Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 1, February 2025, pp. 726~732
ISSN: 2302-9285, DOI: 10.11591/eei.v14i1.8568

An opinionated sentiment analysis using a rule-based method

Mareye Zeleke Mekonen[1], Tsehay Admassu Assegie[2], Shamik Palit[3], Angati Kalyan Kumar[4], Chandrima Sinha Roy[5], Chandi Priya Kompala[6], Komal Kumar Napa[4]
[1]Department of Information Technology, College of Engineering and Technology, Injibara University, Injibara, Ethiopia
[2]School of Electronics Engineering, Kyungpook National University, Daegu, Republic of Korea
[3]Department of Computing Science and Software Engineering, University of Stirling RAK Campus, Al Dhait South, Ras Al Khaimah, United Arab Emirates
[4]Department of Computer Science & Engineering (Data Science), Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India
[5]Department of Computer Science and Engineering, Dr. Sudhir Chandra Sur Institute of Technology & Sports Complex, Kolkata, West Bengal, India
[6]Department of Science and Humanities (General Engineering), R.M.K. College of Engineering and Technology, Gummidipoondi, Thiruvallur, India

Article Info

Article history:

Received Apr 7, 2024
Revised Oct 5, 2024
Accepted Oct 17, 2024

Keywords:

Fake news
Hate speech
Kafi-noonoo
Opinionated text
Sentiment analysis

ABSTRACT

The categorization of opinions into positive, negative, or neutral facilitates information gathering, pinpointing individual weaknesses, and streamlining the decision-making process. Precision in opinion classification enables decision-makers to extract valuable insights, make well-informed decisions, and execute suitable actions. Sentiment analysis is language-specific due to the distinct morphological structures unique to each language, distinguishing them from one another. This study implemented a rule-based sentiment analysis approach for Kafi-noonoo opinionated texts, leveraging a rule-based system tailored for smaller datasets that operate based on a predefined set of rules. The rule-based mechanism calculates the overall polarity of a given sentence by applying a set of rules and categorizes it into positive, negative, or neutral sentiments upon identifying sentimental terms from a dedicated file. While the analysis utilized 1,500 words sourced from Facebook and music review samples, the modest sample size yielded satisfactory results. Performance evaluation metrics such as precision, recall, and F-measure were employed, indicating positive word scores of 91%, 86%, and 88.4%, and negative word scores of 80%, 75%, and 77%, respectively.

Corresponding Author:

Tsehay Admassu Assegie
School of Electronics Engineering, Kyungpook National University
41566, Daegu, Republic of Korea
Email: tsehayadmassu2006@gmail.com

1. INTRODUCTION

Individuals constantly attempt to convey ideas, views, feelings, and other information through language. People can now use their language for a variety of online purposes, such as topic modeling, information extraction, sentiment analysis (also known as opinion mining), and information retrieval (IR), thanks to technological advancements. The practice of examining product evaluations or other issues to ascertain the general perception or emotion of a product, public services, issues, subjects, and other specific items is known as sentiment analysis [1]-[3]. Reviews are an example of what is referred to as user-generated content, and they are gaining popularity as a valuable tool for marketing departments, sociologists, psychologists, and other professionals who may be interested in public opinion, viewpoints, and general or personal sentiments. Opinions and the opinions of others play a very important role in our decision-making process [4]. In recent years, research on the intersection of IR, artificial intelligence, and natural language processing has rapidly expanded to include sentiment analysis [5]. It is the subfield of natural language processing, and it classifies the given text into three categories: positive, negative, or neutral [6]-[8]. One way of doing this is through a lexicon of sentiment-laden words, each annotated with its respective polarity [9]. Lexicon-based methods use a word stock dictionary with opinion words and match a given set of words in a text to find polarity [10]. A sentiment analysis method that analyzes how opinions, reactions, impressions, emotions, and perspectives are expressed in a language [11]. Sentiment analysis plays a very important role in finding out the opinion of the user [12].

