

# Emoticons and Text Analysis as context to Sentiment Analysis

V.Kavya, Mr.M.Swami Das

<sup>1</sup>M. Tech Student, Department of CSE, MallaReddy Engineering College, Maisammaguda, Dhulapally, RangaReddy, Telangana, India.

<sup>2</sup>Associate Professor, Department of CSE, MallaReddy Engineering College, Maisammaguda, Dhulapally, RangaReddy, Telangana, India.

**Abstract:** - Twitter is an online person to person communication benefit where overall clients distribute their feelings on an assortment of subjects, talk about current issues, and express constructive or antagonistic assessment for items they use in day by day life. Along these lines, Twitter is a rich wellspring of information for supposition mining and notion examination. In any case, opinion examination for Twitter messages (tweets) is viewed as a testing issue since tweets are short and casual. This paper around this issue by the investigating of images called feeling tokens, including feeling images. As indicated by perception, these feeling tokens are generally utilized. They specifically express one's feelings paying little heed to dialect, subsequently they have turned into a valuable flag for assessment examination on multilingual tweets. The paper way to deal with performing assumption examination, that can decide positive, negative and impartial slants for a tried theme items.

**Keywords:** social media, emoji, analysis.

## I INTRODUCTION:

Text Analysis as context to Sentiment Analysis Technique is an effective means of discovering public opinions. various companies often use online or paper based surveys to collect customer comments. Microblogging sites, for example, Twitter (www.twitter.com) have developed to end up an incredible well spring of different sorts of data. This is because of the idea of micro blogs on which individuals post ongoing messages with respect to their suppositions on an assortment of subjects, talk about current issues, grumble, and express positive or negative feeling for items they use in day by day life. As the social occasion of individuals of micro blogging stages and casual groups builds up every day, data from these sources can be used as a piece of supposition mining and conclusion.

As the get-together of individuals of microblogging stages and casual groups builds up every day, data from these sources can be used as a piece of feeling mining and conclusion examination endeavors. For example, creating associations may be enthused about the going with request:

What do individuals think about our item (benefit, company etc.)

Political gatherings might be intrigued to know whether individuals bolster their program or not. social associations may ask individuals' gotten from interpersonal organizations, as their clients post regular what they like, and their sentiments on numerous parts of their life.

Suppositions and its related ideas, for example, assumptions and feelings are the subjects of investigation of notion examination and conclusion mining. The beginning and fast development of the field harmonize with those of the online networking on the web. Most NLP based techniques perform without specific accomplishment in web-based social networking. All types of online networking are exceptionally loud and loaded with a wide range of spelling, syntactic ,and accentuation blunders. This article proposes method of giving conclusion investigation utilizing such information as twitter has htags, e.g. #happy, #fail, feelings.

## II. RELATED WORK

The main examination is a developing region of the Natural Language Processing assignment at numerous levels of granularity. Beginning from being a record level arrangement assignment [1],[2] it has been handles at the sentence level [3],[4] and all the more as of late at the expression level [5], [6] or even extremity of words and expressions (e.g.,[7],[8]).

In any case, the casual and particular that is utilized as a part of tweets, and in addition the idea of the microblogging space make conclusion examination in twitter an altogether different assignment. with the developing populace of online journals and informal organizations ,feeling mining and estimation examination have turned into a field of enthusiasm for some inquires about. An extremely expansive reviews of the current

work was exhibited in [9]. J.Read in [10] utilized emojis, for example, " :- )" and ":- (" to frame a preparation set for the assessment order. For this reason, the creators gathered writings containing emojis from usenet newsgroups. The dataset was separated into "positive"(texts with glad emojis) and " negative" (writings with miserable or furious emojis) tests.

