International Journal of Advanced Research in IT and Engineering

ISSN: 2278-6244

STUDY OF STUDENTS' PERFORMANCE USING DATA MINING MODEL WITH EXCEL 2007

Mrs. Atiya Khan*

Abstract: *Data Mining is used to extract meaningful information and to developed significant relationship among variables stored in large data warehouse. Knowledge Discovery and Data Mining (KDD) is a multidisciplinary area focusing upon methodologies for extracting useful knowledge from data and there are several useful KDD tools to extracting the knowledge. Indian education sector has a lot of data that can produce valuable information which an be used to increase the quality of education. Educational data mining (EDM) provides a set of techniques which can help educational system to overcome this issue in order to improve learning experience of students as well as increase their profits[3].*

In this paper we present the educational data mining process with ***"Data Mining Client for Excel 2007"*** *and explain how actuaries can use Excel to build predictive models, with little or no knowledge of the underlying SQL Server system. The Students' past performance data is generate to produce data mining model. The students' performance is evaluated by considering factors which include PSM marks and Current semester internal marks such as ATT, CAT, SSM, CDC and PW. The various data mining techniques are used to predict the students' performance. This study will help the teacher to reduced drop-out ratio to a significant level and improve the performance of students.*

Keywords: *Data Mining, Educational Data Mining, Clustering, Classification, Knowledge Discovery in Database (KDD)*

*Assistant Professor, Department of MCA, Priyadarshini College of Engineering, Hingna Road, Nagpur.

1. INTRODUCTION

Data mining in educational environment is called Educational Data mining, concern with developing new methods to discover knowledge from educational database in order to analyze student's trends and behaviors towards education. The students' performance plays an important role in producing the best quality graduates and post-graduates who will become manpower for the country's economic and social development. Academic achievement is one of the main factors considered by the company in recruiting workers especially the fresh graduates. Thus, students have to place the greatest effort in their study to obtain a good grade in order to fulfill the company demand.

The data collected from different applications require proper method of extracting knowledge from large repositories for better decision making. Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful information from large collections of data*(Mannila,1996).* The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data (*U. Fayadd, Piatesky,1996)*. Data mining tools predict patterns, future trends and behaviors, allowing businesses to effect proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analysis of past events provided by retrospective tools typical of decision support systems.. There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments (*J. Han and M. Kamber, 2000)*.

The main objective of this paper is to use data mining methodologies to study students' performance in the MCA course. Students' academic achievement is measured by the Aggregate Marks scored in an academic session. Aggregate Marks shows the overall students' academic performance where it considers the average of all examinations' grade for all semesters during the tenure in university. Data mining provides many tasks that could be used to study the students' performance. In this research, the classification task is used to evaluate student's performance. Student's information like previous semester marks, Attendance, Class test, Seminar and Assignment marks were collected from the student's database system, to predict the performance at the end of the semester examination.

Clustering is one of the basic techniques often used in analyzing data sets. This study makes use of cluster analysis to segment students into groups according to their characteristics.

2. BACKGROUND AND RELATED WORKS

Although, using data mining in higher education is a recent research field, there are many works in this area. Baradwaj and Pal [10] applied the classification as data mining technique to evaluate student' performance, they used decision tree method for classification. The goal of their study is to extract knowledge that describes students' performance in end semester examination. Han and Kamber[3] explained that k-means is a well known clustering algorithm tends to uncover relations among variables already presented in dataset. Hijazi and Naqvi [5] conducted as study on the student performance from a group of colleges affiliated to Punjab university of Pakistan. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance. Khan [6] conducted a performance study on 400 students from the senior secondary school of Aligarh Muslim University. A sample of data clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socioeconomic status had relatively higher academic achievement in general.

El-Halees [4], gave a case study that used educational data mining to analyze students' learning behavior. The goal of his study is to show how useful data mining can be used in higher education to improve student' performance. Galit, [8] gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams. Ayesha et al. [11], used k-means clustering algorithm as a data mining technique to predict students' learning activities in a students' database including class quizzes, mid and final exam and assignments. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students. Z. J. Kovacic [7] presented a case study on educational data mining to identify up to what extent the enrolment data can be used to predict student's success. The algorithms CHAID and CART were applied on student enrolment data of information system students of open polytechnic of New Zealand to get two decision trees classifying successful and unsuccessful students. The accuracy obtained with CHAID and CART was 59.4 and 60.5

respectively. Pandey and Pal [12] conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of association rule they find the interestingness of student in opting class teaching language. Al-Radaideh et al. [9] applied the data mining techniques, particularly classification to help in improving the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. The extracted classification rules are based on the decision tree as a classification method, the extracted classification rules are studied and evaluated. It allows students to predict the final grade in a course under study. Bray [13], in his study on private tutoring and its implications, observed that the percentage of students receiving private tutoring in India was relatively higher than in Malaysia, Singapore, Japan, China and Sri Lanka. It was also observed that there was an enhancement of academic performance with the intensity of private tutoring and this variation of intensity of private tutoring depends on the collective factor namely socioeconomic conditions.