Additionally, governments, organizations, public services, various offices, and businesses all benefit greatly from sentiment analysis since it gives them a general idea of how consumers and customers feel about their offerings [13], [14]. For sentiment research, there are different levels; document level analysis focuses on understanding the general view of the document, it is categorized as good, bad, and ok instead of providing the insights or actual sense or purpose of the document. Sentence level; sentences are treated as separate parts and given labels denoting positive, negative, or neutral sentiment. Sentence-based analysis is a useful technique to detect changes in a document's opinion. It provides a more complex interpretation of the message by highlighting variations and fluctuating emotions throughout the text. Aspect/feature level; this finds and assesses the sentiment connected to certain topics or sections of a document rather than examining the general emotions, it divides the document [15].

In this study, a rule-based sentiment analysis approach was proposed for the Kafi-noonoo language to address the relevance of sentiment analysis in this context. Rule-based methods typically involve defining specific rules in a particular scripting language to identify subjectivity, polarity, or opinions [16]. Compared to other approaches like Naïve Bayes, rule-based methods are considered more efficient and offer better interpretability of the results [17]-[19].

Although sentiment analysis can be applied to any human language, certain methods may be language-specific, especially when employing rule-based techniques [20]-[22]. Sentiment analysis, being an amalgamation of linguistics and computer science, has predominantly focused on advanced languages such as English, Arabic, and others, with techniques that are not easily transferable to other languages. Even with local languages like Amharic, Afaan Oromo, and Tigrigna, there have been limited attempts at sentiment analysis. To the best of our knowledge, there has been no such trial conducted on the Kafi-noonoo language. For decision-makers, having valuable insights or indicators derived from sentiment analysis can aid in making informed decisions and taking appropriate actions. However, as of now, there is a lack of mechanisms specifically tailored for sentiment analysis in the Kafi-noonoo language. This highlights the significance and potential for further exploration and research in this language domain.

Spoken in the Kaffa Zone, Kafi-noonoo is one of the Omotic languages. The Kaffa Zone is currently found in Ethiopia's southwest. As per the CSA (2008) data, 870,213 individuals speak it. Kafi-noonoo was classified as belonging to the Gonga Gimojan branch of the northern Omotic. The Kafa people refer to themselves as Kafecho and Kafa is their self-name. To be more precise, the word "Kafa" designates the location or homeland of the Kafa people, whilst the terms "Kafecho" and "Kafechi" designate the male and female Kafa people, respectively. The word "Noonoo" in Kafa means "the tongue," which is another language word [23].

While sentiment analysis methods have advanced significantly in recent years, analyzing opinionated writing in languages like Kafi-noonoo, known for its intricate morphological structures, remains a challenge. These languages often pose difficulties due to their heavy reliance on statistical models or machine learning algorithms trained on vast corpora. As a result, existing sentiment analysis techniques may struggle to capture the subtle nuances of opinion expression in languages rich in linguistic variables [24].

Moreover, the current majority of sentiment analysis techniques tend to oversimplify the complex subjective beliefs and sentiment intensities often present in opinionated literature. They often prioritize simple polarity categorization over nuanced understanding. Therefore, given the unique grammatical structures and cultural nuances of languages such as Kafi-noonoo, there is an urgent need for specialized sentiment analysis techniques tailored specifically to these linguistic contexts. These techniques should effectively identify and interpret subjective sentiments articulated in textual content [25]. This study aims to bridge this gap by introducing a rule-based approach specifically designed for analyzing opinionated sentiment in the Kafi-noonoo language. By advancing sentiment analysis methodologies for underrepresented languages, this approach promises a deeper insight into public opinion within diverse linguistic environments.

2. METHOD
2.1. Data collection

The numerous social media sites are the source of the data used in this study. We were able to gather information from Kaffa Zone, YouTube, and Facebook. We obtained a great deal of information from YouTube since, in comparison to other social media platforms, it provided a greater representation of a variety of music, speeches, and other emotions. For our evaluation, we gather 500 sentences totaling 1,500 words from various fields, and we manually identify 354 good and 176 negative words from various sources. Because there is not much online content for Kafi-noonoo, data extraction was challenging and time-consuming.

