

A Parent-Student Perspective View in Selecting the College in Suburban Area

C. Victoria Priscilla, PhD
Associate Professor & Head
Department of M.Sc (CS)
SDNB Vaishnav College for Women
Chromepet

M. Mahadevi
Assistant Professor
Department of M.Sc (CS)
SDNB Vaishnav College for Women
Chromepet

ABSTRACT

Nowadays, the choice of colleges for higher studies plays an important role in decision making for parents and students. The views of parents and students differ because of updated technology and societal interferences amongst them. Also the choice differs based on the social life led by them. This study analyses the factors that influence the admissions in colleges in suburban areas. The study projects the view of the parents and students in selecting the college in suburban areas. A real time data is surveyed separately from parents and students. Separate statistical inferences from the responses are tabulated. A comparative study is made between the responses received from parents and students using the statistical descriptive frequencies. A decision tree is drawn which helps to identify the factors that commonly influence both parents and students by using data mining tool.

Keywords

Survey, suburban, statistical descriptive frequencies, decision tree, data mining

1. INTRODUCTION

With prevalent online courses for higher education and ease of getting educated the choice for collegiate studies among the students has more prerequisites. The preference to college admissions is influenced by factors that are comfortable for a student to study in that college. That too, student's preference of college in a suburban area poses more prerequisite from the student. To analyse such factors, this study uses a survey research method.

Among several research method, the survey research method [1] the researcher selects a sample of respondents from a population and administers a standardized questionnaire to them. The questionnaire, or survey, can be a written document that is completed by the person being surveyed, an online questionnaire, a face-to-face interview, or a telephone interview. To interpret the results of the survey method several statistical methods and data mining techniques can be used. Statistical methods like frequency analysis, correlation and regression are used to find the relationship between the factors. Data mining is the process of analysing data from different perspectives and summarizing it into useful information. Data mining methods has its foundation in statistics and it helps in more extraction of information by analysing the data. Data mining also involves other process such as data cleaning, data integration, data transformation, pattern evaluation and data presentation. Technically, data mining is used to find correlations or patterns among the dozens of fields in large dataset. Data mining technique is used to develop prediction models which help to identify the factors that affect the problem domain.

Many researchers have used different data mining methods to extract information and evaluate their dataset. Amirah Mohamed Shahiri et.al [3] have used decision tree, neural network and Naïve Bayes prediction models to evaluate student's performance for a specific locality. Dubey Pushkar et.al [4] have used factor analysis to evaluate affecting choice of engineering college in Odisha. J.F.Superby et.al [5] have used Decision tree, Random forest, neural networks analysis to evaluate the achievement of first year university student. Saurabh pal et.al [6] have used classification methods. Eden Osmanbeoric, Mirza suljic[7] have used JEL and L86 classification to evaluate prediction of student performance. Jai Ruby et.al [8] have predicted the performance of students in higher education using data mining algorithms. Umi Jensen [9] had done student retention in higher education. Anal Acharya et.al [10] has used machine learning techniques to predict early the student's performance.

The problem that is focused in this research work is to study the factors that influence the admission in suburban colleges, the objective of the study is data collection method, converting the responses into dataset and to analyse the data. To choose the college for higher studies there are different perspective view by both parents and students. The work projects the view of parent and student in selecting the college. In this study survey of data is collected using questionnaire method. The factors that are related to the questions are considered as attributes of the dataset. Two methods are used to analyse the data namely statistical descriptive frequencies method and decision tree method which is a data mining technique.

2. DESCRIPTION OF THE RESEARCH WORK

The research work consists of three stages namely

1. Data collection
2. Creation of dataset
3. Analysis and interpretation.

2.1 Data collection

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions and evaluate outcomes.

This study uses questionnaire method to collect information about the interest of parents and students while selecting the college. A set of questions reasoning to different factors are given to the parents and students of suburban areas to know their choice in selecting college. The questions were related to factors like location of the college, favourite course, self-

finance or government, infrastructure, look of college, goodwill, relationship, placement, alumni, gender, sports, hostel, extension of studies career development, safety, surrounding of campus.

