Decision Support System for Admission in Engineering Colleges based on Entrance Exam Marks

Miren Tanna
B.E. Student
Dept. of Computer Engineering
Thakur College of Engineering
&Technology, Mumbai, India

ABSTRACT
Making a wise career decision is very important for everyone. In recent years, decision support tools and mechanisms have assisted us in making the right career decisions. This paper attempts to enable a student who wishes to pursue Engineering, make up good decisions, using the help of a Decision Support System. Last 3 years' information has been obtained from the website of Directorate of Technical Education, India (DTE) which makes it freely available. Using Decision Rules, results are computed from which a student can choose which stream and college he/she can opt for on the basis of Entrance Exam marks he/she has scored. To make the results more relevant, a search in the already created decision system is performed. A student has to enter his/her Entrance Exam scores and the stream he/she wishes to opt for. Based on the entered information, the decision system will return colleges and streams categorized as Ambitious, Best Bargain and Safe.

General Terms
Data Mining, Decision Support System.

Keywords
Prediction, Result prediction.

1. INTRODUCTION
Universities possess large amount of demographic data about the students and colleges. This data is present without any form of analysis. Informal analysis requires one to read through each line of the data. Such a form of analysis is not economical.
Studies have been conducted in similar area such as understanding student data [1]. There they apply and evaluate a decision tree algorithm to university records, producing graphs that are useful both for predicting graduation, and finding factors that lead to graduation. Another study has been conducted which uses student data to predict which branch a student has high chances of being placed into [2]. In this study, they make use of adjacency list, information gain theory and confusion matrix.
This proposal deals with the two problems. Using the data available, to predict in which college a student has high chances of getting an admit, and also which stream. It also deals with providing relevant results by learning from previous system states and revising itself each time.
Decision Support System are mostly interactive systems, often required by humans to provide necessary information based on specific inputs; and they are also adaptable computer based information system. Such a decision system not only utilizes decision rules, models, and a comprehensive database but also the decision maker's own insights, leading to specific, implementable decisions in solving problems that would be difficult for a human to take alone [3].

2. PROBLEM DEFINITION
Admission into professional colleges for engineering degree course is based on scores of the Common Entrance Test (CET). Students are allotted colleges based on these scores. Seats are allotted on the basis of availability of seats in CAP rounds. The lowest score accepted in a college for a certain CAP round is known as the cut-off score.
Universities under DTE collect data about CET scores and admissions from each college under that particular university. Analyzing this extensive data provides us with an opportunity to predict the admission pattern for a particular score, branch and even a CAP round. Presently there are no such resources to sort out colleges based on the parameters of marks, branches and CAP rounds. Due to absence of such resources a student would be less informed regarding the colleges he is eligible in. Here we propose a technique to make use of Decision Support System to assist in providing a student with such decisions. The decisions taken by the system should not only focus on present decisions, but also should take past decisions into account.
The proposal mainly discusses about the use of DSS for finding the most appropriate colleges forstudents based on their CET scores. However, the scope of the project can be extended to includethe common entrance exam that is being envisioned to bring about uniformity and fairness in thecurrent admission system. The algorithm that has been developed can be modified accordinglyso that it will function properly for the new pattern as well, for obtaining an admission to differentcolleges. The main focus here has been given to the Engineering field and the data has beencollected accordingly. So, the students opting for the engineering field may enter their marks inorder to get an appropriate result for the colleges suitable for them. Similarly, this system can beused for several other fields too such as Medicine, Pharmacy, etc.

3. DATA MINING
Data mining is a process that analyses (often large) observational data sets to find relationships within it and to summarize this data in a way that can be used by humans for various purposes [4]. Techniques such as Bayes' theorem, neural networks and decision trees are an integral part of the data mining process.
Data mining is the process of extracting out knowledge from a set of data. It discovers new patterns from data sets using various methods of artificial intelligence, machine learning and database systems. It is one of the steps of the Knowledge

Discovery in Databases process. In the Knowledge Discovery Process the uncovered hidden knowledge can be identified as relationships or patterns. The relationships may be between two or more different objects which may change over a period of time. Discovery of relationships is a key result of data mining [5].

If knowledge discovery is one aspect of data mining, prediction is the other. Here we look for a specific association with regard to an event or condition. Pattern discovery is another outcome of data mining operations. The data mining tools mine the usage patterns of thousands of users and discover the potential pattern of usage that will produce results.

The most common example of data mining would be analysis of shopping trends amongst shoppers. Products bought most commonly by shoppers, are placed next to each other for greater sales.

4. DECISION SUPPORT SYSTEM

DSS are designed specifically to facilitate decision processes. It should support rather than automate decision making, and should adapt quickly to the changing requirements of decision makers [6]. Decision Support Systems are found most useful as they couple human decision making skills along with the computational capability of a computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semi-structured problems [7].