2.2. Tools and techniques

We selected the Python programming language as our primary coding tool for its robust natural language processing capabilities, particularly in sentiment analysis, making it a powerful choice. Before delving into sentiment analysis, we diligently conducted several preprocessing steps to optimize our techniques. These steps encompass stop-word removal, normalization, and tokenization. Our workflow commenced with tokenization, where input data was segmented into distinct tokens. The core task involved categorizing genuine feedback into positive and negative classes. By leveraging a predefined set of delimiters, such as spaces, the text input was efficiently parsed and tokenized. In the context of the Kafi-noonoo language, normalization played a pivotal role in standardizing text by addressing variations in letter cases, ensuring consistency throughout. In our approach, stop words were systematically excluded from the analysis since they bear no relevance in the evaluation of sentences.

3. RESULTS AND DISCUSSION

As this study focuses on the sentimental analysis of sentences using the Kafi-noonoo language, our task is to begin by taking a sentence as input. The pre-processing steps (tokenize, normalize, and stop word removal) are then applied, and we take further steps to ensure that each word in the sentence is present in the sentiment lexicon. Polarity assignment is used if a sentiment word also known as a word that expresses an opinion about an object or a "polarity word" exists in that sentence. In other words, if the term is positive, its polarity is +2, and if it is negative, its polarity is -2. Based on a set of rules that the researchers designed, choose whether to apply a rule. Lastly, the review polarity value is assessed based on a set of rules designed by the researchers. Since our study is sentence-based, we must begin by accepting a sentence for review according to the following criteria. Total polarity is positive if it is greater than zero and negative unless it is zero.

3.1. Rules

We designed six rules based on different criteria. The rule-based method depends on the grammatical accuracy of sentences [25] so we design with the support of linguistics. The first rule focuses, on if any polarity term is followed by a negative term, the initial polarity value will be reversed. The second rule applied, if a positive sentiment term is preceded by an overstatement term, then the initial value of that term is increased from +2 to +3. The third rule depicted that; if a negative sentiment term is preceded by an overstatement term, then the initial value of the term is decreased by -1 from -2 to -3. Rule four says that; if a positive sentiment term is followed by an understatement term, the initial value of that term is decreased from +2 to +1. The fifth rule states that, if a negative sentiment term is followed by an understatement term, the initial weight of that term is increased by +1. The final rule is set like this, if a sentiment term is not linked to any contextual valence-shifting term, the initially assigned weight is considered for further process.

3.2. Result discussion

The experiment aimed to evaluate the sentiment analysis model's overall performance. To do this, we feed every sentence from the corpora we gathered from various social media platforms especially music reviews as input evaluation for our model. We used the music reviews mostly because there were not many easily accessible reviews in the Kafi-noonoo language. Consequently, 108 music reviews in total are gathered. Out of them, 57 are positive, 40 are negative, and the remaining 15 are neutral; an impartial party labeled these manually for the experiment.

The results were crosschecked with those from the manually classified evaluations. The number of reviews that are successfully and wrongly classified is compared as an evaluation measure for the prototype system. The reviews classified by the suggested prototype method are usually compared to the manually labeled (categorized) reviews. Text classifications use the IR assessment metrics of precision, recall, and F-measure.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Re\,c\,all = \frac{TP}{TP+FN} \quad (2)$$

$$F - score = \frac{2*(precision + recall)}{precision+recall} \quad (3)$$

The general-purpose Kafi-noonoo sentiment terms, which are the standard vocabulary of sentiment phrases, were employed in this investigation. Table 1 depicts the accuracy of the rule-based method based on the three well-known metrics. Facebook reviews and music are the two domains in which the experiment is run. Results are expressed in terms of precision, recall, and F-measure for every class and domain.

Table 1. Positive and negative results

Class	Precision	Recall	F-measure
Positive	0.91	0.86	0.884
Negative	0.80	0.75	0.77

The precision value of 91% indicates that out of all the text segments classified as positive sentiment. This means that the rule-based method is effective in accurately identifying positive sentiment in the text. However, it is essential to note that precision alone may not provide a comprehensive picture of the model's performance. The recall value of 86% suggests that the model successfully identified 86% of all actual positive sentiment instances in the text. This indicates that the rule-based method is relatively good at capturing positive sentiment but may miss some positive statements in the text. The F-measure of 88% is the harmonic mean of precision and recall, providing a balanced evaluation metric that considers both precision and recall. An F-measure of 88% indicates a good overall performance in identifying positive sentiment in the text.

