

# Insight to Emotional tones in WhatsApp Through Sentiment Analysis

<sup>1</sup>Ritinder Kaur

<sup>1</sup>Asst. Professor

<sup>1</sup>IT Department,

<sup>1</sup>SGTBIMIT(GGSIPU), India

**Abstract:** Every day the social media networks (SMN) are generating massive amounts of data that may be structured, unstructured or semi-structured. The data may range from a normal text to a graphic image or an audio or video. Analysing this varied and ever-growing data is a big challenge. This paper focusses on extracting and analyzing data from a much used online text message application WhatsApp through the process of Sentiment Analysis. Sentiment Analysis or also known as opinion mining is contextual mining of data to identify, extract and analyse the underlying sentiment in messages and classify them to be positive, negative or neutral. R language has been used in this paper to understand the different emotions in a WhatsApp chat.

**IndexTerms** - Semantic analysis, opinion mining, Social Media Network, WhatsApp, R Language, Analytics

## I. INTRODUCTION

The burst of Internet and mobile devices has reformed the way information is originated and distributed. Microblogs such as Twitter, social networks such as Facebook, and instant messengers such as Skype and WhatsApp are now commonly used to share opinions and ideas about anything happening in the world. Opinions are central to just about all human activities. Our views and insights of reality, and the selections we make, are, to a considerable degree, conditioned upon how others see and evaluate the world. Therefore, when we need to make a decision we often seek out the opinions of others. This is not only true for individuals but also true for organizations. This has given rise to the massive volume of social media content, consequently producing new prospects to study public opinion at a scale that was never possible before. Unsurprisingly, this abundance of data has quickly attracted business and research interest from various fields including marketing, political science, and social studies, among many others which are interested in the study of opinions of people expressed in social media. This has escalated the growth of the research in the field of sentiment analysis in social media. Twitter is the most popular medium for research study due to its scale, diversity of topics discussed, short size of tweets which simplifies linguistic analysis and the ease of public access to its data. Also, the streaming nature of Twitter allows for real-time analysis and its social aspect encourages people to share opinions.

Sentiment Analysis is based on the concept of Natural Language processing (NLP), Text Analysis and computational Linguistics. It recognizes, extract, measure and study different states in a subjective information. Natural language processing (NLP) is the capability of a computer program to understand human language as it is spoken. Human speech, however, is not always precise -- it is often ambiguous and the linguistic structure can depend on many complex variables, including slang, regional dialects and social context. It uses a list of words and expressions used to express people's subjective feelings and sentiment or opinions.

## II. ANALYSIS

**2.1** Based on the level of granularities (Liu B.,2012), sentiment analysis has been investigated mainly at three levels:

1. Document level: The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment (Pang, Lee and Vaithyanathan, 2002; Turney, 2002). From a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is generally known as document-level sentiment classification. This level assumes that each document expresses opinions on a single entity. Therefore, it is not pertinent to documents which evaluate or compare multiple entities.

2. Sentence level: At this level, the task goes to the sentence level and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification (Wiebe, Bruce and O'Hara, 1999), which differentiates sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. Though, it must be noted that subjectivity is not correspondent to sentiment as many objective sentences can imply opinions.

3. Entity and Aspect level: Aspect level achieves finer-grained analysis. Aspect level was earlier called feature level (feature-based opinion mining and summarization) (Hu and Liu, 2004). Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion). Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better. In many applications, opinion targets are described by entities and/or their different aspects. Thus, the goal of this level of analysis is to discover sentiments on entities and/or their aspects.

### 2.2 CLASSIFICATION OF SENTIMENT ANALYSIS

Sentiment Analysis - "It's the computational study of people's opinions, appraisals and emotions toward entities, events and their attributes. Opinions are important because whenever we need to make a decision, we listen to other's opinions" {Liu B, 2012). The sentiment classification techniques are illustrated below in fig.1

a. Lexicon Based Approach- The lexical or lexicon approach is a method to calculate orientation of a document from semantic orientation of words and phrases. The elementary idea of this approach is to understand and produce lexical phrases as chunks. This sentiment analysis method uses a list of words and expressions used to express people's subjective feelings and sentiment or opinions. It uses words, phrases and idioms. The types of sentences are furthered analyzed using 3 approaches: Manual approach, dictionary-based approach and corpus-based approach.

