**A Novel Approach to Detect Sentiments Based on Social Media Using Machine Learning**

Neetesh kumar Nema1, Bharat Choudhary2, Dileshwar Patel3

1Research Scholar Dr. C. V. Raman University Bilaspur C.G.

2,3Department of Computer Science and Engineering, LCIT Bilaspur

hi.bharat2002@gmail.com2

**Abstract**

Sentiment analysis of some product or service can be favorable in predicting future scope of it. However, manually analyzing a large number of documents in a limited time can be a monotonous and challenging task. Hence, several attempts have been made in the writing to solve this problem and several sentiment analysis techniques have been proposed. There is variety number of machine learning algorithms existing using supervised or semi supervised technique. These algorithms be appropriate unigram, bigram, n-gram or hybrid approach. Semi supervised learning is being used for this research paper. In this work, different approach are combined together to form novel model that uses variety of approach and results were found. This work of fiction approach gave a better result. People all the time more share a lot of content on social media intend to find out memes that convey their feelings. Image and textual Sentiment Analysis (SA) is in advance pace as a result of the rising predisposition of expressing sentiments through images and writtendescriptions. Photographs and videos are ever more being used by social media users to express themselves and share their experiences.

***Keyexpressions***- Sentiment-Analysis***,*** Emoticons, Social Media, OCR, Image, Natural language Processing, Machine Learning

**1 Introduction**

Nowadays in field of computer science and technology the hottest research area is sentiment analysis. Sentiment analysis can be defined as technique/method of identifying the view of people, given in the form of text regarding to a specific object (event, individual, decision, change etc.). Other synonyms of sentiment analysis are opinion mining, confidence analysis, people attitude towards an object, deriving opinion etc. The main reason behind the popularity of sentiment analysis it gives us overview of wide spread public opinion/thinking related to a topic.

Sentiment associated to a particular object is categorized in one of the following category:

* Positive
* Negative
* Neutral

Sentiment analysis is used at multilevel, it can be used to identify sentiment hidden in a document, to be more précised the analysis can be used to calculate the sentiment associated with each paragraph or may be each line.

The basic method adopted to identify the overall sentiment score is tokenization of each sentence in document into expressions, further each expression is categorized into positive, negative and neutral expressions. In next step further, these expressions are categorized according to associated impact (for example: extremely happy is having more impact than happy), finally summation of numbers of positive and negative expression is done to find out overall sentiment score.

The major challenges in sentiment analysis are:

* Multilingual
* Sarcasm
* Emoticons handling
* Natural language processing overheads

To increase the accuracy and overcome challenges of sentiment analysis, there are advance sentiment analysis mechanism which incorporates the sarcasm and emoticons handling technique. Moreover, nowadays natural language processing software’s (such as Open NLP by Apache) are also used in the process of semantic analysis.

**Example**: I am very glad with India’s succeed in hockey match.

If we tokenize the given expression there are two positive expressions: glad and succeed, and there is no negative expression, so it will produce overall positive sentiment score

Sentiment analysis is a involvement of Natural Language Processing (NLP) which guarantees with the computational actions of belief, emotion, unbiased and disinterest inside the given text. The term sentiment includes feelings, conclusions, activities, emotions and others. As the size of number of users and uses of social media are promptly expanding, Social Network Analysis (SNA) has become a massive instrument for specialists and analysts in social computing. Sentiment classification refers to the classification of a text based on the sentiment polarities present in it. Sentiment analysis of some product or service can be valuable in predicting future scope of it. However, manually analyzing a large number of documents in a limited time can be a monotonous and demanding task. Memes on social media sites such as Facebook, Instagram, and Twitter have become a subject of great interest in recent years as the Internet has full-fledged in esteem. Memes have become prevalent internet content over the last decade, often via social media platforms and especially for entertaining causes. This trend has aided in the analysis of their thoughts, feelings on a specific topic, and opinions on a explicit topic through Pictures shared on social media utilizing Machine Learning (ML), SA, emotion analysis, or opinion mining. The process of significant and categorizing opinions in a given piece of text as positive, negative, or neutral is known as SA, also known as opinion mining.

One of the most important variables for human expression beings to interpret their sentiments and motives in contact is facial expression recognition Facial expressions have become one of the most important information networks in interactive communication, encompassing a wide range of information and carrying emotional value. Visual SA tries to deduce the emotion elicited by images.

Sentiment analysis is the opinion mining used on the web for identify the text. It is nothing but to get the real voice of people for specific item for consumption, services, movies, news, issues etc. Sentiment analysis can be done at various levels as, sentence level, document level and entity or attribute level. The attitude of a particular person may be his/her judgment for the particular invention. Opinion or Feedback is very important for consumer as well as producer because most of the people sale or purchase the products online. Individual consumers may want the opinions of already existing users for the product before purchasing it.