The precision value of 80% for negative sentiment implies that when the model classifies a text segment as negative. This indicates that the rule-based method is relatively accurate in identifying negative sentiment in the text. The recall value of 75% suggests that the model captured 75% of all actual negative sentiment instances in the text. This indicates that while the model is decent at identifying negative sentiment, there is room for improvement in capturing all negative statements. Additionally, the F-measure of 77% for negative sentiment represents a balanced evaluation metric that considers both precision and recall. An F-measure of 77% indicates a moderate overall performance in identifying negative sentiment in the text. 80% precision, 75% recall, and 77% F-measure system performance for negative classes. Apart from the experiment, we also performed a manual check by contrasting the results of our suggested model created with the manually classified texts. Linguists who specialize in classification assist with manual classification. Despite the research being in its early stages, the trials' outcomes, which included precision, recall, and F-measure metrics in addition to manual experimentation, are very good as a first work in this area within Kafi-noonoo.

4. CONCLUSION

Given the vast volume of online data, an efficient mechanism is essential to organize and present data clearly and practically. While extensive research has been conducted on complex languages like English, there is a notable scarcity of studies on Ethiopian local languages, particularly Kafi-noon. In the current digital age, individuals utilize the internet for product introductions, communication, idea exchange, debates, and emotional expression in their native language. This trend underscores the capability of users, akin to speakers of other sophisticated languages, to perform a diverse range of tasks using natural language processing tools tailored to their specific language. Consequently, individuals derive a sense of connection to their cultural heritage through their mother tongue. Motivated by this intrinsic need, we embarked on a pioneering study titled "Sentiment analysis for the Kafi-noonoo language." This groundbreaking initiative represents the inaugural attempt to harness the potential of the Kafi-noonoo language within the realm of sentiment analysis, marking a significant advancement in linguistic research and technology integration.

In this research, sentiments extracted from Facebook posts and music video views are utilized to develop a monitored sentiment analysis method. Thanks to their easier accessibility, the researchers utilized reviews in this language that encompass various perspectives from these sources. By applying a rule-based approach, we categorize sentences into positive, negative, and neutral sentiments. The results underscore the encouraging performance of our model across all three metrics assessed. This initial trial effectively enables our approach to categorize texts with strong opinions in the Kafi-noonoo language. Moving forward, we recommend addressing the following aspects for further enhancement: prioritizing subjective analysis as a foundational step, exploring alternative strategies to achieve more favorable outcomes, and potentially integrating a morphological analyzer to supplement the labeling process.

ACKNOWLEDGEMENTS

This research has not received any funding from any organization or institution.

REFERENCES

[1] A. Braoudaki, E. Kanellou, C. Kozanitis, and P. Fatourou, "Hybrid data data driven driven and rule rule based based sentiment sentiment analysis analysis on Greek Greek text text," *Procedia Computer Science*, vol. 178, no. 2020, pp. 234–243, 2020, doi: 10.1016/j.procs.2020.11.025.

[2] H. Rahab, H. Haouassi, and A. Laouid, "Rule-based Arabic sentiment analysis using binary equilibrium optimization algorithm," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 2359–2374, 2023, doi: 10.1007/s13369-022-07198-2.

[3] P. Kaur, "Sentiment analysis using web scraping for live news data with machine learning algorithms," *Materials Today: Proceedings*, vol. 65, pp. 3333–3341, 2022, doi: 10.1016/j.matpr.2022.05.409.

[4] P. Chikersal, S. Poria, and E. Cambria, "SeNTU: sentiment analysis of tweets by combining a rule-based classifier with supervised learning," *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 647–651, doi: 10.18653/v1/s15-2108.

[5] A. M. Alnasrawi, A. M. Alzubaidi, and A. A. Al-Moadhen "Improving sentiment analysis using text network features within different machine learning algorithms," *Bulletin of Electrical Engineering and Informatics,* vol. 13, no. 1, pp. 405-412, Feb. 2024, doi: 10.11591/eei.v13i1.5576.