The questions were circulated for both parents and students separately in and around kancheepuram district and the answers were recorded. Around 400 records were collected from both parents and students.

The lists of questions related to factors with description and option that can be selected are given in Table 1.

Table 1 Survey questions with descriptions

Question	Factors related to questions
Preferring the college based on location	Location
Preferring the college based on government or self-finance	Type
Preferring the college based on favourite course	Favourite course
Preferring the college based on infrastructure	Infrastructure
Preferring the college based on look of college	Look of college
Preferring the college based on goodwill	Goodwill
Preferring the college based on alumni	Alumni
Preferring the college based on placement	Placement
Preferring the college based on gender	Gender
Preferring the college based on sports	Sports
Preferring the college based on hostel	Hostel
Preferring the college based on extension of studies	Extension of studies
Preferring the college based on career development	Career development
Preferring the college based on safety	Safety
Preferring the college based on surrounding of campus	Surrounding of campus

2.2 Creation of dataset

In this stage, the collected information from the questionnaire is converted as dataset. The dataset is created in the form of table where the factors are considered as variables or attributes or columns. The response is tabulated as another column or attribute. Each of the survey questions has answers as 'Yes' and 'No' options. The 'Yes' option takes a quantitative value as '1' and 'No' option takes a quantitative value as '0'. The recorded responses were converted in the

form of dataset named parent dataset for parents and student dataset for students.

2.3 Analysis and interpretation

The datasets were investigated using two methods namely

1. Descriptive frequencies method
2. Decision tree method.

2.3.1 Descriptive frequencies method

In this method both the student data and the parent data is analysed separately. To find the best factors, the collected data were analysed using SPSS tool. For analysis, the frequency of each option in the variable is calculated and the option with higher percentage is selected as the best response for that variable. In similar manner the best responses of all the variables are separately tabulated for student dataset and parent dataset and are given in table 2 and table 3 respectively. In table 2 and 3 the percentage column indicates the percentage of response for the particular option.

Table 2 Descriptive frequencies for student data

Variable	Option	Percentage
1	0	32.1
	1	67.9
2	0	3.2
	1	96.2
3	0	15.4
	1	84.6
4	0	35.9
	1	64.1
5	0	43.6
	1	56.4
6	0	5.1
	1	94.9
7	0	89.7
	1	10.3
8	0	3.8
	1	96.2
9	0	14.1
	1	85.9
10	0	62.8
	1	37.2
11	0	65.4
	1	34.6
12	0	53.8
	1	46.2
13	0	28.2
	1	71.8
14	0	7.7
	1	92.3
15	0	16.7
	1	83.3

Table 3 Descriptive statistics for parent data

Variable	Option	Percentage
1	0	12.8
	1	87.2
2	0	69.2
	1	30.8
3	0	24.4
	1	75.6
4	0	3.8
	1	96.2
5	0	23.1
	1	76.9
6	0	1.3
	1	98.7
7	0	38.5
	1	61.5
8	0	9.0
	1	91.0
9	0	21.8
	1	78.2
10	0	26.9
	1	73.1
11	0	33.3
	1	66.7
12	0	20.5
	1	79.5
13	0	14.1
	1	85.9
14	0	3.8
	1	96.2
15	0	9.0
	1	91.0

In both tables the top five ranking options are selected as the best factors that influence the admissions. According to this ranking, the factor that influences in student view is given in the table 4 and parent view is given in the table 5.

Table 4 List of best factors in students view

Factors	Percentage
Placement	96.2%
Good will	94.9%
Safety	92.3%
Relationship	89.7%
Gender	85.9%

Table 5 List of best factors in parents view

Factors	Percentage
Good will	98.7%
Safety	96.2%

Infrastructure	96.2%
Placement	91%
Surrounding	91%

Tables 4 and 5 show that the factors safety, Good will and placement are common in both the parents and students view. To analyse more about the common factors, decision tree method is used.

