

Research Article

A Study Analyzing an Innovative Approach to Sentiment Analysis with VADER

Raman Chadha¹, Aryan Chaudhary²

¹Professor, Computer Science & Engineering Department, UIE, CHANDIGARH University, Punjab, India.

²Chief Scientific Advisor, Bio-Tech Sphere Research, India.

I N F O

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>

How to cite this article:

Chadha R, Aryan C. A Study Analyzing an Innovative Approach to Sentiment Analysis with VADER. *J Engr Desg Anal* 2023; 6(1): 23-27.

Date of Submission: 2023-04-10

Date of Acceptance: 2023-05-20

A B S T R A C T

Sentiment Analysis is the process of determining whether a written piece conveys a positive, negative, or neutral tone. A sentiment analysis system for text analysis combines Natural Language Processing (NLP) and machine learning techniques to assign weighted sentiment scores to entities, topics, themes, categories within a sentence or phrase. Sentiment analysis aids data analysts in large enterprises to gauge public opinion, conduct nuanced market research, monitor brand and product reputation, understand customer experiences.

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool designed specifically for analyzing sentiments expressed in social media. VADER utilizes a sentiment lexicon, which is a list of lexical features (such as words) categorized based on their semantic orientation as positive or negative. VADER not only provides a Positivity and Negativity score but also offers insights into the intensity of the sentiment. When analyzing a piece of text, VADER checks if any of the words in the text are present in the lexicon. It can identify elements that indicate the writer's sentiment, whether positive or negative, including informal language, multiple punctuation marks, acronyms, even emoticons. Additionally, VADER accounts for intensifiers and negations in the text.

Keywords: Sentiment Analysis, Polarity, VADER

Introduction

In this study, we delve into the examination of an inventive method for conducting sentiment analysis through the utilization of VADER, the Valence Aware Dictionary, Sentiment Reasoner. Sentiment analysis is the practice of determining whether a piece of text conveys a positive, negative, or neutral sentiment. It involves the application of natural language processing (NLP) and machine learning techniques to assign weighted sentiment scores to various entities, topics, themes, categories within a given text. This

analytical process is invaluable for data analysts within large enterprises as it enables the assessment of public opinion, the execution of detailed market research, the monitoring of brand and product reputation, the comprehension of customer experiences.^{1,2}

VADER, as a lexicon and rule-based sentiment analysis tool, is tailored to the examination of sentiments expressed in the context of social media. It operates through a sentiment lexicon, which consists of a comprehensive list of lexical features, such as words, labeled according

to their semantic orientation as positive or negative. What sets VADER apart is its ability not only to provide a measure of positivity and negativity but also to offer insights into the degree of sentiment intensity. When VADER analyzes a piece of text, it scans for the presence of words contained within its sentiment lexicon. This method allows VADER to pinpoint elements in the text that indicate the writer's emotional tone, whether it's positive or negative. Furthermore, VADER is equipped to handle informal language, multiple punctuation marks, acronyms, even emoticons in its analysis. It also takes into account additional components such as intensifiers and negations to provide a more nuanced sentiment analysis.

This study aims to explore the effectiveness and innovation of VADER in the realm of sentiment analysis, shedding light on its capabilities and potential applications.^{3,4}

Today Internet has turned into a vital part of the life of individuals. Individuals frequently utilize sites, discussions, e-news, social systems administration locales like Facebook and Twitter to express their perspectives and opinions. The web is turning into a basic leadership source. Extensive measure of substance is created on the web each day, along these lines mining data and extracting client sentiment, is a vital errand.⁵

Microblogging 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 hashtags (beginning with #) which bunch the tweets on a comparative subject containing that specific hashtag. At the point when a specific hashtag is utilized by a greatly substantial number of individuals, at that point it turns into a pattern that pulls in more individuals to take an interest in the discussion.⁶

Opinion Mining, another name for sentiment analysis, is a kind of text mining that orders the text into various classes. Opinion Mining utilizes some methods to categorize opinions into classes like positive, negative, nonpartisan. This division of classes is called the extremity of text. The principal target of opinion mining is a classification of sentiments into various classes.⁷

