



Sentiment Analysis of Top Colleges in Pune Using Social Media Data

Pravin Katkar, Hrushikesh Bachakar, Sandhya Gude, Prof. Amruta Patil

Dept. of Information Technology, Zeal college of Engineering and Research, Narhe, Pune, India

ABSTRACT: In Social media Twitter, Facebook are different is one of the most usually utilized stages for imparting insights, communicating sees. Sentiment Analysis on social data like review can enable clients to comprehend the suppositions communicated in tweets and characterizing them in positive or negative classes. The associations can utilize slant investigation to get a thought of the client's reviews of different colleges in Pune. Automation of this undertaking makes it simpler to manage the huge proportion of data being created by social sites like twitter, Facebook as well as personal college websites on an on-going premise. In this proposed system we can sentiment analysis on review various college in Pune. According to classification users review suggest the top colleges in Pune also view shortest distance route on map .In our system also performance comparison of the Naïve Bayes and Support vector Machine algorithm.

KEYWORDS: College, Support Vector Machine, Sentiment Analysis

I. INTRODUCTION

A college might be a degree-granting tertiary instructive establishment, a piece of a university or Government College, an organization offering professional training, or an auxiliary college. Right now, system can gather the continuous audit information of various universities in Pune via web-based networking media like twitter, Facebook just as the official website of different colleges'. Presently a day's distinctive school's data accessible on web like evaluations and reviews accessible to us are one of the most fundamental factors in orchestrating our perspectives and affecting the achievement of a brand. Existing framework isn't acceptable or enough for propose top universities. With the help of Naïve Bayes and SVM calculation estimation investigation of survey of those colleges Assessment examination arranged all audits in positive and negative class. Clients search various universities utilizing Stop word expulsions calculation in Pune just as framework can show the top colleges in Pune based on audits. Show the most brief separation course on map from current spot to various college. Presently a day's number of designing, restorative and other stream colleges an in pune are quickly increment for consistently. Every school is giving the different offices to the understudy just as the representative of the school. Concurring the giving offices just as encouraging standard of school is choosing the confirmation of universities. In the event that school instructing staff is acceptable and offices are useful for understudy, at that point understudies take affirmation for that school. In the event that any understudy needs to take confirmation for any school, at that point understudy first pursuit that school and read all the audits for that school from different sites just as internet based life like Facebook and twitter at that point take the affirmation for the school. Right now gather the diverse school surveys from different sites, specific school sites, and school Facebook post from Facebook just as tweets of school from twitter. This gathered school survey information supposition examination the information and grouping of audit of universities. On the off chance that this framework client can look through the changed school from Pune, at that point client can see all the subtleties of school just as view all the surveys of that specific school. Utilizing opinion investigation audit utilizing Naïve Bayes and SVM framework proposes the top colleges as indicated by the best survey for school. Client additionally sees the briefest separation of course from current spot to that specific school on map. In the current framework take a shot at just twitter dataset. In proposed framework take a shot at different web based life information, sites. System can presenting new methodology for expel the disadvantage of existing framework just as improve the precision this framework.



||Volume 9, Issue 4, April 2020||
II. PROBLEM STATEMENT

Now a days different colleges information available on internet like opinions and reviews accessible to us are one of the most critical factors in formulating our views and influencing the success of a brand. Existing system is not good or enough for suggest top colleges.

III. GOAL AND OBJECTIVE

Goal

Using Naïve Bayes and Support vector machine algorithm sentiment analysis of reviews of colleges and identify top colleges in pune.

Objective

- ³⁵₁₇ To search different colleges in pune.
- ³⁵₁₇ Sentiment analysis on various college review.
- ³⁵₁₇ View the college details in the system.

IV. LITERATURE SURVEY

Dr.S.Gomathi and Rohini, Punitha [1] proposed sentiment analysis has lead to advancement of better items and great business the executives, surveys. In this work system proposed Max entropy for feebly administered Deep Embedding for school survey sentence notion grouping. The strategy max entropy performs well for finding the outcomes as positive. The precision of Max entropy produces 96% for school surveys. Max entropy is progressively fit for displaying the long-term conditions in sentences, however it is high productive than Random timberland, SVM, Naïve Bayes, Tree. We likewise find that our strategy can improve general content order.