3. DATA MINING PROCESS

The Data Mining Client is designed to walk you through the data mining process. The basic process in any data mining project is shown in Figure 1 (a).

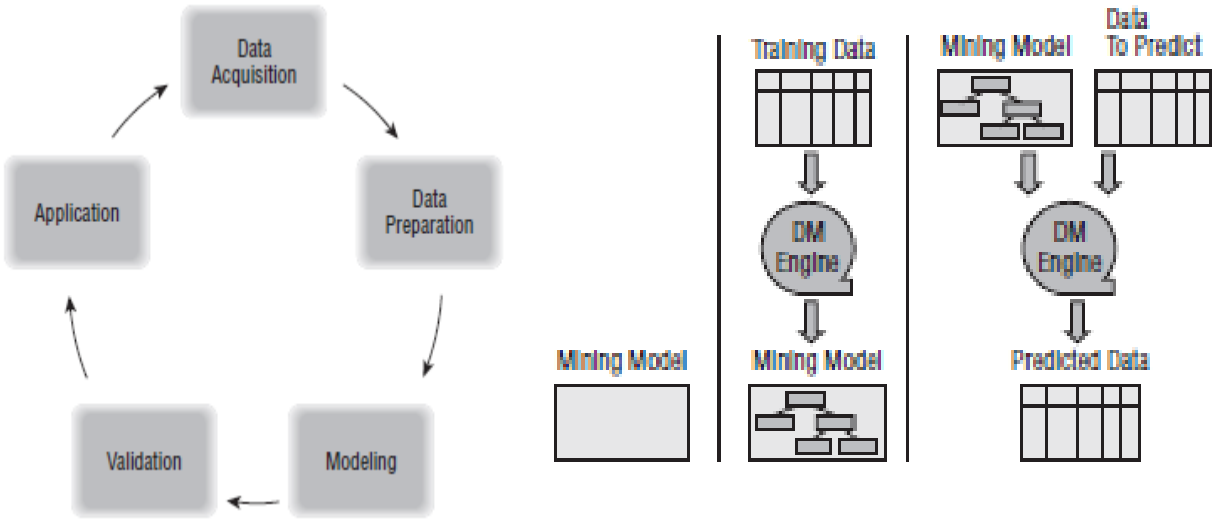


Figure 1 Data mining process for Data mining client Add-In [17]

Data acquisition tools are provided natively by Excel with the Data Mining Client. The Data Mining Client adds an additional method beyond those. Many users already have data in Excel-accessible formats, and Excel has tools for data importation. The other pieces of the data mining process are supported directly from the Data Mining Client ribbon, as shown in Figure 2. Each chunk of the ribbon indicates a step in the process.

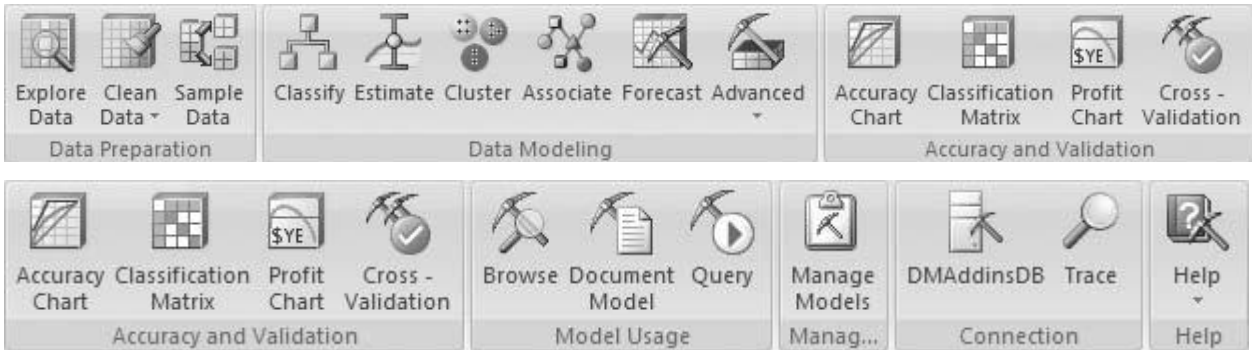


Figure 2 Data Mining Client ribbon [17]

4. APPLICATION OF DATA MINING TECHNIQUE

4.1 Application Software:-

In this study, data is collected from college students were analyzed using a data mining technique like classification and clustering. The data set used in this study was obtained from department of Master of computer Application (M.C.A.), Priyadarshini college of Engineering, in Mar-2013. SQL Server 2008 Data Mining Add-Ins for Microsoft Office 2007 is a freely downloadable package that allows you to unleash the power of SQL Server Data Mining. The Data Mining Client is available as a free download on the Microsoft website. The programming environment use for application was Excel 2007 with data mining Add-Ins for building data mining model and SQL Server 2008 is used to store the data. Data Mining Client for Excel This add-in enables advanced users to go through the full development life cycle for the data mining model within Excel by using either worksheet data or external data from SQL Server Analysis Services.

