A Novel Approach to Analyze Students' Expectation from Colleges using Data Mining Technique

D. Radha
M.O.P. Vaishnav College for Women
Chennai

R. Jayaparvathy
SSN College Of Engineering
Chennai

A. Jyothirmayi Bhargavi
M.O.P. Vaishnav College For Women
Chennai

ABSTRACT
College being home away from home for students, expecting multiple facilities is quite demanding. The following paper aims at building an intelligent Mathematical model to categorize and identify the amenities expected by students from a college using a data mining technique called as classification. Data mining, a vital step in 'Knowledge Discovery in Database' is a practice of transforming raw data into constructive information from large data repositories. A data mining tool is employed to analyze the dataset. The knowledge gained thus helps institutions to deploy the facilities that students expect.

Keywords
Data mining, KDD, model, classification.

1. INTRODUCTION
Institutions to gain popularity among their prospective students and to establish their own success must necessarily understand the needs of students. Though certain basic wants of students are paid attention to, many other facilities wanted by the students goes unnoticed. The introduction of such wanted facilities will catalyze the institution's growth. Thus, understanding the students' wants plays a vital role.

Paper concentrates to bridge the gap between the students' needs and institutions assumptions.

Data mining is said to be a technology that blends traditional data analysis methods with sophisticated algorithms for processing large volumes of data. The objective of data mining is to discover useful and interesting patterns that go unnoticed. Data mining is the process of inspecting large pre-existing datasets to produce useful information. It is a technology that automatically extracts knowledge. The knowledge is used to take strategic decisions. There are basically two types of data mining tasks. They are predictive analysis and descriptive analysis.

Predictive analysis is a task that predicts the value of a dependent attribute (target) based on the value of independent attribute. Descriptive tasks derive patterns that summarize the underlying relationship in data. The data mining technique used to analyze the data is classification. Classification is a task where a model or classifier is constructed to predict class labels. It is a two-step process consisting of learning and a classification step. The learning step focuses on constructing a classification model and classification step is used to predict class labels for given data. A decision tree is a flowchart-like structure which consists of an internal node, branch, and leaf node. The internal node denotes a test on an attribute, branch represents an outcome of the test and leaf node holds a class label.

2. RELATED WORK
Educational data mining (EDM) is an emerging research field with a concern to better understand students, to predict student performance, to detect undesirable student behavior, Student modeling, to identify students' pattern etc.

Mining Of Students' Satisfaction Their College In Theni – This paper by s.Roobini and R.Uma studied the changes in student expectations and institutional performance. There analysis found that the PG students are highly satisfied with teaching Method, attitude, library and also practical training than UG students.

Data Mining Applications in Higher Education – Jing Luan, proposed solutions for challenges such as "What do institutions know about their students?", How to accurately predict academic outcomes in order to facilitate timely academic intervention ? How data mining helps universities focus on the alumni most likely to make pledges?

Mining Educational Data to Analyze Students" Performance - The decision tree method, a classification task was used by Brijesh Kumar Baradwaj, Saurabh Pal to justify the capabilities of data mining techniques in context of higher education by offering a data mining model for higher education system in the university. The paper proposed several techniques to identify those students which needed special attention to reduce fail ration and taking appropriate action for the next semester examination. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester.

Mohammed M. Abu Tair, Alaa M. El-Halees showed how data mining can be used in higher education particularly to improve graduate students' performance.

Mining Students Data to Analyze Learning Behavior: Alaa El-Halees analyzed learning behavior. In his case study, he collected students' data from Database course. The data collected was preprocessed. Association, classification, and clustering and outlier detection rules were discovered using data mining techniques. In each of the four tasks, Alaa extracted knowledge that describes students' behavior.

Most of the research papers in the field of educational data mining concentrate on analyzing the students but not on understanding the students' needs. This paper focuses on that perspective.

3. SYSTEM MODEL
The classification technique is used to understand the students' expectation. The data was collected from students.

In this work, the data set is organized in the form a table consisting of rows and columns where rows correspond to the values for each attribute of an object and columns represents the attributes. The attributes considered are Infrastructure, Wi-Fi, Placement Cell, Canteen, Teaching Skills, Extracurricular Activities, Library, Parking, Hostel, Location, Photocopy Shop, and Stationery Shop. The target variable is the **selection** attribute which determines whether a student will join the college for the facilities available. The independent attributes are binary attributes that take the values 0 or 1.

