

# Application of Data Mining Techniques in Designing Knowledge Base on Student Competency at the Post Graduate level



M. Selvalakshmi\*, J. Arunadevi\*\*, Dr. K. Ravi Chandran\*\*\*

© Institute of Management Studies, Noida  
Online access at [www.publishingindia.com](http://www.publishingindia.com)

## Abstract

The most valuable asset of any organization is its Knowledge and the ability to use this knowledge appropriately to enhance the output. For an Educational Institution, the core strength would be the indepth knowledge of the predictors of a student's performance. Many independent variables such as Gender, Educational background, Work experience might influence a candidate's performance in the Post Graduate level. Information generated about the candidate can be converted into fruitful knowledge about future trends.

A very promising tool to attain this objective is the use of appropriate Data Mining technique. The purpose of this study was to identify valid predictors and measures of the academic performance of MBA students. With the use of an effective data mining tool, an institution can determine relationships among "internal" factors such as Gender, Educational background, Work experience, and "external" factor such as student performance in their Post graduation.

Association rules identify collections of data attributes that are statistically related in the underlying data. Given a database of transactions, a minimal confidence threshold and a minimal support threshold, find all association rules whose confidence and support are above the corresponding thresholds.

This paper aims at using Association Rule Mining technique for classifying students based on various parameters such as their school board exam performance, specialization in the their under graduation level, Gender and their previous work experience if any. This is done by comparing their performance in the internal and external levels in their Post Graduate education.

**Keywords :** Association Rule Mining, Students Performance

## 1. Introduction

Maintenance of academic standards is the prime concern of any educational institution. Education at the Masters level, involves the conscious involvement of time, money and effort by the candidates. With the industrial downturn and reduction in the demand for graduates at the job market due to recession, there may be many who vow to equip themselves for the futures through higher education.

However without a clear and well thought through approach, institution may fail to identify the right candidate who would support the cause of maintaining high academic standards. Infact, it is much likely that many companies may prefer to recruit from B schools which give considerable importance to scientific student admission procedures.

The most valuable asset of any organization is its Knowledge and the ability to use this knowledge appropriately to enhance the output. For an Educational Institution, the core strength would be the indepth knowledge of the predictors of a student's performance.

Many independent variables such as Gender, Educational background, Work experience might influence a candidate's

\*Senior Lecturer, Department of Management Studies, Thiagarajar School of Management, Madurai, India

\*\*Lecturer, Department of Computer Applications, Thiagarajar School of Management, Madurai, India

\*\*\*Professor, Department of Management Studies, , Madurai Kamaraj University, Palkalai Nagar, Madurai, India

performance in the Post Graduate level. Information generated about the candidate can be converted into fruitful knowledge about future trends.

## 2. Review of Literature

M Kotze and L Griessel in their research work "Predicting academic Performance of MBA students – A South African Study" showed that numerical and verbal abilities, as well as personal attributes, play an important role in the prediction of MBA academic success. Academic researchers have developed several statistical models like discriminant analysis, multiple regression, stepwise regression to predict an applicant's success in the MBA program (Pharr et. Al, 1993; Wright and Palmer, 1997)

Adams and Hancock (June 2000) in their research analyzed the role of work experience in predicting MBA academic success. Their study resulted in the finding that prior work experience accounted for only a small proportion of variance in the first semester grades and was unrelated to the academic performance in the second semester.

Adams and Hancock (2000) also found that there existed a negative correlation between graduate grade point average and the student holding an undergraduate business degree. Braunstein (2002) found similar results in his study of MBA students at a private college in the Northeast.

Gump (2003) in his study determined that the business background of the students in the undergraduate does not result in higher academic performance at the MBA level and Better performers were more likely to be in the non business undergraduate background which he attributed to their attitude to work harder for their lack of undergraduate preparat

A study from the GISMA Business School in Hannover, Germany, has found little correlation between GMAT scores and the performance of its full-time MBA class.

## 3. The Study

The purpose of this study was to determine valid predictors and measures of the academic performance of the MBA students especially in their first term. The data for the study was collected from a B-school for a period of three consecutive years from a total sample of 276 about the student profile and their academic performance. Analysis was done on the data collected to arrive at some worthwhile conclusions.

