

## SYSTEM BASED REVIEW ANALYSIS

<sup>1</sup>Anant Saini, <sup>2</sup>Manpreet Singh Bajwa

<sup>1</sup>Software Engineer Trainee Volkswagen IT services India

<sup>2</sup>Assistant Professor, Dept of Computer Sci and Engineering , Shree Guru Gobind Singh Tricentenary University, Gurugram , Haryana, India

### Abstract:

The present paper relates to a review analysis system. In brief it is the method or the system which is used for collecting, analyzing and distributing system's positive and negative reviews. This constitutes an acquisition module used for acquiring a sentence entered by the customer, a filtering module used for removing stop words, a pre-processing module used for stemming the filtered sentence, a features extraction module used for tokenizing the stemmed sentence, a training module used for generating a review and user platform for displaying the display. The method of review analysis system constitutes of the steps of acquiring sentence entered from the user, removing stop words, stemming, tokenization of stemmed, training stemmed sentence and generating a review and displaying a generated review.

**Keywords:** *Review Analysis System, Acquisition Module, Filtering Module, Pre-Processing Module, Features Extraction Module, Training Module, Stemming, Tokenization.*

### Introduction:

Emotional analysis is nothing other than a division of opinion on a given content or document in "positive", "negative" or "neutral" [1]. Every company needs to update, modify or improve its product according to customer requirement. The main focus of a product review is to identify the adjective in a sentence and to identify the feeling behind it. Jan Luo [2] suggests that the final points of the review be the result of a collection of all the adjectives in the review where Shobabi Ahmed [3] suggests that the number of words of opinion obtained is divided into seven categories which may be strong, strong-positive, weak-positive, neutral, weak- neutral, negative, strong-negative. Emotional analysis is done with the help of these score statistics.

Each and every customer review give new task to the production company. By the help of these reviews company come to know about their product quality, strength, weakness services in the field of work. However, classifying each and every review is bit of a tedious and time-consuming job.

It will take lot of time and manpower if each and every customer review is monitored by human. Also, there can a situation where some of the review can be missed while monitoring. Even the possibilities are there of having error in making decision about the positive and negative review.

Therefore, for the fast and accurate online product analysis the machines are employed as they are more budgetary and reliable in comparison to humans.

Text mining is another process similar to data mining, with the exception of data mining tools designed to manage structured data and text mining can be used to manage random or structured data settings such as emails, HTML files and complete text documents etc. [4]. Text data builds on that, it's worth exploiting. Therefore S. Cherfi, A. Nepoli, Y. Toussaint has suggested the use of information retrieval from the text database, or in short, document extraction (TM), is an important and difficult challenge due to the ambiguity of the richness of the original language (widely used of available texts) [5]. Since many ideas are available to us in text format and their functionality is simpler than other formats, emotional analysis has now emerged as a small field of textual mining [6].

The user interface and other enhancements for the native language acquisition system and method [patent], agree on the techniques used to generate the basic representation of the content of both queries and the retrieval system using natural language processing (NLP) techniques to represent and retrieve texts at multiple levels (e.g. morphological, lexical, syntactic, semantic, discourse and pragmatic levels) where people take the concept of writing.

The question is entered by the user and processed by the system to produce other representations, including one word, text structure, complex word, proper noun (nouns are considered as punctuation equals verb and adjectives can also be punctuation but difficult to identify [8]), conceptual exclusion and logical structure of question, add mandatory words. After processing the question, information to the user is displayed in the program, which indicates the program interpretation and represent the content of the question. The user is given the opportunity to provide input in response when the customization system is a different representation of the query.

Currently a number of different programs are available in the online review analysis market. However, these programs focus more on the analysis of each review rather than providing flexibility according to their type (positive or negative review) according to user needs.

The information contained in such reviews is subject to the maximum power or availability of the information. One of the basic functions in sensory analysis is to predict the magnitude of a given sentence, to determine whether it expresses a positive or negative emotion on a particular topic [9]. Therefore overcoming the above-mentioned limitations and risks where necessary to build an online review system for evaluating positive and negative reviews based on Mechanical Learning and Natural Language Processing (NLP) according to user needs.

