

## Features based Analysis of Online Reviews using Sentiment Analysis

<sup>1</sup>Kamaxidevi Raol, <sup>2</sup>Nimesh Patel

<sup>1</sup>PG Scholar, Kamaxidevi P. Raol, LDRP ITR, Gandhinagar, Gujarat, India

<sup>2</sup>Professor, Mr. Nimesh Patel, LDRP ITR, Gandhinagar, Gujarat, India

---

**Abstract:** Sentiment Analysis (SA) and Summarization is another and arising field of research which manages data extraction and information disclosure from text utilizing Natural Language Processing and Data Mining method, which help to follow the mind-set of public about specific items and social or political occasion. Sentiment of people are amazingly helpful for individuals and friends proprietor for settling on a few choices. Anyway decision dependent on a portion of the online audit among the huge arrangement of survey are not simple for grouping the supposition. This report present new Hybrid Polarity Detection System for SA of short casual content for example Twitter post to contrast and condition of-workmanship strategy which utilized as investigation of Sentiment Summarization. Also proposed Hybrid Polarity Detection System infers elite with new arrangement of highlights.

**Keyword:** Feature Extraction, Machine Learning Method, Opinion Mining, Sentiment Analysis, Sentiment Classification, Subjectivity Classification, Twitter

---

Date of Submission: 20-03-2021

Date of acceptance: 04-04-2021

---

### I. INTRODUCTION

Today, high proportions of casual emotional content substance are open online with the developing openness of social and miniature contributing to blog locales. These substance or explanations are portrayed in a few arrangements, for instance news stories, study and audit.

Sentiment Analysis (SA) has as of late become the focal point of numerous analysts in view of its application and various fields. As it examines suspected and thought, sentiments, mentality, and assessment of individual, investigation of this sort of online survey is useful and requested for advertising research inspecting, general assessment following, item exploring, business research, political audit, improving of web shopping puts together, etc <sup>[6]</sup>.

Sentiment Analysis is the system, utilized for programmed removing the extremity of public's emotional assessments from plain normal language articulation. Sentiment Analysis is otherwise called Opinion Mining (OM). In view of assessment of people groups, anybody can settle on a good decision prior to procuring any items or things. Opinion Analysis has a broad assortment of utilization in e-business, which assists with comprehending answer of a few requests like, What do customers think about our things, Which of our clients are unsatisfied with the help, What highlights of our things or item are the most exceedingly awful, Who and what means for our picture, What is individuals overall response to some occasion or some individual <sup>[6]</sup>.

Opinion can be accumulated from any individual on the planet about anything through audit locales, studies, sites and conversation bunches and so on <sup>[10]</sup>. Associations and item proprietors who desire to improve their things/administrations may insistently profit by the rich input of customers or clients. The most usually utilized hotspots for discovering assessment are Blogs, survey destinations, crude dataset, and Micro-publishing content to blog sites <sup>[8]</sup>.

Online messages that are posted by individual in World Wide Web are for the most part casual. Investigation or treatment of this sort of substance is routinely more problematic whenever contrasted and formal works <sup>[4]</sup>. The rule contrast among formal and casual content is in information preprocessing is formal content regularly require less preprocessing while casual content frequently contains emojis, use of awful language structure, mockery, and non dictionary standard words <sup>[9]</sup>. Consequently, extraction of casual substance is routinely more problematic.

Individuals as regularly as conceivable ask their family members, companions and field aces for proposal during the dynamic framework, and their suppositions and perspective depend on encounters and discernment. One's point of view around a subject can either be positive or negative, which is term as the extremity discovery of the assessment. At the hour of estimation examination measure, it obliges extremely quick and succinct information so individual can settle on fast and precise decision <sup>[6]</sup>. In notion examination, the data accumulated from the audits has been explored fundamentally at three opinion investigation level <sup>[11]</sup>:

### **1.1 Document Sentiment Level**

The assignment at this level is to perceive whether an entire supposition archive communicates a positive or negative slant. For instance, given a thing or item audit, the framework recognizes whether the surveys of that thing or item convey a general positive or negative estimation about any things. This errand is fundamentally term as record level opinion order.

### **1.2 Sentence Sentiment Level**

The undertaking at this level goes to the sentences and sorts out if each sentence communicated a positive, negative, or impartial notion. Unbiased as a rule portrays no assessment. This investigation is firmly identified with subjectivity order, which sees sentences as target sentences, that express genuine or verifiable data about the world and abstract sentences that express some individual perspectives, convictions and feelings. This assignment of grouping whether a sentence is abstract or objective is terms as subjectivity characterization.

### **1.3 Entity and Aspect Sentiment Level**

Above depicted both the record level and the sentence level don't examine what precisely people enjoyed and didn't care for. Viewpoint level serves to infer extremity (positive or negative) and an objective of assessment. A supposition without its objective being perceived is of restricted use. Discovering the objective of assessment assists with understanding the assumption investigation issue better.