In this study, we investigate an innovative approach to sentiment analysis using VADER, the Valence Aware Dictionary, Sentiment Reasoner. Sentiment analysis is a crucial tool for assessing the emotional tone of written text, categorizing it as positive, negative, or neutral. It combines techniques from natural language processing (NLP) and machine learning to assign weighted sentiment scores to various components within a given text. This process is invaluable for data analysts, aiding them in understanding public opinion, conducting in-depth market

research, monitoring brand and product reputation, gaining insights into customer experiences.^{8,9}

VADER stands out as a lexicon and rule-based sentiment analysis tool that is specially designed to analyze sentiments expressed in social media and informal text. It leverages a sentiment lexicon, which is a comprehensive list of lexical features (such as words) categorized according to their semantic orientation as either positive or negative. VADER not only provides overall positivity and negativity scores but also offers granularity by indicating the intensity of the sentiment. When VADER analyzes a piece of text, it examines the presence of words found in its sentiment lexicon, enabling it to identify elements that convey the writer's emotional tone, whether it's positive or negative. VADER also handles informal language, multiple punctuation marks, acronyms, even emoticons during its analysis. Additionally, it takes into consideration intensifiers and negations, providing a more nuanced understanding of sentiment.^{10,11}

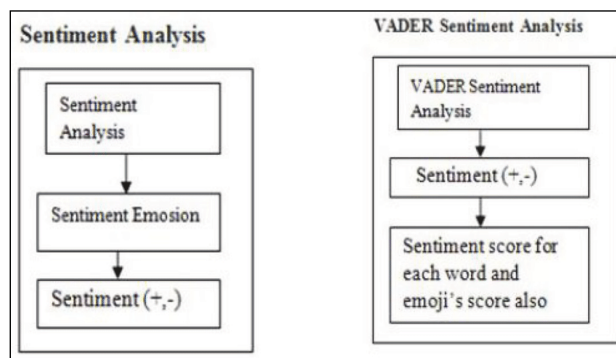


Figure 1. Analysis of Sentiments in VADER Statements

Example

To illustrate the power of VADER in sentiment analysis, consider the following sentence:

“Despite the rainy weather, the outdoor event was a huge success! “

VADER's analysis of this sentence would involve identifying words like “success” and the positive emoticon.” It also takes into account the contrast in the sentence (“Despite the rainy weather”), understanding that it may introduce some negativity. However, the overall sentiment score generated by VADER for this sentence would likely be positive due to the presence of strong positive indicators.

This example demonstrates how VADER can effectively capture nuanced sentiments even in the presence of contrasting elements, making it a valuable tool for sentiment analysis in various contexts.^{12,13}

Literature Survey

M. Wongkar and A. Angdresey:¹⁴ Sentiment analysis, as

defined by Wongkar and Angdresey in their study, is the process of evaluating public sentiments and opinions regarding various entities, including products, services, political figures, celebrities. Their research focused on conducting sentiment analysis for the 2019 Republic of Indonesia presidential candidates using Python. The methodology included steps such as data collection using Python libraries, text processing, testing training data, text classification using the Naïve Bayes method. The study found that the Jokowi-Ma'ruf Amin pair received a positive sentiment score of 45.45% and a negative score of 54.55%, while the Prabowo-Sandiaga pair had a positive sentiment score of 44.32% and a negative score of 55.68%. The combined data achieved an accuracy of approximately 80.1%. The study also compared the Naïve Bayes, SVM, K-Nearest Neighbor (K-NN) methods using RapidMiner, resulting in accuracy values of 75.58%, 63.99%, 73.34%, respectively.

L. Mandloi and R. Patel:¹⁵ Sentiment analysis, also referred to as material polarity or opinion mining, involves determining whether the sentiment expressed in text is positive, negative, or neutral. The authors highlighted the significance of sentiment analysis in the context of social media platforms, particularly Twitter, where users share thoughts within the constraints of 280 characters. Twitter, with its vast user base and daily posting of 550 million tweets, offers a platform for analyzing the sentiments of society in a general sense. The study aimed to compare various machine learning methods, including the Naïve Bayes Classification method, Support Vector Machine Classification Method, Maximum Entropy Classification method.