[6] M. I. Liaqat, M. A. Hassan, M. Shoaib, S. K. Khurshid, and M. A. Shamseldin, "Sentiment analysis techniques, challenges, and opportunities: Urdu language-based analytical study," *PeerJ Computer Science*, vol. 8, 2022, doi: 10.7717/PEERJ-CS.1032.

[7] B. Saju, S. Jose and A. Antony, "Comprehensive study on sentiment analysis: types, approaches, recent applications, tools and APIs," *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, Cochin, India, 2020, pp. 186-193, doi: 10.1109/ACCTHPA49271.2020.9213209.

[8] S. Hamed, M. Ezzat, and H. Hefny, "A review of sentiment analysis techniques," *International Journal of Computer Applications*, vol. 176, no. 37, pp. 20–24, 2020, doi: 10.5120/ijca2020920480.

[9] V. A. Kharde and S. S. Sonawane, "Sentiment analysis of Twitter data: a survey of techniques," *International Journal of Computer Applications*, vol. 139, no. 11, pp. 5–15, 2016, doi: 10.5120/ijca2016908625.

[10] S. Kumar, "Sentiment Analysis," *Python for Accounting and Finance*, pp. 265–300, 2024, doi: 10.1007/978-3-031-54680-8_17.

[11] A. Abdelrazeq, D. Janßen, C. Tummel, S. Jeschke, and A. Richert, "Sentiment analysis of social media for evaluating universities," *Automation, Communication and Cybernetics in Science and Engineering 2015/2016*, pp. 233–251, 2016, doi: 10.1007/978-3-319-42620-4_19.

[12] A. Yenkikar, N. Babu and S. Sangve, "R-SA: a rule-based expert system for sentiment analysis," *2019 IEEE Pune Section International Conference (PuneCon)*, Pune, India, 2019, pp. 1-7, doi: 10.1109/PuneCon46936.2019.9105682.

[13] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Computational Linguistics*, vol. 35, no. 2, pp. 311–312, Jun. 2009, doi: 10.1162/coli.2009.35.2.311.

[14] A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *Journal of Big Data*, vol. 10, no. 1, jun. 2023, doi: 10.1186/s40537-023-00781-w.

[15] M. B. M. A. Busst, K. S. M. Anbananthen, and S. Kannan, "Aspect-level sentiment analysis through aspect-oriented features," *HighTech and Innovation Journal*, vol. 5, no. 1, pp. 109–128, 2024, doi: 10.28991/HIJ-2024-05-01-09.

[16] M. S. Solanki, "Sentiment analysis of text using rule based and natural language toolkit," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 12S, pp. 164–168, Oct. 2019, doi: 10.35940/ijitee.l1049.10812s19.

[17] Y. Bidulya and E. Brunova, "Sentiment analysis for bank service quality: A rule-based classifier," *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, Baku, Azerbaijan, 2016, pp. 1-4, doi: 10.1109/ICAICT.2016.7991688.

[18] H. Liu and M. Cocea, "Fuzzy rule based systems for interpretable sentiment analysis," *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)*, Doha, Qatar, 2017, pp. 129-136, doi: 10.1109/ICACI.2017.7974497.

[19] N. Saraswathi, T. S. Rooba, and S. Chakaravarthi, "Improving the accuracy of sentiment analysis using a linguistic rule-based feature selection method in tourism reviews," *Measurement: Sensors*, vol. 29, Oct. 2023, doi: 10.1016/j.measen.2023.100888.

[20] C. Kumaresan and P. Thangaraju, "Sentiment analysis in multiple languages: a review of current approaches and challenges," *REST Journal on Data Analytics and Artificial Intelligence*, vol. 2, no. 1, pp. 8–15, 2023, doi: 10.46632/jdaai/2/1/2.

[21] L. I. Tan, W. S. Phang, K. O. Chin and A. Patricia, "Rule-based sentiment analysis for financial news," *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Hong Kong, China, 2015, pp. 1601-1606, doi: 10.1109/SMC.2015.283.