M. S. Usha [2] states that sentimental analysis, a sub discipline inside data mining and computational derivation, suggests the computational method for mining, comprehension and looking over the suppositions imparted in various end rich resources like web diaries, trade discourses, etc. The objective of feeling investigation is to recognize enthusiastic states in online content. More often than not classifiers prepared in one area don't perform well in another space. Likewise the issue in existing methodologies isn't to recognize opinion and themes. Assumptions may shift with subjects. This system introduced another model called Combined Sentiment Topic (CST) model to identify assumptions, points at the same time from content. This model depends on Gibbs sampling algorithm. In addition, dissimilar to administered ways to deal with supposition mining which frequently neglect to deliver great execution when moving to different areas, the unaided idea of CST makes it exceptionally convenient to different spaces. CST model performs better contrasted with existing semi-managed approaches.

Mrs. S. Yamini et.al [3] introducing that sentiment analysis is a two level assignment. The first is Identifying Topic and the second is, grouping assessment identified with that theme. Slant Analysis begins with "What other individuals thinks?" Conclusion Extraction manages the recovery of the sentiment or state of mind passed on in square of Unstructured content in connection to the space of the record being broke down. But a lot of research has gone in the NLP, AI and web mining system on expelling composed data from unstructured sources; most by far of the4 proposed techniques depend upon grimly named unstructured data. The World Wide Web has been overpowered by unstructured substance and glancing through the web has been established on frameworks from Information Retrieval. Managed learning count researches the arrangement data and produces an assembled limit which is called characterization.

Arora D et.al [4] states that sentiment investigation of the content information accessible on the web either as online journals or at web based life destinations, for example, Twitter, Facebook, and LinkedIn, offers data through which to survey individuals' viewpoint of items and administrations that are important to them. Customers routinely scour the Internet to survey other client's audits for an item or an administration before settling on their own choice. Similarly, these equivalent information can possibly give organizations a depiction in time of the clients' reaction to their items/benefits and even the patterns after some time. This data increased through conclusion examination would then be able to be utilized by

||Volume 9, Issue 4, April 2020||

organizations to settle on choices about improving their items and administrations, and addition an expanded edge over their rivals. Specifically, user will probably look at whether the tweets open on the web are sufficient to build accommodating comprehension about the show of noticeable propelled cell denotes, their battery life, screen quality, and on the obvious introduction of the phones working structures. Our outcomes demonstrate that in spite of the fact that the Twitter information provides some data about clients' slants to the mainstream advanced mobile phone brands and their hidden working frameworks, the measure of information accessible for various brands differs significantly.

Kanakaraj M.[5] et.al Mining opinions what's more, separating presumptions from casual association information help in different fields, for example, even guess, inquiring about all around disposition of open on a specific social issue, and so on. This structure consolidates analysing the way of the general populace on express news from Twitter posts. The key thought of the structure is to manufacture the exactness of depiction by including Natural Language Processing Techniques (NLP) particularly semantics and Word Sense Disambiguation. The mined substance data is introduced to Ensemble depiction to isolate the end. Troupe solicitation joins blending the impact of different free classifiers on a specific solicitation issue. Evaluations drove show that outfit classifier beats ordinary AI classifiers by 3-5%.

Prakruthi V et.al [6] state that Supposition investigation alludes to the application for preparing normal language, content examination, computational phonetics, what's more, biometrics to efficiently perceive, separate, evaluate, and learn full of feeling states and abstract data. Twitter, being one among a few well known online life stages, is where individuals frequently decide to communicate their feelings and suppositions about a brand, an item or a help. Dissecting slants for tweets is useful in deciding individuals' supposition as positive, negative or on the other hand impartial. This paper assesses the individuals' supposition about an individual, pattern, item or brand. Twitter API is utilized to get to the tweets straightforwardly from twitter and construct a notion grouping for the tweets. The result of the examination is delineated for positive, negative and unbiased comments about their assessments utilizing perception procedures, for example, histogram and Pie outline.

