



Sentimental Analysis based on Text and Emoticons

A.S.C.S. Sastry, K. Veera Venkata Anudeep, T. Namratha Sai, T.S.N. Dheeraj

Abstract: Long range informal communication has bit by bit become an everyday practice for individuals to post their feelings, perspectives and remarks on any item or individual. Individuals share their sentiments online in an exceptionally casual language. Along these lines, it is exceptionally troublesome errand to break down precise slants connected with that common language. Estimation Analysis is an investigation of individuals' mentality, assessments, and feelings to arrange whether it is certain, negative or impartial. Utilization of emojis via web-based networking media has expanded quickly as of late. Subsequently, we have concentrated more on how emojis assume a significant job in opinion examination. Different elements that influence supposition examination are talked about quickly in this paper. Likewise, different issues like mockery identification, multilingualism, taking care of abbreviations and slang language, lexical variety and dynamic word reference dealing with are talked about. A few examinations were then performed to break down the impact of emojis on tweets utilizing AI strategies. At long last, the general execution investigation of the classifier was figured utilizing the NLTK stemming system.

Keywords: Social Network, Emoticon, Sentiment, Support Vector Machine.

I. INTRODUCTION

Emojis, for example, :) :- and :(, are as often as possible utilized online in internet based life, IM (e.g., Skype), sites, discussions, and different sorts of online social cooperations. Since they are usually utilized in online interchanges and they are frequently immediate sign of notion, emojis in content were generally utilized by NLP specialists in assignments, for example, assessment examination as highlights to AI calculations or as passages of conclusion dictionaries for rule-based methodologies. Diverse online networks and devices may evoke changed degrees of emoji use. Twitter, a microblogging webpage, is one of most well-known internet-based life. For specialists and organizations, approaching its immense measure of client produced information is basic for understanding client conduct and the slant communicated.

Natural Language Processing (NLP) has immense commitment in the territory of information mining. It is investigation of programmed examination and portrayal of human language.

Revised Manuscript Received on May 30, 2020.

* Correspondence Author

Dr. A. S. C. S. Sastry*, Professor, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India Email: asessastry@kluniversity.in.

K. Veera Venkata Anudeep, Student, Bachelor of Technology, Department of ECE, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India Email: anudeepkanakam2@gmail.com

T. Namratha Sai, Student, Bachelor of Technology, Department of ECE, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India Email: namrathatanneeru@gmail.com

T.S.N. Dheeraj, Student, Bachelor of Technology, Department of ECE, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India. Email: dheerajsurya62@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

NLP comprises of different undertakings, for example, Word Sense Disambiguation, Coreference goals, Part-of-Speech (POS) Tagging, Information extraction and Sentiment Analysis.

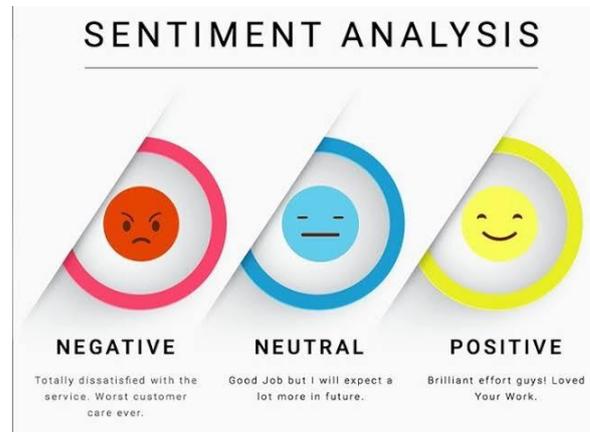


Figure.1: Sentiment Analysis

Sentiment Analysis is an investigation of individuals' frame of mind, conclusions, and feelings to group whether it is certain, negative or neutral. It very well may be applied on different types of information, for example, content, emojis, pictures, sound, and video. Individuals post their perspectives, feelings and feelings on person to person communication locales, for example, Facebook and Twitter with wide utilization of emojis. Twitter clients broadly use hashtags and smileys for accentuating their perspectives behind any thought or an individual. This paper for the most part centers around Sentiment Analysis dependent on content and emojis and how emojis assumes pivotal job in dissecting suppositions. Opinion Analysis can be applied on three levels: Document Level, Sentence Level and Entity Level.

