College Recommendation System

Leena Deshpande
Nilesh Dikhale,[‡] Himanshu Srivastava,[§] Apurva Dudhane, [¶] Umesh Gholap[k].
[‡],dikhalenilesh@gmail.com, [§] himanshu.ozil7@gmail.com,
[¶] apurvamdudhane@gmail.com, [k]
Vishwakarma Institute of Information Technology, Pune. 411 048

Abstract—Educational organizations are one of the important parts of our society and playing a vital role for growth and development of any nation.For that getting appropriate college is of foremost importance.We are proposing a system which involves data analysis and data mining techniques.Recommendation system is a part of data mining techniques which we have used.It includes machine learning and data mining techniques for filtering data and presenting the required information.Our system is made to target students, parents and educationalist who search and need counseling for getting admission in engineering colleges.This paper aims at presenting a framework of system which can recommend best engineering colleges on merit of students.

Keywords—Data mining, Recommendation system, Decision tree(J48) algorithm

I. INTRODUCTION

Data Mining techniques are of great importance in present education and business system . A data mining task can be specified in the form of a data mining query, which is input to the data mining system. A data mining query is defined in terms of data mining task primitives. These primitives allow the user to interactively communicate with the data mining system during discovery in order to direct the mining process, or examine the findings from different angles or depths A Decision tree approach is proposed which may be taken as an important basis of selection of college during any course program. Machine learning techniques are employed as a paradigm in the modeling of college recommendation.It is a tedious job searching the best deserving college due to lack of articulate information about them.It also consumes lot of time in it. Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user. The suggestions relate to various decisionmaking processes. We are proposing a system where student can shortlist colleges based on their merit,interest,fees and locality.We have collected interest rating from a survey using google form. Interest fields includes attributes like infrastructure,cultural and technical activities,sports,etc .Based on the feedback given by the students of various colleges we have averaged their ratings to get a mean value.

II. MOTIVATION

In present scenario there are different college recommendation apps and websites present but using them is a tedious job due to of lack of articulate information regarding colleges and lot of time is consumed in searching the best deserving college. The problem statement for best college recommendation system is achieved from it and to give an insight to college administration for overall rating,cutoff of their colleges, overall admission intake and preferences of students, so that such analysis can be further used by any college system for their advertisement and publicity. Also to avoid spending time and money on counsellor and stressful research related to finding cut-off of each college.

III. LITERATURE SURVEY

Data mining is extracting or mining knowledge from huge amounts of data. . Data mining should have been more appropriately named knowledge mining from data. Knowledge mining may not reflect the emphasis on mining from large amounts of data.

[1]   Recommender systems solve the problem of information overload by searching through large volume of dynamically generated information to provide users with personalized content and services. There are different characteristics and potentials of different prediction techniques in recommendation systems in order to serve as a compass for research and practice in the field of recommendation systems.

[2]   In this paper, a Decision tree approach is proposed which may be taken as an important basis of selection of student during any course program. Data Mining could be used to improve business intelligence process including education system to enhance the efficacy and overall efficiency by optimally utilizing the resources available.

[3]   This paper describes that the Real World data set from a high school is taken and filtration of desired potential variables is done using WEKA an Open Source Tool. The dataset of student academic records is tested andapplied on various classification algorithms such as Multilayer Perception,Nave Bayes, SMO, J48 and REPTree using WEKA an Open source tool.

[4]   In this paper Naive Bayes and J48 algorithm are compared. Naive Bayes is based on probability and J48 algorithm is based on decision tree. Accuracy is checked using WEKA tool. The experiment results shown in this paper are about classification accuracy and cost analysis. The results in the paper on this

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

dataset also show that the efficiency and accuracy of J48 and Naive Bayes is good.

[5]   This paper discusses data mining techniques to process a dataset and identify the relevance of classification test data .It shows the process of WEKA analysis of file converts and selection of attributes to be mined and comparison with Knowledge Extraction of Evolutionary Learning.It shows each Decision tree achieve a high rate of accuracy. It classify the data into the correctly and incorrectly instance.

The data mining application usually has 3-tier architecture.

1. Front End(GUI)
2. Application Layer(Recommender System)
3. Database The architecture diagram shows the overview ofthe application.

