

# Sentiment Analysis : On Product Review

<sup>1</sup>Ugandhara Nikumbh, <sup>2</sup>Priti Nikumbh, <sup>3</sup>Bhagyashree Gaikwad

<sup>1,2,3</sup>BE Student, Late G. N. Sapkal College of Engineering, Savitribai Phule Pune University, Maharashtra, India.

<sup>1</sup>yugandharanikumbh786@gmail.com, <sup>2</sup>pritinikumbh786@gmail.com, <sup>3</sup>bhagyashreegaikwad111@gmail.com

**Abstract :** The Web has rapidly changed the way that people express their views and opinions. Now if one wants to purchase a product, people are no longer limited to asking their friends and families because there are many product reviews on the Web which give opinions of existing users of the product. Here we present the system which provides us information about such products and services in summarization form. Finding opinion sources and monitoring them on Web can still be a difficult task because there are large number of different sources, and each source may have a huge volume of text with opinions or sentiments. In most cases, opinions are hidden in conference posts and blogs. It is complicated for a human reader to find relevant sources, extract related sentences with suggestions, read them, summarize them, and manage into usable forms. Thus, automated summarization systems are needed. Using this summarization we can identify the importance, quality, popularity of product and services. In this system we can make summarization for product but we can use this system anywhere, where text analysis is required. Sentiment analysis is also known as opinion mining, grows out of this need. It is challenging natural language processing or text mining problem. Due to its tremendous value for practical applications, there has been an excessive growth of both research in academia and applications in industry.

**Keywords —** Natural language processing, Machine Learning, Sentiment Analysis, Opinion Mining

## I. INTRODUCTION

Our world has changed drastically in the last 10 years. A person's opinions are no longer shared only with his or her immediate family and friends, but instead have capability of influencing the decisions of thousands or millions of people the individual has never even met. The Internet has given the platform to broadcast grievances and recommendations that can reach across the world to an individual. And the existence of public networks gives these opinions the potential to snowball into a viral frenzy that can make your organization's products or services a worldwide boon or a whole catastrophe in just a matter of days. The savvy marketer checks and evaluates related web content continuously to understand consumer sentiment toward products or services from his organization – and toward his competitors. This attention to Web content allows the company to respond immediately to customer point of view. The volume of references related to your company's products or services makes automating this task necessary. resources such as blogs, product reviews, forums and news articles can all be monitored, scored for relevance against your topics of interest, and then summarized according to sentiment. Sentiment analysis is an automatic method that gives feedback to you regarding the opinions and attitudes of your customers.

## II. PROBLEM DEFINITION

Finding opinion sources and monitoring them on the Web can still be a difficult task because there are a large number of diverse sources, and every source may also have a big volume of opinionated text (text with opinions or sentiments). Selecting an attributes for sentiment classification using feature relation networks. In many cases, opinions are hidden in long forum posts and blogs. It is complex for a human reader to find relational sources, extract relational sentences with opinions, read them, understand them, and organize them into usable forms. Thus, automated summarization systems are needed. Using this summarization we can recognize the importance, quality, popularity of product and services. In this system we make summarization for movie. But, we can use this system anywhere, where text analysis is required. Sentiment analysis, also known as opinion mining, grows out of this need. It is a challenging natural language processing or text mining problem. Due to its tremendous value for practical applications, there has been an explosive growth of both research in academia and applications in the industry

We are designing one system that can help any organization for improving their products, analyze thousands of feedback

and provide generalized opinion for the product. The design can also be extended to other product-review domains easily.

This System is specifically made for-When an organization wanted to find opinions of the general public about its products and services, it conducted surveys and focus groups. Such that, with the explosive growth of the social media content on the Web in the past few years, the world has been transformed. People can now post reviews of products at merchant sites and express their views on almost anything in discussion forums and blogs, and at social network sites. Now if one wants to buy a product, one is no longer limited to asking ones friends and families because there are many user reviews on the Web. For a company, it may no longer need to conduct surveys or focus groups in order to gather consumer opinions about its products and those of its competitors because there is a plenty of such information publicly available.