Excel is the business intelligence platform of choice for most actuaries, its statistical modeling capabilities are limited. Neural networks, classification and regression trees, and other data mining algorithms simply are not available in standard Excel installations. There is a good reason for this; most data mining algorithms require fast processors and large

amounts of memory, which are typically available only on servers. The Data Mining Client acts as a link between Excel (which is typically installed on a laptop or desktop computer) and a server running Analysis Services.

4.2 Data Preparations

The data set used in this study was obtained from Nagpur University, Nagpur (Maharashtra), India on the sampling method for MCA (Master of Computer Applications) course from session 2009 to 2012. Initially size of the data is 48. In this step data stored in different tables was joined in a single table after joining process errors and missing values were removed.

4.3 Data Selection and Transformation

In this step only those fields were selected which were required for data mining. The data values for some of the variables were defined for the present analyses which are as follows:

PSM – Preceding Semester Marks are obtained in MCA course. It is divided into five grades values: Distinction>=75, First >=60% , Second > =50% ,Third > = 40% , Fail < 40%.

ATT – Attendance of Student. Minimum 75% attendance is compulsory to participate in University examination. Students with low attendance can also give university exam on ground of genuine medical issue. Attendance is divided into different classes as :

Poor <45%, Average > = 45% and < 65%, Good > = 65% and<75, Excellent >=75 .

CAG – Class Assignment grade obtained. In every semester two class tests i,e unit test are conducted and average of two class test are used to calculate total class assignment marks. CAG is split into three classes: Poor < 30%, Average >= 30% and <= 50%, Good >=50% and <70%, Excellent>=70%

CDC – Career Development Carrier Performance are obtained. In each semester different CDC activity are planned like Group discussion, seminar, Mock interview, Aptitude test and extra curriculum activity to check the performance of students. Each student is supposed to take part for at least two extracurricular activities such as paper presentation, technical, cultural event, etc. The performance of students are evaluated according to the performance which can be categorized as 1 – Unsatisfactory, 2 – Satisfactory, 3 – Good, 4 – Exceptional

PW – Practical Work / Project Work. Practical work is divided into two classes: Yes – student completed practical list or project assigned by teacher, No – student not completed practical or project assignment work.

SSM – Semester Sessional Marks obtained. In every semester Sessional exam are conducted which is based upon complete syllabus. It is divided into five grades values: Distinction>=75, First >=60% and <75, Second > =45% and < 60%, Third > = 36% and < 45%, Fail < 36%.

4.4 Data Set

The data set used in this study was obtained from Nagpur University, Nagpur (Maharashtra), India on the sampling method for MCA (Master of Computer Applications) course from session 2009 to 2012. Initially size of the data is 48.

Name	PSM	ATT grades	CAT Grades	SSM	CDC Grades	PW
ABHIJIT GINAGULE	First	Excellent	Good	Poor	Good	YES
ABHISHEK Dinesh KUMAR	First	Excellent	Good	Average	Satisfactory	YES
ABHISHEK Shankar KUMAR	Third	Good	Good	Poor	Satisfactory	YES
ANAS M. KHAN	First	Excellent	Good	Average	Good	YES
ANUP KUMAR JHA	Second	Excellent	Poor	Poor	Satisfactory	NO
ASHVANI K. BHARADWAJ	First	Excellent	Good	Poor	Good	YES
ATUL SUROSHE	First	Good	Good	Poor	Satisfactory	YES
BHIMRAO GAYAKWAD	First	Excellent	Good	Poor	Good	YES
DINESH UDAYPURE	Third	Average	Poor	Poor	Satisfactory	NO
GHOUSIYA FARHEEN SHEIKH	First	Excellent	Good	Average	Good	YES
GITESH CHARPE	Second	Poor	Poor	Poor	Satisfactory	NO
GOPAL BADHE	Second	Excellent	Good	Poor	Good	YES
GULSHAN TRIVEDI	First	Excellent	Good	Poor	Good	YES
HARSHA NANDURKAR	First	Excellent	Good	Average	Good	YES
JAYASHRI ZADE	First	Excellent	Good	Average	Good	YES
KANCHAN KARWATKAR	Second	Good	Good	Poor	Good	YES
KRANTI SHRIKHANDE	First	Excellent	Good	Poor	Good	YES
MAHENDRA K.RATNAKAR	First	Good	Good	Poor	Satisfactory	NO
MAYUR TIRARMARE	Second	Excellent	Good	Poor	Good	YES
MAYURI NAGPURKAR	First	Excellent	Good	Average	Good	YES
MEGHA GANER	Second	Excellent	Good	Poor	Satisfactory	YES
NEELAM WASNIK	Third	Excellent	Good	Poor	Satisfactory	NO
NEHA SHINGNAPURKAR	Second	Excellent	Good	Poor	Good	YES
PIYUSH ITKHEDE	Second	Good	Good	Poor	Good	NO
PRACHITI MESHRAM	Second	Average	Average	Poor	Good	NO
PRAVIN KUMAR	Second	Good	Good	Poor	Satisfactory	NO
PRIYANKA BORKAR	Fail	Poor	Poor	Poor	Good	YES
REWATI BOREKAR	First	Excellent	Good	Poor	Good	YES

