project Reports
e. Periodical Test and
f. University Examinations

Literature Review of Sentiment Analysis with Special Reference to Education

Conrad Tucker and Pursel (2014) proposed a methodology for employing robust natural processing and data mining algorithms to quantify temporal changes in students sentiments relating to course topics and instructor clarity.

Agathe Merceron (2015) discussed in detail the aspects relating to educational data mining / learning analytics. He has stated that data scientists should reestablish the formulation of models for studying performance indicators.

Cott Crossley et al. (2015) examined the potential for Natural language Processing (NLP) tools to predictive success in Educational data mining related aspects. It is further noted that the results of the study are both significant and extendible to data sets involving educational data mining.

Carly Robinson et al. (2016) have pointed out the potential of Natural Language Processing for predicting students success in Massive Open Online Courses (MOOCs).

Sujata Rani and Prateek Kumar (2017) developed Sentiment Analysis System for improving teaching and learning activities in universities. They have analyzed sentiment, emotion and satisfaction parameters for helping the educational administrators and teachers for understanding the problematic areas and the needs for the corrective actions.

Baidal et al. (2018) have reviewed the approaches used in Sentiment Analysis. The have noted Naïve Bayes technique is the most suited one for sentiment analysis. They have also mentioned the main benefits of using sentiment analysis in education domain for improvement of students performance.

Sentiment Analysis in the Context of the Present Study

Sujata Rani et al. (2017) have stated that teaching and learning outcomes can be evaluated directly (assessment of student’s work such as assignments, quizzes, exam etc.) and indirectly (student’s learning experience and teaching quality). They have also mentioned three approaches (i) Machine Leaning approach (ii) Lexicon – based approach and (iii) Hybrid approach.

In the present study we have examined in detail the following:

(1) Is there a significant difference in the percentage of passes between the four types of subjects?

(2) Is there a significant difference in the percentage of passes between the engineering disciplines?

We have assessed the outcome namely percentage of passes based on teaching and learning outcome of the students.

A Description of Sampled Data for its Analysis

The sampled data are collected from private cloud based storage system. Private cloud is a computing model that provides an exclusive environment committed to a single business entity. The use of cloud storage is to store information on the Web through a cloud computing platform. This gives you nimbleness, worldwide scale and strength, with "anytime, anywhere" information to get.

Cloud administrations permit universities to cost-effectively update communication and learning systems without enormous investments in infrastructure. Higher education faces the challenge of overseeing and gaining insight from enormous and growing quantities of data - from student and faculty information to modern research analytics. Besides, this information requires high levels of security and administration in order to connect both security and Intellectual Property necessities.

The data has been extracted from third semester examination (2019-2020) by the SRM Institute of Engineering and Technology. The data in respect of percentage of passes in four subjects of different disciplines namely Professional Core (PC), Basic Science (BS), Engineering Science (ES) and Mandatory Course (M) are considered for the analysis. In addition to the above an attempt has also been made to find the significant difference (if any) between engineering disciplines.

Engineering educational experts are of the firm opinion that one professional core subject, one basic science subject, one engineering subject and one mandatory course subject are of crucial importance for effective strengthening of the engineering curriculum.

These sentimental aspects of engineering educational experts have motivated the authors of this research article to examine in detail using a quantitative model. The Friedman Two-way analysis of variance model incorporating non-parametric assumptions formed the basis for this research work.

Methodological Aspects of the Study

Freidman (1937, 1940) suggested a test of the hypothesis that the s treatment effects are equal which does not require the assumption of analysis of variance to be satisfied by the data. Instead of using the observed value in the i^{th} row he suggested the use of ranks of the observations in each of the rows.

We have 'k' columns and 'r' rows.
The ranks are assigned for each of the observations in a row according to the magnitude of the observation.
Let R_{ij} denote the rank of the j^{th} observation in the i^{th} row (i = 1, 2 . . . r; j = 1, 2…k).
Sum of the ranks in each row = k (k + 1) / 2.
Hence sum of ranks in all rows = r. k (k + 1) / 2.