### III. LITERATURE SURVEY

When we consider literature survey there are many drawbacks from existing systems as follows :

1. *Polarity shift problem* : In this when sentence contain positive as well as negative sentiments than system will confuse to give result or give wrong result.
2. *Dual sentiment analysis* : In this system will give positive and negative sentiments
3. *Dual Expansion and Dual Training algorithm* for reverse review : Dependency on external dictionary for review reversion
4. *BOW Problems* : BOW is now the most popular way to model text in statistical machine learning approaches in sentiment analysis. However, the performance of bag of words sometimes remains limited due to some fundamental deficiencies in handling the more complex polarity shift patterns such as transitional, subjunctive and sentiment-inconsistent sentences in creating reversed reviews[4].

### IV. SYSTEM MEHODOLOGY

#### A. METHODS OF SENTIMENTS



Fig A : The Process of Sentiment Analysis.

1. *Data Collection* : Consumers usually express their sentiments on public forums like the blogs, discussion boards, product reviews as well as on their private logs Social network sites like Facebook and Twitter. Opinions and feelings are expressed in different way, with different vocabulary, context of writing, usage of short forms and

slang, making the data huge and disorganized. Manual analysis of sentiment data is virtually impossible. Therefore, special programming languages like R are used to process and analyze the data.

2. *Text Preparation* : Text preparation is nothing but filtering the extracted data before analysis. It includes identifying and eliminating non-textual content and content that is irrelevant to the area of study from the data.

3. *Sentiment Detection* : At this stage, each sentence of the review and opinion is examined for subjectivity. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded. Sentiment analysis is done at different levels using common computational techniques like Unigrams, lemmas, negation and so on.

4. *Sentiment Classification* : Sentiments can be broadly classified into two groups, positive and negative. At this stage of sentiment analysis methodology, each subjective sentence detected is classified into groups-positive, negative, good, bad, like, dislike.

5. *Presentation of Output* : The main idea of sentiment analysis is to convert unstructured text into meaningful information. After the completion of analysis, the text results are displayed on graphs like pie chart, bar chart and line graphs. Carrying out sentiment analysis is an important task for all the product and service providers today.

#### B. DIFFERENT MODULES IN SENTIMENT ANALYSIS SYSTEM

##### 1. User profiles

*Users*: users with no particular knowledge needed, users who are interested to use the tool looking for knowing peoples thoughts about a de-sired topic.

*Advanced end users*: advanced users are those who have valuable in-put and feedbacks. Users who are more familiar with informative sites and can use our features efficiently. These valuable feeds will lead to enhancement of users satisfaction.

##### 2. System Operators:

Maintains for the functional interface of the application and troubleshooting issues. Suggest possible updates and identifying renewal application needs to Co-ordinate with service providers and infrastructure vendors Coordinate and communicate with system administrators

##### 3. System Administrators:

Develop and maintain installation and configuration procedures and operational requirements. Perform

weekly/monthly backup operations, ensuring all required files and data are successfully backed up Repair and recover from hardware or software failures Coordinate and communicate with system operators.

**C. WORKFLOW OF SYSTEM**

1. When user or admin logs in to view a product he will be able to see details of product
2. Each user comments and these comments are stored in the database against the product along with the users mail id.
3. These comments are then passed to the GATE Processor through which the gate processor segregates these comments into verb, adverb, noun, adjective etc. The corpus taken an important role in this section.
4. Co referencing able to find 'and', 'but' etc clause in the process we can correlate.

For example Iphone is costly but it has amazing picture quality.

Here Iphone is noun, but is clause which derive the meaning

5. JAPE rules are basically used to define the negative and positive word dictionary that we used in our system.
6. The user is able to derive meaning from these words by assigning scores and we have implemented methods to add and subtract the below mentioned score [V]
7. Based on derived score we display the output in n point format ie. Positive, Negative, Neutral etc. Show result in graphical format.

the Internet is now a vast resource of opinion rich textual data. The goal of Sentiment Analysis is to harness this data in order to obtain important information regarding public opinion, that would help make smarter business decisions, political campaigns and better product consumption.