Sahar A et.al [7] proposed that these days, individuals from all around the globe utilize internet based life locales to share data. Twitter for instance is a stage where clients send, read posts known as 'tweets' and connect with various networks. Clients share their day by day lives, post their suppositions on everything, for example, brands and places. Organizations can profit by this enormous stage by gathering information identified with sentiments on them. The point of this paper is to show a model that can perform assessment investigation of genuine information gathered from Twitter. Information in Twitter is profoundly unstructured which makes it hard to dissect. Be that as it may, our proposed model is not the same as earlier work right now it joined the utilization of administered and unaided AI calculations. The way toward performing estimation examination as follows: Tweet removed straightforwardly from Twitter API, at that point cleaning and disclosure of information performed. After that the information were nourished into a few models to prepare. Each tweet separated ordered dependent on its feeling whether it is a positive, negative or nonpartisan. Information were gathered on two subjects McDonalds and KFC to show which eatery has greater notoriety. Diverse AI calculations were utilized. The outcome from these models was tried utilizing different testing measurements like cross approval and f-score. Besides, our model exhibits solid execution on mining writings separated straightforwardly from Twitter.

IV.ALGORITHM APPROACH

1. SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) states as it is a particular classifier which is officially described by a confining hyper plane. System can say that the given marked preparing information (supervised learning). SVM is a quick and reliable grouping calculation which performs best with a little and restricted measure of information. The thought behind the SVM calculation is easy to consider and portray, and applying it to its natural language classification which doesn't get most of the complex stuff. And that's the basics of Support Vector Machines. A SVM allows us to classify data that is actually linearly separable. If not then system can use the kernel view of point to make it work.

Algorithm Steps:

Step1: SVMs maximize the margin around the separating hyperplane.

³⁵/₁₇ Assume linear separability for now:



-in 2 dimensions, can separate by a line

-in higher dimensions, need hyperplanes

³⁵₁₇ Can find separating hyperplane by linear programming (e.g. perceptron):

-separator can be expressed as $ax + by = c$

Step2: The decision function is fully specified by a subset of training samples, the support vectors.

Step3: Quadratic programming issue

Step4: Text grouping technique

2. NAIVE BAYES

Naive Bayes is a fundamental structure for making classifiers: models that select class names to give occasions, tended to as vectors of highlight respects, where the class engravings are drawn from some obliged set. There is truly not a solitary figuring for preparing such classifiers, yet a social event of calculations subject to a common guideline: all Naive Bayes classifiers expect that the estimation of a specific segment is self-decision of the estimation of some other section, given the class variable. For instance, a trademark thing might be viewed as an apple on the off chance that it is red, round, and around 10 cm in estimation. A Naive Bayes classifier considers these highlights to contribute self-governing to the likelihood that this characteristic thing is an apple, paying little notice to any potential associations between's the disguising, roundness, and width highlights.

This algorithm, summarized as follows.

$$P(W_k/\text{class}) = (n_k + 1) / (n + \text{Vocabulary})$$

Where, n = total no. of words with specified class

n_k = no. of times word occurred with the specified class Vocabulary = size

Algorithm Steps:

1. Calculate n_k i.e., no. of times the word occurred with class.
2. Calculate n i.e., total no. of words for given class.
3. Calculate $p(w_k/v_j) = n_k/n$
i.e. the probability of word for the given class .
4. Calculate the probability of each class
5. Calculate the vocabulary i.e., the total no. of unique words
6. Repeat the process.

3. K-Nearest Neighbour

In example acknowledgment, the k--nearest neighbour's (KNN) is a non-parametric technique used for request and backslides. In the two cases, the data involves the k nearest planning models in the component space. In case $k = 1$, by then the thing is basically consigned to the class of that single nearest neighbour.