S. Dhawan, K. Singh, P. Chauhan:¹⁶ Sentiment Analysis, described by Dhawan, Singh, Chauhan, involves computationally determining whether a piece of text exhibits positive or negative sentiment. The proposed method in this study calculates the polarity of each tweet to distinguish its sentiment as positive or negative. Polarity, in this context, represents the emotions conveyed by the user, such as anger, sadness, happiness, joy. The mechanism was implemented using the Python programming language.

C. Chauhan and S. Sehgal:¹⁷ Chauhan and Sehgal discussed the analysis of product reviews as a means of understanding public sentiment. They emphasized the need for summarizing voluminous product reviews, focusing on extracting feature-specific summaries from the reviews. Their research concentrated on extracting meaningful reviews and shaping genuine product reviews. Future work includes expanding the scope to more product review websites and employing advanced methods or tools for improved results, excluding keywords not present in the dataset.

C Zhao, C Hu, T Peng:¹⁸ Zhao, Hu, Peng explored sentiment analysis, specifically in the context of the Chinese language. Their research focused on summarizing product reviews and classifying them into positive, neutral, or negative categories. They extended general sentiment dictionaries to include specific ones for different subjects, taking into account degree verb modifiers and negations. Their approach achieved higher accuracy compared to previous methods based on general sentiment dictionaries. The study utilized the K-means clustering algorithm to classify the positive and negative sets, word cloud images were generated to visualize the keywords within each category. The findings offer insights for marketing strategies based on the words and sentiments depicted in these images.

Benefits of VADER

VADER (Valence Aware Dictionary and sentiment Reasoner) is a sentiment analysis tool that is specifically designed for social media text, which often contains informal language, slang, emoticons. It was developed by researchers at the Georgia Institute of Technology and is a lexicon and rule-based sentiment analysis tool. VADER has several advantages, including:

1. **Designed for Social Media:** VADER is well-suited for sentiment analysis of text from social media platforms like Twitter, Facebook, other platforms where people often use informal language, abbreviations, emojis to express their sentiments. It can effectively handle these unique aspects of social media text.
2. **Valence-Based Analysis:** VADER uses a valence-based approach, which means it not only classifies text as positive, negative, or neutral but also assigns a sentiment intensity score. This allows for a more nuanced understanding of sentiment, as it can differentiate between strong and weak sentiments.
3. **Built-in Emoticon and Slang Handling:** VADER has a built-in lexicon that includes a wide range of emoticons, slang terms, other informal expressions commonly found in social media text. This helps it accurately interpret the sentiment of such expressions.
4. **Fast and Lightweight:** VADER is a relatively lightweight sentiment analysis tool, which means it can process text quickly. This makes it suitable for real-time sentiment analysis applications, such as monitoring social media trends.
5. **Pre-trained Model:** VADER comes pre-trained on a large dataset of social media text, making it a convenient choice for sentiment analysis tasks without the need for extensive custom training on specific datasets.
6. **No Training Data Required:** Unlike some machine learning-based sentiment analysis approaches that require extensive training on labeled data, VADER relies on a predefined lexicon and rule-based heuristics.

This makes it easier to implement and use without the need for training data.

- 7. **Open Source:** VADER is an open-source tool, which means it is freely available to the public and can be easily integrated into various applications and programming languages.

While VADER has several advantages, it's important to note that it may not perform as well on certain types of text, such as formal documents or texts with complex and nuanced sentiments. For those cases, machine learning-based sentiment analysis models might be more appropriate. Additionally, VADER's lexicon might not cover all possible slang and neologisms, so it may require occasional updates to stay accurate in the ever-evolving world of social media language.

Proposed Work

The phrase "Analytical Examination of a Crowdfunding Application Employing Blockchain Technology + Proposed work" implies that there is a comprehensive review and study of a crowdfunding application that utilizes blockchain technology.

This analysis likely involves investigating the mechanisms and features of the crowdfunding platform, especially in how it leverages blockchain technology for its operation.¹⁹

Additionally, the term "Proposed work" suggests that there is an intention to conduct further research or activities related to this subject. This could involve identifying areas for improvement, developing new features, enhancing the existing platform, or exploring the potential for innovation and advancement within the context of crowdfunding and blockchain technology.