For instance, "although the camera quality is not too much great, I still love this mobile"

This assertion is sure about the portable (element), yet negative about its camera quality (viewpoint). Along these lines, the objective of this degree of assessment is to find suppositions on substances or potentially their angles.

## **II. RELATED WORK**

A ton of exploration has been completed by means of analysts in the assumption examination zone. A portion of the procedures used for supposition characterization are examined here.

### **2.1 Naive Bayes Approach**

It is a straight forward and most ordinarily used classifier model concentrated around bayes decide that figures post-earlier likelihood of a class focused on dissemination of words in records and utilized for report order. This procedure work with Bag of Words (BOW) include extraction which overlook position of words in archives.

The grouping approach can be gotten together with a choice principle, a typical standard being, to pick the speculation that is no doubt which is known as the best a back model or the MAP choice guideline<sup>[7]</sup>.

There are two first request probabilistic models for Naïve Bayes order are Bernoulli model and the Multinomial model<sup>[7]</sup>. The Bernoulli model is a Bayesian Network with no word conditions and parallel word highlights; it moreover delivers a Boolean pointer for every one term of the jargon relying on its quality or nonappearance; in this way how, the Bernoulli model likewise considers words that don't show up in the archive into record<sup>[7]</sup>. The Multinomial model is a unigram language model with whole number word tallies and when the recurrence of a word happening in a record checks; along these lines, a binarized form Of the Multinomial model is used which just considers the presence of a word yet not its recurrence<sup>[7]</sup>. It is break down that the multivariate Bernoulli performs well with little jargon sizes, anyway the multinomial model essentially performs far and away superior at bigger jargon sizes, giving on a normal 27% abatement in blunder over the multivariate Bernoulli model at any jargon size<sup>[7]</sup>.

### **2.2 Maximum Entropy**

Maximum entropy grouping (MaxEnt, or ME) is a component based<sup>[5]</sup> likelihood conveyance assessment model and an elective procedure which has demonstrated viable in various normal language handling applications. Guideline of greatest entropy is if very little is thought about the information or data, circulation ought to be pretty much as uniform as conceivable<sup>[7]</sup>. Essentially, dissimilar to Naive Bayes, MaxEnt makes no presumptions about the connections among highlights, thus may possibly performs better when restrictive freedom suppositions are not met<sup>[3]</sup>. This suggests it ought to permit adding highlights like bigrams and expressions to MaxEnt without agonizing over its element covering<sup>[5]</sup>

### **2.3 Support Vector Machine (SVM)**

Support Vector Machine (SVM) is another famous high edge factual order strategy proposed for slant examination and exceptionally successful for text classification<sup>[3]</sup>.

The primary head basic SVM for supposition arrangement is to find a hyper plane what isolates the records according to the assumption, and the edge between the classes being just about as high as could really be

expected; it additionally engaged around the Structural Risk Minimization rule <sup>[7]</sup>. Highlight determination is a significant undertaking in AI strategies; there are various highlights that should be considered for text grouping, to avoid over fitting and to build general exactness <sup>[7]</sup>. SVM can possibly deal with huge element spaces with high number of estimations.

To manage an enormous number of highlights, customary content arrangement strategies accept that a portion of the highlights are immaterial, anyway even the most minimal positioned highlights as per include determination techniques contain impressive data; considering these highlights as unimportant regularly bring about a deficiency of information <sup>[7]</sup>. Accordingly how the data misfortune can be limited as SVMs doesn't needs at the hour of making a supposition. Despite the fact that SVM outflanks every one of the conventional procedures for estimation arrangement, it is a discovery strategy <sup>[7]</sup>. It is hard to investigate the model of characterization and to recognize which words are more significant for order. This is one of the disservices of using SVM as a strategy for archive grouping <sup>[7]</sup>.

### III. A HYBRID POLARITY DETECTION SYSTEM

Modules contain in the Existing Hybrid Polarity Detection System are demonstrated as follow <sup>[9]</sup>:

#### 3.1 Pre-Processing of Data set :

Several Pre-Processing Steps for the Sentiment Summarization of the given data set is taken, this several Steps are:

- @username – removed the username and because these are not considers for sentiments.
- URLs – delete all string that describes links or hyperlinks.
- #hashtag - hash tags can give some helpful information, so it is helpful to replace them with the actually same word without the hash. E.g. #Dissertation replaced with Dissertation.
- The target (of sentiment) word is replaced by “TARGET”
- Lower Case – changed over all the content in a string to lower case.
- Stop words - a, an, is, the, with and so on, that don't demonstrate any sentiment and can be removed.
- Punctuations and additional white spaces – removed punctuation at the begin and closure of the tweets. E.g.: „today is my presentation.!, Replaced with 'today is my presentation'.
- Words must begin with an alphabet – deleted each one of those words which don't begin with an alphabet, for example 24th, 7:45pm

#### 3.2 Sentiment Feature Generator Module <sup>[9]</sup>

Several Features include in the Hybrid Polarity Detection System are as shown in the table, Measurements of all this features are required for further calculation.