1. Load the information
2. Initialise the estimation of k
3. For getting the anticipated class, emphasize from 1 to add up to number of preparing information focuses
4. Figure the division between test data and each line of planning data. Here we will use Euclidean partition as our Euclidean separation since it's the most notable methodology. Various estimations that can be used cosine system, etc.
5. Sort the decided divisions in rising solicitation reliant on partition regards
6. Get top k lines from the masterminded bunch
7. Get the result

4. Stopwords Removal Algorithm

A word reference based methodology is been used to expel stopwords from record. A conventional stopword rundown containing 450 stopwords made utilizing half and half approach is utilized

The calculation is actualized as underneath given advances.

Stage 1: The objective report content is tokenized and individual words are put away in exhibit.

Stage 2: A solitary stop word is perused from stopwords list.

Stage 3: The stop word is contrasted with objective content in type of exhibit utilizing successive pursuit strategy.

Stage 4: If it matches, the word in exhibit is evacuated, and the correlation is proceeded till length of cluster.

Stage 5: After evacuation of stopword totally, another stopword is perused from stopword list and again calculation pursues stage 2. The calculation runs consistently until all the stopwords are thought about.

Stage 6: Resultant content without stopwords is shown, additionally required insights like stopword expelled, no. of stopwords expelled from objective content, all out include of words in objective content, include of words in resultant content, singular stop word include found in objective content is shown.

V.PROPOSED SYSTEM APPROACH

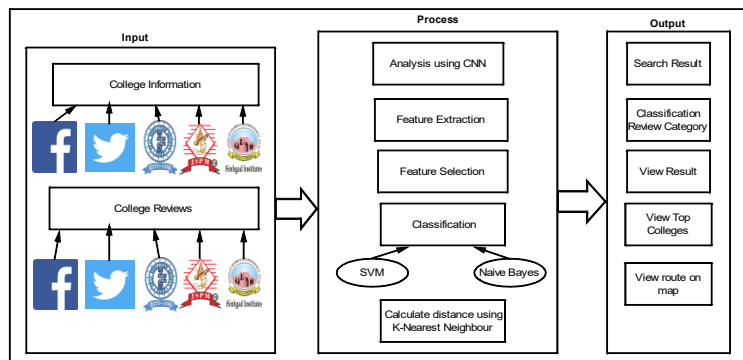


Fig.1 Block Diagram of Proposed System

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. In proposed system do the following functionality

1. View different college of list according to different streams like engineering, medical etc.
2. Sentiment analysis of review on different college.
3. View top colleges in pune after sentiment analysis of review.
4. Evaluate the performance of Naïve Bayes and SVM sentiment analysis.
5. Show the shortest distance route from current place to different college address on map with the help of K-Nearest Neighbour.

VI.COMPARATIVE RESULTS

In our experimental setup, as shown in table 6.1, the total numbers of review of different colleges were 1000. In this system consist 20 engineering colleges of different types of review .Each engineering college have 50 review. Using Naïve bayes algorithm system can classify the review into mainly 3 categories. i.e. positive, negative and neutral. Following table shows that zeal college number of review according to naïve bayes algorithm.

Sr.No	College Name	Positive Review	Negative Review	Neutral Review
1	Zeal	43	3	4



	college			
--	---------	--	--	--

Table1 6.1: Classification of review using naïve bayes

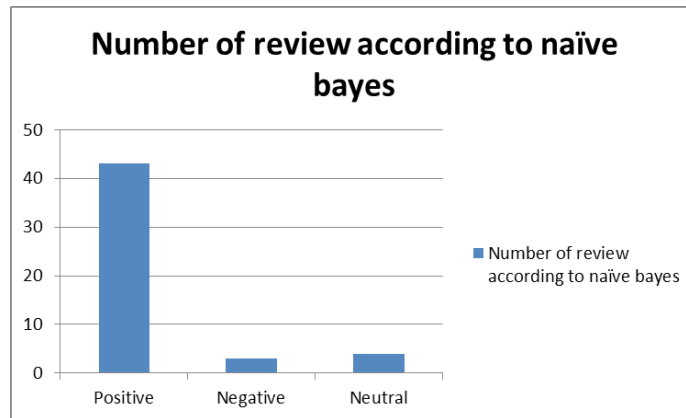
Using support vector machine algorithm system can classify the review into mainly 3 categories. i.e. positive, negative and neutral. Following table shows that zeal college number of review according to support vector machine (SVM) algorithm.

