International Journal of Control and Automation
Vol. 13, No. 2, (2020), pp. 358 – 367

Educational Data Mining to Identify Relationship Between Technical Knowledge and Academic Performance

Dr. Vikas Goel
Associate Professor, Dept. of IT, KIET Group of Institutions, Ghaziabad, U.P.

Dr. Pragati Goel
Associate Professor, Dept. of MCA, NCRD's Sterling Institute of Management Studies, Mumbai, Maharashtra

Dr. Amit Kumar Gupta
Associate Professor, Dept. of MCA, KIET Group of Institutions, Ghaziabad, U.P.

Abstract

Now-a-days educational data is becoming very crucial to analyze students' performance in their ability to get absorbed in the industry. Educational Data Mining (EDM) is the field of study concerned with mining instructive information to discover fascinating examples and learning in instructive associations. It demonstrates to obtain information from collection of student data from academic Institution databases. This paper highlights the importance of using student data to drive improvement in recruitment or placement of a student. Further, it will enable faculty members to identify, predict and classify students based on academic performance measured using Cumulative Grade point average (CGPA) grades. This investigation investigates different components accepted to influence advanced education and technical skills to finds a subjective model which predicts the capability in view of related individual and specialized elements.

Keyword: *Data Mining, Educational Data Mining, Prediction, Correlation.*

1. Introduction

In order for the software companies to cope with frequently shifting conditions and the raising quality demands, they need to constantly improve the quality of their work. Now-a-days educational data is becoming very crucial to analyze students' performance in their ability to get absorbed in the industry. Education Data Mining (EDM) is the field of concentrate worried about mining instructive information to discover fascinating examples and learning in instructive associations. It demonstrates to obtain information from collection of student data from academic Institution databases [1,4,9]. This paper highlights the importance of using student data along with other criteria during the

ISSN: 2005-4297 IJCA
Copyright ⓒ 2020 SERSC

placement. It aims at analyzing the relationship between technical knowledge and academic performance. Is there a clear correlation between the two and as the result success achieved? The research is based on both literature reviews and empirical studies. The quantitative research conducted at NCRD's Sterling Institute of management Studies located at Nerul, Navi Mumbai.

Education Data Mining (EDM) is utilized to perform research and help in building models in a few regions impacting web based learning frameworks. This educational research may be used to focus on developing new tools to discover various patterns in data. Specialists in this field center around finding valuable information to enable the instructive foundations to deal with their training and expectations better and upgrade their execution. Analyzing student's data helps to classify students, or to make decision trees or association rules, to form better decisions. [2,3,5]

2. Objectives of the study

The main purpose of this paper is to explore the relationship between technical knowledge and academic performance to enter in the industry. The main objective of this study is to identify which of the two areas i.e. technical knowledge or academic performance are effective or helpful in getting a student placed. The research objectives of exploring the relationship between technical knowledge and academic performance are:

- To understand the role of technical knowledge in getting placement in industry
- To identify what does academic performance weighs in industry
- To know the importance of academic grades and technical knowledge in industry
- To analyze clear relationship between technical knowledge and academic performance

3. Literature Review

According to A Review on Educational Data Mining (International Journal of science and research). It is important to review and analyze educational data especially students' performance. Educational data processing (EDM) is that the field of study concerned with mining educational data to seek out out interesting patterns and knowledge in educational organizations. This study is equally concerned with this subject, specifically, the students' performance. This study explores multiple factors theoretically assumed to affect students' performance in education , and finds a qualitative model which best classifies and predicts the students' performance supported related personal and social factors. [3,6,7]

Data mining is used to find out new and useful information from large amount of data. Techniques of data mining are useful in various application areas like fraud detection, Businesses, banking and telecommunications. The major application of Data

ISSN: 2005-4297 IJCA
Copyright ⓒ 2020 SERSC

Mining Technique is educational data mining in order to extract useful information from educational data. [8,10,11,12]

Analyzing students' data and knowledge to classify students, or to make decision trees or Association rules, to form better decisions or to reinforce student's performance is a stimulating field of research, which mainly focuses on analyzing and understanding students' educational data that indicates their educational performance, and generates specific rules, classifications, and predictions to assist students in their future educational performance. [13,14,15]

4. Methodology of the Study

4.1 Research Design: Experimental Study

Experimental studies are those where the researcher tests the causal relationship between variables. Such studies require procedures that will not only reduce bias and increase reliability, but will permit drawing inference about causality. So to analyze this quantitative research is used for this study.

4.2 Database

This study is based on primary data. A structured questionnaire was formed to collect the primary data from the students of NCRD's Sterling Institute of Management Studies.