[22] I. Paramonov and A. Poletaev, "Adaptation of semantic rule-based sentiment analysis approach for Russian language," *2021 30th Conference of Open Innovations Association FRUCT*, Oulu, Finland, 2021, pp. 155-164, doi: 10.23919/FRUCT53335.2021.9599992.

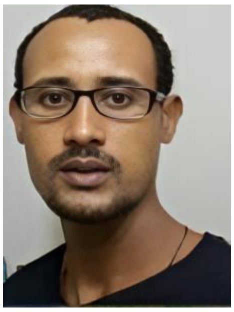
[23] M. Bordoloi and S. K. Biswas, "Sentiment analysis: A survey on design framework, applications, and future scopes," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 12505–12560, 2023, doi 10.1007/s10462-023-10442-2.

[24] T. Mulugeta and G. Mengistu, "The phonology of Kafi Noonoo ideophones," *Macrolinguistics and Microlinguistics*, vol. 3, no. 1/2, pp. 15–38, 2022, doi: 10.21744/mami.v3n1/2.29.

[25] P. Ray and A. Chakrabarti, "A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis," *Applied Computing and Informatics*, vol. 18, no. 1–2, pp. 163–178, Aug. 2022, doi: 10.1016/j.aci.2019.02.002.

BIOGRAPHIES OF AUTHORS

Mareye Zeleke Mekonen graduated with a bachelor's degree in 2015 from Hawassa University with a Department of Information Technology in Ethiopia and an advanced standing master's degree in information technology in 2019 from Jimma University in Ethiopia. He is currently employed with the College of Engineering and Technology as a Lecturer in the Department of Information Technology. His areas of interest in the study are computer vision, deep learning, NLP, artificial intelligence, machine learning, bioinformatics, and IR. He is the author of two papers and one book. He can be contacted at email: mareye132@gmail.com.



Tsehay Admassu Assegie holds a master of science degree in computer science from Andhra University, India 2016. He received his B.Sc., in computer science from Dilla University, Ethiopia in 2013. Currently, he is pursuing Ph.D. at the Department of Electronic and Electrical Engineering, Kyungpook National University, Daegu, Republic of Korea. His research interests include machine learning and biomedical image processing. He has published over 60 articles in reputed international journals and international conferences. Tsehay is an active member of the International Association of Engineers (IAENG), with membership number: 254711. He is an active reviewer of different MDPI journals. He has reviewed many research articles in MDPI, IEEE Access, and other reputed international journals verified by the Web of Science. He can be contacted at email: tsehayadmassu2006@gmail.com.

Dr. Shamik Palit is currently working as an Associate Professor in the Department of Computing Science and Software Engineering at the University of Stirling RAK Campus, United Arab Emirates. His areas of interest in research are enterprise application design and development, artificial intelligence, machine learning, enterprise resource planning, and software system engineering. He can be contacted at email: shamik1980@gmail.com.

Angati Kalyan Kumar is currently working as an Assistant Professor in the Department of Computer Science and Engineering (Data Science) at Madanapalle Institute of Technology and Science (MITS), Madanapalle, Andhra Pradesh, India. His research interests include machine learning and data mining. He can be contacted at email: kalyankumara@mits.ac.in.

Chandrima Sinha Roy is currently working as Assistant Professor in the Department of Computer Science and Engineering at Dr. Sudhir Chandra Sur Institute of Technology and Sports Complex, India. Her area of interest in research are artificial intelligence, machine learning, and application programming using Java. She can be contacted at email: mail2chandrima@gmail.com.

Chandi Priya Kompala is currently working as an Assistant Professor in the Department of Science and Humanities (General Engineering) at R.M.K. College of Engineering and Technology, Puduvoyal. Her research interests include machine learning, IoT, and data mining. She can be contacted at email: chandipriyamca@gmail.com.

Komal Kumar Napa is currently working as an Assistant Professor in the Department of Computer Science and Engineering at St. Peter's Institute of Higher Education and Research, Avadi, Chennai. His research interests include machine learning, data mining, and cloud computing. He can be contacted at email: komalkumarnapa@gmail.com.