F1	Document (or tweet) overall sentiment score using the unsupervised polarity detection algorithm
F2	Number of positive words
F3	Number of negative words
F4	Number of negation words
F5	Number of negation words followed by a positive word
F6	Number of negation words followed by a negative word
F7	Inverse sentiment
F8	Number of positive words followed by target
F9	Number of negative words followed by target
F10	Number of negation words followed by target
F11	Number of positive words followed by a negative word
F12	Number of negative words followed by a positive word
F13	Number of target words followed by a positive word
F14	Number of target words followed by a negative word

#### 3.3 Machine Learning Classifier <sup>[9]</sup>

Sentiment Summarization of a linear SVM that takes as input the feature set described in the previous subsection that contain opinion about some entity of interest and accordingly classifies tweets (documents) and generate summary of all input tweets. Now, the proposed approach from the above three module is done by adding two more feature with doing sentiment analysis on live twitter data set. This proposed approach is as shown here:



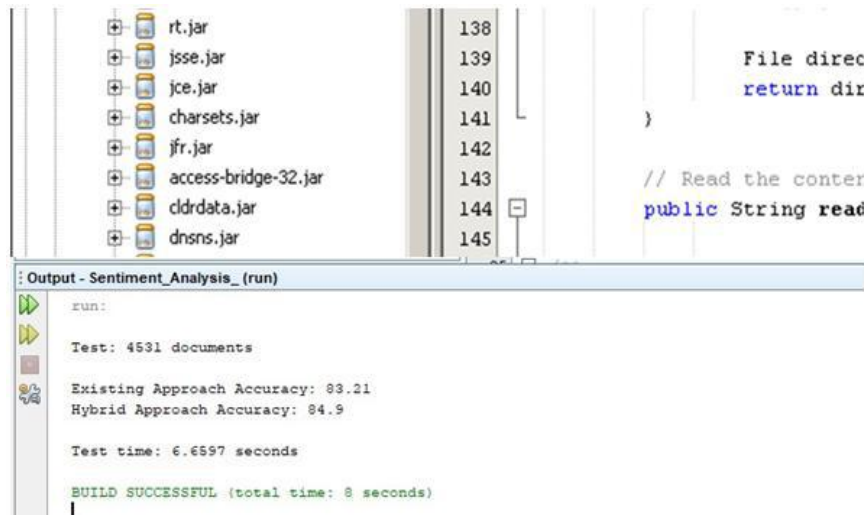


Fig: 2 Result Analyses

## V. CONCLUSION

As Sentiments of individuals are extremely useful for people and company owner for making several decisions, introduced proposed Hybrid Polarity Detection System for Sentiment Analysis and summarization that uses new set of features, tries to improve the accuracy compare to state-of-the-art techniques to get the clear idea about the marketing research auditing, public opinion tracking, product reviewing, business research, political review, enhancing of web shopping bases, and so on. As per our experiment, we believe that as the part of Sentiment Analysis, Moving towards Sentiment Features rather than manual text processing would be a promising outcome to these issues. Now, finding more features set that could help to improve the accuracy and also detection of sarcasm would be future work of this study.

## REFERENCES

- [1]. S. Uma Maheswari, Dr. S. S. Dhenakaran, "Opinion Exploration of Tweets and Amazon Reviews", International Journal of Scientific & Technology Rersearch, Jun 2020.
- [2]. Benito Alvares, Nishant Thakur, Siddhi Patil, —Sentiment Analysis Using Opinion Miningl, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278- 0181, Vol. 5 Issue 04, April-2016.
- [3]. Bo Pang, Lillian Lee and Shivakumar Vaithyanathan, Thumbs up Sentiment Classification using Machine Learning technique, proceedings of EMNLP,2002.
- [4]. Khairullah Khan, Baharum Baharudin, Aurnagzeb Khan and Ashraf Ullah, Mining opinion components from unstructured reviews: A review, Journal of King Saud University – Computer and Information Sciences, 2014.
- [5]. Alec Go, Richa Bhayani and Lei Huang, Twitter Sentiment Classification using Distant Supervision, CS224N project report, Stanford, 2009.
- [6]. Kamaxidevi Raol and Nimesh Patel, Survey on Opinion Mining on Online Reviews, International Journal of Research in Engineering and Science (IJRES), Feb 2020.
- [7]. Sagar Bhuta, Uehit Doshi, AvitDoshi and Meera Narvekar, A Review of Techniques for Sentiment Analysis of Twitter Data, Issues and Challenges in Intelligent Computing Technique (ICICT), 2014.
- [8]. Blessy Selvam and S.Abirami, A survey on Opinion Mining Framework, International Journal of Advanced Research in Computer and Communication Engineering, 2013.
- [9]. Seyed-Ali Bahrainian and Andreas Dengel, Sentiment Analysis using Sentiment Features, IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), 2013.
- [10]. Khairullah Khan, Baharum B. Baharudin, Aurangzeb Khan and Fazal-e-Malik, Mining Opinion from Text Documents: A Survey, 3rd IEEE International Conferences on Digital Ecosystem and Technology, 2009.
- [11]. Bing Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.