### **Objective:**

The purpose of the current paper is to reduce the time required to analyze positive or negative reviews as per customer instruction. One of the purposes of this paper is to introduce a system that reviews manufacturing firms and their products as user reviews used to further the company's growth. Disposable features and other items, items, feature, and benefits of the current paper will be easily reflected in the ongoing review of this detailed description of the selection situation as shown in the following diagrams.

**Short Description:**

The current paper describes the system or method used for the review of reviews that update products and services on the basis of automated customer feedback monitoring. In addition, this program is also helpful in providing real-time comment analysis to eliminate errors and shortcomings.

According to the archetype of the current paper, an online review system, consisting of a discovery module embedded in the system collects a sentence from the user, and then a filter module with a detection module that removes the sixth word in the sentence, and the pre-processing module is connected to the filter module in the pre-processing module to make a formal penalty mark, and then the training module attached to the release features trains the token sentence and analyzes the sentence to produce a review, these reviews are then displayed by the user platform connected to the training module. The output of the model is only as good as the fed data, so special emphasis is placed on the preliminary processing of the text [10].

Also according to the archetype of the present paper, a method for automatic review analysis, consisting the steps of acquiring sentence which are framed by the user, used for removing the stop words from the sentence, stemming the filtered sentence and tokenizing the stemmed sentence, training the tokenized sentence is used to verify the tokenized sentence and further generating a review. The generated review is then displayed upon the detection of the verification. However, variations might be possible that would fall within the scope of the present paper that has been described and displayed with particular reference to the preferred embodiment.

**Detailed Description:**

The following includes the excellent mode of the current paper from which it will be removed that the paper not only covers the display but also the paper includes various modifications. Therefore, the current paper should be descriptive, analogous and ambiguous. Since paper is affected by all different modifications and constructions, it should be understood, that there is no intention to limit the paper's ability to reduce what is presented but the opposite of paper than all other modifications or other formations and equilibrium failures within the air and paper width as described in the claims.

In any of the actions described here, the open words "contain", "construct" and the like (such as "include", "have a" respect ") can be replaced by closed sentences that include" basically include "; "including ", and parallel or closed phrases that "include", are made up of "similar. As used here, art forms "a," an, and "the" placing both singular and plural, unless it is clearly stated to refer only to unity.

To update data using artificial intelligence the term "NPL (Natural Language Processing)" means a major process of language data analysis. The word "Tokenizing" means to separate a sentence / string from a list of tokens (such as phrases, symbols, words and more). Emotional token is a word or phrase that conveys emotion. Given those emotional nouns raised by the token of words that contain specific, negative words and part of the speech tag [11] as shown in the four figs.