Architectural Diagram
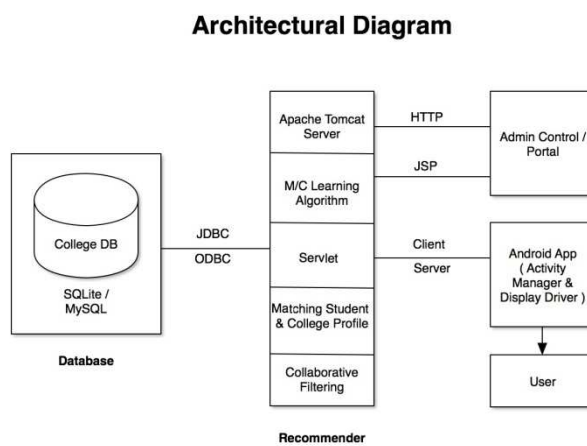
IV. EXPERIMENTAL ANALYSIS

We have used WEKA tool for comparing algorithms based on inputs.We had given 160 entries as input which include cuttoff of 5 colleges namely COEP, VIT, PICT, VIIT and SCOE. We considered 6 attributes for classification including College, Branch, Gender, University, Caste, MeritNumber. After giving training data we passed testing data for classification.Here we took college's names as class labels for classification. And we found that Naive Bayes gave an accuracy of 50 percent that means it classified 5 out of 10 colleges correctly.While for J48 we found the accuracy to be 80 percent.

The text around the Confusion Matrix is arranged so as row labels are on the left instead on the right but we read it just the same. The row indicates the true class, the column indicates the classifier output.Each entry then gives number of instances of row that were classified as column. All correct classifications are on the top-left to bottom-right diagonal. Everything off that diagonals is an incorrect classification.

A. *Several standard terms used in comparison*

- True positive (TP)
  Refers to the positive tuples that were correctly labeled.

- False positive (FP)
  Refers to the negative tuples that were incorrectly labeled.

- False negative (FN)
  Refers to the positive tuples that were incorrectly labeled.

- True negative (TN)
  Refers to the negative tuples that were correctly labeled.

- Precision and recall
  Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance.

- Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity. High recall means that an algorithm returned most of the relevant results. High precision means that an algorithm returned more relevant results than irrelevant.

Precision and Recall is stated as precision=tp/(tp+fp)
recall=tp/(tp+fn)

Fig. 1.Naive Bayes

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
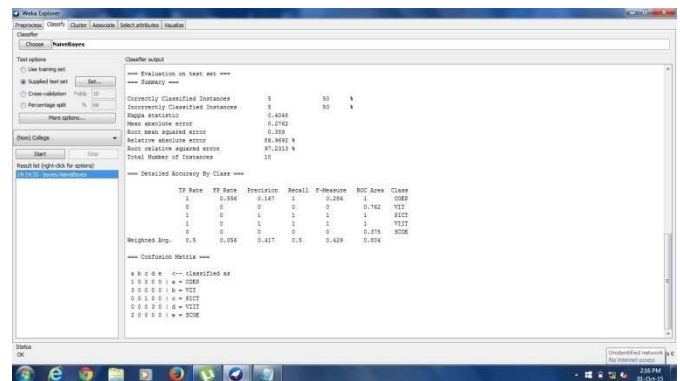*National Conference "NCPCI-2016", 19 March 2016*
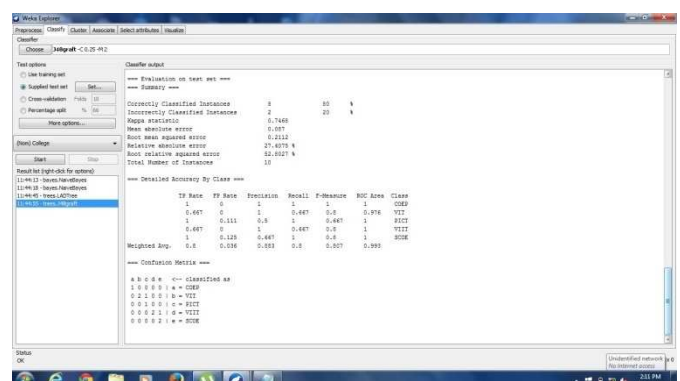*Available online at www.ijrat.org*

