Review Article

Innovative Method for Analyzing Sentiments using VADER

Raman Chadha[1], Aryan Chaudhary[2]

[1]Professor, Computer Science & Engineering Department, UIE, CHANDIGARH University, Punjab, India.
[2]Chief Scientific Advisor, Bio-Tech Sphere Research, India.

INFO

Corresponding Author:
Raman Chadha, Computer Science & Engineering Department, UIE, CHANDIGARH University, Punjab, India.
E-mail Id:
dr.ramanchadha@gmail.com
Orcid Id:
https://orcid.org/0009-0003-7030-4691

ABSTRACT

The method of figuring out whether a piece of writing is positive, negative, or neutral is called sentiment analysis. In order to assign weighted sentiment ratings to the entities, topics, themes and categories contained inside a sentence or phrase, sentiment analysis systems for text analysis integrate natural language processing (NLP) and machine learning approaches. Sentiment analysis aids data analysts at major corporations in understanding consumer experiences, doing complex market research, monitoring brand and product reputation and gauging public opinion. A vocabulary and rule-based sentiment analysis tool called VADER (Valence Aware Dictionary and Sentiment Reasoner) is customised precisely to the sentiments expressed in social media. VADER makes use of a variety of A sentiment lexicon is a collection of lexical elements (such as words) that are often classified as either positive or negative depending on their semantic orientation. VADER not only informs us of the positivity and negativity scores, but also of the sentimentality of each score. When VADER analyses a text, it looks to see if any of the terms are recognised by the lexicon. We can see that all of the features in this text (in the blue boxes) that suggest that the author is dissatisfied are actually informal writing, including a number of acronyms, punctuation marks and an emoticon. Additional components for negations and intensifiers are handled collectively.

Keywords: Sentiment Analysis, Polarity, VADER

Introduction

Today Internet has turned into a vital part in the life of individuals. Individuals frequently utilize sites, discussion, e-news and social systems administration locales like Facebook, twitter to express their perspectives and opinions. Web is turning into a basicleadership source. Extensive measure of substance is created on web each day, along these lines mining data and extracting client sentiment, is a vital errand.[1]

Micro blogging websites like twitter are well known for imparting of insights on a wide assortment of points by millions of clients. Along these lines such locales are rich wellsprings of data for sentiment analysis. Twitter clients utilize hashtag (beginning with #) which bunches the tweets on a comparative subject containing that specific hashtag. At the point when a specific hashtag is utilized by greatly substantial number of individuals, at that point it turns into a pattern which pulls in more individuals to take an interest in the discussion.[2]

Opinion Mining, another name for sentiment analysis, is a kind of text mining which orders the text into various classes. Opinion Mining utilizes some method to categorize

Chadha R et al.
J. Engr. Desg. Anal. 2023; 6(1)

24

opinions into classes like positive, negative, nonpartisan. This division of classes is called extremity of text. Principle target of opinion mining is classification of sentiments into various classes.[4]

Sentiment Analysis manages investigating feelings, emotions, the state of mind of a speaker or an essayist from a given bit oftext. "Sentiment analysis or opinion mining alludes to the use of normal language handling, computational phonetics and text examination to recognize and extract abstract data in source materials" (Source: Wikipedia). Sentiment Analysis includes catchingof client's conduct, different preferences of a person from the created web content. There is no solid meaning of "Sentiments", however by and large they are considered as considerations, perspectives and state of mind of a man emerging for the most part based on the feeling rather than a reason. Sentiments are considered as the sign of our sentiments and feelings. This field of software engineering manages breaking down and foreseeing the shrouded data put away in the text. This concealed data give profitable bits of knowledge about client's goals, taste and likeliness. Sentiment Analysis center on ordering the text at the level of emotional and target nature. Subjectivity demonstrates that the text contains/bears opinion content while Objectivity shows that the text is without opinion content.[3]

Some Examples

Subjective- This movie by Amir Khan and Kajol is superb.