4.3 Instrumentation

For the purpose of the study a structured questionnaire was designed with 15 closed ended questions each for five technical areas and their marks for Secondary, Higher Secondary, Graduation and Post Graduation are considered for Academic performance.

4.4 Population and Sample Size

The population is 1200 students of three different years and sample size of 30 respondents are shown in this paper.

4.5 Sampling Techniques

Purposive sampling technique was used for the study. Under this sampling technique, the students are selected with different caliber and are simply given questionnaire to fill from different strata.

4.6 Statistical Techniques

Appropriate but simple statistical method like mean, standard deviation, and correlation (Karl Pearson's) analysis are employed to analyze and interpret the data collected.

4 Data Analysis

Data is collected from survey where the student's information according to their knowledge in technology is analyzed. A question set of technical subject was prepared and then circulated amongst students divided into all class sections.

- The first table, Table-1 represents the sample data collected from students for five technologies identified as the latest ongoing technologies in market.
- The second table, Table-1 represents the sample academic achievement of the students, based on the marks in Secondary, Higher Secondary, Graduation and Post Graduation.

ISSN: 2005-4297 IJCA
Copyright ⓒ 2020 SERSC

- The third table, Table-2 presents the mean value of the result gathered from the questionnaire research. The standard deviation is also presented under the S.D. column.
- The correlations between the values are presented in Table 3.

Table: 1 Survey Data of Students along with educational marks

Sr. No	Technology					Total (75)	Percentage %	Education Percentage				Avg. %
	C++ (15)	Java (15)	C# (15)	Aptitude (15)	DBMS (15)			10^{th}	12^{th}	UG	PG	
1	15	13	13	12	12	65	86.67	57.65	73	74	79	70.91
2	5	9	1	5	10	30	40.00	65	59	66	60	62.50
3	6	12	6	16	9	49	65.33	66.00	48	60.00	65	59.75
4	13	10	11	9	12	55	73.33	58	62	70	62.00	63.00
5	11	11	12	7	10	51	68.00	61.45	59.58	60	65	61.51
6	11	10	11	9	12	53	70.67	65	59	66	60	62.50
7	11	11	14	12	7	55	73.33	76.43	62	70	60	67.11
8	8	14	12	13	10	57	76.00	62	58	66	63	62.25
9	8	11	11	12	12	54	72.00	72	63	68	70	68.25
10	6	10	10	12	12	50	66.67	61.45	59.58	60	65	61.51
11	6	12	10	10	14	52	69.33	74.63	56	65.67	74	67.58
12	5	6	8	6	9	34	45.33	60	59	66	60	61.36
13	10	10	12	7	10	49	65.33	58	64	60	60	60.50
14	13	14	13	9	12	61	81.33	86.33	70	72	69	74.33
15	6	8	5	10	8	37	49.33	65	59	66	60	62.50
16	4	8	4	11	7	34	45.33	80	74	75	70	74.75
17	5	9	7	16	12	49	65.33	58	64	60	60	60.50
18	7	9	8	8	14	46	61.33	74.63	56	65.67	74	67.58
19	8	10	9	9	14	50	66.67	65	59	66	60	62.50

Copyright ⓒ 2020 SERSC

20	10	10	12	6	13	51	68.00	61.45	59.58	60	65	61.51
21	6	4	7	5	11	33	44.00	58	64	60	60	60.50
22	9	8	7	9	10	43	57.33	86.33	70	72	69	74.33
23	11	13	13	9	12	58	77.33	76	64	60	60	65.00
24	7	9	9	9	12	46	61.33	74.63	56	65.67	74	67.58
25	9	10	10	7	10	46	61.33	76	64	60	60	65.00
26	10	9	9	11	10	49	65.33	61.45	59.58	60	65	61.51
27	10	12	12	13	13	60	80.00	58	64	60	60	60.50
28	14	10	13	10	11	58	77.33	80	74	75	70	74.75
29	12	11	10	9	9	51	68.00	51	59.70	60	65	58.93
30	9	11	6	12	8	46	61.33	86.33	70	72	69	74.33

Table: 2 Analysis table from Technical percentage and Educational Percentage

Sr. No.	Technology (%)	Education (%)	S.D. (Technology)	S.D. (Education)	Final Data Analysis
1	86.67	70.91	1.22	9.22	Good
2	40.00	62.50	3.61	3.51	Average
3	65.33	59.75	4.27	8.26	Good
4	73.33	63.00	1.58	5.03	Good
5	68.00	61.51	1.92	2.46	Good
6	70.67	62.50	1.14	3.51	Good
7	73.33	67.11	2.55	7.57	Good
8	76.00	62.25	2.41	3.30	Good
9	72.00	68.25	1.64	3.86	Good
10	66.67	61.51	2.45	2.46	Good
11	69.33	67.58	2.97	8.73	Good
12	45.33	61.36	1.64	3.20	Average

