

Analysis of College by using Data Mining

Vaibhav Fale¹ Shamal Deshmukh² Mayur Burange³

^{1,2,3}Department of Computer Science & Engineering

^{1,2,3}P. R. Pote (Patil) College of Engg & Mgnt, Amravati,, India

Abstract— Since its been few years social media has captured the attention of the entire world as it is thundering fast in sending thoughts across globe, user friendly and free of cost requiring only a working internet connection. Only limit is internet connection to the users side. Opinion and reviews are the most critical factor in formulating views and influencing the success product or services. Though it is difficult to analyze these information based on opinion and reviews because of humongous or disoranzged nature. With rapid growth in user of Social Media in recent years, the research get attracted towards the use of social media data for social media data for sentiments analysis of people or particular product or person or event. Many a times, people unknowingly misspell words which may take away the meaning of the sentence. The ubiquitous approach of replacing more than two occurrences of a letter with two occurrences of the same letter is not a complete solution as misspellings may occur from the users's finger slipping to a nearby letter or the user's spelling the word phonetically. Therefore the words have been corrected to the best possible effort by applying a probabilistic model based on Bayes' theorem which shows roughly 86.67% accuracy. During this process enhancement of data has been done like removing of duplicate comment based on Bayes Theorem is used for spelling correction which is overlooked in other research studies.

Key words: Sentiment Analysis, Opinion Mining

I. INTRODUCTION

Since its been few years social media has captured the attention of the entire world as it is easy to communicate and reach our thought across the world with any people at within certain amount of time. A social media is a critical factor reviewing any product or brand as the user basic requirement. Data mining is simply we can say that one the motivating research factor and increasingly popular day by day that helps us out finding meaningful information from huge data.

Data Mining provides benefits by providing detection of fraud or any repeatative information that has lead down the decrement in the performance of the system. This project proposed the mechanism of system which review and rate the college as per the students requirement and recommend them. This project will contain the opinion mining technique which work is to analyze the review giving by the students.

This proposed system contains three main module Administrator, student and review analysis. Student will register and will give his opinion necessarily. Admin will create college admin login and review college performance .Review Analysis panel is where the student will get recommendation of the colleges per three requirement.

A. Motivation

From the past few years colleges rankings is the rankings of institutions in higher education ordered by various

combination of vaarious factors.colleges rankings have most often been conducted by governments and academics.So some students are not getting the area expertise of their colleges and are not compatible with their trade.so main intension to develop this system to provide a decision support for student requirement colleges and let them have their own compatible trade.

B. Aim

To an effective software solution of decision supporting system that helps in finding out the college as per the student recommended area of expertise and compatibility.

C. Objectives

- To analyze the colleges using data mining.
- To introduce the recommendation system that recommend the college base on student given data & preferences.
- To enrich the students with different Training and placement activity by providing panel with Dynamic multimedia content.

D. Application uses

The completion of this proposed system owns a great deal to the help, advice and assistance given by numerous students along the way to the end user.

II. PROPOSED SYSTEM

A. System Architecture Design

1) Sentiment analysis:

Text preprocessing of data is the process of preparing and cleaning the data of dataset for classification. Here is the hypothesis of having the data properly pre-processed to reduce the noise in the text should help improve the performance of the classifier and speed up the classification process, thus aiding in eal time sentiment analysis.

2) Tokenization:

Given input as character sequence, tokenization is a task of chopping it up into pieces called tokens and at the same time removing certain characters such as punctuation marks. A token is an instance of sequence of characters that are grouped together as a useful semantic unit for processing.

3) Stop word removal:

A stop-list is the name commonly give to a set or list of stop words. It is typically language specific, although it may contain words. A search engine or other natural language processing system may contain a variety of stop-list one per language, or it may contain a single stop-list that is multilingual. Some of the more frequently used stop words for English include "A", "of", "the", "I", "it", "you", and "and" these are generally regarded as 'functional words' which do not carry meaning.

4) Stemming:

It is the process for reducing derived words to their stem, or root form. Stemming programs are commonly referred to as

stemmers algorithms. A simple stemmer looks up the inflected form in a lookup table, this kind of approach is simple and fast.

B. Working of Proposed System

This project is proposing the mechanism for providing better result to the students as per their recommended college. Working flow of proposed system is as below.

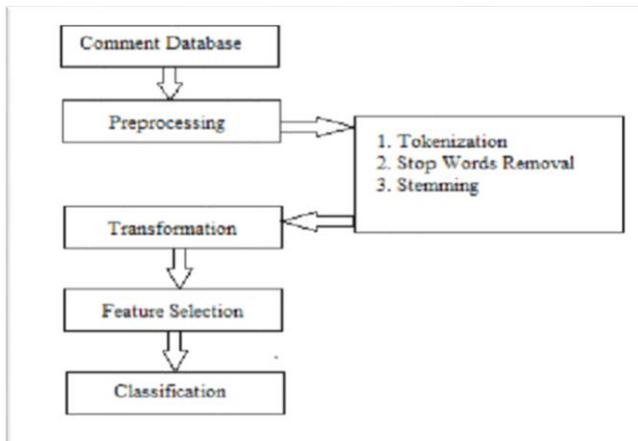


Fig.1 Steps and techniques used in sentiment classification.

Fig. 3.1: Steps and technique used in sentiment classification

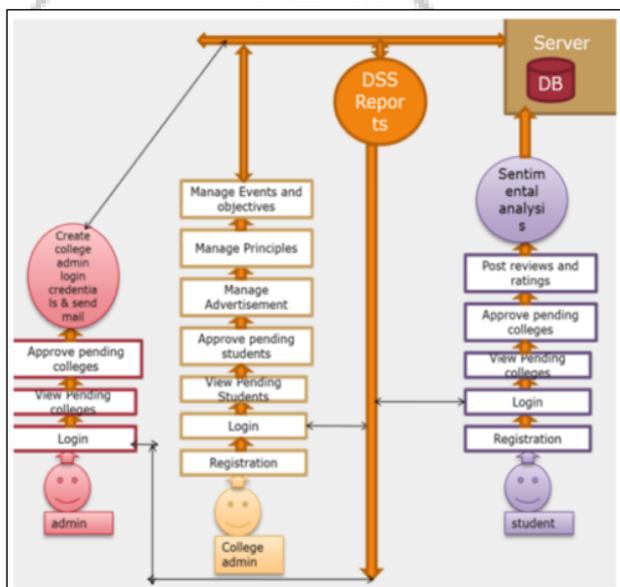


Fig. 3.2: Working of proposed system

III. CONCLUSION

There have been many different methods or ways to analyze the data for influencing the brand, product and services, but still there is some hurdles in those methods like for example analyzing the twitter data may never know who is the end user can be anyone or might not at all.

This contain a proposed system which will provide security by so called user has to approve a login request to college admin for the identification. The analysis of the college will be based on the current student and passed out student review which will be given to the end user based on his recommendation.

REFERENCES

- [1] Arora D.,Li K.F. and Neville S.W.,” Consumers’ sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study”,29th IEEE International Conference on Advanced Information Networking and Applications.
- [2] Kanakaraj M., Guddeti R M.R.,” Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques” 9th IEEE International Conference on Semantic Computing.
- [3] Bahrainian S.-A., Dengel A.,” Sentiment Analysis and Summarization of Twitter Data”,16th IEEE International Conference on Computational Science and Engineering.
- [4] Shahheidari S., Dong H., Bin Daud M.N.R.,” Twitter sentiment mining :A multidomain analysis “,7th IEEE Internataional Conference on Complex,Intelligent and Software Intensive Systems.