If the value of an attribute is 1 for a particular object, it denotes that the student expects that facility from the institution. If the value is 0, it indicates that the student considers it as a trivial facility.

Each attribute in this work is assigned a weight. The value of each attribute of an object is multiplied with its corresponding weight assigned and are summed up and assigned to the total attribute t.

$$t = \sum_{i=1}^{n} V_i * W_i$$

t represents the total, n refers to the number of attributes, V_i refers to the value of the attribute and W_i is the weight assigned to the attribute.

The threshold value for the model is fixed and the records whose **t** value is less than the threshold indicate that the institution fails to meet the students' expectation.

4. PERFORMANCE ANALYSIS AND RESULTS

As explained in the system model, the attributes are infrastructure, Wi-Fi, Placement Cell, canteen, teaching skills, extracurricular activities, library, parking, hostel, location, photocopy shop, Stationery Shop.

The abbreviations are used in this technique for the attributes. The abbreviations used are listed below.

Table 1. Abbreviations

INF	Infrastructure
WF	Wi-Fi
PC	Placement Cell
CAN	Canteen
TS	Teaching Skills
EXC	Extra-Curricular Activities
LIB	Library
PAR	Parking
HOS	Hostel
LOC	Location
PS	Photocopy Shop
SS	Stationery Shop

The weights attached for each value of the attribute are given in the following table.

Table 2. Weights for corresponding attributes

Attributes	Weights
Infrastructure	0.5
Wi-Fi	0.25
Placement Cell	1
Canteen	0.75
Teaching Skills	1
Extracurricular activities	.75
Library	1
Parking	0.25
Hostel	0.25
Location	1
Photocopy Shop	0.5
Stationery Shop	0.5

Applying the assigned value to the collected data objects, the following table is obtained. The table below is a subset of the data set.

Table 3. Subset of the dataset

INF	WF	PC	C	TS	EXC	LIB	PAR	HOS	LOC	PS	SS	t	selection
1	1	0	1	1	1	1	0	1	1	1	1	6.5	yes
0	0	1	0	0	0	0	1	0	1	1	1	3.25	no
0	0	1	0	1	1	1	0	1	0	1	1	5	no
1	1	0	1	0	0	0	1	0	1	1	1	3.75	no
1	1	1	1	1	0	1	1	0	1	1	1	6.75	yes
0	0	0	0	0	1	0	0	1	0	1	0	1.5	no
1	0	1	1	1	0	1	1	0	0	1	0	5	yes
0	1	0	0	0	1	0	0	1	1	1	0	2.75	no
1	0	1	1	1	1	1	0	1	1	1	1	7.25	yes
0	1	0	0	0	0	0	1	0	1	1	0	2	no