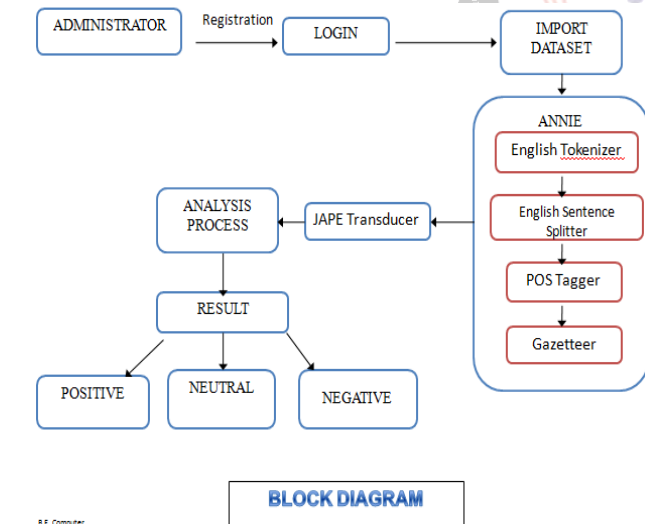
Sentiment Analysis focuses on identifying whether a given piece of text is subjective or objective and if it is subjective, then whether it is negative or positive. The recent trends in Sentiment Analysis techniques have moved towards building generative models that can capture complex contextual phenomena. Conversely, due to the unavailability of annotated data, the focus is moving towards unsupervised approaches that use the power of co-occurrence to solve the problem. Since, the web has a huge amount of opinionated data, in the form of blogs, reviews, etc., the unsupervised approaches flourish.

**E. APPLICATIONS OF SENTIMENTS SYSTEM**

1. Sentiment Analysis has many applications in various fields Tools that help summarize the sentiment regarding a product or service help users in identifying their product of choice.
2. Applications to Review-related Websites
3. Applications as a Sub-component Technology
4. Applications in Business Intelligence
5. Applications across different Domains such as business, Studies in sociology, human emotions especially on social networks
6. Applications in smart homes: Smart homes are supposed to be the technology of the future

**F. FUTURE SCOPE**

Sentiment Analysis has been more than just a social analytic tool. Its been an interesting field of study. But it is a field that is still being stud-ied, although not at great lengths due to the intricacy of this analysis. That is this field has functions that are too complicated for machines to understand. The ability to understand sarcasm, hyperbole, positive feelings, or negative feelings has been difficult, for machines that lack feelings. Algorithms have not been able to predict with more than 60 percent accuracy the feelings portrayed by people. Yet with so many limitations this is one field which is growing at great pace within many industries. Companies want to accommodate the sentiment analysis tools into areas of customer feedback, marketing, CRM, and ecommerce



**Fig C: Block Diagram of Sentiment Analysis**

**D. NECESSITY OF SENTIMENTS**

Sentiment Analysis (SA) is one of the most widely studied applications of Natural Language Processing (NLP) and Machine Learning (ML). This field has grown tremendously with the advent of the Web 2.0. The Internet has provided a platform for people to express their views, emotions and sentiments towards products, people and life in general. Thus,

There is too much potential in machine learning, overtaking some of the manual labor of some lexicon based tasks that are labor intensive. For ex-ample, lexicon sentiment creation is labor intensive and there are already unsupervised methods to create them. This is where machine learning will play a crucial role. Such algorithms will also have to understand and analyze natural text concept-wise and context-wise. Time will also be a

crucial element looking at the amount of data that is being generated on the Web today. Collecting opinions on the web will still requires processing that can filter out un-opinionated user-generated content and also to test the trustworthiness of the opinion and its source

Despite all the challenges and potential problems that threatens Sentiment analysis, one cannot ignore the value that it adds to the industry. Because Sentiment analysis bases its results on factors that are so inherently humane, it is bound to become one the major drivers of many business decisions in future. Improved accuracy and consistency in text mining techniques can help overcome some current problems faced in Sentiment analysis. Looking ahead, what we can see is a true social democracy that will be created using Sentiment analysis, where we can harness the wisdom of the crowd rather than a select few experts. A democracy where every opinion counts and every sentiment affects decision making.