(This sentence has an opinion, it talks about the movie and the writer's feelings about same "superb" and hence it'ssubjective)

Objective- This movie stars Amir Khan and Kajol.

(This sentence is a fact, general information rather than an opinion or a view of some individual and hence its objective)

The subjective text can be further categorized into 3 broad categories based on the sentiments ex-pressed in the text.[5]

- Positive- I love to watch Star TV series

- Negative- The movie was awful

- Neutral- I usually get hungry by noon. (This sentence has user's views, feelings hence it is subjective but as it doesnot have any positive or negative polarity so it is neutral.)

Literature Survey

M. Wongkar and A. Angdresey,[3] Sentiment analysis is a process used to gauge how the general public feels or thinks about products, services and even a person, including politicians and celebrities. In this study, contenders for the 2019 Republic of Indonesia presidential election were the subject of a sentiment analysis application for Twitter analysis using the Python programming language. This sentiment analysis is carried out in a number of processes, including data collection using Python libraries, text processing, testing training data and Naive Bayes text categorization. To categorise social classes or the intensity of societal attitudes, the Nave Bayes approach is employed. According to the study's findings, the relationship between Jokowi and Ma'ruf Amin had a positive sentiment polarity score of 45.45% and a negative sentiment score of 54.55%, while the relationship between Prabowo and Sandiaga had a positive sentiment score of 44.32% and a negative sentiment score of 55.68%. The accuracy was 80.90% 80.1% when the combined data was examined using the training data utilised for each presidential contender. In this study, naive bayes, svm and K-Nearest Neighbour (K-NN) approaches were compared. These methods were tested using RapidMiner, which produced accuracy values for naive bayes of 75.58%, svm of 63.99% and K-NN of 73.34%.

Mandloi and R. Patel,[4] The process of determining whether a text's sentiments are favourable, negative, or neutral is called sentiment analysis. It is often referred to as mining of opinions or material polarity. Numerous people were interested in social media platforms' expansion and development. A social media site that allows users to write tweets with 280 characters. Tweets only have a certain number of characters, making sentiment analysis simple. Every day, 550 million tweets are published on Twitter. Twitter also fairly portrays both genders and all age groups of people. As a result, the social sentiments revealed by the sentiment analysis of Twitter data are fairly generic. The Naive Bayes Classification approach, the Support Vector Machine Classification method and the Maximum Entropy Classification method are just a few of the different Machine Learning techniques that the authors will compare in this work.

S. Dhawan, K. Singh and P. Chauhan[5] Sentiment analysis is a process for computationally determining if a piece of writing is positive, negative, or neutral. The process of determining a user's emotion or frame of mind is also known as assumption mining. An effort has been made in this research to provide an analysis approach for the sentiment of the Twitter dataset. The proposed method determines the polarity of each tweet to determine whether it is positive or negative. Anger, sadness, happiness and joy are examples of user emotions that might have a polarity. Python code has been used to implement the suggested technique.

C. Chauhan and S. Sehgal[6] The Internet is a huge storage facility of ordinary language. People share their contemplations andexperiences which are unique in nature. Various a period, getting fitting information about a thing can wound up dull for customers. Associations may not be totally aware of customer necessities. Thing reviews can be

25

Chadha R et al.
J. Engr. Desg. Anal. 2023; 6(1)

poor down to appreciate the sentiment of the all-inclusive community towards a particular subject. In any case, these are voluminous; in this manner an outlineof positive and negative reviews ought to be delivered. In this paper, the rule spotlight is on the review of counts and procedures used for concentrate highlight savvy summary of the thing and inspected them to shape a true blue review. Future work will join all the more thing reviews destinations and will focus on higher level standard language getting ready errands. Using best and newmethods or instrument for progressively exact result in which the structure except for simply those watchwords which are in dataset rest of the words are cleared out by the system.