In summary, this phrase indicates that a detailed evaluation of a specific crowdfunding application that incorporates blockchain technology is being conducted, there are plans to undertake additional work or research in this domain to potentially enhance or extend the platform's capabilities.

Implementation and Result Analysis

The Implementation is created using the writing of the code in Python containing the statements for the VADER-based sentiment analysis with the improvement using the multi-intensifier and negation handling.^{20,21}

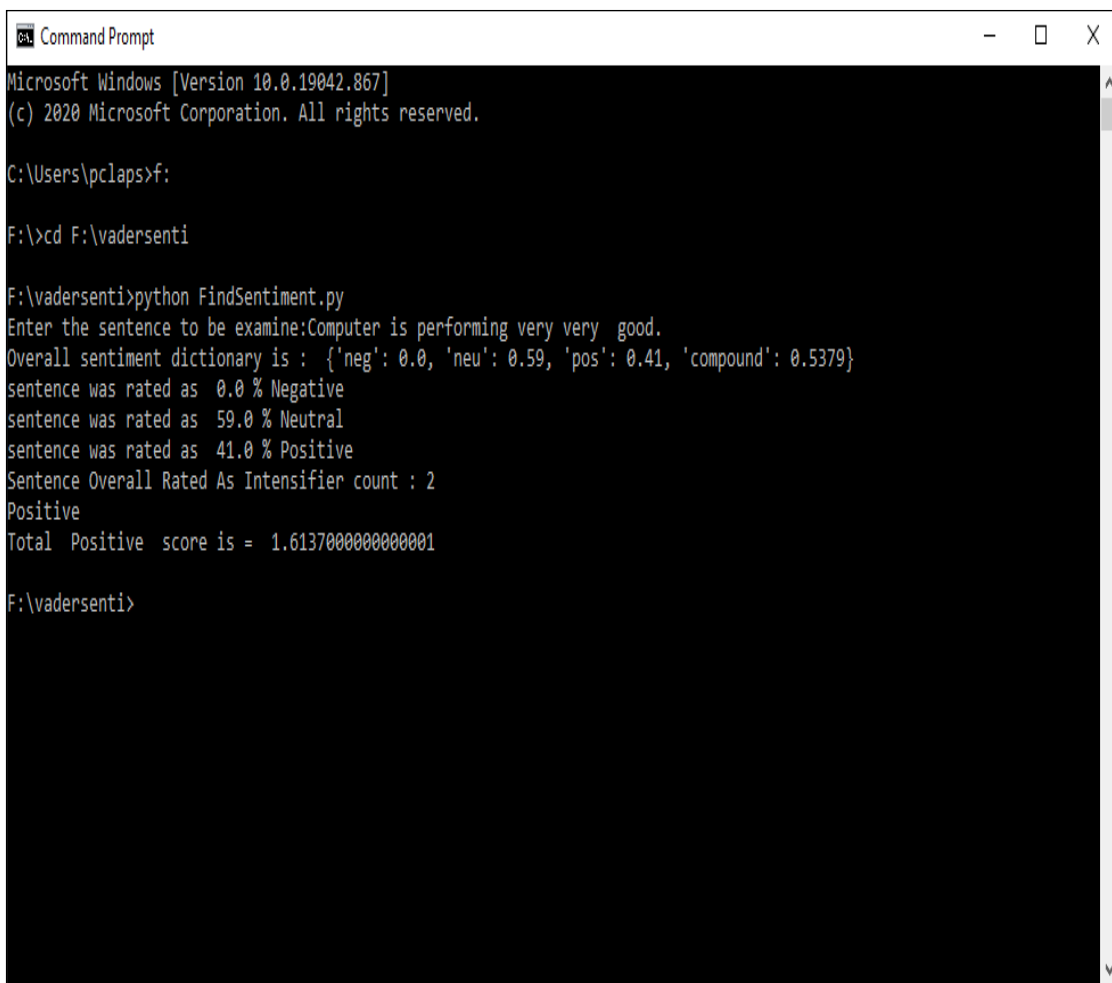


Figure 2.Implimentation

A system for analyzing sentiment in text combines natural language processing (NLP) and machine learning techniques to allocate weighted sentiment scores to various aspects like entities, topics, themes, categories within a sentence or phrase. Sentiment analysis is valuable for large enterprises, enabling data analysts to understand public sentiment, conduct intricate market research, monitor brand and product reputation, gain insights into customer experiences. VADER (Valence Aware Dictionary and Sentiment Reasoner) is a sentiment analysis tool based on a lexicon and rule-based approach, specially designed for capturing sentiments expressed in social media. VADER employs a sentiment lexicon, which is essentially a list of words categorized based on their positive or negative orientation. Notably, VADER provides not just a Positivity and Negativity score but also offers a measure of the sentiment's intensity. When VADER evaluates a piece of text, it checks if any of the words in the text are present in the lexicon. Additionally, it handles informal language, multiple punctuation marks, acronyms, even emoticons in its analysis. Furthermore, VADER considers intensifiers and negations as additional components in its sentiment analysis.

References

1. NM Shelke, V Thakre S. Deshpande conducted a study titled "Identification of scope of valence shifters for sentiment analysis of product reviews" at the 2016 Sixth International Symposium on Embedded Computing and System Design (ISED) in Patna.
2. Hegde Y, SK. Padma presented a paper on Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada at the 2017 IEEE 7th International Advance Computing Conference (IACC) held in Hyderabad.
3. Valdivia, Ana & Luzon, Maria & Herrera, Francisco. published Sentiment Analysis in TripAdvisor in IEEE Intelligent Systems in 2017.