**V. DATASET DESCRIPTION**

In the languages where lexical resources are abundant, a straightforward way is to get the antonym dictionary directly from the well-defined lexicons, such as Word Net4 in English. Word Net is a lexical database which groups English words into sets of synonyms called syn sets, provides short, general definitions, and records the various semantic relations between these synonym sets. Using the antonym thesaurus it is possible to obtain the words and their opposites. The Word Net antonym dictionary is simple and direct. However, in many languages other than English, such an antonym dictionary may not be readily available. Even if we can get an antonym dictionary, it is still hard to guarantee vocabularies in the dictionary are domain-consistent with our tasks.

**A. DATASET**

The Multi-Domain Sentiment Datasets5 are used as the English datasets. In this project we used amazon online shopping dataset. from Their website : <http://www.amazon.com> They contain product reviews including different domains: such as Book, Mobile, Camera, TV etc Electronics and Kitchen appliances. Each of the reviews is rated by the customers from score Zero to Hundred.

**Score calculations**

1. Positive word : +0.5
2. Negative Word : -0.5
3. Hot list of words: e.g. very/ mostly : 0.25
4. Sentence evaluation as based on the logic: ++,--,+,-,+-

Method to add and subtract these score as follows :

++ = Positive
-- = Positive
+- = Negative
-+ = Negative

**FORMULAE :**

$$T1 = \frac{\sum p}{n} \quad (1)$$

$$T2 = \frac{\sum N}{n} \quad (2)$$

$$TT = T1 + T2 \quad (3)$$

Where,

T1 = Total analysis of positive words

T2 = Total analysis of negative words

$\frac{\sum p}{n}$  = Number of positive words/Number of Sentences

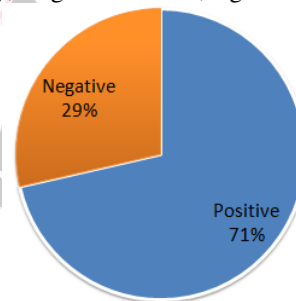
$\frac{\sum N}{n}$  = Number of Negative words/Number of Sentences

TT = Total Analysis

**VI. RESULT REVIEWS**

Manually given comment example:

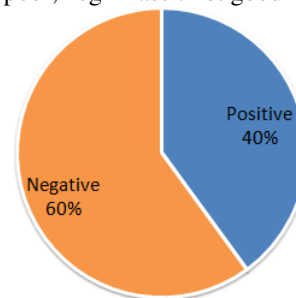
- Apple iphone 5 is very nice phone its functions and performance is excellent and good picture quality  
Positive Review->Score: 2.25 Pos Words: nice, excellent, good, quality, Neg Words: Null,negPhrase : null



**Fig VI (a):-Graph for Positive feedback**

- Apple iphone 5 is not good phone its functions and performance is not so good and also picture quality is poor.

**Negative Review->** Score: 0.0 Pos Words: good, good, quality, Neg Words: poor, negPhrase : not good



**Fig VI(b) :-Graph for Negative feedback**

- Apple iphone 5 is nice phone its functions and performance is also good but its picture quality is poor so it made me unhappy.

**Neutral Review** -> Score: 0.75 Pos Words: nice, good, quality, Neg Words: poor, unhappy, negPhrase : null

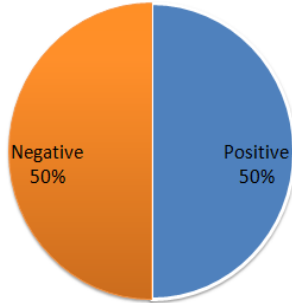
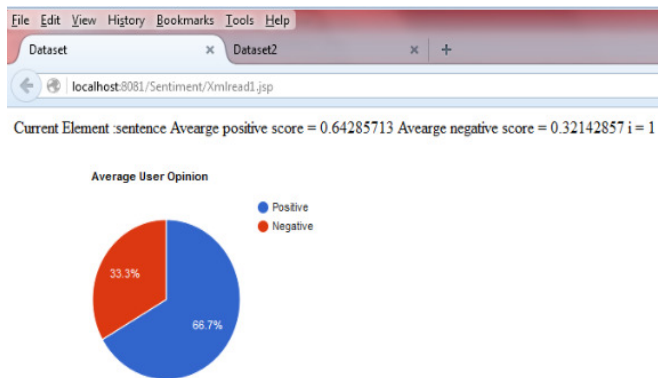