Sr.No	College Name	Positive Review	Negative Review	Neutral Review
1	Zeal college	44	4	2

Table2 6.2: Classification of review using support vector machine

VII.RESULTS

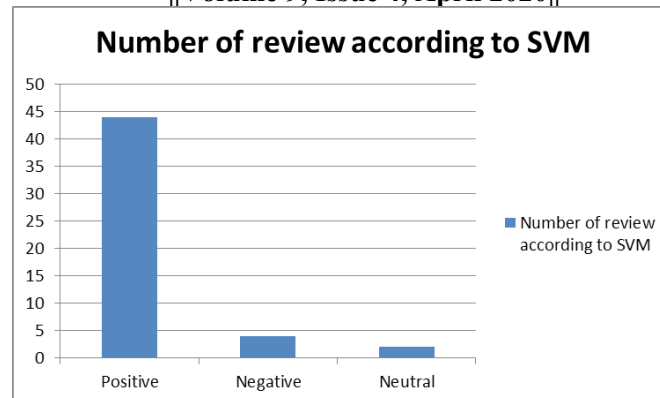
From above data, as shown in table 6.1, the total numbers of review of different colleges were 1000. In this system consist 20 engineering colleges of different types of review .Each engineering college have 50 review. Using Naïve bayes algorithm system can classify the review into mainly 3 categories. i.e. positive, negative and neutral. particular college review shown in graph 7.1, the numbers of positive review found to be Positive were 43, Negative review were 3, Neutral review were 4.



Graph 7.1: Number of review according to naïve bayes

From above data, as shown in table 6.2, the total numbers of review of different colleges were 1000. Using support vector machine algorithm system can classify the review into mainly 3 categories. i.e. positive, negative and neutral. particular college review shown in graph 7.2, the numbers of positive review found to be Positive were 44, Negative review were 4, Neutral review were 2.

||Volume 9, Issue 4, April 2020||

**Graph 7.2: Number of review according to SVM**

VIII. CONCLUSION

In this work, sentiment analysis is a powerful method for ordering the conclusions figured by individuals in regards to any subject, administration or item. Robotization of this assignment makes it simpler to manage the enormous measure of information being delivered by social sites like twitter, Facebook on an ongoing premise. In this proposed system, with the help of sentiment analysis on review various college in Pune According to classification users review suggest the top colleges in Pune. In our system also performance comparison of the Naïve Bayes and Support vector Machine algorithm.

ACKNOWLEDGMENT

This work is supported in a sentiment analysis on top colleges system of any state in India. Authors are thankful to Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune for providing the facility to carry out the research work.

REFERENCES

- [1] Dr.S.Gomathi , Rohini, Punitha. R Sentiment Analysis of Google Reviews of a College International Conference on Computing Intelligence and Data Science (ICCIDS 2018)
- [2] M. S. Usha Analysis of Sentiments using Unsupervised Learning Techniques KLN College of Information Technology Tamilnadu 2017
- [3] Mrs. G. L. Swathi Mirthika, Mrs. S. Yamini Supervised Approach to Extract Sentiments from Unstructured Text International Journal of Engineering Inventions 2013
- [4] Arora D., Li K.F. and Neville S.W. Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter 29th IEEE International Conference on Advanced Information Networking and Applications, pp. 680-686, Gwangju, South Korea, March 2015
- [5] Kanakaraj M., Guddeti R M.R. Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques IEEE International Conference on Semantic Computing, pp. 169-170, Anaheim, California, 2015
- [6] [V. Prkruthi](#) ; [D. Sindhu](#) ; [Dr. S. Anupama Kumar](#) "Real Time Sentiment Analysis Of Twitter Posts" [2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions \(CSITSS\)](#)
- [7] [Sahar A. El Rahman](#) ; [Feddah Alhumaidi AlOtaibi](#) ; [Wejdan Abdullah AlShehri](#) "Sentiment Analysis of Twitter Data" [2019 International Conference on Computer and Information Sciences \(ICIS\)](#)