

Analysis of College by Using Data Mining and Security

Sonali Harel, Yogita Patange

Department Of Computer Science And Engineering
P.R.Pote(Patil) College Of Engg and Mgnt
Amravati, India
{harelsonali9,yogitapatange24396}@gmail.com

Mayur Burange

Department Of Computer Science And Engineering
P.R.Pote(Patil) College Of Engg and Mgnt
Amravati, India
Mayurmsb123@gmail.com

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 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 opinions and reviews because of humongous or disorganized nature. With rapid growth in user of Social Media in recent years, the researcher get attracted towards the use of social media data for sentiments analysis of people or particular product or person or event .

Keywords- Sentiment Analysis , Machine Learning , Opinion Mining.

I. INTRODUCTION

Data Mining is simply we can say that one the motivating research factor and increasingly popular day by day that helps us out finding the meaningful information from huge data. This project proposed the mechanism of System which review and rate the college as per the students requirement and recommend them. This project 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 view college performance. Review Analysis panel is where the student will get recommendation of the colleges per there requirement .The main intension of this project to develop the web application of decision supporting system that helps in finding out the college as per the student recommended area of expertise and compatibility.

1.1 Motivation : From the past few years colleges rankings is the rankings of institutions in higher education .Colleges rankings have most often been conducted by governments and academics .So some students are not getting the area expertise of their 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.

1.2 Aim : “To design and developed a decision support system that analyzed and recommend a college to the student on the basis of evaluating student feedback by implementing data mining concept”.

1.3 Objectives :

1. To implement a recommender system which recommend appropriate college as per the student's requirement.
2. To implement opinion mining technique to analyze reviews given by college students.
3. To develop a Decision support system for admin.

1.4 Scope : The completion of this project owes a great deal to the help, advice and assistance given by numerous students along the way to the end user.

II. PROPOSED SYSTEM

Text classification is to classify data into pre defined classes . Here they are positive and negative classes. Text classification is supervised learning problem .First step in text classification is transforming document which is in string format into format suitable for learning algorithm and classification task .In informal retrieval it is found that word stem works well as representation unit . This leads to attributed value representation of text. Each word corresponds to feature with , number of times word occurs in document, as its value. Words are considered as features only if they are not stop words (like "and" , "or" ,etc).

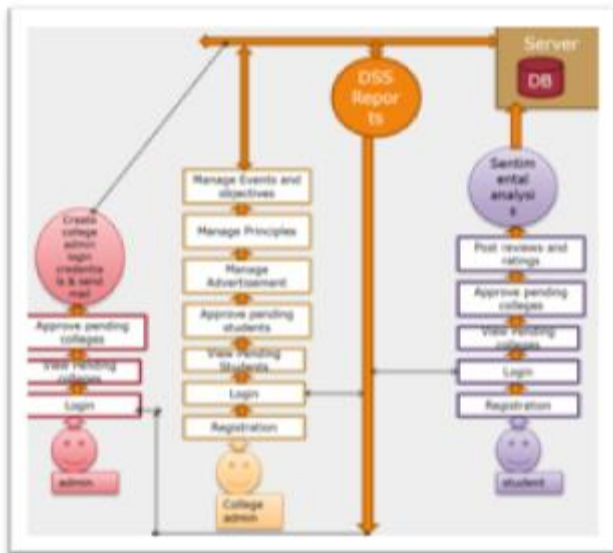


Fig : Working of proposed system

System Architecture Design

Sentimental 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 real time sentiment analysis.

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.

Stop word removal: A stop-list is the name commonly given 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-lists, 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. When assessing the contents of natural language, the meaning can be conveyed more clearly by ignoring the functional words. Hence it is practical to remove those words which appear too often that support no information for the task.

Stemming: It is the process for reducing derived words to their stem, or root form. Stemming programs are commonly referred to as stemmers or stemming algorithms. A simple stemmer looks up the inflected form in a lookup table, this kind of approach is simple and fast. The disadvantage is that all inflected forms must be explicitly listed in table. eg. "developed", "development", "developing" are reduced to the stem "develop".

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.

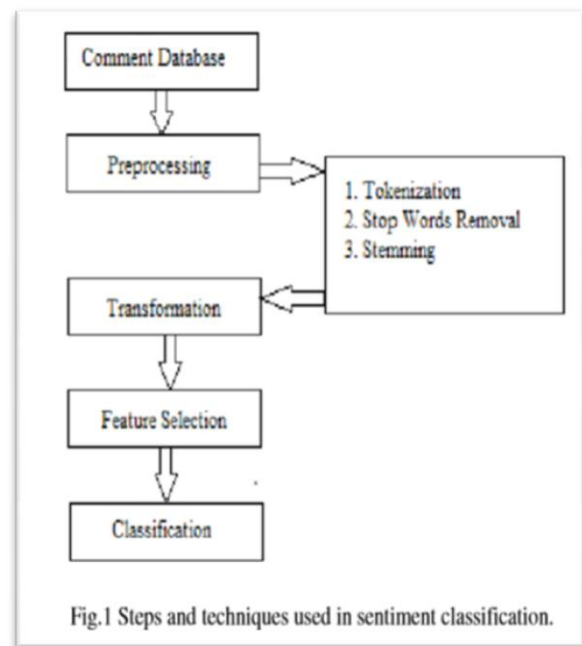


Fig.1 Steps and techniques used in sentiment classification.

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 you may never know who is the end user he can be anyone or might not at all. There is a limitation out there in this method.

This project 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.] Sentiment Analysis of Top Colleges in India Using Twitter Data Nehal Mamgain¹, Ekta Mehta², Ankush Mittal⁴ Department of Computer Science Graphic Era University Dehradun, India {nehal.mamgain1,ektamehta1412,dr.ankush.mittal4}@gmail.com
- [2.] 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*, pp. 680-686, Gwangju, South Korea, March 2015
- [3.] 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*, pp. 169-170, Anaheim, California, 2015
- [4.] Shahheidari S., Dong H., Bin Daud M.N.R.,” Twitter sentiment mining: A multidomain analysis”, *7th IEEE International Conference on Complex, Intelligent and Software Intensive Systems*, pp.144-149, Taichung, Taiwan, July 2013
- [5.] Bahrainian S.-A., Dengel A.,”Sentiment Analysis and Summarization of Twitter Data”, *16th IEEE International Conference on Computational Science and Engineering*, pp. 227-234, Sydney, Australia, December 2013
- [6.] Shahheidari S., Dong H., Bin Daud M.N.R.,” Twitter sentiment mining: A multidomain analysis”, *7th IEEE International Conference on Complex, Intelligent and Software Intensive Systems*, pp.144-149, Taichung, Taiwan, July 2013
- [7.] Neethu M. S. and Rajasree R., “Sentiment Analysis in Twitter using Machine Learning Techniques”, *4th IEEE International Conference on Computing, Communications and Networking Technologies*, pp. 1-5, Tiruchengode, India, 2013
- [8.] King R. A., Racherla P. and Bush V. D.,” What We Know and Don't Know about Online Word-of-Mouth: A Review and Synthesis of the Literature”, *Journal of Interactive Marketing*, vol. 28, issue 3, pp. 167183, August 2014
- [9.] Choi C., Lee J., Park G., Na J. and Cho W., “Voice of customer analysis for internet shopping malls”, *International Journal of Smart Home: IJSH*, vol. 7, no. 5, pp. 291-304, September 2013.
- [10.] Jotheeswaran J. and Koteeswaran S., “Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis”, *Journal of Engineering and Applied Sciences*, vol. 10, issue 14, pp. 5883-5894, January 2015.