C. Zhao, C. Hu and T. Peng,[7] In the previous decades, the advancement of Internet has produced huge data trades. The market procedure advantage a great deal from breaking down sentiment inclination covered up in short messages. Notwithstanding, it appears to be difficult to investigate the sentiment inclination covered up in short messages in Chinese in viewof the unpredictability of this language. In this paper, a methodology for abridging the item assessments is displayed. A versatile strategy for sentiment analysis is connected to do sentiment characterization which partitions the assessments into positive set, nonpartisan set and negative set. In this strategy, the general sentiment word references are reached out to be extraordinary ones for explicit subjects.

Also, the degree verb modifiers and different nullifications are mulled over. Reproduction results demonstrate this new technique achieves higher precision contrasted and past strategy dependent on general sentiment word references. At that point K-implies bunching calculation is utilized to group the positive set and negative set. For an unmistakable articulation, authors attract word cloud pictures to demonstrate the watchwords of each grouping. By dissecting the words appeared on the photos, marketing experts can alter their market techniques.

Z. Singla, S. Randhawa and S. Jain[8] Big Data business has given a major jump to online business. It has opened up the roads to more astute and educated basic leadership for enormous enterprises just as the purchasers. Online reviews on internet business mammoths like Amazon, Flipkart are one such worldview which can be utilized to touch base at increasingly gainful choices. They are valuable for the shoppers as well as for the item producers. Online reviews can possibly give a knowledge to the purchasers about the item like its quality, execution and suggestions; in this way giving a reasonable image of the item to the future purchasers. The handiness of online reviews for producers to acknowledge client necessities by breaking down supportive reviews is one such hidden potential. Both positive and negative reviews assume a major job in

deciding the client prerequisites and extricating shopper's criticism about the item quicker. Sentiment Analysis is a computational report to separate abstract data from the content. In this exploration, information analysis of a huge arrangement of online reviews for cell phones is led. Authors have not just arranged the content into positive and negative sentiment yet have additionally included sentiments of displeasure, expectation, disturb, dread, bliss, misery, shock and trust. This depicted order of reviews is useful to assess the item comprehensively, empowering better-basic leadership for buyers.

P. R. Mala and S. S. Devi,[9] Sentiment analysis is one of the real assignments of NLP. In this paper, it means to handle the issue of sentiment extremity order,that is one of the key issue of sentiment analysis. Information utilized in this paper are online item reviews gathered from Facebook. com. It very well may be an objective for the marketing and publicizing industry to accumulate opinions about a specific item or an administration. This paper proposes to evaluate sentiments in Facebook by dissecting the clients' feelings toward an issue dependent on the remarks in a post. The feeling (upbeat, trouble and aloof) examples will evaluate the sentiments.

In their framework, this paper will investigate the quantity of emoticon responses, number of remarks and extremity of remarks moreover. Remark, checks, likes and quantities of preferences of the post are recovered utilizing Graph API from facebook. Recovered information is put away in MONGO db. A product device is created by Phython language to play out the sentiment characterization by utilizing NLTK and result can be appeared in diagram arrangement utilizing D3js.

N. M. Shelke, V. Thakre and S. Deshpande[10] Sentiment Analysis gives much all-encompassing and sensible methodology for self-evaluation of brands/items. Contributory work has been done on recognizing the item features from reviews, however deal with deciding the extension and treatment of logical shifters like refutations and valence shifters or diminishers or intensifiersis in outset organize. Proposed work is planned for this basic undertaking and oddity of the proposed framework is that it is area autonomous.

