College Guesstimate (A Web Portal for College Prediction)

Rahul Sathawane[1], Rohan Battulwar[2], Prasheel Fuley[3], Ananiya Mahajan[4], Shivam Joshi[5], Roshan Chaturpale[5]

1. Professor, Dept of IT, RGCER, Nagpur Maharashtra, INDIA

2. Student, Dept of IT, RGCER, Nagpur Maharashtra, INDIA

3. Student, Dept of IT, RGCER, Nagpur Maharashtra, INDIA

4. Student, Dept of IT, RGCER, Nagpur Maharashtra, INDIA

5. Student, Dept of IT, RGCER, Nagpur Maharashtra, INDIA

6. Student, Dept of IT, RGCER, Nagpur Maharashtra, INDIA

ABSTRACT

The ease of making better choices and making better decisions in terms of selecting colleges is our aim. Our analysis on colleges for the students makes easier for them to make accurate decision about their preferred colleges. For such analysis, it requires future possibilities from the past record data from DTE which can potentially make the predictions and recommendation for students. Our analysis with the data mining methods would help giving probable accuracy and this requires analytical methods for predicting future recommendation. Today, most students make mistakes in their preference list due to lack of knowledge, improper and incorrect analysis of colleges and insecure predictions. Hence repent and regret after allotment. Our project will solve the general issue of the student community by using technology.

Keywords - *DTE, Analysis, Data Analysis, Random forest, big data, Prediction Analysis*

1. INTRODUCTION

There are more than 400 engineering colleges in Maharashtra, for which admission is governed by DTE (Directorate of Technical Education). DTE carries out the admission through CAP (Centralized Admission Process).The process is done through the cap rounds and is very confusing for student to analyze the perfect college. The student's needed to verify the documents at Facilitation Centre and are supposed to give their preference list of colleges. Then based on their marks, Category, Home University and other attributes, college is allotted to them in three consecutive forms.

It's very difficult for the students to and out suitable colleges for them based on their MHCET Score, MHCET Rank, Category, Home University, etc. Various colleges provide degree in engineering in various branches (IT, Computer, Mechanical, Electrical, civil, etc). Though analysis of colleges and their cut offs is required in order to get the most correct preference list.

It is very tedious job for a student to understand about the suitable colleges which provides preferred branch and to analyses it's last three years cut offs in order to predict whether that he can get one of those colleges in CAP. Most of the students make mistakes in their preference list due to lack of knowledge, improper and incorrect analysis of colleges and insecure predictions. Hence those students regret after what they get the college after allotment. Our project will solve the issue of the student community by using a technology. We are designing a web application for predicting the college through an algorithm for students who aspires for taking admission in engineering through CAP. The graphical representation would be shown to students and the

probabilities will be given for better and easy understanding to them. Our system will generate the most suitable preference list for the student.

2 TECHNOLOGYS UESD

1. HTML (FOR FRONT END)

The most basic form of web development includes designing page with the help of the most prominent and consistent web technology which not only contributes in designing the front end which is in this case a web page but also helps in organizing and it with different forms of styles using CSS (cascading style sheets). HTML also allows developers to work along with other database applications that are necessary to store the details of the students.

Hypertext Markup Language (HTML) is the standard markup language for creating web pages and web applications. With Cascading Style Sheets (CSS) and JavaScript, it forms a triad of cornerstone technologies for the World Wide Web. Web browsers receive HTML documents from a web server or from local storage and render them into multimedia web pages. HTML describes the structure of a web page semantically and originally included cues for the appearance of the document.

HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects, such as interactive forms, may be embedded into the rendered page. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items.

HTML elements are delineated by tags, written using angle brackets. Tags such as <imp /> and <input /> introduce content into the page directly. Others such as ... surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.