- i) Manual Approach: This approach is not feasible as it is time consuming.
  - ii) Dictionary based approach: This approach uses 'Word Net' to find appropriate words of the sentiment word to carry out the analysis.
  - iii) Corpus-based approach: This is used to create a domain-specific sentiment lexicon to carry out the analysis.
- b. Machine Learning Approach-Machine learning is a subset of computer science that developed from the study of pattern acknowledgement and computational study in artificial intelligence. Machine learning give computer the learning ability with explicit program which not only explores the study and creation of algorithms hat can make a prediction on data. The algorithms construct a model from an example training set of input classifications in order to make datadriven predictions as outputs. It as a strong mathematical optimization. There are two different types of machine learning
- i) Supervised Learning -Supervised learning is the machine learning task of deducing a function from labelled training data. The training data contain a set of training examples.
  - ii) Unsupervised Learning -Unsupervised learning is a type of machine learning algorithm, which is used to draw inferences from datasets containing input data without labelling reactions. The most common unsupervised learning method is cluster analysis, which is used for data analysis to determine hidden patterns or group in data.
- c. Hybrid approach hybrid approach is used for both combinations of machine learning and lexicon-based approach. encompassing the use of semantic rules, fuzzy sets, unsupervised machine learning techniques and a sentiment lexicon improved with the support of WordNet. A Hybrid Standard Classification is first conducted, and then further enhanced into a Hybrid Advanced approach integrating linguistic classification of semantic polarity modelled using fuzzy sets.

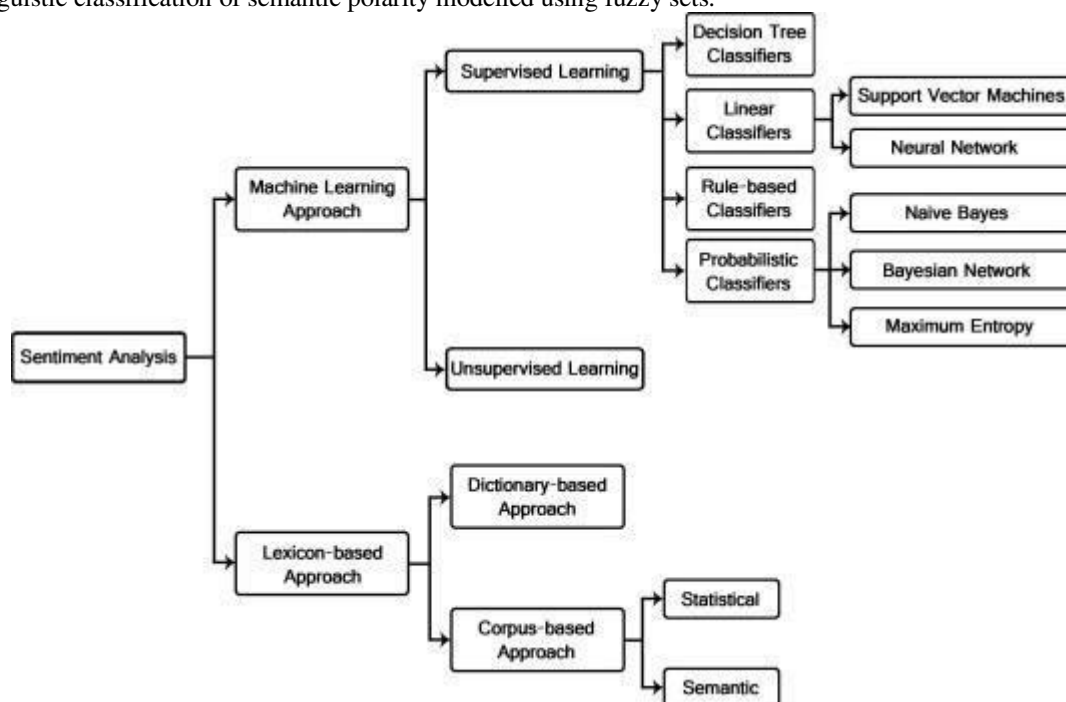


Fig 1 Sentiment Analysis Classification

### III. TOOLS AND TECHNIQUES

Before Sentiment analysis is used for many applications, especially in business intelligence [6]. Some examples of applications for sentiment analysis include:

- Analyzing the social media discussion around a certain topic
- Evaluating survey responses
- Determining whether product reviews are positive or negative

The paper uses R Language to analyze sentiments extracted from a popular text message application WhatsApp. It can help in understanding the overall sentiment of the text messages in a chat. R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R provides a wide variety of statistical time-series analysis, classification, clustering techniques and graphical techniques, and is highly extensible.

#### 3.1 THE R ENVIRONMENT

R is an integrated suite of software facilities for data manipulation, calculation and graphical display [6]. It includes

- an effective data handling and storage facility,
- a suite of operators for calculations on arrays, in particular matrices,
- a large, coherent, integrated collection of intermediate tools for data analysis,
- graphical facilities for data analysis and display either on-screen or on hardcopy, and
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

#### 3.2 THE PROCESS OF SENTIMENT ANALYSIS

There are different ways to implement Sentiment Analysis, but the most general idea is as follows:

- Create or extract a list of words associated with strongly positive or negative sentiment.
- Count the number of positive and negative words in the text.