Y. Hegde and S. K. Padma[11] Sentiment Analysis(SA) for Kannada reports has been researched starting late. In the continuous examination, the sentiment analysis for Kannada substance is researched using Naive Bayes classifier. The objective of this work is to improve the introduction of the past examination on the sentiment analyzer for Kannada language explored in the paper. In this work, creators propose the outfit of classifier with unpredictable timberland strategy to perceive the limit of the sentiment andtest the show of the equal. Furthermore in this work,

Chadha R et al.
J. Engr. Desg. Anal. 2023; 6(1)

26

a segment of the imperatives of, for instance, dealing with multi class marks, unmistakable confirmation of sentiment limit of relative and unexpected clarifications have been tended to. The overall precision is improved from 65% to 72 %, demonstrating their technique subject to Random Forest procedure is logically powerful for SA for Kannada.

Proposed Work

Sentiment analysis helps data analysts within large enterprises gauge public opinion, conduct nuanced market research, monitor brand and product reputation and understand customer experiences. VADER (Valence Aware Dictionary and Sentiment Reasoner)is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER usesa combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labelled according to their semantic orientation as either positive or negative. VADER not only tells about the Positivity and Negativity score but also tells usabout how positive or negative a sentiment is. VADER analyses a piece of text it checks to see if any of the words in the text are present in the lexicon. We can see that all of the elements of this text that indicate that the writer is unhappy (in the blue boxes) areactually informal writing multiple punctuation marks, acronyms and an emoticon. Together additional components for intensifier and negations are handled.[12]

VADER

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.

It is used for sentiment analysis of text which has both the polarities i.e. positive/negative. VADER is used to quantify how much of positive or negative emotion the text has and also the intensity of emotion.[13]

Advantages

Here are the advantages of using VADER which makes a lot of things easier:

- It does not require any training data
- It can very well understand the sentiment of a text containing emoticons, slangs, conjunctions, capital words,punctuations and much more
- It works excellent on social media text
- VADER can work with multiple domains

Proposed Algorithm

Algorithm for the Negation Handling

Step 1: Read WORD,NFILE.

Step 2: Set N := LENGTH(NFILE).Step 3: Set FOUND :=FALSE.

Step 4: Repeat For I := 1 to N By 1 do :

Step 5: If NFILE(I)==WORD then :

Set FOUND: = TRUE.

[End of If structure][End of For Loop]

Step 6: If FOUND=FALSE then :

Return FALSE.

Else

Return TRUE. [End of If structure]

Step 7: Return

Algorithm for Multi-Negation Handling

Step 1: Read WORD, PWORD1 , PWORD2 Step 2: If NEGATION(WORD) then :

If INTENSIFIER(PWORD2) then :

If INTENSIFIER(PWORD2) then :Set SCORE=3

Else

Set SCORE =2

[End of If structure]

Else

Set SCORE =1

[End of if structure] Step 3: Return Score.Step 4: Stop.

Algorithm for the Intensifier Handling

Step 1 : Read WORD,IFILE.

Step 2: Set N := LENGTH(IFILE).Step 3: Set FOUND :=FALSE.

Step 4: Repeat For I := 1 to N By 1 do :Step 5: If IFILE(I)==WORD then :

Set FOUND: = TRUE.

[End of If structure][End of For Loop]

Step 6: If FOUND=FALSE then :

Return FALSE.

Else

Return TRUE. [End of If structure]

Step 7: Return

Algorithm for Multi-Intensifier Handling

Step 1: Read WORD, PWORD1 , PWORD2 Step 2: If INTENSIFIER(PWORD1) then :

If INTENSIFIER(PWORD2) then :Set SCORE=3

Else

Set SCORE =2

Else

[End of If structure]

27

Chadha R et al.
J. Engr. Desg. Anal. 2023; 6(1)

Set SCORE =1

[End of if structure] Step 3: Return Score.

Step 4: Stop.