ROSHANI WASEKAR	Second	Excellent	Good	Average	Satisfactory	YES
ROSHNI BAJIRAO	First	Average	Good	Poor	Good	YES
RUSHIKESH PATTEWAR	First	Excellent	Good	Average	Satisfactory	YES
SANKET KAPSE	Second	Good	Good	Poor	Good	YES
SHAGUFTA ANJUM	Second	Excellent	Poor	Poor	Unsatisfactory	NO
SHEET KUMAR	Second	Excellent	Good	Average	Good	YES
SHILPA DAKHORE	Second	Excellent	Good	Poor	Good	YES
SHRUTIKA POTWAR	First	Excellent	Good	Average	Good	YES
SHUBHAM NEMA	Third	Poor	Average	Poor	Unsatisfactory	NO
SUJAY NERKAR	Second	Excellent	Good	Poor	Good	YES
SWAPNIL KAMBE	First	Good	Good	Poor	Good	YES
TEJAS BHAGWATKAR	First	Good	Good	Poor	Good	YES
VARUN KUMAR	First	Excellent	Good	Average	Satisfactory	YES
VINOD K. PUSHPATODE	Third	Poor	Poor	Poor	Unsatisfactory	NO
YOGITA HEDAOO	First	Excellent	Good	Poor	Satisfactory	YES

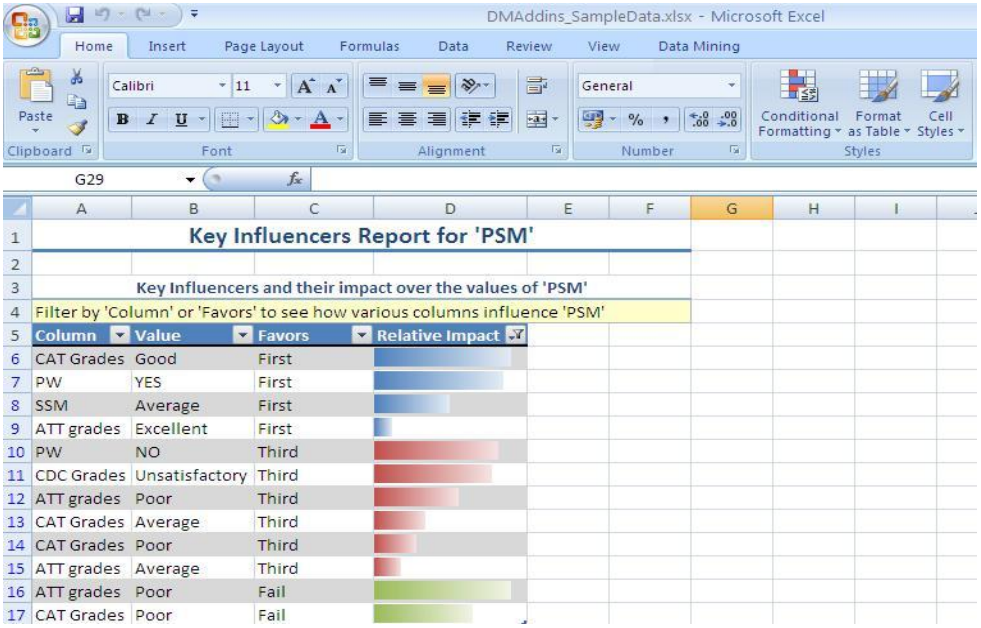
Table 1:- Sample Records of Students' Data Set

4.5 Implementation of Data Mining Model and result discussion

With the release of Excel 2007 and SQL Server 2008, it is possible to build complex statistical models directly in Excel. The visualizations offered for the Microsoft Naıve Bayes algorithm suggest a different kind of application: analyzing the key influencers for a specific target.

A. Analyze Key Influencers

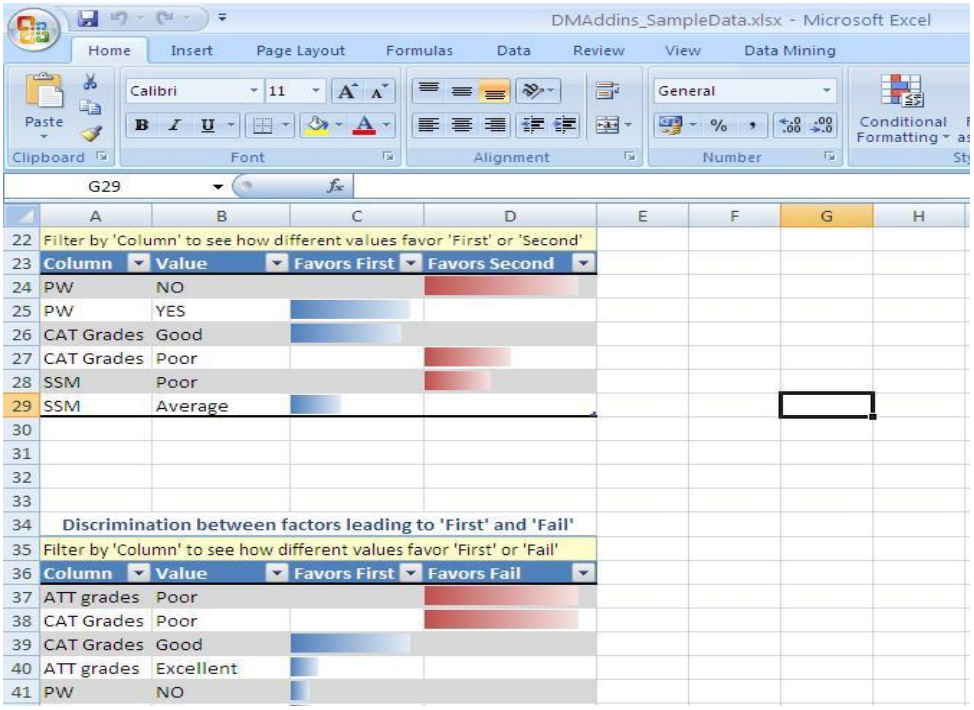
The Analyze Key Influencers tool analyzes the correlation between all columns in your table and a specified target column. The result is a report that identifies the columns having significant influence on the target and explains in detail how this influence manifests itself. The Analyze Key Influences tool will create a report as shown in figure 3, that shows how strongly PSM, affect the CDC, ATT and PW grades.



