



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 5      Issue: VII      Month of publication: July 2017**

**DOI:**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Data Analysis Based on Online Reviews

Atiq Patel<sup>1</sup>, Prof. Nilambari Joshi<sup>2</sup>, Prof. Aarti Sahitya<sup>3</sup>

<sup>1,2</sup>Department of Computer Engineering, Mumbai University

**Abstract:** Users share their view about product, hotel, news and topic on web in the form of reviews, blogs, comments etc. Many users read review information given on web to take decisions such as buying products, watching movie, going to restaurant etc. Reviews contain user's opinion about product, event or topic. It is difficult for web users to read and understand contents from large number of reviews. Here our proposed system will be for the college students who want to pursue engineering after their junior college. System will use database and will match the review with the keywords in database and will rank the review. System will rate the College based on the rank of review and give the suitable suggestion or recommend the option.

**Keywords:** C4.5 algorithm, Decision tree, College Selection, Students.

## I. INTRODUCTION

The students in today's world are very practical in life. Some are sure what they want to pursue in their life for further studies but there are students who are not completely aware of the world. This System basically designed for the 12<sup>th</sup> Pass out students and they want to get admission in engineering college. Those who achieve the highest marks in their board and entrance examinations get admission in good colleges but those students get the average marks, it is difficult for them to find the best suitable colleges for themselves.

Our system is beneficial for those students who want to seek admission in the appropriate colleges on based on the marks they have obtained. Students using this application will be able to find out which College is best and suitable for them based on their own criteria such as Location, Faculty, Infrastructure and Placements.

## II. LITERATURE SURVEY

Currently there are no existing systems which work for the student admissions for the colleges. So we had to start from the scratch and build the project. We took help of two papers which have helped us shaping the project. They are

### A. Mining and Summarizing Customer Reviews

M. Hu and B. Liu [1] For each discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations. A count is computed to show how many reviews give positive/negative opinions to the feature. All features are ranked according to the frequency of their appearances in the reviews. Feature phrases appear before single word features as phrases normally are more interesting to users. Other types of rankings are also possible. For example, we can also rank features according the number of reviews that express positive or negative opinions. This paper helped to get information about the comments which helped us to analyze the comments given by the students who have posted on the specific college website. It uses the concept of opinion sentences where each word is given a particular value. The value can be positive or negative or even neutral in some cases. The values inside each positive or negative varies, it can be slightly positive or negative or to the extremes. We have used this concept.

### B. Cross-Domain Extraction of Sentiment and Topic Lexicons

F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, Extracting sentiment and topic lexicons is important for opinion mining. Previous works have showed that supervised learning methods are superior for this task. However, the performance of supervised methods highly relies on manually labelled training data. In this paper, we propose a domain adaptation framework for sentiment- and topic-lexicon co-extraction in a domain of interest where we do not require any labelled data, but have lots of labelled data in another related domain.

## III. PROPOSED SYSTEM AND DESIGN

The Opinion Mining for College Review system that detects hidden sentiments in feedback of the user and rates the feedback accordingly. The system uses opinion mining methodology in order to achieve desired functionality. Opinion mining for college reviews is a web application which gives review of the feedback that is posted by various users. The System takes review of various users, based on the opinion, system will specify whether the posted college is good, bad, or worst. System will rate the college based

on the rank of review for this we use the C4.5 Algorithm. The role of the admin is to post new college and add keywords in database Following are the main components of the proposed system:

- A. Login
- B. fill the form with 12th marks and JEE Main Marks
- C. Show the recommended college to user
- D. Sentiment word addition
- E. College registration system

*F. Proposed algorithm*

1) *C4.5 Algorithm:* C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

Fig. 1-C4.5 Algorithm

```

Algorithm 1.1 C4.5(D)


---


Input: an attribute-valued dataset  $D$ 
1:  $Tree = \{\}$ 
2: if  $D$  is "pure" OR other stopping criteria met then
3:   terminate
4: end if
5: for all attribute  $a \in D$  do
6:   Compute information-theoretic criteria if we split on  $a$ 
7: end for
8:  $a_{best}$  = Best attribute according to above computed criteria
9:  $Tree$  = Create a decision node that tests  $a_{best}$  in the root
10:  $D_v$  = Induced sub-datasets from  $D$  based on  $a_{best}$ 
11: for all  $D_v$  do
12:    $Tree_v = C4.5(D_v)$ 
13:   Attach  $Tree_v$  to the corresponding branch of  $Tree$ 
14: end for
15: return  $Tree$ 


---