Conjunction Handling: The conjunction is used for handling the multiple sentences which are joined together. We have takenthe following conjunctions in consideration

And , or , either

We have stored these conjunctions in the file and when a sentence contains either of these words we will split the sentence intotwo clauses on the basis of the conjunction find.[14]

Implementation and Result Analysis

The Implementation is created using the writing of the code in python containing the statements for the VADER based sentimentanalysis with the improvement using the multi-intensifier and negation handling.[15]

Figure 1.Implementation

Result Analysis

The reviews are taken on the basis of the tweets on the twitter and the random sites and collected and checked using theimplementation which we have performed and the result obtained in presented in the form of tables.

Multi-Negation Handling

Multi-Negation Handling refers to handling the negation words which reverts the meaning of the sentence and also the presenceof the multiple negative words. The sentence which is positive will become negative.[16]

In the below mentioned Table 1, the collection of the reviews which contains both the reviews containing the negation work are taken up and the same review is then analyzed without the negative words. N means negative review and

P means the positive review. And the scoring the in Table 1, means the number of the positive words , negative words, positive emotions and negative emotions summation and on the basis of the result from the summation we decides which result is better and also check the polarity of the review.[21]

Table 1.Comparison Test Results For Negation Handling

ReviewText	Review Type	Review Score
the camera was not very good (Review1)	Negative	-0.773
thecamerawasverygood	Positive	0.9854
the camera was not very bad(Review2)	Positive	0.9416
the camera was very bad	Negative	-1.1698
the cloth stitching is not extremely very good (Review3)	Negative	-.854
the cloth stitching is extremely very good	Positive	1.0758
the penis notvery bad quality (Review4)	Positive	0.9416
the penis verybad quality	Negative	-1.1698

Multi-Intensifier Handling

Multi-Intensifier Handling refers to handling the intensifier words which push up the meaning of the sentence and also the presence of the multiple intensifiers. The sentence score get emphasized by making use of the intensifier words.[19]

In the below mentioned Table 2 the collection of the reviews which contains both the reviews containing the intensifiers words are taken up and the same review is then analyzed without the intensifier words. N means negative review and P means the positive review.

And the scoring the in Table 2, means the number of the positive words, negative words, positive emotions and negativeemotions summation and on the basis of the result from the summation it is decided which result is better and check the polarityof the review.[20]

Table 4.Comparison Test Results For Intensifier Handling

Review Text	Review Type	Review Score
camera quality is extremely very good(Review 1)	Positive	1.0758
camera quality is very good	Positive	0.9854
The packing of the product supplied is very good (Review 2)	Positive	0.9854

Chadha R et al.
J. Engr. Desg. Anal. 2023; 6(1)

28

Review Text	Review Type	Review Score
The packing of the product supplied is good	Positive	.4404

Random Reviews

Table 3.Comparison test results for Random reviews

Review Text	Review Type	Review Score
acting, direction, story, music, dialogues everything is fabulous.	Positive	1.0534
rajesh khanna as anand is absolutely brilliant this is his career bestperformance.	Positive	0.8516
this movie is one of the best movies that Hrishikesh Mukherjee gave.	Positive	.6369
hrishikesh mukherjee was always brilliant but in this his direction is exceptional.	Positive	.34
the acting is superb with Amul Palekar , Dina Pathak , Utpal Dutt, Manju Singh and a whole lot of special appearances.	Positive	.7783
this is probably actor best role in hindi movies.	Positive	.6369
the most fabulous music by r d burman.	Positive	1.1418
the most delightful thing to me about this light comedy is the invention ofplayback singing as a seduction strategy for Sunil Dutt.	Positive	1.529
excellent performances by complete cast.	Positive	.5719
great performances by actor sanjeev kumar.	Positive	.6249
good supporting acting by the rest of the cast.	Positive	.7003
songs are not bad and the song pritam aan milo makes movie superb andworth to watch.	Positive	.8338

Conclusion

In order to assign weighted sentiment ratings to the entities, topics, themes and categories contained inside a sentence or phrase, sentiment analysis systems for text analysis integrate natural language processing (NLP) and machine learning approaches. Sentiment analysis aids data analysts at major corporations in understanding consumer experiences, doing complex market research, monitoring brand and product reputation and gauging public opinion.