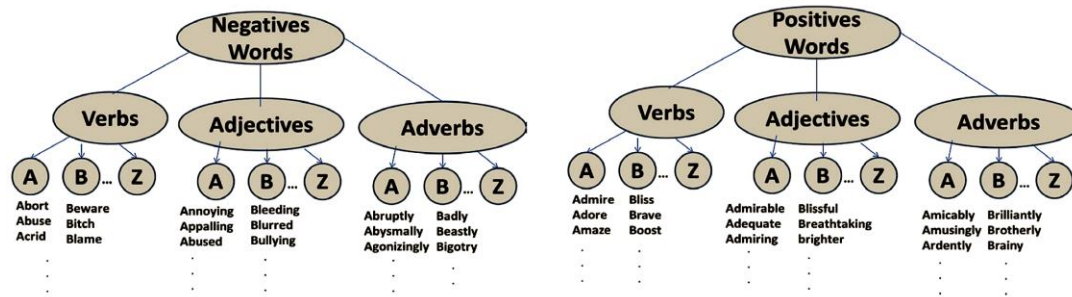


Fig.4 Word Opinion Ideology

The current paper uses machine learning and NLP (Natural Language Processing) related to the system and method that can provide real-time analysis. Indigenous Language processing can be used in a variety of areas such as machine translation, spam detection, extracting information, summarizing, answering questions etc. [12]. Additionally, this helps the service company grow assets based on customer reviews. A systematic machine learning method is designed to collect opposing words within a review sentence. Automatic detection was performed using a part-of-speech (POS) marker [13]. The POS tag is used to filter the emotionless word. Emotional analysis works better in a submissive text than does an objective text only. This is because when a text context has a particular theme or expectation, the text often expresses certain statements or facts without expressing feelings, emotions or feelings [14] [15].

The system constitutes of the acquisition module installed in the system that retrieve the sentences entered by the customer on the user platform. The user comments on the online products, items and the services. The sentence is further transferred to the filtering module.

The filtering module is connected to the acquisition module in order to remove a stop words from the sentences. The stop words are removed by using NLTK (Natural Language Toolkit) in python protocol as illustrated in fig.1. The examples of stop words are a, an, the, will, you, may etc. The filtered sentence is further retrieved by using pickle in the python which is used to serializes and DE-serializes a python object structure and save it on the disk. The retrieved sentence is further transferred to the pre-processing module.

The pre-processing module is connected to the filtered filter filter module. Stem means to remove / reduce the derived or curved words (such as suffix, prefix and similar words) in the title stem. Snowball Stemmer is used in the system to make degrees. Disability and disability procedures are selected for employees to see the common term for words such as "accepted", "considered". These words are deliberate and are reduced to "acceptance", "think" [16]. The stem removes words that attach to the morphological, leaving only the word stem. For example, the words "connect, connect, connect, connect can all be included

in the word" connect "[17]. The purpose of this method is to remove various suffixes, to have similar titles, to save time and space in memory. A limited sentence is also transferred to the feature module.

Features of the output module and the pre-processing module are used to make basic sentence token Assignment refers to the act of breaking a string sentence into pieces (such as sentences, keywords,