Specialists have additionally started to explore different methods for consequently gathering preparing information. A few analysts have depended on emojis to characterize preparing information [11],[12]. Barbarosa and feng [13] misused existing twitter supposition locales for gathering preparing information. davidov,tsur, and Rappoport [14] likewise utilized hashtags to make preparing information, however they restricted their examinations to senti 2. Information portrayal and accumulation

Twitter has its own traditions that renders it unmistakable from other printed information. The Twitter messages are called tweets. There are some specific highlights that can be utilize dto create a tweet.

Figure 1. Example of a tweet



Figure 1. Example of a tweet

The principal snippets of data "UE in katowice retweeted" implies the tweet was sent from a past post. "@RadioKatowice" and "@UE\_Katowice" are a twitter name of Radio Katowice and college of Economics in Katowice. utilizing twitter names in tweet sends data the client about say of them. #Katowice,#Studenci and #Koncert are a labels gave by the client to this message, alleged hashtags and "bit.ly/lzpnJmc" is a connection to some outside source. Length of tweets is constrained, along these lines long connections are regularly shortended utilizing unique sites like bitly.com.

Clients of Twitter utilize the "@" image to different clients. Alluding to different clients in this way consequently cautions them. clients for the most part utilize hashtags to check subjects. This is primarlily done to expand the perceivability of their tweets.these images gives simple approach to distinguish twitter client names and points and in this way permits seeking and separating of data regarding any matter.

Twitter messages have numerous interesting properties, which separates twitter examination from different fiels of research. in the first place is length. The most extreme length of a twitter message is 140 characters. Normal length of tweet is 14 words [15].

This is not quite the same as the spaces of other research, which were generally centered around audits which comprised of numerous sentences. The second property is accessibility of information. With the Twitter API or different apparatuses, it is significantly simpler to gather a huge number of tweets for preparing.

### III EMOTICONS:

The rise of popularity of emoji is due to its being incorporated into sets of characters available in mobile phones.Emoji can be categorized into same categories as emoticons.There are two principal information mining undertakings that can be considered in conjunction with Twitter information: content investigation and image examination. Because of the idea of this microblogging administration (snappy and short messages), individuals utilize acronyms, commit spelling errors, utilize emojis and different characters that express exceptional implications. Emojis are metacommunicative pictorial portrayal of an outward appearance pictorially spoke to utilizing accentuation and letters or pictures; they express the client's state of mind.

The utilization of emojis can be followed back to the nineteenth century. The principal reported individual to have utilized the emojis :- ) and :- ( on the Internet was Scott Fahlman from Carnegie Mellon University in a message dated 19 September 1982.

A few emojis as a characters are incorporated into the Unicode standard – three in the Miscellaneous Symbols piece, and more than sixty in the Emoticons square [16]. Emojis can be ordered as:

- Happy emoticons : :- ) :) :D :o) :] :3 :c) :> =] 8), etc.

Sad emoticons: >:[ :-(: (:-c :c :-< :>C <:-[ :[ :{, etc.

- Neutral emoticons: >:\>:/ :-/ :-./ :\ =/ =\ :L =L :S >.<, etc.

More symbols and meanings like angry, crying, surprise can be found on Wikipedia site [17], which can be used to determine their emotional state. The top 20 of emoticons collected from 96 269 892 tweets is presented in [18].

### IV EMOJI IDEOGRAMS:

Emoji were originally used in Japanese electronic messages and spreading outside of Japan. The characters are used much like emoticons, although a wider range is provided. The rise of popularity of emoji is due to its being incorporated into sets of characters available in mobile phones. Emoji can be categorized into same categories as emoticons. Emoji can be even translated to English using <http://emojitranslate.com/>.

**Data collection:**

The main problem is how to extract the rich information that is available on Twitter and how can it be used to draw meaningful insights. To achieve this, first we need to build an accurate sentiment analyzer for tweets, which is what this solution aims to achieve. As a software to data analyze can be used SAS Text Miner, SAS Visual Analytics or other tools. The challenge remains to fetch customized Tweets and clean data before any text or symbol mining. SAS Visual Analytics allows direct import of Twitter data, but to use SAS Text Miner and other tools, data have to be downloaded and converted.