```

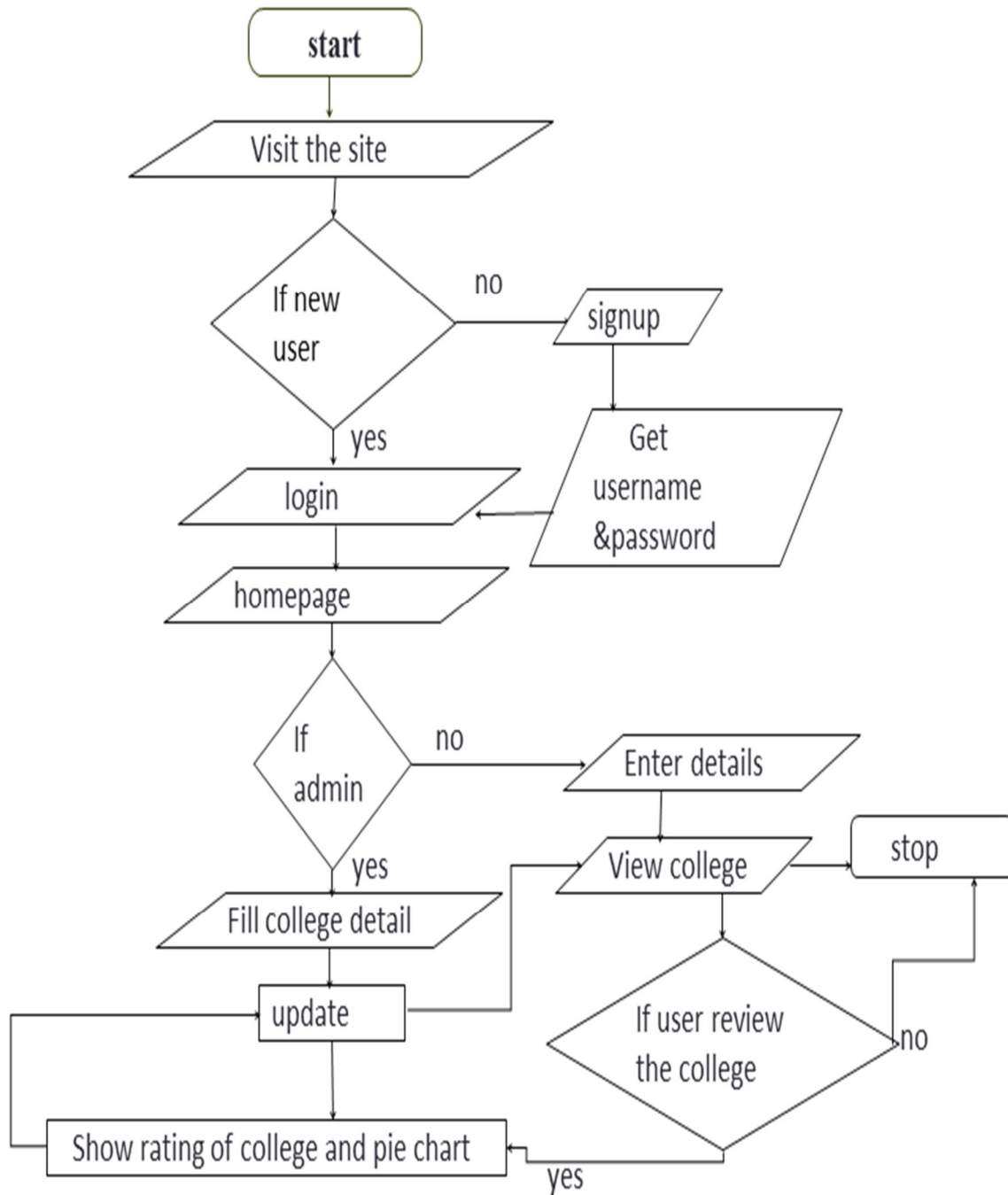
*G. Rating or Classification using C4.5 Algorithm:*

The whole Classifications process involves steps

- 1) *Authentication Module:* By using the Username and password of the user we authenticate the user. It is for the security.
- 2) *Upload Require Database Module:*
  - a) All college related data is not available easily, It is very hectic work to find the all the required data of the college.
  - b) For building the required database we take the help of DTE (Directorate of Technical Education) site which gives us the required data related engineering college.
  - c) We find out the required attributes in the database like (Cut-off, subject code).
- 3) *Construction of the Decision Tree Module:*

- a) By using the the help of the C4.5 algorithm we rate the colleges and as per the requirements of the user or student we show the suitable options to student for choose from them. The inputs of the users are the decision nodes of the decision tree.
- b) For the build the decision tree, our inputs are the marks of the student.
- 4) Show the Result to the User Module: After building the decision tree the system will recommend user the appropriate college.