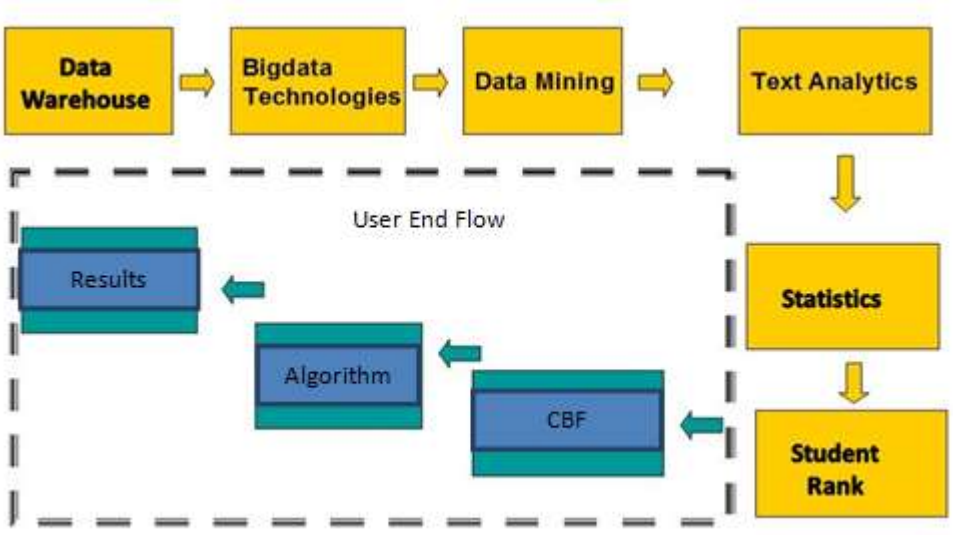

Fig-1- The Methodology

2. PHP (embedded code)

PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. Originally created by Rasmus Lerdorf in 1994, the PHP reference implementation is now produced by The PHP Group. PHP originally stood for Personal Home Page, but it now stands for the recursive acronym PHP: Hypertext Preprocessor.

PHP code may be embedded into HTML code, or it can be used in combination with various web template systems, web content management systems, and web frameworks. PHP code is usually processed by a PHP interpreter implemented as a module in the web server or as a Common Gateway Interface (CGI) executable.

He web server combines the results of the interpreted and executed PHP code, which may be any type of data, including images, with the generated web page. PHP code may also be executed with a command-line interface (CLI) and can be used to implement standalone graphical applications.

PHP is used for various proposes in our project -

1. Read data from CSV files

PHP is used to import csv files from the existing pool of csv files that include the data of college.

2. Export user data from fields to MySQL

This application will be able to connect to the database and take the input through Graphical User Interface. The application will be able to navigate the user properly. The application will also be able to create new database to store CET marks and rank which would b used as a training set for Rank predictors. The application will be able to display the generated preference list and will be able to receive the feedback from user and store it in a text form.

Fig-2 HTML page with embedded PHP code. (Front End)

The fig.[2] Shows an HTML page the various field that user inputs in the page are the name, MHCET Rank and the Score as well as Category, Branch and Region in which the candidate is interested.

2. METHEDOLOGY

Predictive analytics is a form of advanced analytics that uses both new and historical data to forecast activity, behavior and trends. It involves applying statistical analysis techniques, analytical queries and automated content based filtering algorithms to data sets to create datasets that have a likelihood of a particular event happening.

Predictive analytics software applications use variables that can be measured and analyzed to predict the likely behavior of individuals, machinery or other entities. For example, an insurance company is likely to take into account potential driving safety variables, such as age, gender, location, type of vehicle and driving record, when pricing and issuing auto insurance policies.

Multiple variables are combined into a predictive model capable of assessing future probabilities with an acceptable level of reliability. The software relies heavily on advanced algorithms and methodologies, such as Context Based Filtering, high value ranks.

Predictive analytics has grown in prominence alongside the emergence of big data system. As enterprises have amassed larger and broader pools of data in Hadoop clusters and other big data platforms, they have created increased data mining opportunities to gain predictive insights. Heightened development and commercialization of machine learning tools by IT vendors has also helped expand predictive analytics capabilities.

1. PHP is a server-side scripting language used to develop static WebPages and dynamic WebPages. PHP scripts can be interpreted on a server that has PHP installed.

2. Several issues have to be considered when implementing a content-based filtering system. First, terms can either be assigned automatically or manually.

3. Data visualization is the presentation of data in a pictorial or graphical format. It enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns. With interactive visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see how it's processed. Fig III (i) shows the output.

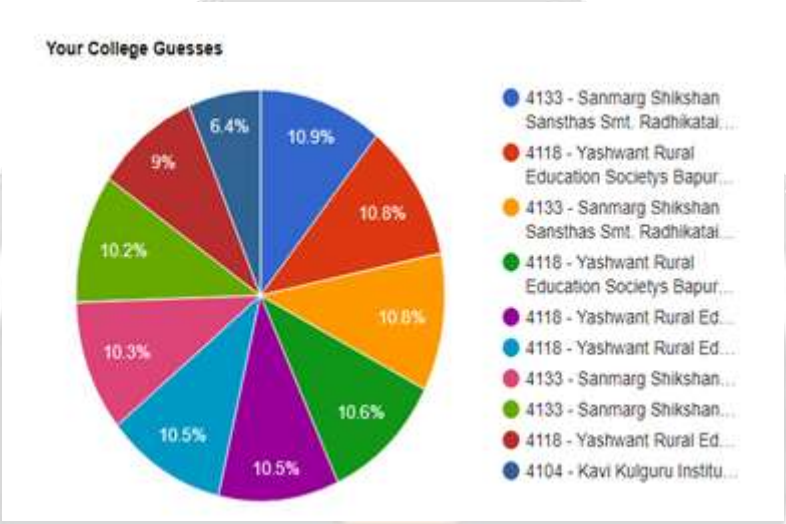

Chart -1: Google Chart API (Representation of rank of cut off among various categories.