- Analyze the mix of positive to negative words. Many positive words and few negative words indicate positive sentiment, while many negative words and few positive words indicates negative sentiment.

#### IV. RESULTS

R Language generates results from different lexicons embedded in packages. In this paper the sentiments in chats had been classified into 10 different categories to analyse the orientation of the group. The text was extracted using tools and then parsed to yield most appropriate words. Classification of text was done by comparing them with lexicons in embedded libraries. Figure 2 shows the data classified into 10 different categories:

S.No	Sentiment	Total Count
1.	Anger	2
2.	Anticipation	37
3.	Disgust	3
4.	Fear	13
5.	Joy	12
6.	Sadness	6
7.	Surprise	10
8.	Trust	30
9.	Negative	25
10.	Positive	56

Table 1. Output of Experiment on sample data

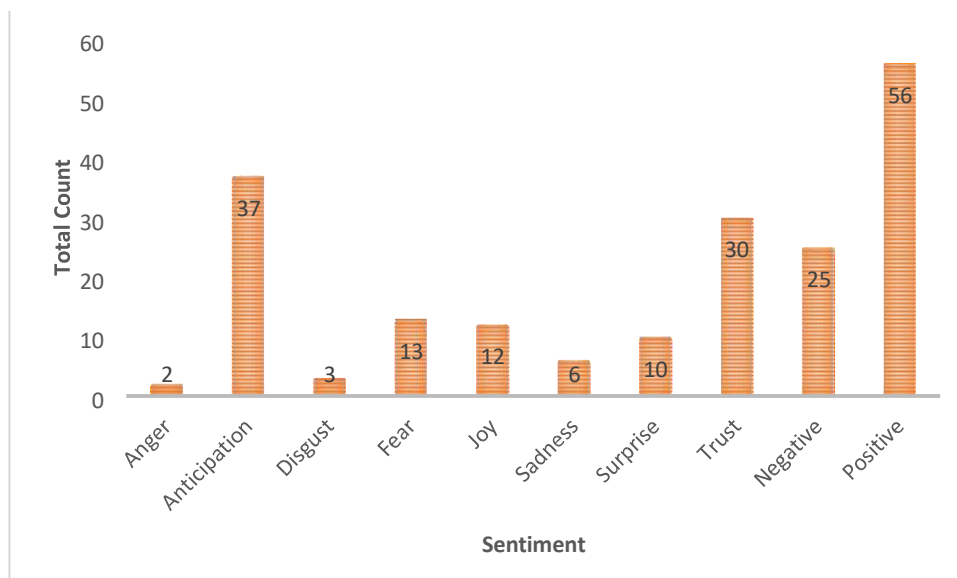


Fig.2 Data Analysis Report of Result Set

D From Figure 3, it can be concluded that the overall polarity of the Text message application WhatsApp group taken as sample data is positive.

#### V. CONCLUSION

This paper has explored the concepts of Sentiment Analysis and Data Analytics. Despite the wide variety of knowledge sources explored so far in the literature, sentiment polarity lexicons remain the most commonly used resource for the task of sentiment analysis. The concepts can be further explored to know more issues and challenges in the area of Sentiment Analysis. The study can be further extended to Machine Learning approaches to classify the text and make predictions through Supervised or Unsupervised Learning.

## REFERENCES

- [1] Ali, Liu B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. 2012 May 22;5(1):1-67.
- [2] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002). 2002. asu, S. 1997. The Investment Performance of Common Stocks in Relation to their Price to Earnings Ratio: A Test of the Efficient Markets Hypothesis. *Journal of Finance*, 33(3): 663-682.
- [3] Wiebe, Janyce, Rebecca F. Bruce, and Thomas P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. in Proceedings of the Association for Computational Linguistics (ACL-1999).1999.
- [4] Hu, Minqing and Bing Liu. Mining and summarizing customer reviews. in Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004). 2004.
- [5] Lewis M. Pedagogical implications of the lexical approach. *Second language vocabulary acquisition: A rationale for pedagogy*. 1997:255-70.
- [6] <https://www.r-project.org/about.html>
- [7] Saif H, He Y, Alani H. Semantic sentiment analysis of twitter. In International semantic web conference 2012 Nov 11 (pp. 508-524). Springer, Berlin, Heidelberg.
- [8] Edison, M., & Aloysius, A. (2016). Concepts and Methods of Sentiment Analysis on Big Data. *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, 16288-16296.
- [9] Soni, R., & Mathai, K. J. (2015). Improved twitter sentiment prediction through cluster-then-predict model. *arXiv preprint arXiv:1509.02437*.
- [10] Tauch, C. and Kanjo, E., 2016, September. The roles of emojis in mobile phone notifications. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct* (pp. 15601565). ACM.