

# Tracking Of Employees' Feedback Of An Organization Using Sentimental Analysis

Karthikeyan C, D. Poojitha, P. Rukmini

**Abstract:** The Goal of this project is to analyze the employee evaluations of any specific organization, It's difficult to hone in on employee sentiment no matter what size the organization, however large, complex establishments have a unique challenge. The Opinion analysis software program takes a look at all worker survey responses and shortly determines the "why" behind the engagement ratings. People discuss and talk about the most diverse topics in the social media platform, including their jobs. This results in a stream of employee-related data, and organizations are increasingly interested in making sense of this data on an ongoing basis to access key factors such as employee engagement, retention, and satisfaction. In this project, we are going to estimate such factors from different sentiments that are implicit in employee communication in an online platform. We collected a large set of employee reviews from the social platform called glass doors on a particular organization, here in the "glass doors" website the reviews of the employees are easily available but we cannot directly download it into some file, for that we need to scrap the data i.e by using web scraping methodology we can easily scrap, here instead of developing a program to scrap we used a readymade tool called "web harvy" then we took whole data into sentence-level type, where each sentence is analyzed. In our implementation, the reviews are collected and it is prepared as text corpus and is given as input to our program. Since it is the text corpus and comes under unlabeled data, we found it more comfortable in solving this through a lexicon-based approach where an external dictionary is being used in finding the polarity of the text. The text corpus includes positive, negative and neutral feedback of an organization "COGNIZANT". This approach uses word matching based on lexicon.

**Keywords:** Employees feedback of an organization, sentimental analysis, lexicon-based approach.

## I. INTRODUCTION

Sentiment Analysis one of extremely valuable strategy of recognizing a sentence is negative, positive or impartial. A slant assessment machine for content examination joins regular language handling (NLP) and registering gadget becoming acquainted with systems to dole out weighted feeling scores to the elements, points, issues, and classes inside a sentence or expression. Assessment helps insights investigators inside huge organizations check general conclusion, direct nuanced statistical surveying, screen maker and item notoriety, and perceive customer encounters. Moreover, insights investigation bunches routinely coordinate outsider conclusion assessment APIs into their own special customer trip the executives, online networking observing, or individual examination stage, to supply valuable experiences to their clients. The article will give a clarification for how central estimation examination functions, assess the endowments and disadvantages of rules-based assumption investigation, and blueprint the situation of work area acting in conclusion examination. At long last, we'll investigate the top elements of feeling examination before closing with some useful assets for likewise learning. Assessment can likewise be applied to sound, photos and, recordings. Today the web has come to be the basic area of our life. The vast majority utilize web based blogging sites or person to person communication destinations to unmitigated their suppositions on specific things. They likewise utilize these destinations to comprehend what various individuals' sentiments are. Accordingly mining of these insights and assessment extraction has risen as a basic subject. Notion assessment is the strategy of removing contemplations or suppositions from a bit of literary substance for a given point. It empowers us to secure the demeanors, sentiments, and feelings in the content. In it, clients' preferences are caught from net substance. It includes anticipating or breaking down the concealed certainties current in the content. These concealed records are gainful to get experiences into clients' preferences. The objective of assumption assessment is to choose the mentalities of a creator or a speaker for a given subject.

## Sentiment classification:

Two approaches are mainly used-

**Subjective lexicon:** Emotional dictionaries are arrangement of words where each expression has a rating demonstrating the positive, negative, nonpartisan and target nature of the corpus. In this methodology, two for a given bit of content, collection of evaluations of emotional expressions is done for example positive, negative, unprejudiced and target word scores are summarized independently. For this situation, there are 4 scores. The best evaluating gives the typical extremity of the literary substance. **Dictionary-based approach-** In this methodology, a lot of supposition words are physically accumulated and a seed list is readied. At that point we scan for word references and dictionary to find equivalent words and antonyms of content. The newfound equivalent words are added to the seed list. This method proceeds until no new words are found. **Detriment:** subject in deciding setting or territory situated supposition words. **Corpus-based approach –** Corpus is a series of writings, frequently on a particular topic. In this approach, seed list is arranged and is extended with the help of corpus content. Consequently, it unravels the issue of restricted territory situated content. It tends to be done in two different ways. **Statistical approach:** This methodology is utilized to discover co-event words in the corpus. Thought is that on the off chance that the word appears to be normally ineffectual content, at that point its extremity is certain. In the event that it basically shows up in terrible content, at that point its extremity is negative. **Semantic approach:** This procedure ascertains slant esteems by the utilization of the guideline of similitude between words. Wordnet can be utilized for this reason. Equivalent words and antonyms of a given word can be resolved the utilization of this and estimation expense can be determined.