The study used Association Rule Mining technique for classifying students based on various parameters such as their school board exam performance, under graduation performance, Gender and their previous work experience if any. This was done by comparing their performance in the Viva on all subjects and external University exam levels in the I term of Post Graduate education. This has been used to identify the factors contributing to the success of the students in their higher education and to discover potential student groups with similar characteristics and the degree of association to basic variables or scores.

## 4. Tool For The Study – Data Mining

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. The purpose of this study was to identify valid predictors and measures of the academic performance of MBA students. With the use of an effective data mining tool, an institution can determine relationships among "internal" factors such as Gender, Educational background, Work experience, and "external" factor such as student performance in their Post graduation.

### 4.1 Levels Of Data Mining

Based on the technology and statistics perspective, there could be different levels of data mining.

1. Directed data mining allows business users to infer using classes and clusters.
2. Undirected data mining uses clusters to identify pure statistical patterns.
3. Association helps in discovering links between variables. A supermarket inferred that customers bought beer and diapers on certain days of the week. And thereafter the retail store started placing expensive beers right after diapers.
4. At the final level, data is mined to predict behavior patterns and trends on consumption. The mined data could be analyzed using special statistical tools for fruitful business insights

Some of the widely used tools of analysis include: Artificial Neural networks, Genetic Algorithms, Decision trees, Nearest neighborhood method, Rule Induction, Data Visualization Methods

### 4.2 Associative Data Mining

The current study employed Associative data mining as its major tool to infer meaningful association among variables. Association rules identify collections of data attributes that are statistically related in the underlying data. An association rule is of the form  $X \Rightarrow Y$  where  $X$  and  $Y$  are disjoint conjunctions of attribute-value pairs. The confidence of the rule is the conditional probability of  $Y$  given  $X$ ,  $P(Y|X)$ , and the support of the rule is the prior probability of  $X$  and  $Y$ ,  $P(X \text{ and } Y)$ . Here probability is taken to be the observed frequency in the data set. The traditional association rule mining problem can be described as follows. Given a database of transactions, a minimal confidence threshold and a minimal support threshold, finding all association rules whose confidence and support are above the corresponding thresholds. XLMINER demo edition 3.2.4 has been used for generating the association rules.

In this case, data base comprised of the profile of the students for three consecutive years from 2005, 2006 and 2007 in terms of their  $X$  standard( $X$ ), Higher Secondary (H), Undergraduate (U), Post Graduate University Exam(P) and Post Graduate Internal all subject Viva (V) Performance being banded, MAT / TANCET entry to the course, Gender(M/F), Prior Work Experience(Yes/No). Thousands of associative rules were generated through the XLMINER which can be interpreted as follows:

Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data

Example :

Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support (a U c)	Lift Ratio
100	MAT, U4, V5=>	No, P6, X5	6	9	6	6.444444

and, unlike the if-then rules of logic, association rules are probabilistic in nature.

In addition to the antecedent (the "if" part) and the consequent (the "then" part), an association rule has two numbers that express the degree of uncertainty about the rule. In association analysis the antecedent and consequent are sets of items (called item sets) that are disjoint (do not have any items in common).

The first number is called the **support** for the rule. The support is simply the number of transactions that include all items in the antecedent and consequent parts of the rule. (The support is sometimes expressed as a percentage of the total number of records in the database.)

The association rules also reveal the confidence levels for the associations generated. **Confidence** is the ratio of the number

of transactions that include all items in the consequent as well as the antecedent (namely, the support) to the number of transactions that include all items in the antecedent.

**Lift** is one more parameter of interest in the association analysis. Lift is discusses the ratio of Confidence to Expected Confidence. Lift is a value that gives information about the increase in probability of the "then" (consequent) given the "if" (antecedent) part.

### 5. Results of the Study :

Following were the inferences drawn from the Association rule generated from mining the database :

Across three years, it is inferred that Student selected through MAT and their performance in the Viva (range -60- 100%) serves as an antecedent to their PG Performance (P6, P5 – range

Table 5.1. MAT Antecedent to PG Performance :

Year	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support (a U c)	Lift Ratio
2005	100	MAT, U4, V5=>	No, P6, X5	6	9	6	6.444444
2006	85.71	MAT, No, U4, V5=>	P5	14	62	12	1.548387
2007	100	M, MAT, V3=>	P5	14	72	14	1.527778

70-90%)

From the data, it can be inferred that Gender has an association with Work experience and more male students are found to have prior work experience than the female counterparts.