The decision tree for the data set is given below generated by the most efficient data mining tool, RapidMiner.
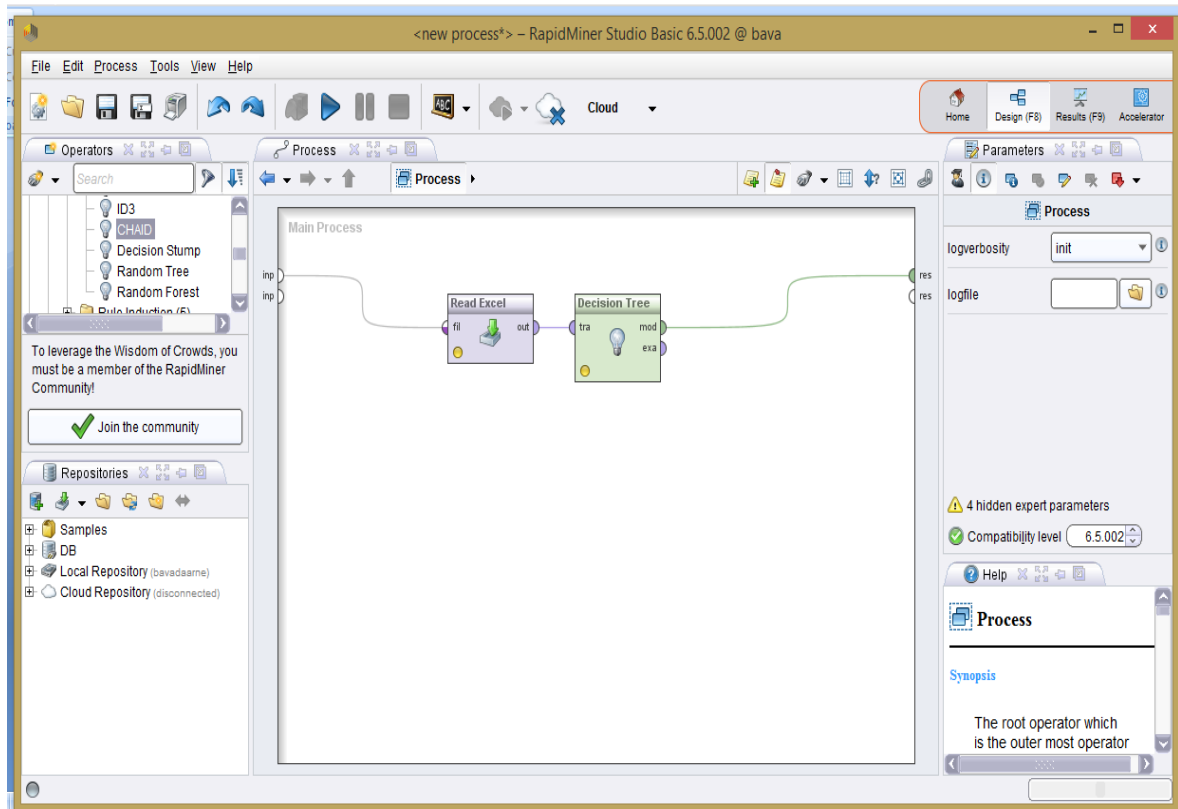
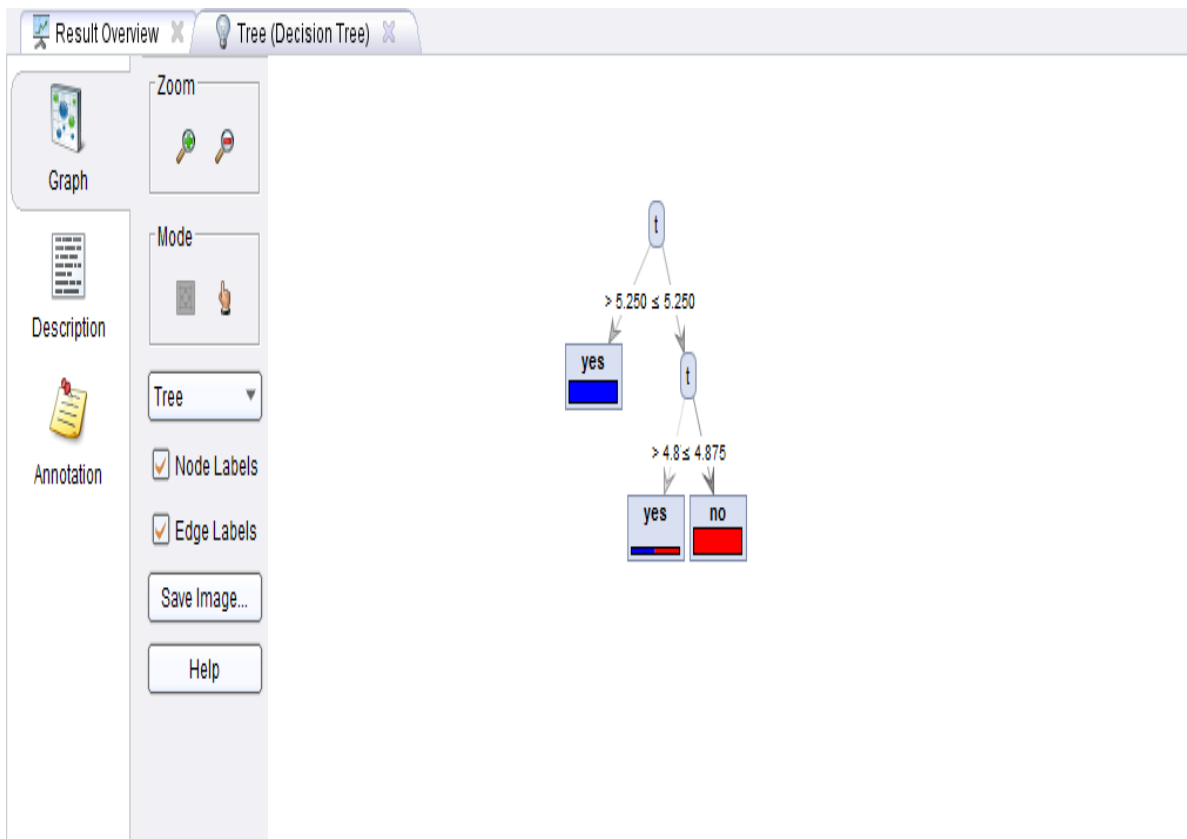**Figure 1 Decision tree creation process**