## II. RELATEDWORK

With the quick increment of IT modernization and internet business net destinations, expanding patterns in individuals to posting on the web audits. Assumption vocabularies need to outrage used to looking at the huge degree of on line review realities available and achieve useful aptitude

from it. The greater part of the assessment dictionary are issue base, utilizes reliance parsing for separating the expression which is never again fit to characterize the nostalgic word so precisely. Attempt to help a strategy which consolidates feeling vocabulary and shallow parsing. Which decide issue and region base estimation investigation and afterward relegate extremity to a vocabulary. The significant benefits of proposed techniques are that it is particularly right and mechanically producing organized to warding off the pace of physically naming information. The shallow parsing used to investigate sentences as well and get the expressions of the parts. It will now not pondering about the two inward shapes two of constituent two words, nor indicating their cost in a sentence. At that point the utilization of the extremity of expressions incredible or terrible advancement of the item finishes up. In [1] creator proposed the example based strategy which detects the skepticism on twitter. To find mockery they utilized an example principally based methodology with the assistance of Parts of Speech (POS) and for grouping, AI approach. The component extricated by utilizing them was once delegated (i) Sentiment essentially based (ii) accentuation based (iii) syntactic and semantic basically based (iv) design based. This grouping helps in the disposal of uproarious or vain information. They saw whether the literary substance used to be wry or not, in which the exactness has acquired as 83.1% with accuracy 91.1%. In 2016, a sentiment analysis was once performed on the product feedback got from a web website online called Amazon with Naïve Bayes classifier, Logistic regression analysis and, Sent-Word Net dictionary-based evaluation strategies by using Kumar K.L. et al. As a result a sentiment evaluation and classification methods have been carried out with an accuracy ranging from 60% to 85% [2]. In the study performed with the aid of Rajan and Victor in 2014, supervised and unsupervised mastering and lexicon-based tactics had been used on a total of 114,500 tweets on Twitter for 5 particular companies. A method based totally on the formation of the rating values using splendid coefficients relying on the fantastic/terrible sentiment intensity of the phrases and determining of the rating values of the associated tweets used to be proposed. Based on the feedback on the Twitter account of the stated company, the sentimental adjustments of the customers have been monitored every month [3]. In [4] writer beat the problem of extremity move identification by utilizing proposing a model known as "Double Sentiment Analysis (DSA)". The DSA utilized the pair of audits, unique assessments and turned around surveys. These turned around audits were made through records development approach which used to be the arrangement of every training and evaluating surveys. The regulated methodology was utilized for arrangement with the assistance of a lexicon, which used to be region versatile as appropriately as language-free. They dispose of the reliance on the outside antonym word reference which upgrades the exhibition. In [5] stresses the sentiment evaluation of internet based life locales records like Twitter, Myspace, and Digg. They anticipated a dictionary based, significantly less region explicit, unconstrained and solo acing calculation to get a higher outcome. The appropriate response given by means of them was once relevant for subjectivity identification and extremity classification. The

addition of a given methodology used to be that it shows a vigorous and solid arrangement. In Semi-Supervised learning, the features are removed using a total of managed and solo learning. In [6] creator achieved the sentiment analysis through using the laptop getting to know strategies in different dialects (English, French and Dutch Languages). The intention was to group the sentiments given by method for the clients on the items utilized by utilizing them. Since they have been removing the feelings of people they teach the arrangement of conclusions that used to be as of now improved utilizing labeling the words into "constructive", "antagonistic" and "unbiased" class. This used to be completed physically. They executed 83% exactness on account of the English language,

Neutral	>-0.05 and <0.05
---------	------------------

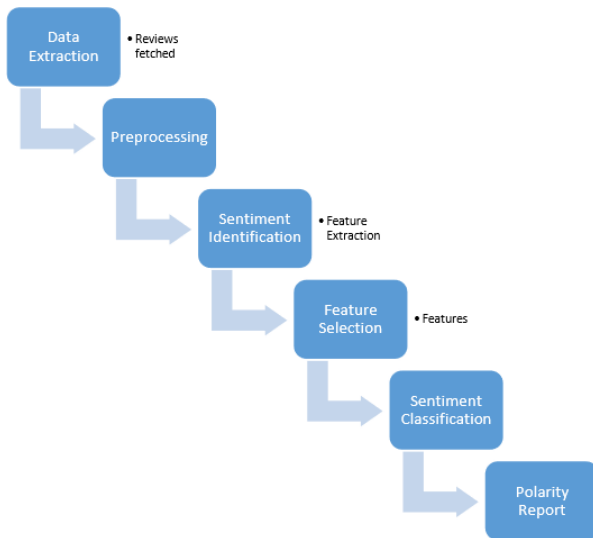
70% for Dutch content and 68% in the French language. In [7] proposed the methodology for assessment of the estimations of tweets. They focus on realities mining classifiers like k-closest neighbor, irregular backwoods, Naive Bayes and Bayes Net classifiers and are assessing the precision of these classifiers by considering stop words and excepting end words. In [8] creator proposed a novel device for Hindi vernaculars given utilizing the individual on different motion pictures. This machine is recognized as Hindi Opinion Mining System (HOMS). They utilize the Naive Bayes classifier which furthermore incorporates the blend of Parts of Speech (POS) labeling and PC picking up information on procedures for arranging the dataset into "positive", "negative" "neutral" class. On account of POS labeling, just expressions that come underneath the descriptive word zone are taken into picture. The downside of