DMAddins_SampleData.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I
1	Key Influencers Report for 'PSM'								
2									
3	Key Influencers and their impact over the values of 'PSM'								
4	Filter by 'Column' or 'Favors' to see how various columns influence 'PSM'								
5	Column	Value	Favors	Relative Impact					
6	CAT Grades	Good	First						
7	PW	YES	First						
8	SSM	Average	First						
9	ATT grades	Excellent	First						
10	PW	NO	Third						
11	CDC Grades	Unsatisfactory	Third						
12	ATT grades	Poor	Third						
13	CAT Grades	Average	Third						
14	CAT Grades	Poor	Third						
15	ATT grades	Average	Third						
16	ATT grades	Poor	Fail						
17	CAT Grades	Poor	Fail						

Figure 3:- The main output report generated by Analyze Key Influencers tool for 'PSM'



DMAddins_SampleData.xlsx - Microsoft Excel

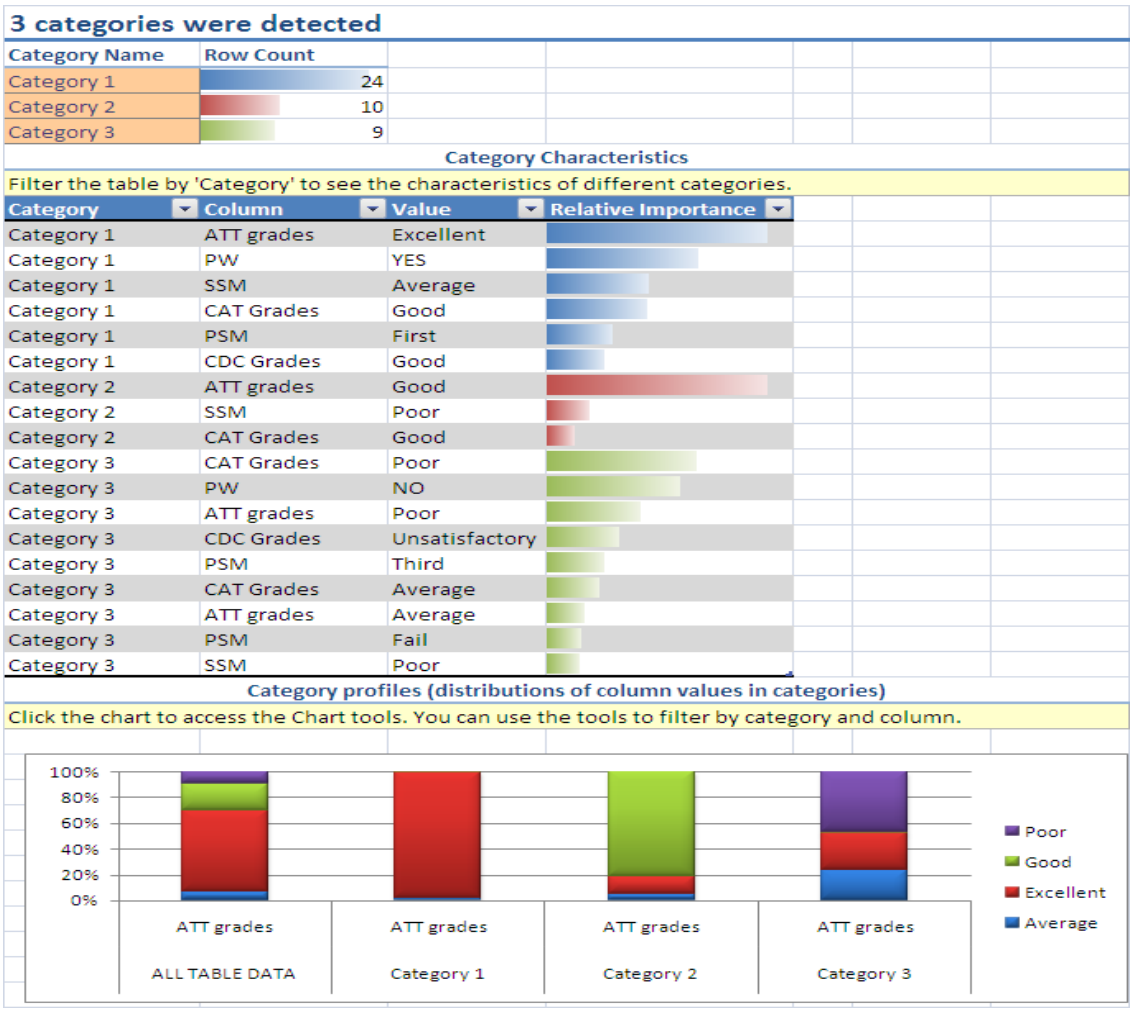
	A	B	C	D	E	F	G	H
22	Filter by 'Column' to see how different values favor 'First' or 'Second'							
23	Column	Value	Favors First	Favors Second				
24	PW	NO						
25	PW	YES						
26	CAT Grades	Good						
27	CAT Grades	Poor						
28	SSM	Poor						
29	SSM	Average						
30								
31								
32								
33								
34	Discrimination between factors leading to 'First' and 'Fail'							
35	Filter by 'Column' to see how different values favor 'First' or 'Fail'							
36	Column	Value	Favors First	Favors Fail				
37	ATT grades	Poor						
38	CAT Grades	Poor						
39	CAT Grades	Good						
40	ATT grades	Excellent						
41	PW	NO						