We propose implementation of combination of Data-Driven as well as Knowledge driven type of Decision Support System. Data-driven decision support systems are based on access and manipulation of series of internal, external and sometimes real-time data of an organization. Simple file systems accessed by query and retrieval tools provide the most elementary level of functionality [8]. Knowledge-driven DSS suggest or recommend actions to users. They use business rules, knowledge bases and also human expertise in form of programmed internal logic. Decisions and tasks which can be taken and performed respectively by a human expert are taken by a Knowledge-driven DSS. The generic tasks include classification, configuration, diagnosis, interpretation, planning and prediction [9].

DSS systems often require user involvement in the construction of problem representation and model verification. They also require direct user involvement in the analysis and evaluation of decision outcomes. These activities involve subjective judgments and, therefore, a DSS should focus on effective support and not on automatic selection. An effective DSS is one which is flexible, adaptable to changing user scenarios, its environment and one which learns from knowledge gained on past user scenarios [10].

5. CENTRAL TENDENCY

Central tendency refers to the number of ways, in which the central value or the median value is calculated. The most common and most effective numerical measure of the "center" of a set of data is the (arithmetic) mean [11].

Let *x*1; *x2*,…, *xN* be a set of *N* values or observations, such as for some attribute, like salary. Equation 1 shows the mean of this set of values.

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \cdots + x_N}{N}. \quad (1)$$

6. IMPLEMENTATION

The CET scores are stored in the database, and it is processed to obtain the relevant results. Colleges are sorted out on the basis of previous year results as well as each CAP round. High and low values are taken into consideration to apply conditions to the entries in the database to categorize them. The results are categorized as Ambitious, Best Bargain and Safe colleges.

$$\text{High} = \text{Score} + \textit{offset} \quad (2)$$

$$\text{Low} = \text{Score} — \textit{offset} \quad (3)$$

where *offset* is the deciding factor which decides the relevance of the results.

Algorithm of the logic is given as follows:

1) Accept input from user.
2) Search database for results pre-existing in the database.
3) If result is found, display them to the user. This method thereby reduces the complexity and provides a quick result to the user. Update low and high values for the existing result using the current input of the user.
4) Else, do the following:
   a) Search entire database.
   b) For entries, with mean greater than input but less than or equal to high, mark them as Ambitious. It means the candidate has slim chances of getting into that college and stream.
   c) For entries, with mean less than input but greater than or equal to low, mark them as Best Bargain. It means the computed result is best possible college-branch result for the candidate and is most likely to get into it.
   d) For entries, with mean less than or equal to low, mark them as Safe. It means the candidate has highest chances of getting into these colleges.
5) Enter the result generated in step 4 into the database for making it available for future users.

Mean is the average of marks of all years for a particular CAP round. At step 4 and 5, the system learns on its own on the basis of marks and previous year results. It revises itself each time the system is used. Therefore, each time, the system is in a new state, except when results pre-exist in the table.

7. TEST RESULTS

Sample data taken from the data collected from the DTE website is used to test the working of this decision system.

Data is compiled by them every year, after the admission process gets over with the help of feedback from the colleges under it. This data is available on its website [12]. Table 1 shows the sample data. The user has to provide his/her CET score. The system will provide the user its decision and will update itself and be ready for future decision queries.

Table 1. Sample data of college cut off scores for particular branch and year

College ID	Branch ID	CAP Round	2009 Marks	2010 Marks	2011 Marks
1	1	1	144	149	126
1	1	2	142	146	0
1	1	3	148	136	101
2	4	1	150	147	128
2	4	2	147	149	120
2	4	3	142	0	0

Decision rules have been tested using the data from the year 2009-2011. MySQL is used for processing of the data.

Table 2. Mean scores of three years for a college for a particular CAP Round

College ID	Branch ID	CAP Round	Mean
1	1	1	139
1	1	2	144
1	1	3	128
2	4	1	141
2	4	2	138
2	4	3	142

Table 3. Result shown to the user

College ID	Branch ID	CAP Round	Type
1	1	1	Best Bargain
1	1	3	Safe
2	4	1	Ambitious
2	4	2	Safe

College ID refers to a college, Branch ID refers to an engineering stream. Mumbai University offer specialization in Computer, Information Technology, Electronics, Mechanical and many more streams. CAP Rounds are the admission rounds undertaken by a college, if there are seats available under a particular stream. 2009 Marks, 2010 Marks, 2011 Marks are the cut off marks for that year, for a particular college, branch and CAP round. 0 marks indicate no seats were available for a particular CAP round in that year. If scores of two years are available, then mean is calculated for three years, else, if it is available for two years, then mean of two years is calculated. Values of mean calculated by the system are shown in Table 2.

Let the *offset* value be equal to 2. If a higher offset value is considered, then the results may not be realistic. Let the user input be 140.

The system will follow the algorithm as follows.