**2. Related Work**

Essentially, Sentiment Analysis is used to articulate individual person’s sentiment. According to current state of the art sentiment analysis is used to categorize sentiments into two categories positive and negative. Some works classified them into as positive, negative and also in one more group as neutral.

**C. Hauff et al.** gives the way how to handle the negation expressions like not, no, neither, couldn’t, etc expressions in the sentence. It may happen that even if the negative expressions are present in the sentence still its meaning is in a positive way.

**A. Neviarouskaya et al**. performs fine grained categorization of sentences using ten categories: nine emotions (‘Anger’, ‘Disgust’, ‘Fear’, ‘Guilt’, ‘Interest’,’ Joy’, ‘Sadness’ (‘distress’), ‘Shame’ and ‘Surprise’) and neutral.

**Anurag P. Jain et al**. in which they extracted the information related the political stories using Twitter API v 1.1.This paper describe the mechanism that can be used to predict overall sentiment inclination of people towards political issues and situations. In the defined methodology two information sets are prepared after preprocessing the raw stories: training information set of stories and testing information set of stories. Further, authors performed the sentiment analysis to build a model to categories the tweet into one of the following category positive, negative and neutral.

**Antonio Teixeira et al.**  are also working on sentiment analysis on the basis the information from Facebook. The main purpose of authors behind this paper is to explain the process of Facebook information extraction, information preparation and sentiment analysis (using open source tools).

**Rincy Jose et al.** proposed an approach for performing sentiment analysis of twitter messages based on lexical resources SentiExpressionNet and ExpressionNet along with Expression Sense Disambiguation. For better accuracy they also implemented the negation handling in the preprocessing step. They have presented their methodology including three steps: Information acquisition, Preprocessing and Sentiment classification.

**J.M. Weibe et al.** the researcher brings out different algorithms in best identification of sentiment analysis.

**M. A. Hearst et al.** had come up with adding intelligence to sentiment analysis. Different machine learning methods are being used by researchers.

**V. Suresh et al.** presented an approach that used stop expressions and gaps between stop expressions as the feature for sentiment analysis.

**Murthy G et al.** made a comparative study on sentences and web context based sentiments.

**Pang et al.** suggested with unigram approaches in their research work.

**Dave et al.** used a tool to synthesize reviews.

**3. Problem Definition**

To understand people's thoughts and feelings based on their proposed text. It gives an overview of the different sentiments classification approaches and tools used for sentiment analysis. The machine learning approach is used for predicating the polarity of sentiments based on trained data sets. In this study automated analysis of social media is accomplished by building predictive model.

**4. Objective of Research**

The fundamental objective of this research topic is to spot out the emotions and opinions of the customers or users via a text basis. Analyze Social media information to detect any factors that may replicate the depression of relevant social users. To identify machine learning methods is most effective at identifying depression in social media comments. Identify the factors to look for depression detection in Social annotations. improve the performance using the essential idea to use another skill to take out more types of emotional features. Improvement of the proposed method use authentication, efficiency and effectiveness.

**5. Proposed Methodology**

In proposed methodology R studio (IDE) will be used for the purpose of information analysis. Proposed methodology for analysis include four major phases-

1. Collection of information of target figure from his/her social media account.
2. Preprocessing of information.
3. Mining the information to find hidden patterns and trends
4. Present the different Knowledge Pattern

**A. Collection of information of target figure from his/her social media account** - Basically, this methodology is focusing on extracting information from two platforms-

1. **Facebook -** The simple way to connect to Facebook API and extracting the information is by using the Facebook APP. Following are the steps involved in this-
   1. Install and load require package in R studio.
   2. Create app on Facebook developer platform
   3. Establish Connection between R studio and Facebook through app authentication key.
   4. After successful establishment of connection, extract the required information.

There are many other methods available for information extraction such as Facebook Graph API and third-party tools.

1. **Instagram**  **-** In order to extract the information from Instagram we can use instagram App. Following are steps involved in this-
2. Install and load required packages in R studio to support instagram connection.
3. Create an app on instagram and get the Consumer key (API key) and Secret key (API Secret).
4. Establish the connection between instagram app and R studio with the help of authentication key mentioned in previous step.
5. After successful authentication extract the required information.

**B. Preprocessing of Information -** Initial phase helps in removal of all other information present in all other languages except the English language. For the next level of preprocessing first we need to load the target file in a user defined object. Further the target information columns are loaded into the information corpus (collection of documents containing (natural language) text. In next step the operations are carried out on the information corpus to clean the information.