II. RELATED WORK

A Novel Approach for Polarity Determination Using Emoticons: Emoticon-Graph [1]

Inferable from the rising fame of person to person communication locales and talk based applications, visual feeling pieces of information, for example, emojis are progressively being utilized in online journals, tweets, games, and item surveys. The current slant examination apparatuses basically center around anticipating the extremity dependent on printed content, and showing the outcomes as diagrams or outlines. In this paper, we propose a framework to represent emojis and shout checks alongside words while performing assessment examination of the information content. The yield of this examination is spoken to on an interesting figure, which we characterize as an 'emoji chart'. An online overview was led to gather item and news audits to dissect the estimation and furthermore to assess the acknowledgment of the 'emoji diagram'.

The discoveries of this study show that powerfully plotted emoji charts could assume a significant job in improving the aftereffects of extremity assurance techniques. The discoveries of this study show that powerfully plotted emoji diagrams could assume a significant job in streamlining the consequences of extremity assurance strategies. The extraordinary element of our proposed framework is the assessment of emojis and shout marks, alongside words, for performing estimation investigation of the info content. Additionally, from the aftereffects of our review, it is apparent that emoji charts are simpler to break down and favored over the ordinary diagrams for considering the consequences of extremity assurance techniques. A full-confirmation strategy with appropriate approvals, intended to produce these emoji charts would demonstrate to be a significant advance in the investigation of immense datasets.

Localized twitter opinion mining using sentiment analysis [2]

Examination of open data from internet-based life could yield intriguing outcomes and bits of knowledge into the universe of general feelings about practically any item, administration or character. Interpersonal organization information is one of the best and precise pointers of open slant. In this paper we have talked about a strategy which permits usage and elucidation of twitter information to decide popular feelings. Investigation was done on tweets about the iPhone 6. Highlight explicit popularities and male-female explicit investigation has been incorporated. Blended conclusions were found however broad consistency with outside surveys and remarks was watched.

In this exploration we examined a procedure by which it is conceivable to decide the prevalence/supposition/conclusion of an item in various areas crosswise over male and female clients. For our examination we picked the iPhone 6 as during the hour of the exploration a sensible measure of tweets dependent on the iPhone 6 was accessible. The quantity of tweets must be huge for exact outcomes. Along these lines, regardless of whether an item doesn't have enormous number of tweets at some random minute, we could gather tweets over a time of a little while or months or buy huge datasets from server farms. For the decision of area we select seven significant urban communities in the United States. The purpose for this is additionally information accessibility. 35 % of the considerable number of universes' tweets are from the USA with the rest of the portion vigorously partitioned among every other nation.

III. FRAMEWORK

As of late, much research has been done on dissecting estimations dependent on traits, for example, sex, age, area, time, and so forth. These ascribes makes sense of right supposition dependent on setting. Slant Analysis on item audit information profoundly relies upon sexual orientation, age and area of individual. Individuals posts their perspectives for an item dependent on their own encounters which fluctuate from area to area [2]. Investigating temperament of a person from his/her interpersonal organization profile is likewise a setting-based notion examination where age and sexual orientation of individual goes about as significant parameters to separate their assumptions [4]. Temperament of an individual may change in hour, day and week by week premise [5]. In this way

examination as far as time is a significant angle for setting-based feeling investigation. In such cases, emojis utilized by individuals of various age, sexual orientation at various area and on various time differs.

In this venture utilizing SVM calculation and NLTK (Natural language Processing Toolkit) we are distinguishing conclusions from content information, NLTK help us in applying significant strategies, for example, word stemming, unigram (help in evacuating exceptional images and additional areas), bigram (help in rehashing of words, for example, 'an apple daily will change over to two words rehashing, for example, 'an apple, apple an, a day') and other content preparing steps.

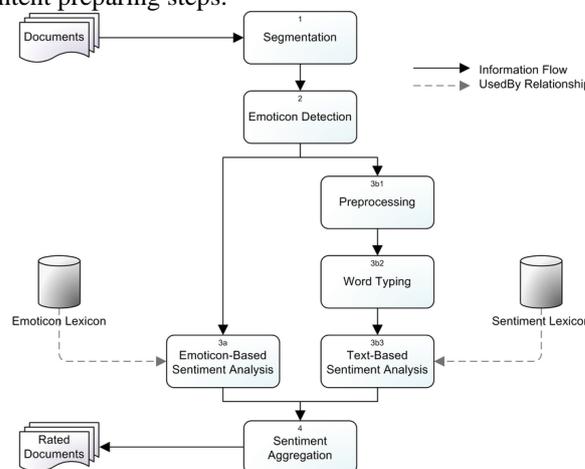


Figure.2: System Model

A large portion of the current framework evacuates exceptional characters and numbers in content pre-preparing step. Because of this, characters, for example, (, - and! which makes up emojis are evacuated and framework group slants just on premise of printed signals given in information. Preprocessing is a significant advance in estimation investigation which improves productivity just as adequacy [15].