ISSN: 2005-4297 IJCA
Copyright ⓒ 2020 SERSC

13	65.33	60.50	1.79	2.52	Good
14	81.33	74.33	1.92	8.09	Good
15	49.33	62.50	1.95	3.51	Average
16	45.33	74.75	2.95	4.11	Average
17	65.33	60.50	4.32	2.52	Good
18	61.33	67.58	2.77	8.73	Average
19	66.67	62.50	2.35	3.51	Good
20	68.00	61.51	2.68	2.46	Good
21	44.00	60.50	2.70	2.52	Average
22	57.33	74.33	1.14	8.09	Average
23	77.33	65.00	1.67	7.57	Good
24	61.33	67.58	1.79	8.73	Average
25	61.33	65.00	1.30	7.57	Average
26	65.33	61.51	0.84	2.46	Good
27	80.00	60.50	1.22	2.52	Good
28	77.33	74.75	1.82	4.11	Good
29	68.00	58.93	1.30	5.82	Good
30	61.33	74.33	2.39	8.09	Average

Table 3 : Correlation between technical knowledge or academic performance

	Academic Performance	**Coefficient**
Technical Knowledge	**0.80**	**r**

Correlation of coefficient between two variables is presented by "r".

Conditions:

1. "r" ranges from -1 to 0 and from 0 to +1.
2. If r = 0.9 then it indicates very high degree of positive correlation between two variables.
3. If "r" is in range from 0.75 to 0.9 then it indicates a reasonable high degree of positive correlation between two variables.

ISSN: 2005-4297 IJCA
Copyright ⓒ 2020 SERSC

In the wake of ascertaining all information of instruction and innovation, we have given evaluations as indicated by following the following equation. Here, as indicated by equation 1, Technical Knowledge contrasted with Educational performance has "good" Grade who got over 60 (top of the line) else have "Average" Grade.

IF (Technical Knowledge > Education performance)	*eq(1)*
THEN	
Student grade="Good"	
ELSE	
Student grade= "Average"	

In our survey, we got information where the educational percentage is not as much as the technical knowledge.
Here, Sr. No1,3,4,5,6,7,8,9,10,11,13,14,17,19,20,23,26,27,28,29 got more rate in specialized subject than Education percentage.

Next, we have used a tool to find data the relation between data and depicted in graphical format. In figure 1, the technical and educational percentage of students are depicted in interval 40-50, 50-60, 60-70, 70-80, 80-90. The maximum student data lie between 60 - 70 for both technical and educational education.

Based on the technical and educational percentage of students, the student grades are categories between good and average and depicted in figure 2.

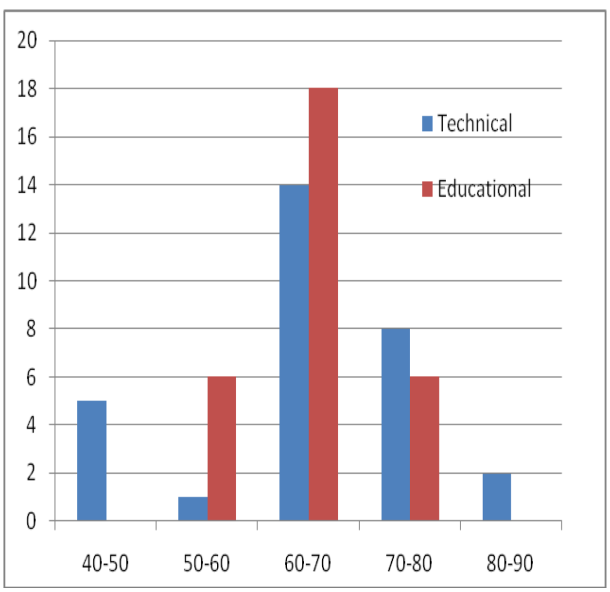

Figure 1: Technical and Educational percentage of students

Copyright ⓒ 2020 SERSC

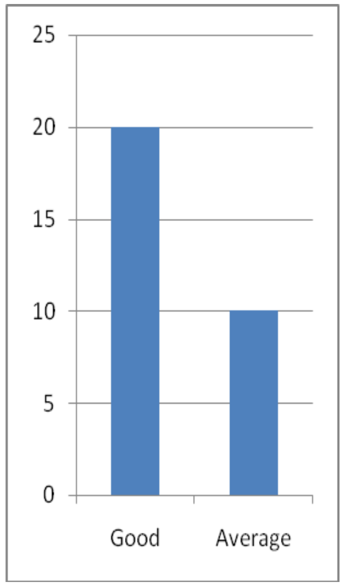

Figure 2: Students Grade

This quantitative research strongly confirmed that there is a high degree of correlation between technical knowledge and academic performance. However, technical knowledge is needed in order to get succeeded. Thus, even though academic performance is a proven necessity in modern software development, it is up to each student to develop their own technical skills in order to get placed.