Figure 4:- Naïve Bayes : Attribute Discrimination of Class

From above figure 4, it is cleared that relative impact of PSM First category is more for those who have done practical/project work and scored good marks in Class assignment test.Relative impact of PSM Second Category is more for those who have not performed Practical/project work and scored Poor grades in CAT.

B. Detect Categories

The Detect Categories tool uses a clustering algorithm. It identifies rows in the table that are similar and assigns the similar rows to a category. The final number of categories will depend on the number of identifiable groups of similar rows.



3 categories were detected

Category Name	Row Count
Category 1	24
Category 2	10
Category 3	9

Category Characteristics

Filter the table by 'Category' to see the characteristics of different categories.

Category	Column	Value	Relative Importance
Category 1	ATT grades	Excellent	
Category 1	PW	YES	
Category 1	SSM	Average	
Category 1	CAT Grades	Good	
Category 1	PSM	First	
Category 1	CDC Grades	Good	
Category 2	ATT grades	Good	
Category 2	SSM	Poor	
Category 2	CAT Grades	Good	
Category 3	CAT Grades	Poor	
Category 3	PW	NO	
Category 3	ATT grades	Poor	
Category 3	CDC Grades	Unsatisfactory	
Category 3	PSM	Third	
Category 3	CAT Grades	Average	
Category 3	ATT grades	Average	
Category 3	PSM	Fail	
Category 3	SSM	Poor	

Category profiles (distributions of column values in categories)

Click the chart to access the Chart tools. You can use the tools to filter by category and column.

Figure 5:- Evaluation of categories based on ATT(Attendance variable)

From figure 5, it is cleared that according to ATT factor the data was categorized into three groups, Category 1 having excellent attendance, Category 2 students having good to average attendance and category 3 students having average to poor attendance.

C. Clustering Algorithms

Classification and estimation algorithms are two classes of supervised algorithms. The term "supervised" is used to describe data mining algorithms that model a pre-selected dependant variable. "Unsupervised" algorithms, such as clustering algorithms, look at all of the available data in order to identify patterns. All patterns, rather than just those affecting the dependant variable, are reviewed and analyzed. Clustering algorithms try to split

records into similar groups. The Data Mining Client for Excel makes it easy to interpret the groups identified by the clustering algorithm.