FLOW DIAGRAM:

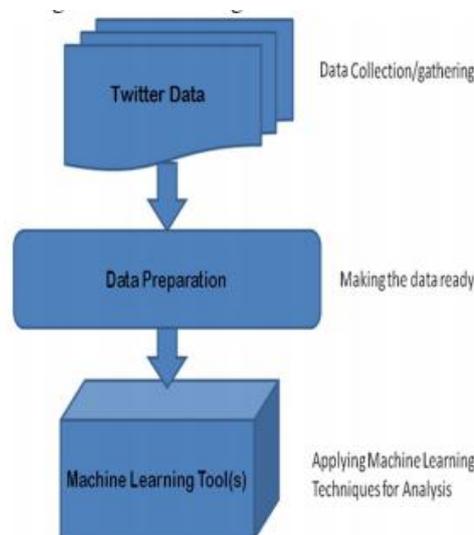


Figure.3: Flow Diagram For Proposed Work

Preprocessing errands incorporates Stop word evacuation, Stemming, Lemmatization, Part-of-Speech (POS) labeling [2], Expand Abbreviations and Chunking [11].

Legitimate preprocessing prompts better outcomes in arranging information into positive, negative and impartial classes. It likewise makes the undertaking lighter and productive to process. Information that don't comprise of any notions are expelled during pre-handling stage itself with the goal that the errand doesn't get cumbersome in later phases of assumption investigation.

IV. EXPERIMENTAL RESULTS

SVM is a machine learning algorithm help us in detecting sentiments from text, first SVM will train with all possible sentiments text and then build a model and this model can be used to detect sentiments from new text sentences, whenever user will give any sentence then application will apply new sentence on that SVM train data and then SVM will look for best matching class for given sentence as positive or negative. Before building model SVM extract features from text such as BAG OF WORDS (BOW) from text sentences, this BOW consists of word and its count and then build a vector. Later this vector will be separate into two different classes such as positive or negative based on class label given in train sentences data. Intuitively, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyperplane as possible, while still being on the correct side of it. So when new testing data is added, whatever side of the hyperplane it lands will decide the class that we assign to it.