1) System will check if results are available for the score of 140. Let us assume that the results are not available.
2) The system will calculate the values of High and Low as 142 and 138 on the basis of (2) and (3) respectively.
3) Now, system will search through the database and perform following updates:
    a) For entries with mean greater than input but less than or equal to High, i.e. 140 and 142 respectively are listed as Ambitious.
    b) For entries with mean less than input but greater than or equal to Low, i.e. 140 and 138 respectively are listed as Best Bargain.
    c) For entries with mean less than or equal to Low, i.e. 138 are listed as Safe.
4) Results of steps a, b and c are updated in the database for future use.
5) If another search on marks of 140 is performed, the system will show values of calculations performed in previous steps. This way it saves resources and does not perform calculation unnecessarily.

Results provided to the user are shown in Table 3. The results shown to the user are the decisions made by the system, which it thinks are optimal for a user having CET marks of 140. A user can also provide with specific branch of engineering in which he/she is interested in getting decisions about. This way, more user-oriented decisions can be generated.

A user can also provide with specific branch of engineering in which he/she is interested in getting decisions about. This way, more user-oriented decisions can be generated.

8. CONCLUSION

The papers referred helped getting a better idea about the technique to be applied in this proposed system. Thus, helping in getting a better idea about the way in which the systemmust work with minimum faults in it. The papers provided me with a clear idea about which techniques may have which advantages and disadvantages with them when implemented.

The advantages of the system proposed in this paper are that it uses past inputs whichenables the user to get a more realistic result as compared to a system generating resultsbased purely on the thresholds assigned. The cut-off marks keep changing year afteryear due to the varying difficulty levels in different sections for different years thus, it isnecessary for the system to be able to update itself. The system proposed in this paperkeeps revising itself after each calculation so as to provide the user with the most possiblyaccurate and up-to-date information regarding the colleges they are eligible for. As thesystem is put into use, more and more data will be collected each year. Thus, with more data, a stronger system can be guaranteed to be made available to the students/users.

Choosing the right career path, coupled with the right institution is extremely important for any student. With large amount of data at hand, it is important that it should be analyzed efficiently. Hence this work demonstrates how data mining technologies can be used to help take wise career decisions. This method can further be extended by using various probability based prediction methods.

If the current system is kept in mind, then all the results from the different entrance exams such as state CETs, IIT-JEE, BIT-SAT, AIEEE can be taken into account and a system can be developed togive proper results to students according to their scores in the different exams. However, differentstudents may score differently in different tests and thus, the system must be robust enough todecide which is the best score amongst all and thus must be taken into consideration. And, not allstudents always appear for all the exams to gain admission to an engineering college. Keeping thisin mind, the options must be made available to students while selecting the scores to be entered forthe different entrance examinations.

9. ACKNOWLEDGMENTS

I would like to thank Prof. Shiwani Gupta for encouraging us to implement this project, Ushang Thakker for assisting me in designing the logic and Lavanya Singh for helping me in preparing this paper.

10. REFERENCES

[1] Elizabeth Murray, Using decision trees to understand student data, Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany,2005, unpublished.

[2] Sudheep Elayidom, Dr. Sumam Mary Idikkula, Joseph Alexander, Anurag Ojha, Applying data mining techniques for placement chance prediction, 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, Trivandrum, Kerala, 2009, published, pp 669-671.

[3] V. S. Janakiraman, K. Sarukesi, Decision Support Systems, Chapter 6, PHI Learning Pvt. Ltd., New Delhi, 2004, p 26.

[4] D. J. Hand, Heikki Mannila, Padhraic Smyth, Principles of data mining, Cambridge, MIT Press, 2001, p 1.

[5] Paulraj Ponniah, Data Warehousing Fundamentals, John Wiley & Sons, 2001, p 402-403.

[6] Daniel J. Power, Decision Support Systems: Concepts and Resources for Managers, Greenwood Publishing Group, 2002, p 6-13.

[7] Keen, P. G. W. and M. S. Scott-Morton. Decision Support Systems: An Organizational Perspective. Reading, MA: Addison-Wesley, 1978.

[8] Frada Burstein, C. W. Holsapple, Handbook on Decision Support, vol. 1, Springer-Verlag, Berlin, 2008, p 127.

[9] Daniel J. Power, Decision Support Basics, Business Expert Press, 2009, p 41.

[10] Gregory E. Kersten, Zbigniew Mikolajuk, Anthony G. O. Yeh, Decision support systems for sustainable development, Kluwer Academic Publishers, Norwell, 2000, p42.

[11] Jiawei Han, MichelineKamber, Data Mining Concepts and Techniques, Chapter 2, Morgan Kaufmann Publishers, San Francisco, 2006, p. 51

[12] Directorate of Technical Education [Online]. Available: http://www.dte.org.in/fe2011/StaticPages/Default.aspx