In this case, TANCET students with no work experience, with UG score in the range of 50- 60% are found to be associated with the gender Male at 100% confidence level across three years.

It has been inferred that undergraduate score of around 50-60%

Table 5. 2. Work Experience and Gender :

Year	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support (a U c)	Lift Ratio
2007	92.86	Yes=>	M	14	73	13	1.399217
2006	86.67	Yes=>	M	15	68	13	1.427451
2006	95.24	F=>	No	42	95	40	1.122807
2005	75	F=>	No	20	45	15	0.966667

serves as an antecedent across three years for the Gender Male. The performance of the candidate in the Higher Secondary level between 70-90% has consistently been an antecedent to

their Viva performance in the range of 80-90%

Table 5.3. Performance of MALE, TANCET students :

Year	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support(a U c)	Lift Ratio
2005	100	M, TANCET, U3=>	P5	7	35	7	1.657143
2006	100	No, P3, TANCET, U3=>	M	12	68	12	1.647059
2007	100	No, P5, TANCET, U3=>	M	12	73	12	1.506849

Table 5.4. Performance in the Under graduation level and Gender :

Year	Rule No:	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support (a U c)	Lift Ratio
2005	649	100	MAT, U3=>	M, No	6	30	6	1.933333
2006	218	100	TANCET, U3=>	M	20	68	20	1.647059
2006	294	92.86	U3=>	M	28	68	26	1.529412
2007	299	100	U3, V3=>	M	13	73	13	1.506849
2007	301	100	TANCET, > U3=	M	15	73	15	1.506849

Table 5. 5. Factors Influencing Viva Performance :

Year	Rule No:	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support (a U c)	Lift Ratio
2005	774	100	H5, yes=>	V4	6	34	6	1.705882
2006	191	91.67	H4, M, U3=>	V4	12	60	11	1.711111
2007	213	91.67	H5, P6, U4=>	V4	12	64	11	1.575521

## 6. Discussion

Performance of the candidate in the Higher Secondary (Table 5.5) proved to be an antecedent to the first class performance in eth All subject Viva at the Post gradual level.

Selection of candidate based on MAT test (Table 5.1) proved to be an antecedent to the first class performance at the Post gradual level.

Male students who have joined MBA course are more likely to have attained prior work experience than their female counterparts (Table 5.2)

Undergraduate score of around 50-60% is found to be an antecedent to the consequent of the candidate being male (Table 5.3 and 5.4)

## 7. Limitations of the Study

The results of this study should be interpreted with consideration of limitations such as the period considered for the study which is only three years and the number of institutions being considered. To ensure greater generalizability, database from more than one Business schools should be enlisted in future studies.

Next level of data mining can also be attempted to arrive at more meaningful inferences.

## 8. Conclusion

Master of Business Administration (MBA) program admission wing is keen in thoroughly examining the admission criteria they can use to admit student into their graduate programs. The results of this study suggest that the Higher Secondary, and Gender serve as valid indicator for students who are likely to perform successfully in a professional course like MBA.

## 9. References

- Galit Shmueli, Nitin R. Patel, Peter C. Bruce, *Data Mining for Business Intelligence : Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, Wiley Publications, December 2006
- Adams, A.J., & Hancock, T. (2000). *Work experience as a predictor of MBA performance*. *College Student Journal*, 30(2), 211-216.
- Ahmadi, Mohammed, et al. (1997). *An examination of the admissions criteria for the MBA programs: a case study*. *Education*, 117, 540-546.
- Braunstein, Andrew W. (2002). *Factors determining success in a graduate business program*. *College Student Journal*, 36(3), 471-477.
- Carver, M.R., & King, T.E. (1994). *An empirical investigation of the MBA admission criteria for nontraditional programs*. *Journal of Education for Business*, 70(2), 94-98.
- Gump, Steven E. (2003). *Defining distinction: characteristics of top MBA students at Cardiff business school*. *International Education*, 32(2), 63-84.
- <http://www.resample.com/xlminer/index.shtml>
- <http://www.buzzle.com/articles/mba-success-undergrad-degree.html>
- <http://www.questia.com>
- <http://whichmba.wordpress.com>