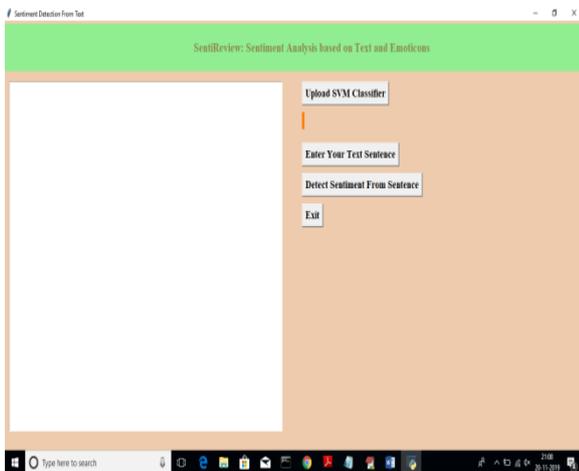


Figure.4: Home Screen

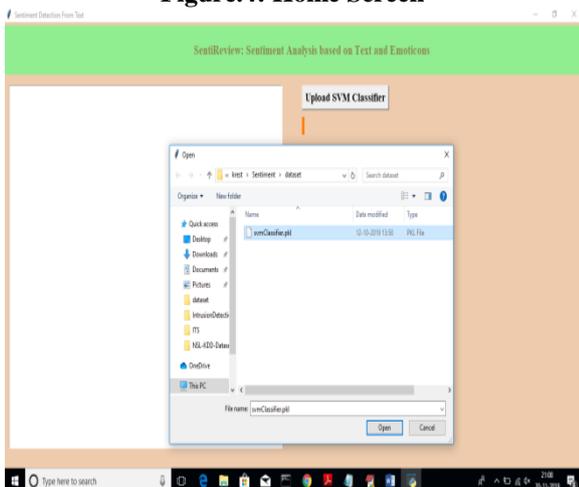


Figure.5: Upload Svm Classifier

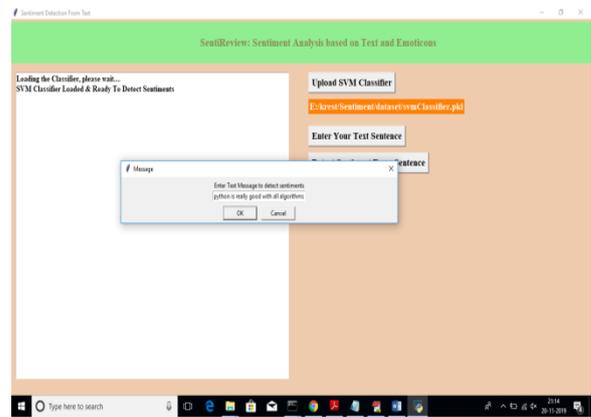


Figure.6: Enter Text Screen

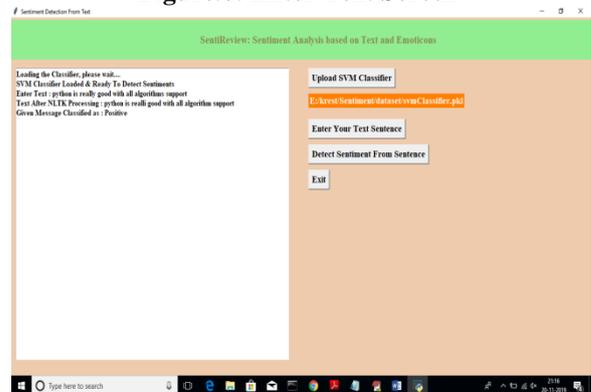


Figure.7: Detect Sentiment Screen

V. CONCLUSION

In this paper, significance of emojis in assumption investigation has been appeared by changed models. Components that influence notion examination are talked about in a nutshell. The paper likewise abridges existing methodologies for notion investigation. Content pre-preparing, highlight extraction and highlight determination assumes a significant job for examining slants productively. Different Machine Learning strategies and dictionary-based systems can be consolidated to shape a mixture approach which may result into increasingly precise opinion investigation.

EXTENSION

For future work, we intend to comment on refreshing the Python-Code with more streamlined and effective code. The present Machine Learning Technique can be supplanted with all the more dominant strategies/calculations to figure, examine and anticipate the outcomes a lot quicker and exact with least mistake rate.

REFERENCES

1. M. Datar and P. Kosamkar, "A Novel Approach for Polarity Determination Using Emoticons: Emoticon-Graph," In Advances in Intelligent Systems and Computing, 2016, pp. 481-489.
2. S. Hridoy, M. Ekram, M. Islam, F. Ahmed and R. Rahman, "Localized twitter opinion mining using sentiment analysis," Decision Analytics, 2015, p.1.
3. A. Hogenboom, D. Bal and F. Frasinca, "Exploiting Emoticons in sentiment analysis," In Proceedings of the 28th Annual ACM symposium on Applied Computing, 2013, pp. 703-710.

4. J. Zhao, L. Dong, J. Wu and K. Xu, "MoodLens: an emoticon-based sentiment analysis system for chinese tweets," In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12, 2012, pp. 1528-1531.
5. R. Rosa, G. Schwartz and I. de Campos Ribeiro, "Monitoring system for potential users with depression using sentiment analysis," In International Conference on Consumer Electronics (ICCE) , 2016, pp. 381-382.
6. M. Boia, B. Faltings, C. Musat and P. Pu, "A :) Is Worth a Thousand Words: How People Attach Sentiment to Emoticons and Words in Tweets," In International Conference on Social Computing (SocialCom), Alexandria, VA, 2013 pp. 345-350.
7. Z. Wang, V. Joo, C. Tong and D. Chan, "Issues of social data analytics with a new method for sentiment analysis of social media data," In 6th International Conference on cloud computing technology and science, 2014, pp. 899-904.
8. G. Solakidis, K. Vavliakis and P. Mitkas, "Multilingual Sentiment Analysis Using Emoticons and Keywords," In IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014.
9. H. Wang and J. Castanon, "Sentiment expression via emoticons on social media," In IEEE International Conference on Big Data, Santa Clara, CA, 2015, pp. 2404-2408.
10. E. Cambria and B. White, "Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]," In Computational Intelligence Magazine, 2014, vol. 9, no. 2, pp. 48-57.

AUTHORS PROFILE



Dr. A. S. C. S. Sastry is currently professor in Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (KLEF). He researched the topics feature extraction, sign language recognition, video signal processing, image motion analysis, feedforward neural nets, image classification, image matching, learning (artificial intelligence), gesture recognition.



K. Veera Venkata Anudeep is currently pursuing Bachelor of Technology in Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (KLEF).



T. Namratha Sai is currently pursuing Bachelor of Technology in Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (KLEF).



T.S.N. Dheeraj is currently pursuing Bachelor of Technology in Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation (KLEF).