3. APPROACH

Using the Content based filtering algorithms which can be used to solve the given problem. The algorithm which is expected to have higher accuracy in recommending the best preference list of colleges is used. This project would prove helpful for students minimizing their time in searching colleges and predicting whether they will be allotted the desired college or not. And also will help in deciding what should be the order of preference.

Our Web Portal comprises of modules which are as follows:

Our E-admission system will make it possible to finish up the admission process in the first attempt itself. And bring students the satisfaction that they are studying in the right college according to the marks obtained. Basically, this system will select and present only those colleges in which the student can get admission.

We have collected the cut off details of all the colleges for creating the database. Using these previous year cut off marks we have calculated the range in which the college may fall. Accordingly, the comparison will be made between the marks entered by the user and the list of colleges will be displayed in which the user may get admission.

With respective to the input received, the content-based filtering algorithm or cognitive filtering is applied on database. The temporary table is created which would contain filtered data. This table will contain the colleges which should be included in preference list in unordered or random order. Proper ordering is done on this data with the function get sort order. The main thing which we are considering is placements of the college. And as we all know, it is quite open secret that the colleges with higher cut offs have better placements. So the colleges

having greater rank cut offs are given first preference. The second thing we are considering is location of the college.

This application will be able to connect to the database and take the input through Graphical User Interface. The application will be able to navigate the user properly. The application will also be able to create new database to store CET marks and rank which would b used as a training set for Rank predictors. The application will be able to display the generated preference list and will be able to receive the feedback from user and store it in a text form.

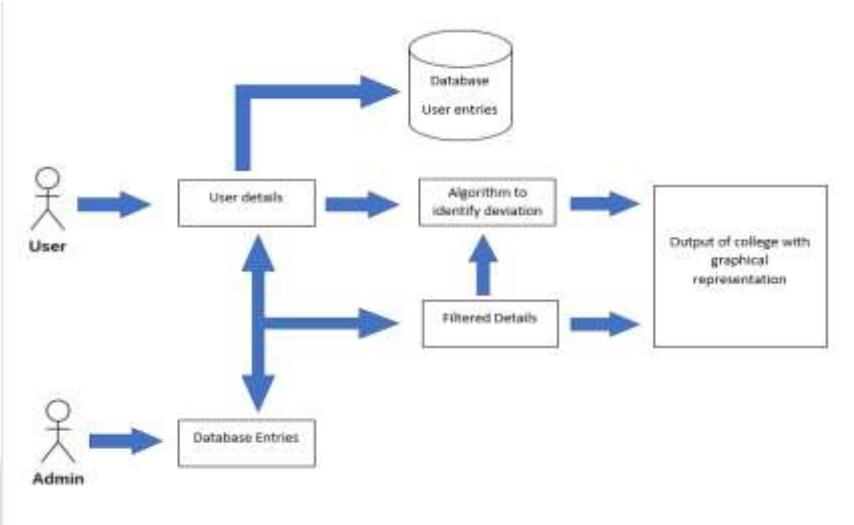

Fig- 2: Use Case Diagram

4. CONCLUSION

A huge amount of data is being generated every single moment, even while you're reading these words, at unbelievably rapid speeds across the globe. According to an estimate, the global annual rate of data production in the year 2015 was 5.6 Zettabytes. That was almost double the rate of growth just three years back in the year 2012.

When it comes to technology management, planning, and decision making, extracting information from existing data sets—or, predictive analysis—can be an essential business tool. Predictive models are used to examine existing data and trends to better understand customers and products while also identifying potential future opportunities and risks.

These business intelligence models create forecasts by integrating data mining, machine learning, statistical modelling, and other data technology.

Benefits –

- Better predictions better results for students.
- Helpful for students to make a choice of engineering college after 12^{th}.
- Making it easier for colleges to know where they stand in attracting students with a cluster of score.
- Google API makes it easier for students to understand the current scenario of admissions by providing graphical representations of college cut-offs.

Limitations

- Attracts only certain section of students
- Possibility of shutting down of MHCET

- Noisy data makes predictions hard as training the model is purely based upon the raw data.

Application

- Makes it easier for students to select better based upon the score and cut off of previous colleges
- Colleges can better understand where to start a campaign.

5 . REFERENCES

The Predictive model has some of the various areas of applications we have studied those application and came up with a variety of methods we can use to perform the task in hand. Below are the websites and books that we have refereed throughout the understanding and creation of the project.

[1]http://www.webopedia.com/TERM/P/predictive_analytics.html

[2]http://blogs.wsj.com/digits/2014/01/17/amazon-wants-to-ship-your-package-before-you-buy-it/

[3]http://www.entrepreneurial-insights.com/predictive-analytics-forecast-future/

[4]https://hbr.org/2014/09/a-predictive-analytics-primer/

[5] Applied Predictive Modelling, Chapter 7 for regression and Chapter 13 for classification.

[6]Data Mining: Practical Machine Learning Tools and Techniques, page 76 and 128

[7] http://iitnit.com.