Twitter allows developers to collect data via Twitter REST API [20] and The Streaming API [21]. Twitter has numerous regulations and rate limits imposed on its API, and for this reason it requires that all users must register an account and provide authentication details when they query the API. This registration requires users to provide an email address and telephone number for verification, once the user account is verified the user will be issued with the authentication detail which allows access to the API.

Unfortunately Twitter API exports data only in JSON format, which need to be translated to readable for databases or analytical software format. A combination of Twitter API, scripts for converting JSON to CSV [22], SAS Macro [23] or Excel Macro [24] can be used to extract information from twitter and create an input dataset for the analysis. The entire process of data acquisition can be fully automated by scheduling the run of Visual Basic for Applications (VBA) or SAS macros. Since opinions have targets, further pre-processing and filtering of collected data can be done using @twitter\_names and #hashtags as a targets in the way described in [20]. This method is more precise and provides better result than other text mining approaches.

**V SENTIMENT ANALYSIS:**

Sentiment Analysis is an Assessment investigation which is otherwise called feeling mining, centers around finding designs in the content that can be broke down to group the slant in that content. The term feeling examination most likely first showed up in [25], and the term sentiment mining initially showed up in [26]. Be that as it may, the examination on slants and feelings seemed before. Liu expressed that "Conclusion investigation is the field of concentrate that examinations individuals' feelings, slants, assessments, evaluations, states of mind, and feelings towards elements, for example, items, administrations, associations, and their traits. It speaks to an expansive issue space. There are additionally numerous names and somewhat unique errands, e.g., estimation investigation, assessment mining, feeling extraction, notion mining, subjectivity examination, influence investigation, feeling examination, survey mining, and so on"

[27, p. 7]. Supposition examination has become a standout amongst the most dynamic research fields in common dialect handling. It is likewise broadly contemplated in information mining, Web mining, and content mining. Truth be told, it has spread from software engineering to administration sciences and sociologies because of its significance to business and society.

Feeling investigation is prevalently executed in programming which can self-sufficiently separate feelings and assessments from a content. It has numerous true applications it enables organizations to examine how their items or brand is being seen by their buyers, legislators might be keen on knowing how individuals will vote in decisions, and so forth. It is hard to characterize conclusion examination as one particular field of concentrate as it fuses various regions, for example, phonetics, Natural Language Processing (NLP), and Machine Learning or Artificial Intelligence. As most of the feeling that is transferred to the web is of an unstructured nature it is a troublesome assignment for PCs to process it and concentrate important data from it. Probably the most viable machine learning calculations, e.g., bolster vector machines, innocent Bayes and contingent arbitrary fields, deliver no human justifiable outcomes.

Feelings are firmly identified with assumptions. Feelings can be characterized as a subjective emotions and contemplations. Individuals' feelings have been sorted into some unmistakable classifications. In any case, there is as yet not an arrangement of concurred essential feelings among scientists. In view of [28], individuals have six essential feelings, i.e., adore, euphoria, astonish, outrage, pity, and dread, which can be sub-separated into numerous auxiliary and tertiary feelings. Every feeling can likewise have distinctive forces. Feelings in virtual correspondence contrast in an assortment of routes from those in eye to eye associations because of the attributes of PC interceded correspondence. PC intervened correspondence may need huge numbers of the sound-related and visual signals regularly connected with the enthusiastic parts of communications. While content based correspondence disposes of sound and visual signs, there are different techniques for including feeling. Emojis, or enthusiastic symbols, can be utilized to show different sorts of feelings.

For reasons for this work, estimation can be characterized as an individual positive, impartial or negative assessment. Grouping is done in managed getting the hang of utilizing dictionary based approach. The assumption dictionary contains a rundown of notion emojis and emoticon ideograms. Assessments can be accumulated via looking Twitter posts utilizing Twitter API. Each tweet can named, utilizing emojis and emoticon symbols, as positive, negative, unbiased or garbage. The "garbage" mark implies that the tweet can't be comprehended. So as to utilize this strategy a suspicion must be made, this supposition is that the emoji in the tweet speaks to the general opinion contained in that tweet. This

presumption is very sensible as the most extreme share of cases the emoji will effectively speak to the general estimation of that tweet. This sort of assessment is usually known as the report level supposition characterization since it thinks about the entire record as a fundamental data unit.