6. Conclusion

This paper highlights the importance of using student data along with other criteria during the placement. This study explores multiple factors theoretically assumed to affect students' performance and finds a qualitative model which best classifies and predicts the students' performance based on academic performance and technical skills. This study is equally concerned with his subject, specifically, the students' performance.

This paper highlighted the importance of using student data to drive improvement in recruitment or placement of a student. Further, it enabled faculty members to identify, predict and classify students based on academic performance measured using Cumulative Grade point average (CGPA) grades.

- In future, the research work may also be carried out using some other mining tool.
- This survey is only done for MCA students. In future it may be carried out for other students of different streams also.

Copyright ⓒ 2020 SERSC

While doing education, this survey may also be done on different programming languages that may help to enhance their skill sets.

References

[1] B. K. Baradwaj and S. Pal, "Mining Educational Data to Analyze Students' Performance", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6 **(2011)**.

[2] A.B.E.D. Ahmed and I.S Elaraby, "Data Mining: A prediction for Student's Performance Using Classification Method", World Journal of Computer Application and Technology, 2(2), pp.43-47 **(2014)**.

[3] J. Fuentes, C. Romero, C. García-Martínez, S. Ventura, "Accepting or Rejecting Students' Self-grading in their Final Marks by using Data Mining", International Conference on Educational Data Mining, London, UK, 327-328 **(2014)**.

[4] S. A. Kumar, M. Vijayalakshmi, "A novel approach in data mining techniques for educational data", In: 3rd International Conference on Machine Learning Computing (ICMLC 2011) A, no. Icmlc, pp. 152–154 **(2011).**

[5] R.S. Baker, P.S. Inventado, "Educational data mining and learning analytics. In: Learning Analytics: From Research to Practice", pp. 61–75. Springer, New York **(2014).**

[6] K. Sin, L. Muthu, "Application of big data in education data mining and learning analytics a literature review", ICTACT J. Soft Computing: Special Issue Soft Computing Models Big Data5(4), 1035–1049 **(2015).**

[7] H. Sun, "Research on student learning result system based on data mining", International Journal of Computer Science Network Security 10(4), 203–205 **(2010)**.

[8] A. Permata Alfiani, F. Ayu Wulandari, "Mapping student's performance based on data mining approach (a case study)", Ital. Oral Surg. 3, 173–177 **(2015).**

[9] R. S. Baker, "Educational data mining: an advance for intelligent systems in education", AI Educ. 78–82 **(2014).**

[10] M. Ramaswami, R. Bhaskaran, "A study on feature selection techniques in educational data mining", Journal of Computing 1(1), 7–11 **(2009)**.

[11] V. Kumar, "An empirical study of the applications of data mining techniques in higher education", International Journal of Advances Computer Science Applications 2(3), 80–84 **(2011)**.

[12] P. Gulati, A. Sharma, "Educational data mining for improving educational quality", International Journal of Computer Science Information Technological Security 2(3), 648–650 **(2012)**.

[13] S. Ayesha, T. Mustafa, A. Raza, M. I. Sattar, "Data mining model for higher education system", European journal science research, 43(1), 24–29 **(2010).**

[14] P. Cortez, A. Silva, "Using data mining to predict secondary school student performance", In: 5th Annual Future Business Technology Conference, vol. 2003, no. 2000, pp. pp. 5–12 **(2008)**.

Copyright ⓒ 2020 SERSC

[15] M. I. Al-Twijri, A.Y. Noaman, “A new data mining model adopted for higher institutions”, Procedia Computer Science 65, 836–844 **(2015)**.

[16] S. Aslam, I. Ashraf, “Data mining algorithms and their applications in education data mining”, Journal of Computer Applications, 7782, 50–56 **(2014)**.

[17] J.F Grafsgaard, J.B.Wiggins, K.E. Boyer, E.N. Wiebe, J.C. Lester, “Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue”, In: Proceedings of 7th International Conference on Educational Data Mining, pp. 122–129 **(2014)**.

[18] M. Swamy, M. Hanumanthappa, “Predicting academic success from student enrolment data using decision tree technique”, International Journal of Applied Information Systems, 4(3), 1–6 **(2012)**.

ISSN: 2005-4297 IJCA
Copyright ⓒ 2020 SERSC