Fig. 2- System flow chart



#### IV. REQUIREMENTS

##### A. Software Specifications

- 1) JDK



- 2) ECLIPSE
- 3) MySQL(Back-End/Database)

### V. EXPERIMENTAL RESULTS

Fig 3- Home Pages



Fig. 4-User Signup

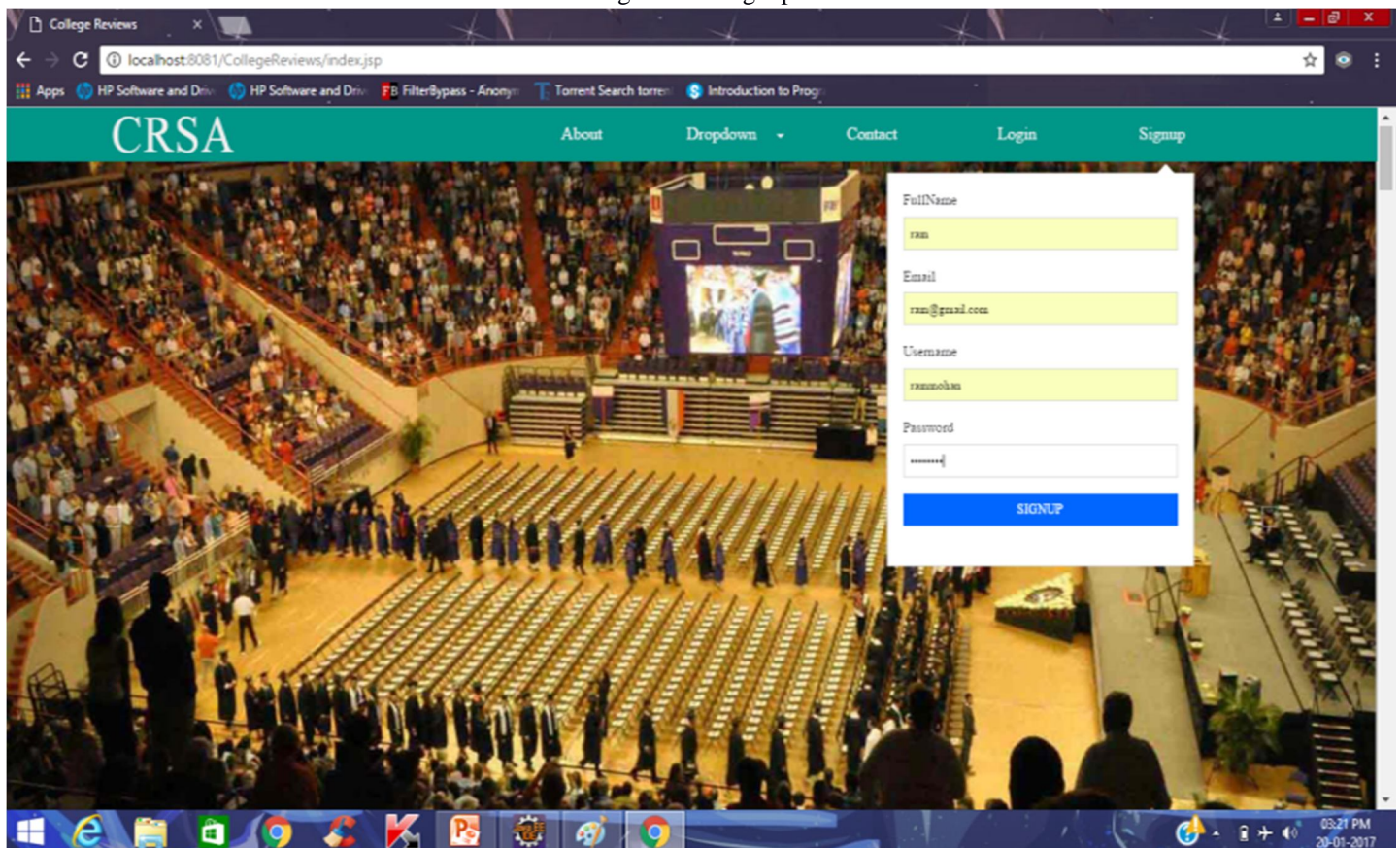
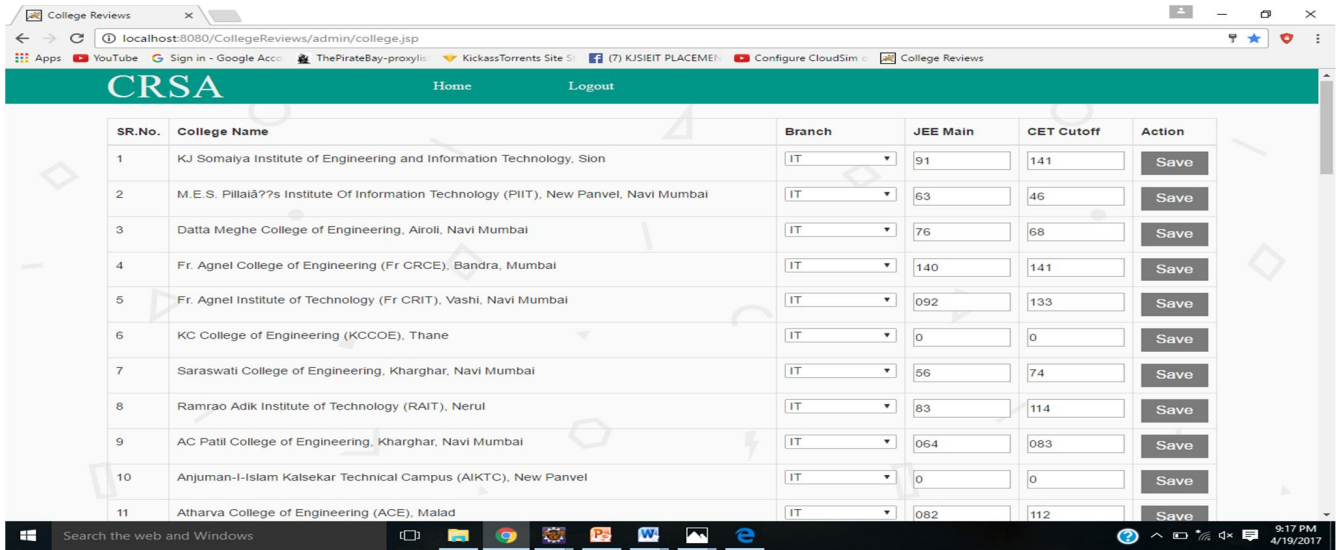
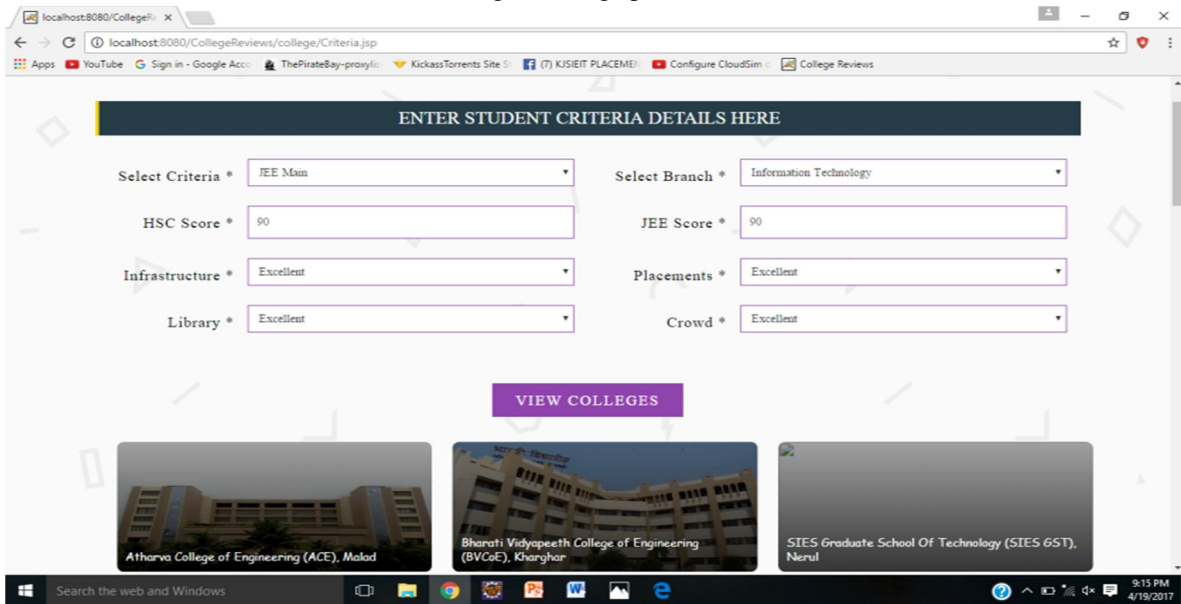


Fig. 5-Admin Workspace.