Figure shown below presents the complete method of preprocessing-

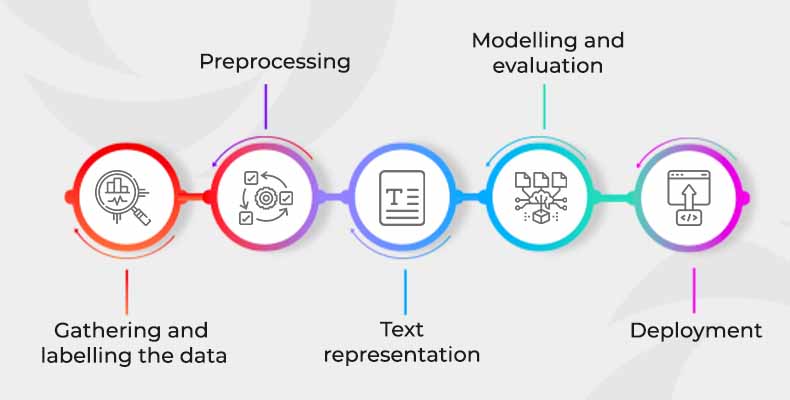
****

Fig- Sentimental Analysis Process

Level 2 of preprocessing of information basically involves:

1. **Removing punctuations**- English language is supported by different punctuation marks such as dot (.), coma (,) etc. However, punctuations are meaningless whenever we have to perform analysis, so it is become important to remove the punctuation marks.
2. **Removing white spaces-** It might be possible that extracted text contain unwanted whitespaces, which may act as noisy information during analysis. For better results, it is advisable to remove the white spaces.
3. **Converting all the text into lower case**- Most of the analysis/mining code treat are case sensitive, so to reduce errors it is advisable to have our all our text in same case.
4. Remove the stop expressions of English-Stop expressions are comprised of general expressions which is to support our sentence such as I, me, my, do, should etc. However, these are not important from the information analysis point of view, so it is advisable to remove such expressions.
5. **Stemming of the expressions-** For the analysis purpose it is important to convert all the expressions to base expressions such as played is converted to play.
6. **Removing other unwanted expressions or symbols-** It might be the case that we want to remove certain targeted expression from the text file for more specific results.

**C. Mining the information to find hidden pattern and trends –**

In our proposed methodology we are basically focusing on analysis for the purpose of feature extraction of target candidate and confidence of people related to the candidate. So for the general analysis we can use the Information Corpus and Document term matrix for presenting the generating the knowledge based on the frequency of expressions such as expression cloud and 'n grams'.

For more depth analysis one need to perform natural language analysis. There are certain open sources libraries are available that we can integrate with R studio such as OpenNLP (provided by Apache). The basic aim of semantic analysis is to generate a plot that represents the score of a candidate on the basis of kind of expression used in his/her posts/stories. additional we can identify confidence of a candidate among people with the help of semantic analysis of people comment and stories information.

To find out the confidence score the people opinion information can be broken down into single expressions (tokens). For the purpose of converting the text document into stream of tokens some natural language processing tool is required such as OpenNLP by Apache. Further with the assistance of NLP tool expression can be categorized into a scale (-5 [very negative expression] to +5 [very positive expression]) as shown in Figure



Fig- Proposed scale for calculate sentiment information for each expression

**D. Presenting the Knowledge -** To present the knowledge different graphical structures can be used-

1. Expression Cloud is an image comprised of expressions; in which size of expression depend on the frequency of expression in the document.
2. N- gram in Corpus is to present the group of expressions which are used together.

**5.1 Phases of the Methodology-**

**1) Information collection**: Information collection is the process of gathering and measuring the information from different resources of the interest in a systematic fashion that will enables us to answer the questions, test hypotheses, and evaluate the result. The information collection will be done from the reviews of the customers.

**2) Identification of information**: It identifies the information according to its value and what we are going to use. After information has been identified it will be given as an input to the system.

**3) Pre-processing:** The pre-processing is done to remove the unnecessary expressions or irrelevant expressions from the customer’s opinion. Our system deals only with the description of part of speech each review, processing means breaking review into expressions to differentiate it.

**4) Part of Speech tagging**: it parses each sentence and yields the part- of speech tag of each expression whether the expression is a noun, verb, adjective, adverb, etc and identifies simple noun and verb groups. If one sentence contains no frequent feature but one or more opinion expressions, find the nearest noun or noun phrase of the opinion expression as an infrequent feature. Speech tagging is used to identify every expression of feedback as either noun, adjective or adverb

**5) Negation detection**: It is also an important element of implementing sentiment analysis by using term scores, since negation in a sentence such as “I did not find this product funny or interesting” would invert the opinion orientation of otherwise positive terms such as “funny” and “interesting”.

**6) Stop Expressions Removal:** We eliminate expressions like prepositions, digits, articles and nouns like name of product etc with the help of Parts of speech tagging method as their existence are meaningless in system. It assists better extraction of opinion phrases/expressions from the tagged file.