Test of Hypothesis

H_o: There is no significant difference in the pass percentage in the categories of subjects against
H_1: There is significant difference in the pass percentage of categories of subjects.
Gibbon (1971) has shown that the following statistic F can be used to test H_o against H_1.

$$F = \frac{12}{rk(k+1)}\sum_{j=1}^{k} R_j^2 - 3r(k+1) \qquad (1.1)$$

Where,
Rj = Rank of the jth column.
k = Number of columns.
r = Number of rows.
If Ho is true F has a χ2distribution with (k – 1) degrees of freedom.
Hence the test procedure for a specified level of significance α is
Reject Ho if F > χ2s-1, α

Data Structure for Analysis

The main objective of the analysis in respect of performance evaluation of the students is explained in detail in the following section. The study is confined to the percentage of pass in respect of the different courses offered under different engineering streams. The following streams are considered for the detailed analysis.

(i) Computer Science and Engineering (CSE)
(ii) Electronics and Communication Engineering (ECE)
(iii) Mechanical Engineering (MECH)
(iv) Biotechnology Engineering (BIO) and
(v) Civil Engineering (CIVIL)

For evaluation purposes the percentage of pass of the above mentioned courses consists of the following:

a) Professional Core (PC)
b) Basic Science (BS)
c) Engineering Science (ES) and
d) Mandatory Course (M)

Empirical Analysis

The educational authorities are of the firm opinion that there does not exist significant differences in the percentage of passes between the categories of courses under different engineering streams. In this study an attempt is made in respect of the above statement and we present empirical evidences for the same using Friedman Two-way analysis of variance model.

Grading Codes and Engineering Streams for the Study

The data in respect of our analysis comprises of the following:

i. Rows representing Grading codes namely O, A+, A, B+, B, C. The different gradings codes relates to the following weights:

$$O \Rightarrow 10$$
$$A+ \Rightarrow 9$$
$$A \Rightarrow 8$$
$$B+ \Rightarrow 7$$
$$B \Rightarrow 6$$
$$C \Rightarrow 5$$

ii. Columns representing the different streams of engineering: CSE, ECE, MECH, BIO and CIVIL.

In the context of the above, we have structured the data and the percentage of pass in respect of category of subjects namely Professional Core, Basic Science, Engineering Science, and Mandatory Course are presented in the following tables Table1, Table 2, Table 3, and Table 4.

Table 1 Percentage of pass in Professional Course

Grading	Departments				
	CSE	**ECE**	**MECH**	**BIO**	**CIVIL**
O	26.72	6.26	7.10	5.34	23.27
A+	32.87	24.17	11.11	33.50	18.87
A	20.26	28.17	14.81	35.92	15.10
B+	13.15	23.83	29.01	18.93	21.38
B	4.74	7.13	15.44	3.88	10.69
C	2.26	10.44	22.53	2.43	10.69
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Table 2 Percentage of pass in Basic Science

Grading	Departments			
	CSE	**ECE**	**MECH**	**CIVIL**
O	37.67	21.62	17.50	24.30
A+	21.97	20.72	15.25	9.72
A	14.69	20.72	16.25	15.28
B+	12.67	17.84	22.25	18.06
B	5.49	8.65	11.25	16.67
C	7.51	10.45	17.50	15.97
Total	**100.00**	**100.00**	**100.00**	**100.00**

Table 3 Percentage of pass in Engineering Science

Grading	Departments			
	CSE	**ECE**	**MECH**	**BIO**
O	14.94	9.71	5.20	19.39
A+	32.68	22.53	10.12	22.96
A	25.97	27.47	22.83	22.96
B+	13.96	21.43	29.20	20.92
B	6.06	7.69	13.29	8.16
C	6.39	11.17	19.36	5.61
Total	**100.00**	**100.00**	**100.00**	**100.00**

Table 4 Percentage of pass in Mandatory Course

Grading	Departments				
	CSE	ECE	MECH	BIO	CIVIL
O	5.19	2.29	1.33	0.98	4.35
A+	25.40	12.70	14.89	10.68	9.94
A	35.21	31.43	34.00	38.83	26.09
B+	24.04	36.81	33.33	36.89	32.92
B	6.77	11.56	11.33	8.74	13.66
C	3.39	5.21	5.12	3.88	13.04
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Assigning ranks to each observation for each grading independently, we get the following data of ranks. The rankings are presented in Table 5, Table 6, Table 7 and Table 8.