Figure 2 Decision Tree

Row No.	selection	prediction(s...	confidence(...	confidence(...	INF	WF	PC	C	TS	EXC	LIB	PAR
1	?	no	0	1	0	1	1	1	1	0	0	0
2	?	yes	1	0	1	1	1	1	1	1	0	0
3	?	yes	1	0	1	0	1	0	1	0	1	1
4	?	no	0	1	1	1	1	1	0	0	0	0
5	?	no	0	1	0	0	0	0	0	0	1	0
6	?	no	0	1	0	1	1	0	0	1	0	0
7	?	yes	0.500	0.500	1	1	0	0	0	1	1	1
8	?	yes	1	0	1	1	1	1	1	1	0	0
9	?	yes	0.500	0.500	1	1	1	1	1	1	1	0
10	?	yes	1	0	1	1	1	1	1	1	1	1
11	?	yes	1	0	1	0	0	1	1	0	1	1
12	?	yes	1	0	1	1	1	1	1	1	1	1
13	?	yes	0.500	0.500	1	0	0	0	1	1	1	1
14	?	no	0	1	1	1	0	1	1	0	0	0
15	?	no	0	1	0	1	1	0	0	1	0	1
16	?	no	0	1	1	1	1	0	0	1	0	1
17	?	yes	0.500	0.500	1	0	1	1	0	0	1	0
18	?	yes	1	0	1	1	1	1	1	1	1	1
19	?	yes	1	0	0	0	1	1	1	1	1	1
20	?	yes	1	0	1	1	1	1	1	1	1	1
21	?	no	0	1	0	0	0	0	0	0	0	0

Figure 3 predicting the values of label attribute for new records (Using RapidMiner)

If the value of t is a greater value then the students' requirements are met. The above model is evaluated for 1000 students. Accuracy of the above proposed model was found to be 95%

5. CONCLUSION

This research work proves that it is mandatory to understand students' to provide a better equipped educational environment with several facilities required by the students. The efficiency of the proposed technique is observed to be 95%. The proposed model considered a specific set of facilities. For future work heterogeneous type of facilities will be taken into consideration. The facilities will be prioritized using a data mining technique to be aware of facilities that top the charts which can aid the institutions to deploy the most expected amenities.

6. REFERENCES

[1] S.Roobini, R.Uma. "MINING OF STUDENTS' SATISFACTION THEIR COLLEGE IN THENI". International Journal of Advanced Technology in Engineering and Science Volume No 03, Special Issue No. 01, April 2015

[2] D.Radha, R.Jayaparvathy. "Content filtering and management using cluster based enrichment of keywords for mobile devices". Advances In Info Sciences,Service Sciences(AISS). Volume 7,number2,April 2015

[3] D.Radha, R.Jayaparvathy. "Enhancing Quality and efficiency of service in network communication using rough set theory." Research India Publications. Volume 10,number 10,2015

[4] Jing Luan. "Data Mining Applications in Higher Education"

[5] Brijesh Kumar Baradwaj , Saurabh Pal ."Mining Educational Data to Analyze Students Performance". (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011

[6] Mohammed M. Abu Tair, Alaa M. El-Halees. " Mining Educational Data to Improve Students' Performance: A Case Study" Volume 2 No. 2, February 2012 International Journal of Information and Communication Technology Research

[7] ALAA EL-HALEES. "MINING STUDENTS DATA TO ANALYZE LEARNING BEHAVIOR: A CASE STUDY"

[8] Manpreet Singh Bhullar Amritpal Kaur Use of Data Mining in EducatProceedings of the World Congress on Engineering and Computer Science 2012 Vol I WCECS 2012, October 24-26, 2012, San Francisco, USAion Sector