**7) Rule-Based approach:** In the Rule-Based approach, rules are to be defined which contains some defined relation which have originator and its associated resultant. In this methodology, certain rules are to be defined and then the sentiments should be viewed or analyzed depending on rules

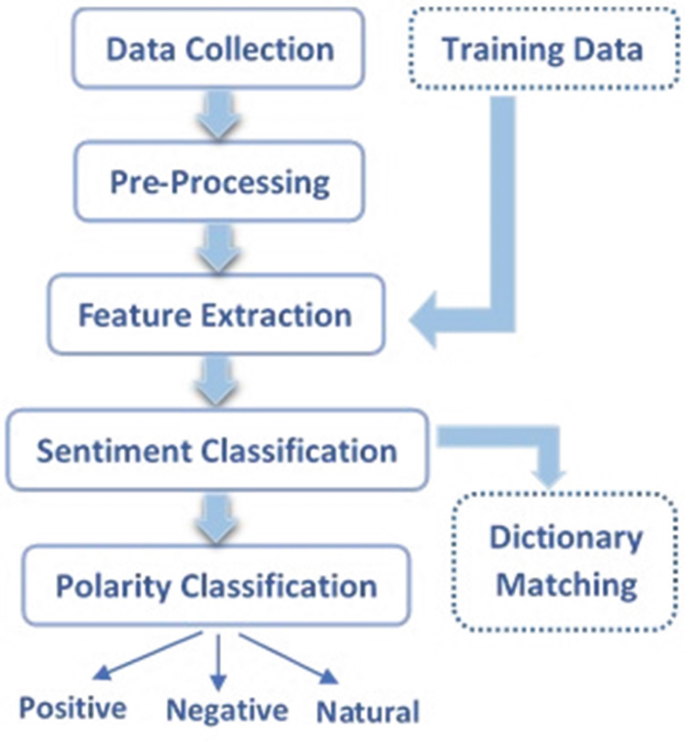
****

Fig- Sentiment analysis Process

1. **Conclusion and Future Works**

The main motivation behind this research is to define the, we determine if the sentiment is good or negative in this study. In addition to its advantages, it aids in social media monitoring and provides public opinion on particular subjects. The prime outcome will be detection of depression in sentiments caused due to Social media. We will improve the performance of the system in sentiment analysis to detect depression. We will build model for finding the sadness, happiness, customer behavior in sentiment analysis. The result will be obtained in lesser time using proposed method. Proposed work can be beneficial for E-Commerce websites, Social media etc, enhancing the customer satisfaction. Since it can deliver more trustworthy signals and information for a number of data analytics activities using digital platforms for prediction,

we believe sentiment categorization on sizable amounts of online user-generated content is advantageous. Because there are many negative memes centered on racism, religion, politics, and terrorism and because a high text score does not always suggest a positive meme, no single model can analyze all genres in a trend.

1. **Reference**

[1] Wiebe Janyce “Identifying subjective characters in narrative, Proceedings of the International Conference on Computational Linguistics (COLING-1990).”1990.

[2] Hearst M., 1992, Direction-based text interpretation as an information access refinement in TextBased Intelligent Systems, P. Jacobs, Editor 1992, Lawrence Erlbaum Associates, 257-274.

[3] V. Suresh 2011, A Non-syntactic Approach for Text Sentiment Classification with Stopwords, WWW 2011,March 28–April 1, 2011,Hyderabad, India.

[4] Murthy G. and Bing Liu, 2008, Mining opinions in comparative sentences, Proceedings of the 22nd international conference on computational linguistics (Coling 2008), Manchester, August 2008, 241248.

[5] B. Pang L. Lee, and S. Vaithyanathan., “sentiment classification using machine learning techniques” 2002.

[6] Matsumoto and Takamura, “syntax-based features construct parse trees”, 2005.

[7] Dave, Lawrence &Pennock, “Opinion extraction and semantic classification of product reviews”, 2003.

[8] Alena Neviarouskaya, Helmut Prendinger, Mitsuru Ishizuka, Affect Analysis Model: novel rule-based approach to affect sensing from text, Natural Language Engineering, Cambridge University, Vol. 17, pp. 95- 135, September 2010.

[9] C. Hauff, Dadvar, Maral and Jong de, Franciska, Scope of negation detection in sentiment analysis, Dutch-Belgian Information Retrieval Workshop, Netherlands, February 2011.

[10] Hao, F., Park, D., Pei, Z.: When social computing meets soft computing: opportunities and insights. Hum. Centric Comput. Inf. Sci. 8, 8 (2018).

[11] A. P. Jain and V. D. Katkar, “Sentiments analysis of Twitter data using data mining,” 2015 Int. Conf. Inf. Process., pp. 807–810, 2015.

[12] A. Teixeira, “Data extraction and preparation to perform a The example of a Facebook fashion brand page.”

[13] R. Jose, “Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Word Sense Disambiguation,” no. November, pp. 638–641, 2015.