Table 5 Ranking of pass percentage for the professional core

Grading	Departments				
	CSE	ECE	MECH	BIO	CIVIL
O	1	4	3	5	2
A+	1	3	5	2	4
A	3	2	5	1	4
B+	5	2	1	4	3
B	4	3	1	5	2
C	5	3	1	4	2
Sum of Ranks	**19**	**17**	**16**	**21**	**17**

Table 6 Ranking of pass percentage for the Basic Science

Grading	Departments			
	CSE	ECE	MECH	CIVIL
O	1	3	4	2
A+	1	2	3	4
A	4	1	2	3
B+	4	3	1	2
B	4	3	2	1
C	4	3	1	2
Sum of Ranks	**18**	**15**	**13**	**14**

Table 7 Ranking of pass percentage for the Engineering Science

Grading	Departments			
	CSE	ECE	MECH	BIO
O	2	3	4	1
A+	1	3	4	2
A	2	1	4	3
B+	4	2	1	3
B	4	3	1	2
C	3	2	1	4
Sum of Ranks	**16**	**14**	**15**	**15**

Table 8 Ranking of pass percentage for the Mandatory Course

Grading	Departments				
	CSE	ECE	MECH	BIO	CIVIL
O	1	3	4	5	2
A+	1	3	2	4	5
A	2	4	3	1	5
B+	5	2	3	1	4
B	5	2	3	4	1
C	5	2	3	4	1
Sum of Ranks	**19**	**16**	**18**	**19**	**18**

Computational Details

The computations in respect of the statistic given in Equation 1.1 are noted below.
The steps involved in the computations are:

i. We calculated R_j^2 for each study variable viz., Professional Core (PC), Basic Science (BS), Engineering Science (ES) and Mandatory Course (M).
ii. χ^2 is calculated for each of the study variables (χ^2_{Cal}).
iii. Table value in respect of each study variables is given (χ^2_e).

The results are presented in the following Table 9.

Table 9 A Comparison of Study Variables

	Study variables			
	PC	**BS**	**ES**	**M**
R_j^2	1636	914	902	1626
χ^2_{Cal}	1.6120	1.4000	0.2000	0.9420
χ^2_e	9.4900	7.8200	7.8000	9.4900
Inference	Accept H_o	Accept H_o	Accept H_o	Accept H_o

Conclusions

The Sentiment Analysis based on the sampled data has revealed the following:

(i) There is no significant difference in the pass percentage of students between the different categories of the courses offered.
(ii) There is also no significant difference in pass percentage of student between the different Engineering streams.

The above conclusions may be helpful for engineering educational planners for more effective policies in respect of quality improvement.

Acknowledgment

The authors of this research article would like to thank the authorities of the SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India for providing the necessary data and facilities for carrying out the research study.

References

Tucker, C., Pursel, B.K., & Divinsky, A. (2014). Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes. *The ASEE Computers in Education (CoED) Journal*, *5*(4), 84.

Merceron, A. (2015). Educational Data Mining/Learning Analytics: Methods, Tasks and Current Trends. *In DeLFI Workshops*, 101-109.

Crossley, S., McNamara, D.S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to Completion: Success in an Educational Data Mining Massive Open Online Class. *Proceedings of the 8th International Conference on Educational Data Mining*, 388-391.

Robinson, C., Yeomans, M., Reich, J., Hulleman, C., & Gehlbach, H. (2016). Forecasting student achievement in MOOCs with natural language processing. *In Proceedings of the sixth international conference on learning analytics & knowledge*, 383-387.

Rani, S., & Kumar, P. (2017). A sentiment analysis system to improve teaching and learning. *Computer*, *50*(5), 36-43. https://doi.org/10.1109/MC.2017.133

Mite-Baidal, K., Delgado-Vera, C., Solís-Avilés, E., Espinoza, A.H., Ortiz-Zambrano, J., & Varela-Tapia, E. (2018). Sentiment analysis in education domain: A systematic literature review. *In International Conference on Technologies and Innovation*, 285-297. https://doi.org/10.1007/978-3-030-00940-3_21

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American statistical association*, *32*(200), 675-701.

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, *11*(1), 86-92.

Gibbons, J.D. (1971). *Nonparametric Statistical Inference*. Mc-Graw Hill, Kogakusha, Tokyo.