2.3.2 Decision tree method

A decision tree is a structure that includes [2] a root node, branches, and leaf nodes. Each internal node denotes a test on attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

The parent dataset and student dataset are combined together as a single dataset. In this dataset one more column named as label is added as to indicate whether the record belongs to parents group or students group. The value 1 in the “label” attribute indicates that the record belongs to students group. The value 2 in the label attribute indicates that the record belongs to parents group. Thus dataset contains 400 records with 16 attributes. The dataset is analyzed using weka tool and decision tree is drawn for this dataset. Decision tree is given in Figure 1.

Figure1 shows that the first selected attribute is type (options are whether the selection is based on Government or self-finance college) which contains more information and it is the selected as first split criteria, it gets splits into two, yes (1) or no (0). The leaf node associated with the first split criteria is the label: 2 which mean the record belongs to parents group. Similarly the leaf nodes determine the classifier label. The set of rules that are interpreted from the decision tree are

1. If type = 0 (Option selected -preference is Government) then the view is parent perspective
2. If type =1 and alumni = 0 (Option selected is Self finance and alumni influence is “No”) then the view is student perspective.
3. If type =1, alumni = 1, Infrastructure = 0 then the view is student perspective.
4. If type =1, alumni = 1, Infrastructure =1, hostel = 1 then the view is parent perspective.
5. If type =1, alumni = 1, Infrastructure =1, hostel =0, sports =1 then the view is parent perspective.
6. If type =1, alumni = 1, Infrastructure =1, hostel =0, sports =0, look of college = 0 then the view is parent perspective.
7. If type =1, alumni = 1, Infrastructure =1, hostel =0, sports =0, look of college = 1 then the view is student perspective.

The problem of this work is to find the factors that influence the admission.