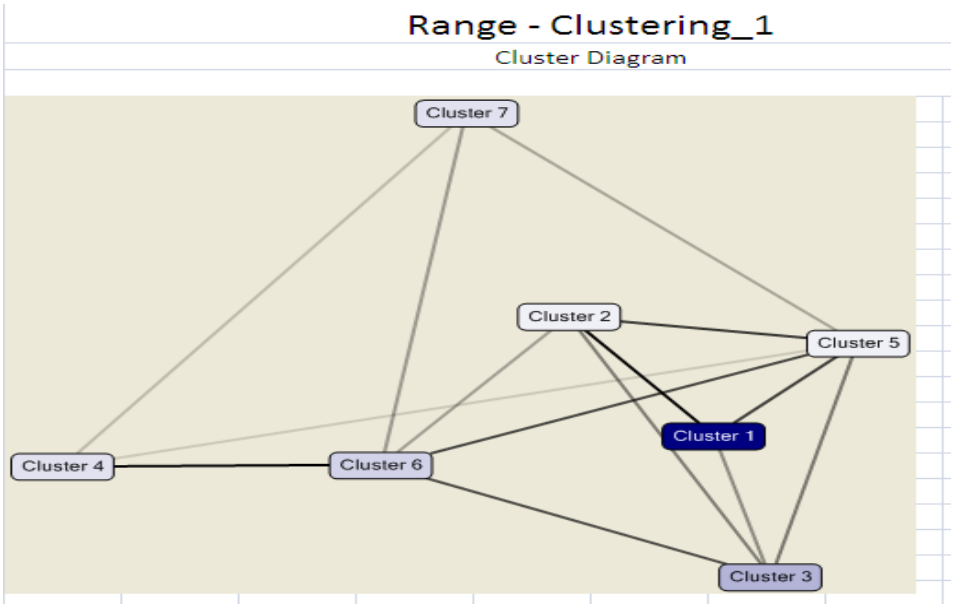


Figure 6:- Cluster Diagram

A cluster diagram allows the analyst to see the relationship between various clusters, based on different attributes. The darkest line in Figure 6 shows that clusters 3, 4, 6 and 7 are most similar, while the various shades of color show how the cluster are related with respect to the "population" variable.

Range - Clustering_1
Cluster Profiles

Variables	States	Population (All)	Cluster 1	Cluster 3	Cluster 6	Cluster 7	Cluster 4	Cluster 2	Cluster 5
Size		31	17	5	3	2	2	1	1
ATT grades	Excellent	10	82 %	54 %	29 %	1 %	15 %	69 %	44 %
ATT grades	Good	5	12 %	46 %	44 %	9 %	28 %	11 %	37 %
ATT grades	Poor	3	0 %	0 %	13 %	52 %	30 %	14 %	4 %
ATT grades	Average	3	6 %	0 %	14 %	38 %	28 %	5 %	15 %
CAT Grades	Good	15	100 %	100 %	73 %	10 %	43 %	86 %	84 %
CAT Grades	Poor	3	0 %	0 %	26 %	1 %	57 %	14 %	4 %
CAT Grades	Average	2	0 %	0 %	1 %	89 %	0 %	0 %	12 %
CDC Grades	Good	12	100 %	2 %	9 %	45 %	0 %	95 %	97 %
CDC Grades	Satisfactory	6	0 %	98 %	79 %	3 %	70 %	5 %	2 %
CDC Grades	Unsatisfactory	2	0 %	0 %	12 %	52 %	30 %	0 %	1 %
PW	YES	12	100 %	77 %	18 %	0 %	2 %	100 %	56 %
PW	NO	8	0 %	23 %	82 %	100 %	98 %	0 %	44 %
SSM	Poor	16	71 %	70 %	93 %	100 %	99 %	74 %	84 %
SSM	Average	4	29 %	30 %	7 %	0 %	1 %	26 %	16 %

Figure 7:- Cluster Profiles

The figure 7 shows "cluster profiles", which are essentially a univariate statistical analysis of each of the variables, for both the overall population and each cluster individually. This view makes it possible to identify the differences between the clusters.

Range - Clustering_1
Cluster Characteristics
Population (All)

Variables	Values	Probability
SSM	Poor	81 %
CAT Grades	Good	77 %
PW	YES	62 %
CDC Grades	Good	58 %
ATT grades	Excellent	50 %
PW	NO	38 %
CDC Grades	Satisfactory	32 %
ATT grades	Good	24 %
SSM	Average	19 %
ATT grades	Poor	13 %
CAT Grades	Poor	13 %
ATT grades	Average	13 %
CDC Grades	Unsatisfactory	10 %
CAT Grades	Average	10 %

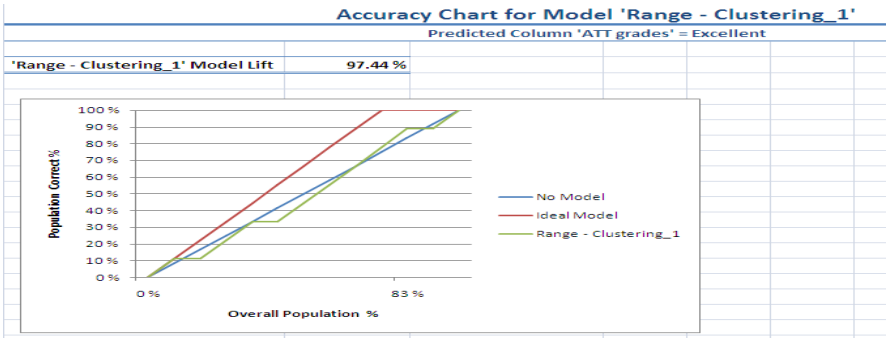
Figure 8- Cluster Characteristics

The figure 8 shows "cluster characteristics" makes it possible for an analyst to understand the nature of a given cluster. Characteristics are values of a given variable that help distinguish one cluster from another.

D. Accuracy and Validation

The Accuracy and Validation section of the data mining toolbar allow analysts to evaluate the quality of a data mining model. Models can be evaluated using accuracy charts, a classification matrix, profit charts or the cross-validation method. Lift chart, this graph shows how a chosen model compares to a perfect model and a model based on random guessing

Accuracy Chart – Evaluates the performance of the model against test data by drawing a lift chart for classification models and a scatter plot for estimation models.