"HOMS" is that it can't deal with "Talk connection" like "yet". In [9] Sentimental Analysis performed by Al-Horabi and Khan in 2016 on the Arabic tweets the usage of Arabic SentiWordNet (ArSenL) dictionary and BBN data set. When we seem at studies in the literature, it is seen that facts sets created from Twitter comments, Facebook comments, IMBD movie opinions and remarks on merchandise from e-commerce websites such as amazon.com are generally used for sentiments. In lexicon primarily based strategies used for sentimental analysis, SentiWordNet seems to be the most favored sentiment dictionary [10].

### III. METHODOLOGY

**Lexicon Based Sentimental Analysis:** The lexicon-based approach involves calculating orientation for a record from the semantic orientation of words or phrases in the document. Dictionaries for lexicon-based procedures can be created manually, as authors describe in this article two or automatically, the usage of seed words to increase the listing of words. Much of the lexicon-based lookup has targeted the use of adjectives as warning signs of the semantic orientation of the text. First, a list of adjectives and corresponding Sentiment Orientation (SO) values are compiled into dictionary. In this paper, we have used Vader sentimental analyzer in python and an algorithm that included emotions using R. Vader Sentimental Analyzer: VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis

tool that is particularly attuned to sentiments expressed in social media. VADER makes use of a combination of A sentiment lexicon is a record of lexical factors (e.g., words) which are generally labeled per their semantic orientation as each gorgeous or negative. VADER now no longer



solely tells about the Positivity and Negativity score on the other hand additionally tells us about how high-quality or poor a sentiment is. Vader is an external dictionary which is used to calculate a compound rating for every character evaluate. TABLE1

RESULT	COMPOUND SCORE
Positive	$\geq 0.05$
Negative	$\leq -0.05$

In this algorithm, an external dictionary Vader is being used as a sentimental analyzer. The text corpus is analyzed based on the above -mentioned compoundscores. Considering emotions: The dataset that is used in this implementation is cognizant employees dataset where the reviews are made in the form of text corpus and final analysis is made. In this algorithm, first text preprocessing is done by removing punctuations, numbers and stop words. The whole text is converted into the lower case where the ambiguity is minimized. ggplot2 is a special package in r where the method get\_nrc\_sentiment() is used in determining various emotions that are involved and graphical representations areanalyzed.

#### IV. IMPLEMENTATION

##### Data Set

We gathered a large set of employee reviews form the social platform called glass doors on a particular organization, here in the glass doors website the reviews of employers are easily available but we cannot directly download it into some file, for that we need to scrap the datai. e. by using a web scraping technique we can easily scrap, here instead of developing a program to scrap we used a readymade tool called “web harvy” then we took whole data into sentence-level type, where each sentence is

Analysis of employee Reviews On cognizant

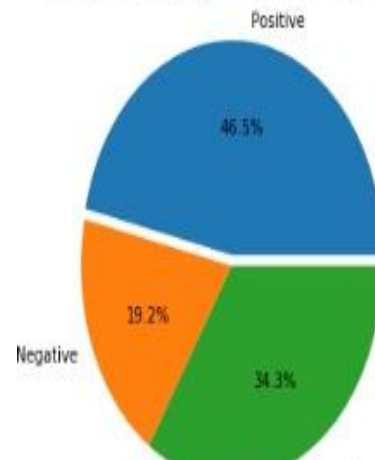


Fig 1

analyzed. In our implementation, the reviews are collected and it is prepared as text corpus and is given as input to our program. Since it is the text corpus and comes under unlabeled data, we found it more comfortable in solving this through a lexicon-based approach where an external dictionary is being used in finding the polarity of the text. The text corpus includes positive, negative and neutral feedback of an organization “COGNIZANT”.