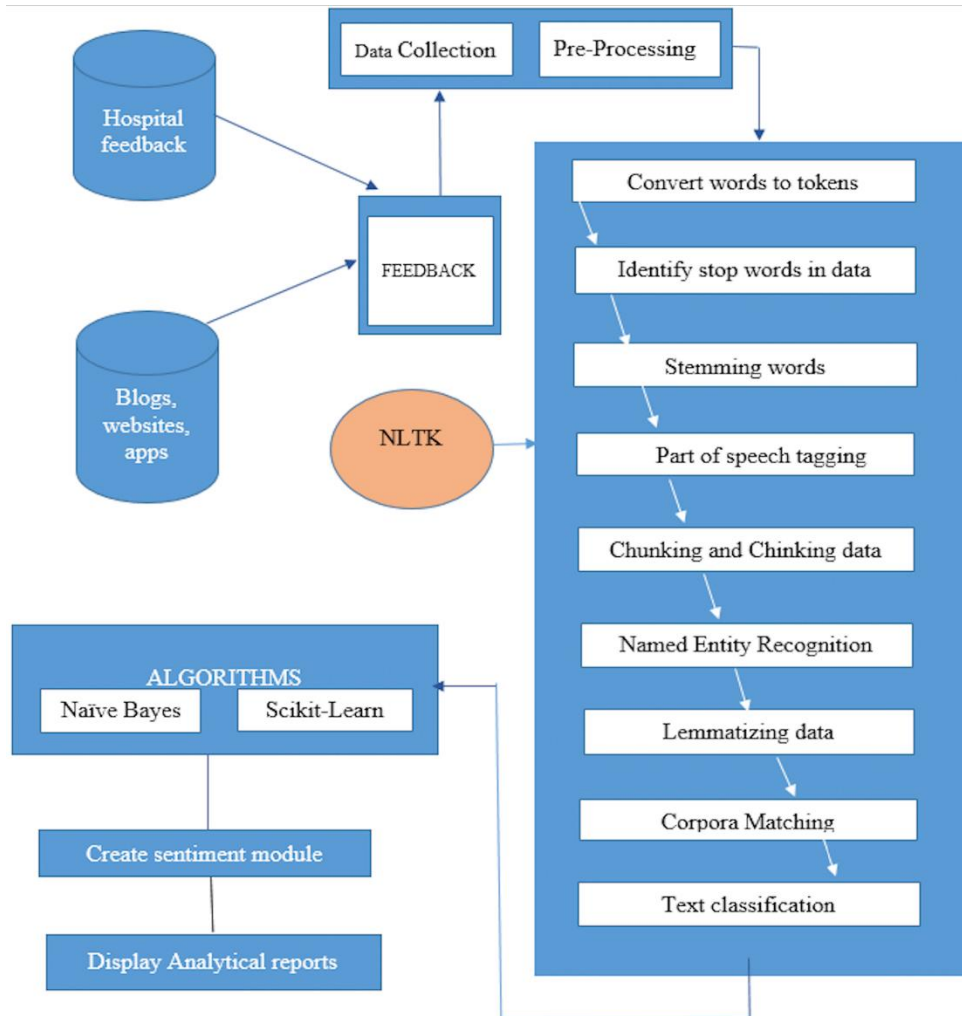


Fig.1.[7] Classification

words, punctuation and more). Token making is done in the python protocol with NTLK (Natural Language Toolkit). The toolkit does not need to be optimized for working time. However, it should be efficient enough that students can use their NPL system to do real work [18]. The verdict made by the tokens is also transferred to the training module.

The training module is connected to a feature removal module that trains token sentences. Training is done with the help of machine learning and artificial intelligence (NTPL). To generate reviews for comments provided by the user the training module analyzes the token sentence on the user's platform during or after online purchases. The user platform is in association with the training module. It displays the review generated by the training module. The generated review is classified in two basic categories i.e. negative or positive on the basis of the comment.

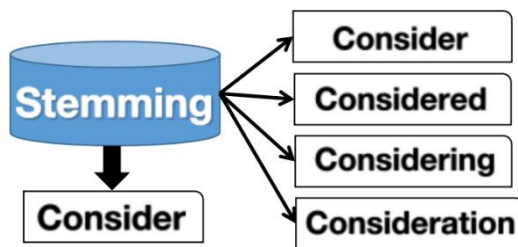


Fig.2 Stemming Process

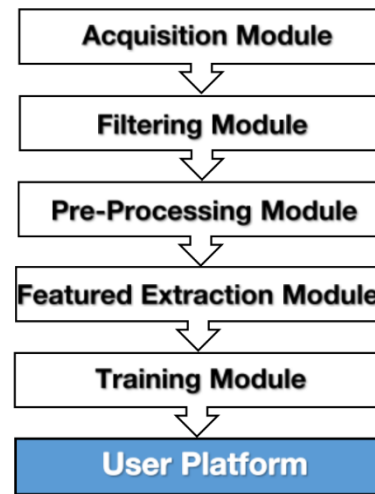


Fig.3 Archetype

Referring to fig 3, an archetype of a review analysis is illustrated. Initially in the first step user enter the acquiring sentence on the user platform. It is possible that the sentence can of vary length i.e. very large or very small. Also, there can be numerous comments given by the customer on the same product/services.

In second step there is the removal of the words from the sentence by employing filtering module. The stop words are removed so as to shorten the length of the sentence (comment) in order to perform fast and accurate analysis. Referring to fig 2. Block diagram of a Stemming process is the third step in which there is the stemming of the filtered sentence by deploying the pre-processing module. This step removes the unwanted words such as suffix, prefix and other similar words from their word stem to short the length of the sentence. These words have negligible role in differentiating the negative and positive sentences. The tokenized word is further retrieved by the pickle on the python protocol.