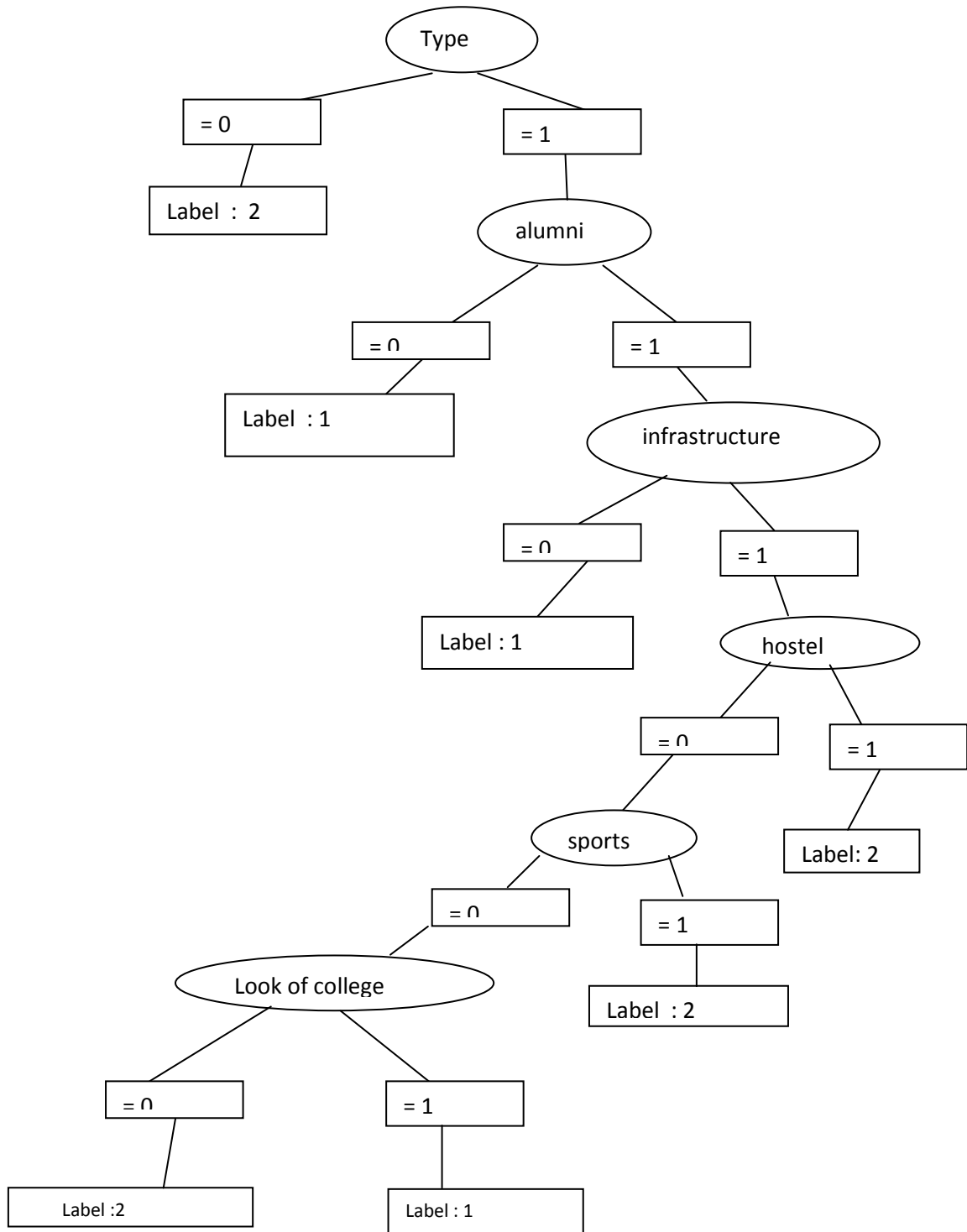


Figure 1 Decision Tree for the Combined Dataset

It is clear from the figure that the common factors that influence the admissions in suburban areas are type, alumni, infrastructure, hostel, sports, and look of college.

Table 6 Common factors in both parents and students view

Best factors of students	Best factors of parents	Common factors of both parents and students view

placement	Good will	type
Good will	Safety	Alumni
Safety	Infrastructure	Infrastructure
Relationship	Placement	hostel
Gender	surrounding	Sports and look of college

Table 6 gives the comparison between the student view of influencing factors, parent view of influencing factors and common influencing factors.

3. CONCLUSION & FUTURE WORK

This research work used questionnaire method for collection of data and created dataset for student responses and parent responses. Using statistical frequencies method and decision tree, the work concludes that the common factors that influence the college admission in sub urban areas are type (Government or self-finance), Alumni influence, hostel facility influence, sports quota and look of college. The work can be extended by collecting real data from the city area and comparing it with the suburban area.

4. REFERENCES

- [1] writing.colostate.edu/guides/research/survey/
- [2] www.tutorialspoint.com/data_mining/dm_dti.htm
- [3] Amirah Mohamed Shahiri, Wahidah Husain Nuraini Abdul Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques", Science Direct Procedia Computer Science, 2015.
- [4] Dubey Pushkar, Sharma Sudhir Kumar, Surenthiran N, "Factors affecting choice of engineering colleges in odisha" ,Research Journal of Management Sciences, April 2013.
- [5] J.F.Superby, J.P.Vandamme, N. Meskens, "Determination of factors influencing the achievement of the first-year university students using data mining methods", Workshop on Educational Data Mining, 2006
- [6] Saurabh pal, Brijesh Kumar Bhardwaj, "A prediction for performance improvement using classification", International Journal of Computer Science and Information Security, April 2011.
- [7] Edin Osmanbeoric, Mirza suljic, "Datamining approach for predicting student performance", Research Gate, May 2012,
- [8] Jai Ruby, Dr David "Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms - A Case Study", International Journal for Research in Applied Science & Engineering Technology, November 2014.
- [9] Umi Jensen "Factors influencing student retention in higher education",Research Gate, February 2011.
- [10] Anal Acharya, Devadatta Sinha, "Early Prediction of Students Performance using Machine Learning Techniques", International Journal of Computer Applications, December 2014.