#### RESULTS AND DISCUSSION

- i. Vader: Our dataset consists of 1000 reviews that are made into a huge text corpus. Python consists of an NLTK package where the Sentiment Intensity Analyzer

package is imported. Polarity\_scores() method is used to find out the polarity scores and graphical representation is made and the percentages of positivity, negativity and neutrality are calculated

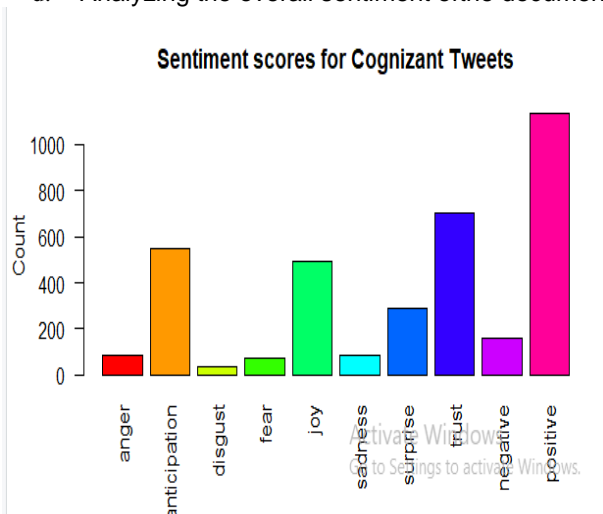
When subset is considered the confusion matrix is as follows:

	Positive	Negative	Neutral
Positive	25	6	0
Negative	8	0	13
Neutral	16	7	14

**Table 1**

ii. Sentimental analysis with emotions: In R, stringr, wordcloud, lubridate, reshape2 are the packages that play a vital role in this algorithm. The steps included are as follows:

- TextPreprocessing.
- Generating a term-documentmatrix.
- Obtaining emotions of eachreview.
- Analyzing the overall sentiment ofthe document.



**Fig 2**

The overall review regarding an organization "cognizant" is positive with more anticipation and less of disgust.

## V. CONCLUSION

The algorithm that is proposed concerning emotions is an efficient algorithm because rather than just declaring whether an overall result is positive, negative or neutral it is really essential for dealing with emotions. There are many applications that are been carried out with emotions. One such application is conducting surveys online where the customers/users' feedback is captured and the final emotion is analyzed. In such applications, this algorithm would be handy and the result is quick and efficient.

Moreover, in any similar applications like twitter and in any online survey this can beutilized. FUTUREWORKIn future enhancement, focused on the lexicon-based approach and compare the result with other machine learning approach to compare the result with the lexicon-based approach. Also, utilize an AI way to deal with contrast the outcome and the vocabulary based methodology. Then consider about emojis, talk words and slang words utilized in surveys while communicating the inclination. Half and half language and complex sentences will be considered as well.

## REFERENCES

- [1] Bouazizi, M., Ohtsuki, T.: A pattern based method for sarcasm detection on Twitter. IEEE Access 4, 5477–5488 (2016).
- [2] K.L.S Kumar, J. Desai and J. Majumdar, "Opinion mining and sentiment analysis on online purchaser review," Computational Intelligence and Computing Research (ICCIC),2016.
- [3] A.P. Rajan and S.P. Victor, "Web sentiment for scoring high quality or bad phrases using Tweeter data," International Journal of laptop Applications, vol. 96, no. 6, June2014.
- [4] Xia, R., Xu, F., Zong, C., Li, Q., Qi, Y., Li, T.: Dual sentiment analysis: thinking about two facets of one review. IEEE Trans. Knowl. Data Eng. 27, 2120–2133 (2015).
- [5] Paltoglou, G., Thelwall, M.: Twitter, MySpace, Digg: unsupervised sentiment evaluation in social media. ACM Trans. Intel. Syst. Technol. (TIST) 3, 66 (2012).
- [6] Boiy, E., Moens, M.F.: A computing device learning strategy to sentiment analysis in multilingual Web texts. Inf. Retrieval 12(5), 526–558 (2010).
- [7] Parveen, H., Pandey, S.: Sentiment evaluation on Twitter Data-set using Naive Bayes algorithm. In: 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp. 416–419. IEEE, July 2016.
- [8] Jha, V., Manjunath, N., Shenoy, P.D., Venugopal, K.R., Patnaik, L.M.: Homs: hindi opinion mining system. In: 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (RETIS), pp. 366–371. IEEE, July2015.
- [9] BBN Blog Posts Sentiment Datasets (2018, May 25). [Online]
- [10] Available ;<http://saifmohammad.com/WebDocs/Arabic-Sentiment-Copora>.
- [11] Twitter4j open-source library(2018,May 22). [Online]:<http://twitter4j.org>.