In fourth step there is the tokenization of the stemmed sentence by means of the feature extraction module. Tokenization is described as the act of breaking up a sentence of strings into pieces (such as phrases, keywords, words, symbols and other elements). This is performed by the help of NLTK (Natural Language Toolkit) on the python protocol. Each and every token have their own significance even the symbol has its own significance as it describes the mood of the customer.

In the final step there is the displaying of the generated review upon the detection of the verification. This review is displayed over the user protocol. The positive/negative review helps to grow the firm/industry by removing the deficiency or service as per the requirement of the user. Although the field of paper described has been herein with limited reference to specific embodiments, this description is not meant to be constructed in a limiting sense. Various modification of the disclosed embodiments, as well as alternate embodiments of the paper, will become apparent to person skilled in the art upon reference to the description of the paper.

**Discussion:**

- An acquisition module installed for acquiring a sentence entered by the user in said system.
- A filtering module connected to said acquisition module for removing stop words from said sentence.
- A pre-processing module for stemming said filtered sentence associated with said filtering module.
- A feature extraction module for tokenizing said stemmed sentence in connection with said pre-processing.
- A training module linked to said feature extraction module for training said system, wherein said training module verifies said tokenize sentence and generates a review into a least two categories.
- A user platform for displaying said category upon verification in association with said training module.

**Conclusion:**

- The system as claimed, wherein said sentence is basically a comment entered into user platform by user in order to rate a product or service.
- The system as claimed, wherein said stop words are selected from but not limited to a, an, the, in, will, can etc.
- The system as claimed, wherein said stemming refers to reducing/removing inflected words or derived words (such as suffix, prefix, or similar words) to their words stem.
- The system as claimed, wherein said tokenizing is act of breaking up a sentence of strings into pieces (such as phrases, keywords, words, symbol and other elements).
- The system as claimed, where training is done through NPL (Natural Language Processing) and machine learning.
- The methods for review analysis as claimed, consisting the steps of:
  - Acquiring sentence entered by the user.
  - Removing stop words from said sentence.
  - Stemming said filtered sentence and tokenizing it after stemming.
  - Training said tokenizing sentence in order to verify said tokenizing sentence and thus generate a review by a means of training module and
  - Displaying a generated review.

**References:**

- [1]. Anjali Ganesh Jivami, “ A Comparative Study of Stemming Algorithm”, International Journal Computer Technology Applications, Vol 2 (6), 1930-38, ISSN:2229-6093.

- [2]. Yan Luo,Wei Huang, “Product Review Information Extraction Based on Adjective Opinion Words”, 2011 Fourth International Joint Conference on Computational Sciences and Optimization.
- [3]. Shoiab Ahmed, Ajit Danti, “Novel Approach for Sentiment Analysis and Opinion Mining based on SentiWordNet using Web Data”, 2015 International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15).
- [4]. Vishal Gupta and Gurpreet S. Lehal, a Survey of Text Mining Techniques and Applications, Journal of Emerging Technologies in Web Intelligence, Vol. 1, August 2009.
- [5]. S.Cherfi, A.Nepoli, Y.Toussaint- “ Towards a text mining methodology using association rule extraction”.
- [6]. V.P.H. Binali and W.Chen. A State Of The Art Opinion Mining and Its Applicaton Domains. In IEEE International Conference On Industrial Technology, Pages 1-6, February 2009.
- [7]. Pradadevi. B, Nandyala Srujana Reddy, Deepa- “Heart rate Encapsulaion and Response Tool using Sentiment Analysis”, International Journal of Electrical and Computer Engineering 9(4):2585, August 2019.
- [8]. Nidhi R. Sharma, Prof. Vidya D, Chitra - “ Opinion Mining, Analysis and its Challenges “, ence IJIACS ISSN 2347 - 8616, Volume 3, Issue 1 April 2014.
- [9]. Pang, B., Lee, L. : Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2(1-2), (2008).
- [10]. Vasileios Athanasiou and Manolis Maragoudakis. A Novel, Gradient Boosting Framework for Sentiment Analysis in Languages where NLP Resources Are Not Plentiful: A Case Study for Modern Greek. Algorithms 10 (2017), 34.
- [11]. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, NY, USA. pp 168-177.
- [12].Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh 1.Department of Computer Science and Engineering Manav Rachna International University, Faridabad-121004, India 2.Accendere Knowledge Management Services Pvt. Ltd., India Natural Language Processing: State of The Art, Current Trends and Challenges.
- [13]. Roth D, Zelenko D (1998), “Part of speech tagging using a network of linear separators”– In: Coling-Acl, The 17th International Conference on Computational Linguistics, 1136-1142.
- [14]. Sarkar, Dipanjan. Text Analytics with Python.Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data. Apress, 2017.
- [15]. Jeon J.H., Xia R., Liu Y. Sentence level emotion recognition based on decisions from subsentence segments. In ICASSP, Lyon, France (2011) 4940-4943.