SR.No.	College Name	Branch	JEE Main	CET Cutoff	Action
1	KJ Somaiya Institute of Engineering and Information Technology, Sion	IT	91	141	Save
2	M.E.S. Pillai's Institute Of Information Technology (PIIT), New Panvel, Navi Mumbai	IT	63	46	Save
3	Datta Meghe College of Engineering, Airoli, Navi Mumbai	IT	76	68	Save
4	Fr. Agnel College of Engineering (Fr CRCE), Bandra, Mumbai	IT	140	141	Save
5	Fr. Agnel Institute of Technology (Fr CRIT), Vashi, Navi Mumbai	IT	092	133	Save
6	KC College of Engineering (KCCOE), Thane	IT	0	0	Save
7	Saraswati College of Engineering, Kharghar, Navi Mumbai	IT	56	74	Save
8	Ramrao Adik Institute of Technology (RAIT), Nerul	IT	83	114	Save
9	AC Patil College of Engineering, Kharghar, Navi Mumbai	IT	064	083	Save
10	Anjuman-I-Islam Kalsekar Technical Campus (AIKTC), New Panvel	IT	0	0	Save
11	Atharva College of Engineering (ACE), Malad	IT	082	112	Save

Fig. 6-College predictor



ENTER STUDENT CRITERIA DETAILS HERE


Select Criteria \*       Select Branch \*

HSC Score \*       JEE Score \*


Infrastructure \*       Placements \*

Library \*       Crowd \*

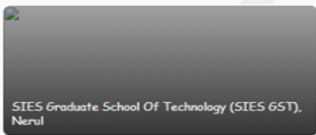
[VIEW COLLEGES](#)



Atharva College of Engineering (ACE), Malad

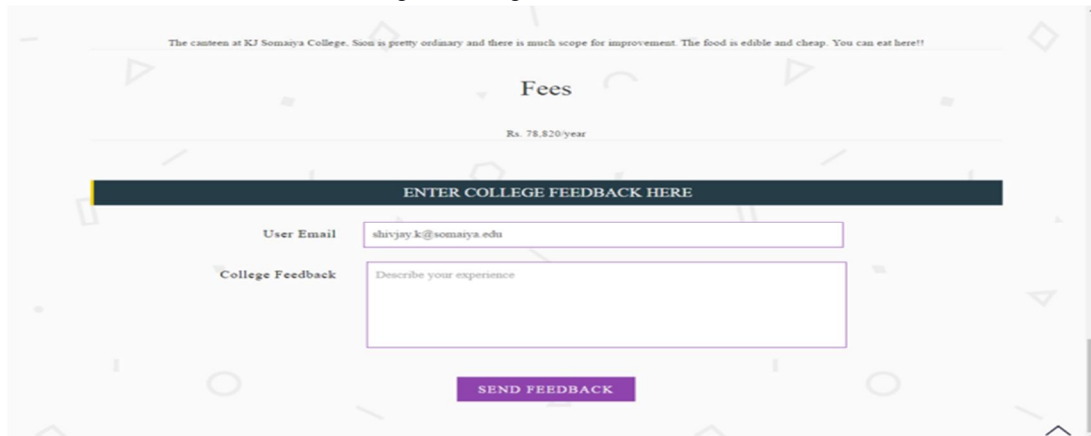


Bhanati Vidyapeeth College of Engineering (BVCoE), Kharghar



SIES Graduate School Of Technology (SIES GST), Nerul

Fig. 7- College Comment Section



The canteen at KJ Somaiya College, Sion is pretty ordinary and there is much scope for improvement. The food is edible and cheap. You can eat here!!

**Fees**

Rs. 78,820/year

ENTER COLLEGE FEEDBACK HERE

User Email

College Feedback

[SEND FEEDBACK](#)

## VI. CONCLUSION

Thus, System is basically built for students to guide them selecting the college according to their marks. System will use classification C4.5 algorithm to show college based on marks and also in this proposed system will match the review with the keywords in database and will rank the review of college .System will rate the College based on the rank of review and give the suitable suggestion or recommend the option of college. The advantages of the system are

- A. More accurate than the human record.
- B. Fast and Accurate results.
- C. Non-Intrusive Technology.
- D. More reliable than human record.

## REFERENCES

- [1] M.Hu and B.Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining,Seattle, WA,USA, 2004, pp. 168–177.
- [2] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-domain extraction of sentiment and topic lexicons," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea,2012, pp. 410–419





10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)