Fig VI(c) :-Graph for Neutral feedback

**Dataset of Computer :**



**Dataset of Router :**

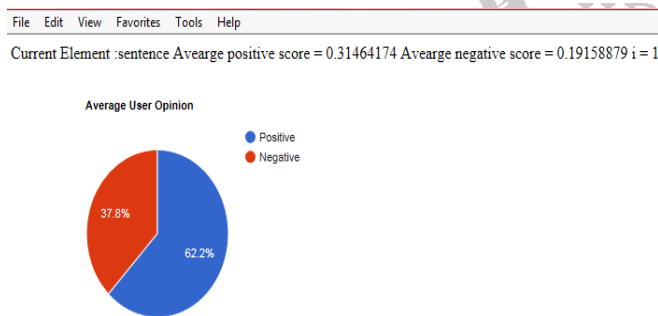


Table VI(a):An Example of Creating Reversed Training Reviews:

Reviews	Review text	Class
Original review	I dont like this movie.It is boring	Negative
Reversed review	I like this movie. It is interesting	Positive

**Reverse Training Review :**

Reviews	Review text	Class
Original review	The room is large. But it is not clean.	Neutral
Reversed review	The room is small. But it is clean	Neutral

Table VI(b) : The dataset in sentiment classification

Dataset	#positive	#negative	#neutral
Mobile	2000	2000	-
TV	1000	1000	-
Laptop	1000	1000	-
Camera	2000	2000	-
Hard Disc	2000	542	-
Pen Drive	540	540	240
Tab	763	763	440

**VII. CONCLUSION**

On real-life applications, to provide a completely automated solution is nowhere in sight. However, it is possible to devise effective semi automated solutions. The key is to fully understand the whole range of issues and pitfalls, cleverly manage them, and determine what portions can be done automatically and what portions need human assistance. In the continuum between the fully manual solution and fully automated solution, we can push more and more toward automation. Till today, the existing system manually analyze the sentiments. By using this system, the analyzing of sentiments will be done automatically.

**ACKNOWLEDGMENT**

It is our immense pleasure to work on this project. It is only the blessing of our divine guide Prof. N. K. Zalte who has prompted and mentally equipped us to undergo the study of this project. We are also grateful to our entire staff of Computer Engineering Department, Family and Friends for their kind co-operation which helped us in successful implementation of project.

**REFERENCES**

- [1] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li, "Dual Sentiment Analysis: Considering Two Sides of One Review", IEEE transactions on knowledge and data engineering, vol. 27, no. 8, august 2015
- [2] A. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Trans. Knowl. Data Eng., vol. 23, no. 3, pp. 447-462, Mar. 2011.
- [3] E. Agirre, and D. Martinez, Exploring automatic word sense disambiguation with decision lists and the web, in Proc. COLING Workshop Semantic Anno-tation Intell. Content, 2000, pp. 1119.
- [4] J. Cano, J. Perez-Cortes, J. Arlandis, and R. Llobet, Training set expansion in handwritten character recognition, in Proc. Struct., Syntactic, Statistical Pattern Recognit., 2002, pp. 548556.
- [5] Y. Choi and C. Cardie, Learning with compositional semantics as structural inference for subsentential sentiment analysis, in Proc. Conf. EmpiricalMeth-odsNatural Language Process., 2008, pp. 793801.
- [6] Sameer N. Kolhe, Swati V. Darade, Pratiksha B. Suryavanshi, Pramod S. Ahire, User Feedback Analysis System Using Natural Language Processing and Artificial Intelligence, IJRITCC, Volume 3, Issue 9,ISSN : 2321-8169 5574-5577.