- [16]. Madhav Singh Solanki, Sentiment Analysis of Text using Rule Based and Natural language Toolkit, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-12S, October 2019.
- [17]. C.Ramasubramanian and R.Ramya, Effective Pre-Processing Activities in Text Mining Using Improved Porter. Stemming Algorithm, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 12, December 2013, ISSN (Online): 2278-1021.
- [18]. Edward Loper and Steven Bird (July 2002), NLTK: The Natural Language Toolkit, Department of Computer and Information Science University of Pennsylvania, Philadelphia, PA 19104-6389, USA.
- [19]. T. Karnan and G. Seenuvasan, Sentiment Analysis In E-Commerce Using Recommendation System, IJCSMC, Vol. 6, Issue . 8 2017, pg.55-62, ISSN 2320-088X.
- [20]. Priyanka Tyagi and Dr. R.C. Tripathi, A Review Towards The Sentiment Analysis Techniques for The Analysis of Twitter Data, 2<sup>nd</sup> International Conference On Advance Computing And Software Engineering (ICACSE-2019).
- [21]. Xing Fang and Justin Zhan, Sentiment Analysis Using Product Review Data, Fang and Zhan Journal Of Big Data (2015) 2-5, DOI 10.1 186/s40537-015-0015-2.
- [22]. Omkar Borade, Kaushik Gosavi, Ajay Shinde and Avinash Gowda, Sentiment Analysis Of College Reviews, 2017 IJEDR, Volume 5 , Issue 2, ISSN:2321-9939.
- [23]. Aashutosh Bhatt, Ankit Patel, Harsh Chheda and Kiran Gawande, Amazon Review Classification and Sentiment Analysis, Aashutosh Bhatt et al./ (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6), 2015, 5107-5110.
- [24]. Gurshobit Singh Brar , Asst. Prof. Ankit Sharma, Sentiment Analysis of Movie Review Using Supervised Machine Learning Techniques, International Journal of Applied Engineering Research ISSN 0972-4562, Volume 13, Number 16 (2018) pp. 12788-12791.
- [25]. P. Sanjay Bhargav, G. Nagarjuna Reddy, R.V. Ravi Chand, K.Pujitha, Anjali Mathur, Sentiment Analysis for Hotel Rating using Machine Learning Algorithm, International Journal of Innovative and Exploring Engineering (IJITEE), ISSN:2278-3075, Volume-8 Issue-6, April 2019.
- [26]. Prachi Sanghvi, Disha Shah, Prof. Bharathi H. N. , Movie Review System Using Sentiment Analysis, International Journal for Research in Applied Science & Engineering Technology (IJRASET), ISSN:2321-9653; IC Value:45.98; SJ Impact Factor:7.177, Volume 7 Issue V, May 2019.
- [27]. Aarti Potdar, Pranav Patil, Raunak Bagla, Rohitashwa Pandey, Prof. Nagesh Jadhav, Sentiment Based Product Review Analysis System, International Conference on Information Security & Privacy (ICISP2015), 11-12 Dec 2015, Nagpur, INDIA, AartiPotdar et al./ Procedia Computer Science 78 (2016) 513 -520.