Model can be produced on an example of information; this can be utilized to group assessments of the tweet. Manual order will be done on an example of tweets. Exactness of model can be tried against approving example. Tweets allocated physically will be partitioned into 2 sections – 80% of information ought to be taken in Model example and 20% of information ought to be taken as approving example. Results acquired will be contrasted and the physically allotted grouping.

## VI CONCLUSIONS:

Microblogging like twitter nowadays became one of the major types of the communication. The large amount of information contained in these web-sites makes them an attractive source of data for opinion mining and sentiment analysis. Most text based methods of analysis may not be useful for sentiment analysis in these domains. To make a significant progress, we still need novel ideas. Using twitter names and hashtags to collect training data can provide better results. Also adding symbol analysis using emoticons and emoji characters can significantly increase the precision of recognizing of emotions. The most successful algorithms will be probably integration of natural language processing methods and symbol analysis.

## VII REFERENCES:

[1] S. Das, M. Chen, Yahoo! for Amazon: Extracting Market Sentiment from Stock Message Boards, "Techniques of the Asia Pacific Finance Association Annual Conference (APFA)" 2001, Vol. 35.

[2] S. Kim, E. Hovy, Determining the Sentiment of Opinions [in:] COLING '04 Proceedings of the twentieth overall gathering on Computational Linguistics, Geneva 2004.

[3] Unicode Emoji, Draft Unicode Technical Report, Contextual Phrase-Level Polarity Analysis Using Lexical Affect Scoring and Syntactic n-Grams, Proceedings, Athens 2009, pp. 24-32.

[4] Level Sentiment Analysis [in:] Proceedings of the Conference Empirical Methods in Natural Language Processing, Association for Computational Linguistics, MIT Press, Cambridge, MA 2005, pp. 399-433.

[5] V. Hatzivassiloglou, K.R. McKeown, A Publicly Available Resource for Opinion Mining, "Strategy of LREC" 2006, Vol. 6.

[6] A. Esuli, F. Sebastiani, Sentiwordnet: Predicting the opinion Orientation of Adjectives [in:] Association for Computational Linguistics, Association for Computational Linguistics, Madrid 1997, pp. 174-181.

[7] "Establishments and Trends in Information Retrieval" 2008, Vol. 2(1-2), pp. 1-135.

[8] B. String, L. Lee, Opinion Mining and Sentiment Analysis, Proceedings of the PFD Student Research Workshop (student '05), Stroudsburg, PA 2005, pp. 43-48.

[9] D. Davidov, O. Tsur, A. Rappoport, Enhanced Sentiment Learning Using Twitter Hashtags and Smileys [in:] Computational Linguistics, Beijing 2010, pp. 241-249.

[10] List of emojis, Wikipedia, [http://en.wikipedia.org/wiki/List\\_of\\_emoticons](http://en.wikipedia.org/wiki/List_of_emoticons) (got to: May 2015).

[11] N. Berry, DataGenetics, <http://www.datagenetics.com/blog/october52012/index.html> (got to: May 2015).

[12] A. Agarwal, F. Biadys, K. McKeown, #51, <http://www.unicode.org/reports/tr51/>(got to: May 2015).

[13] Twitter REST API, The Search API, <https://dev.twitter.com/rest/open/search> for (got to: May 2015).

[14] Twitter Text Mining Using SAS, Social Media Analytics, <http://www.analyticstools.com/2012/06/electronic-frameworks-organization-examination-twitter-text.html> (got to: May 2015).

[15] B. Liu, Sentiment Analysis and Opinion Mining, Morgan and Claypool Publishers, Williston, VT 2012.