4. M Wongkar and A. Angdresey presented Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter at the 2019 Fourth International Conference on Informatics and Computing (ICIC) in Semarang, Indonesia.
5. In Shastri L, Parvathy AG, Kumar, A., Wesley, J., & Balakrishnan, R. explored "Sentiment Extraction" at the IAAI-10 conference.
6. Snow R, O'Connor B, Jurafsky D. Discussed Cheap and Fast - But is it Good? in their research presented at the EMNLP-08 conference in 2008.
7. Socher R, Perelygin A, Wu J. Introduced Recursive Deep Models for Semantic Compositionality Over Sentiment Treebank at the EMNLP-13 conference in 2013.
8. Singla Z, Randhawa S, Jain S. conducted Statistical and sentiment analysis of consumer product reviews at the 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) in Delhi.
9. Mala PR, Devi SS. explored Product response analytics in Facebook at the 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) in Madurai.
10. Mandloi L, Patel R. presented Twitter Sentiments Analysis Using Machine Learning Methods at the 2020 International Conference for Emerging Technology (INCET) in Belgaum, India.
11. Dhawan S, Singh KP. Chauhan worked on Sentiment Analysis of Twitter Data in Online Social Network at the 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC) in Solan, India.
12. Chauhan C, Sehgal S. conducted research on "Sentiment analysis on product reviews" at the 2017 International Conference on Computing, Communication, Automation (ICCCA) in Greater Noida.
13. Zhao C, Hu C, Peng T. explored Analysis of Product Evaluations: An Adaptive Approach Based on Extended Sentiment Dictionaries at the 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) in Hangzhou.
14. Stone PJ, Dunphy DC, Smith MS. Published the General Inquirer in 1966.
15. Strauss AL, Corbin J. authored the Basics of Qualitative Research book in 1998.
16. Surowecki J. wrote The Wisdom of Crowds in 2004.
17. Tumasjan A, Sprenger TO, Sandner PG. explored Predicting Elections with Twitter at the ICWSM-10 conference in 2010.
18. Turney PD, Littman ML. presented Measuring praise and criticism at ACM Trans. Inf. Syst. in 2003.
19. Wang H, Can D, Kazemzadeh A, Bar F. worked on A system real-time Twitter sentiment analysis at ACL-12.
20. Wilson T, Wiebe J, Hwa R. conducted research on Just how mad are you? finding strong and weak opinion clauses" at NCAI-04s.
21. K Zvarevashe, OO Olugbara analyzed "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews" at the 2018 Conference on Information Communications Technology and Society (ICTAS) in Durban, South Africa.