Fig. 2. J48

B. Comparison of Existing Techniques

Naive Bayes algorithm is based on probability and J48 algorithm is based on decision tree. The paper sets out to make comparative evaluation of classifiers Nave Bayes and J48 to maximize true positive rate andminimize false positive rate of defaulters rather than achievingonly higher classification accuracy using WEKA tool. The experiments results shown in this paper are about classification accuracy . The results in the paper on this dataset also show that the efficiency and accuracy of J48 and Naive Bayes is good. [4]

Decision tree thus achieves a high rate of accuracy. It classifies the data into the incorrect and correct instances. We can use these decision tree algorithms in medical , banking , stock market and various area. [5]

V. PROPOSED SYSTEM

As we have studied the algorithms Naive Bayes and Decision Trees which can be useful to solve the given problem. The algorithm which is expected to have higher accuracy in recommending the best college is used. Thus technique would be helpful for students minimizing their time in searching colleges. Our system consists of in all four modules which describe the various aspects for recommending colleges,details of the same,branches and comparison with other colleges.

A. Our Application comprises of four modules

- College Search: In this module user will give name of college and location by which he has to search as input.Accordingly a list of colleges will be displayed to the user.
- Comparison: In this module user will be given a choice to select colleges he wants to compare.By this he will get a clearer idea for the distinguished college.
- Advanced Search: We have collected interest rating from a survey using google form. Interest fields includes attributes like infrastructure,cultural and technical activities,sports,NSS,Edc and other attributes include faculty,hostel,placement and fees. Based on the feedback given by the students of various colleges we have averaged their ratings to get a mean value. In this module user will be asked to give his candidature details and his interests in co-curricular and extra curricular activities.According to his merit and his interests ,colleges will be shortlisted based on their merit,interest,fees and locality.
- Branch Search: In this module user will be asked to give his candidature details and the college name.According to his merit a list of branches will be recommended with the student is most likely to get according to merit in that college.

VI. CONCLUSION

After doing research about the problems faced by students to get admission in Engineering College, we came up with the idea to try and solve their problem with real time Application. Then coming into technicalities of using data mining algorithm in application we compared various algorithm using Weka tool and finally came to conclusion of using Decision Tree Algorithm (J48) due to higher accuracy.Our proposed system consists of in all four modules which describes the various aspects of recommendation sytem including deserving colleges,details of the same,branch-wise search and comparison with other colleges.

REFERENCES

[1] F.O. Isinkaye,Y.O. Folajimi ,B.A. Ojokoh, Recommendation systems: Principles,methods and evaluation, Received 13 March 2015; revised 8 June 2015;accepted 30 June 2015.

[2] Sonali Agarwal, G. N. Pandey, and M. D. Tiwari, Data Mining in Education:Data Classification and Decision Tree Approach International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April2012

[3] ParneetKaur ,Manpreet Singh , Gurpreet Singh Josan. Lakshmi M. Gadhikar, Deepa Vincent, Lavanya Mohan and Megha V. Chaudhari *Classification and prediction based data mining algorithms to predict slow learners in education sector* rd International Conference on Recent Trends in Computing 2015(ICRTC-2015)

[4] Anshul Goyal and RajniMehta , Performance Comparison of Naive Bayes and J48 Classification Algorithms,International Journalof Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012).

[5] Trilok Chand Sharma ,Manoj Jain , WEKA Approach for Comparative Studyof Classification Algorithm,International Journal of Advanced Research inComputer and Communication Engineering Vol. 2, Issue 4, April 2013.

[6] Data Mining : Concepts and Techniques , "Han, Jiawei.; Kamber, Micheline."

[7] Data Mining: Know It All, "Soumen Chakrabarti; Earl Cox; Eibe Frank; Ralf Hartmut Gting; Jiawei Han; Xia Jiang; Micheline Kamber; Sam S. Lightstone; Thomas P. Nadeau; Richard E Neapolitan; Dorian Pyle; Mamdouh Refaat; Markus Schneider; Toby J. Teorey; Ian H. Witten".

[8] Recommender Systems Handbook, "Francesco Ricci   Lior Rokach Bracha Shapira Paul B. Kantor".