A vocabulary and rule-based sentiment analysis tool called VADER (Valence Aware Dictionary and Sentiment Reasoner) is customised precisely to the sentiments expressed in social media. VADER makes use of a variety of A sentiment lexicon is a collection of lexical elements (such as words) that are often classified as either positive or negative depending on their semantic orientation. VADER not only informs us of the positivity and negativity scores, but also of the sentimentality of each score. When VADER analyses a text, it looks to see if any of the terms are recognised by the lexicon. We can observe that all of the characteristics in this text (in the blue boxes) that suggest the author is dissatisfied are truly informal. Using a lot of punctuation, acronyms and emoticons. Additional components for negations and intensifiers are handled collectively.

References

1. K. Zvarevashe and O. O. Olugbara, "A framework for sentiment analysis with opinion mining of hotel reviews," 2018 Conference on Information Communications Technology and Society (ICTAS), Durban, South Africa, 2018, pp. 1-4.

2. Valdivia, Ana & Luzon, Maria & Herrera, Francisco. (2017). Sentiment Analysis in TripAdvisor. IEEE Intelligent Systems. 32. 72-77.

3. M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5,

4. L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learninig Methods," 2020 International Conference for EmergingTechnology (INCET), Belgaum, India, 2020, pp. 1-5.

5. S. Dhawan, K. Singh and P. Chauhan, "Sentiment Analysis of Twitter Data in Online Social Network," 2019 5th International Conferenceon Signal Processing, Computing and Control (ISPCC), Solan, India, 2019, pp. 255-259, doi: 10.1109/ISPCC48220.2019.8988450.

6. Chauhan and S. Sehgal, "Sentiment analysis on product reviews," 2017 International Conference on Computing, Communication andAutomation (ICCCA), Greater Noida, 2017, pp. 26-31.

7. C. Zhao, C. Hu and T. Peng, "Analysis of Product Evaluations: An Adaptive Approach Based on Extended Sentiment Dictionaries," 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, 2017, pp. 148-152

8. Z. Singla, S. Randhawa and S. Jain, "Statistical and sentiment analysis of consumer product reviews," 2017 8th International Conferenceon Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6.

9. P. R. Mala and S. S. Devi, "Product response analytics in

29

Chadha R et al.
J. Engr. Desg. Anal. 2023; 6(1)

Facebook," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 1265-1269.

10. N. M. Shelke, V. Thakre and S. Deshpande, "Identification of scope of valence shifters for sentiment analysis of product reviews," 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), Patna, 2016, pp. 265-269.

11. Y. Hegde and S. K. Padma, "Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada," 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, 2017, pp. 777-782.

12. Shastri, L., Parvathy, A. G., Kumar, A., Wesley, J., & Balakrish-nan, R. (2010). Sentiment Extraction. IAAI-10.

13. Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. Y. (2008). Cheap and Fast - But is it Good?. In Proc. EMNLP-08.

14. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C., Ng, A., & Potts, C. (2013). Recursive Deep Models for Semantic Compositionality Over Sentiment Treebank. In Proc. EMNLP-13.

15. Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M.(1966). General Inquirer. Cambridge, MA: MIT Press.

16. Strauss, A. L., & Corbin, J. (1998). Basics of Qualitative Re-search. Thousand Oaks, CA: Sage Publications.

17. Surowiecki, J. (2004). The Wisdom of Crowds. NY, NY: Anchor.

18. Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M.(2010). Predicting Elections with Twitter:. In Proc. ICWSM-10.

19. Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism. ACM Trans. Inf. Syst., 21(4), 315–346.

20. Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S.(2012). A system…real-time Twitter sentiment analysis. ACL-12.

21. Wilson, T., Wiebe, J., & Hwa, R. (2004). Just how mad are you? finding strong and weak opinion clauses. In Proc. NCAI-04s.