Percentile	Ideal Model	Range - Clustering_1
0 %	0.00 %	0.00 %
8 %	11.11 %	11.11 %
17 %	22.22 %	11.11 %
25 %	33.33 %	22.22 %
33 %	44.44 %	33.33 %
42 %	55.56 %	33.33 %
50 %	66.67 %	44.44 %
58 %	77.78 %	55.56 %
67 %	88.89 %	66.67 %
75 %	100.00 %	77.78 %
83 %	100.00 %	88.89 %
92 %	100.00 %	88.89 %

Figure 9:- Accuracy chart of Clustering model Attendance class

The figure 9 shows how accurate the data mining model is; in this example, the 50% of students whom the model selects as most likely to have ATT grades='Excellent', 44.44% of the total students who will actually have excellent attendance grades.

E. Classification Matrix

Displays a matrix of correct and incorrect classifications by evaluating your model against test data. To create Classification Matrix, click the Classification Matrix button in the Accuracy and Validation Section of the Data Mining ribbon.

Counts of correct/incorrect classification for model 'Range - Clustering_1'

Predicted Column 'ATT grades'
Columns correspond to actual values
Rows correspond to predicted values

Model name:	Range - Clustering_1	Range - Clustering_1		
Total correct:	58.33 %	7		
Total misclassified:	41.67 %	5		

Results as Percentages for Model 'Range - Clustering_1'

	Average(Actual)	Excellent(Actual)	Good(Actual)	Poor(Actual)
Average	**0.00 %**	0.00 %	0.00 %	0.00 %
Excellent	0.00 %	**77.78 %**	100.00 %	0.00 %
Good	0.00 %	22.22 %	**0.00 %**	100.00 %
Poor	0.00 %	0.00 %	0.00 %	**0.00 %**
Correct	**0.00 %**	**77.78 %**	**0.00 %**	**0.00 %**
Misclassified	**0.00 %**	**22.22 %**	**100.00 %**	**100.00 %**

Results as Counts for Model 'Range - Clustering_1'

	Average(Actual)	Excellent(Actual)	Good(Actual)	Poor(Actual)
Average	**0**	0	0	0
Excellent	0	**7**	2	0
Good	0	2	**0**	1
Poor	0	0	0	**0**
Correct	**0**	**7**	**0**	**0**
Misclassified	**0**	**2**	**2**	**1**

Figure 10:- Correct and misclassified Classification of Attendance category

5. CONCLUSION AND FUTURE WORK

In this paper, we gave a case study in the educational data mining. It showed how useful data mining can be used in higher education particularly to improve students' performance. We used post graduate students data collected from the department of MCA, Priyadarshini college Engineering, Nagpur. This study will help the students and the teachers to improve the performance of the students. This study is also helpful for those students who need special attention and will also lower failure ratio by taking proper action for the next

semester examination. The information generated after the implementation of data mining technique may be helpful for a teacher as well as for students.

FUTURE WORK:

Our future work include applying data mining techniques on an expanded data set with more distinct attributes to get more accurate results. The future work can be done using more data mining techniques such as neural nets, genetic algorithms, k-mean, and others data mining model. Some different software's may be exploited using various factors to refine our technique in order to get more accurate outputs.

REFERENCES

[1] Heikki, Mannila, "Data mining: machine learning, statistics, and databases", IEEE, 1996.

[2] U. Fayadd, Piatesky, G. Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases", AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0–262 56097–6, 1996.

[3] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.

[4] El-Halees, A. (2008) 'Mining Students Data to Analyze Learning Behavior: A Case Study', The 2008 international Arab Conference of Information Technology (ACIT2008) –Conference Proceedings, University of Sfax, Tunisia, Dec 15- 18.

[5] 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.

[6] Z. N. Khan, "Scholastic achievement of higher secondary students in science stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005..

[7] Z. J. Kovacic, "Early prediction of student success: Mining student enrollment data", Proceedings of Informing Science & IT Education Conference 2010

[8] 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.

[9] 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.

[10] B.K. Bharadwaj and S. Pal. "Data Mining: A prediction for performance improvement using classification", InternationalJournal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011

[11] Ayesha, S., Mustafa, T. , Sattar, A. and Khan, I. (2010) 'Data Mining Model for Higher Education System', European Journal of Scientific Research, vol. 43, no. 1, pp. 24-29.

[12] 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.

[13] M. Bray, "The shadow education system: private tutoring and its implications for planners", (2nd ed.), UNESCO, PARIS, France, 2007.

[14] An Excel 2007 trial edition is available at: http://www.microsoftstore.com/store/msstore/cat/categoryID.61301000

[15] The SQL Server 2008 version of the data mining client for Excel 2007 is available at: http://www.microsoft.com/en-us/download/details.aspx?id=8569

[16] A trial edition of SQL Server 2008 is available at: http://www.microsoft.com/en-in/download/details.aspx?id=1842

[17] J. MacLennan, Z. Tang, B. Crivat "Data mining with Microsoft SQL Server 2008", 2009.

[18] Spyridon Ganas, "Data Mining and Predictive Modeling with Excel 2007", Casualty Actuarial Society *Forum*, Winter 2009