ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 9, September 2023

Sentiment Analysis of Product Reviews Containing Hinglish Text

[1] Vandana Yadav, [2] Dr. Parul Verma, [3] Prof. (Dr.) Vinodini Katiyar

[1] Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, India
[2] Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow, India
[3] Dr. Shakuntala, Misra Rehabilitation University, Lucknow, Uttar Pradesh, India
Corresponding Author Email: [1] vandanayadav771@gmail.com, [2] pverma1@lko.amity.edu, [3] vkatiyar@dsmnru.ac.in

Abstract*— Due to how easy it is to compare items and read customer reviews from the comfort of one's home, online shopping has become much more popular. Customers frequently post product reviews to share their thoughts on the items they have purchased. Hindi has been found to be used in product reviews in India, and a large amount of Hindi literature contains words with strong opinions, such as "baahut achha," "bakbbas," "peisa wassool," etc. With an emphasis on reviews that include both Hindi and English language, we looked at numerous Hindi texts from Indian e-commerce websites. In order to provide a final evaluation of the features and general quality of the product, our study sought to identify the tone used in each review.

Index Terms*— Sentiment Analysis (SA), Review Analysis, Part-Of-Speech Tagging, Product Summary.

I. INTRODUCTION

Due to widespread use the Internet and e-commerce worldwide, society's way of life is changing. When purchasing a product in the past, advertisements and recommendations from friends were two primary sources of information.

There weren't many recommendations to compare similar products from various brands. As the e-commerce industry has developed, more products are now being offered. The e-commerce companies often ask their customers to provide product reviews that describe their experiences with the items they purchased. Customers can learn important details about the product they intend to purchase from these reviews, and they can also use them to compare goods from various brands. By comparing items based on what other customers think of them, reviews assist shoppers in making the best product selections. By alerting the makers about the benefits and shortcomings of their products, it also helps with product enhancement.

The amount of product reviews has greatly increased as e-commerce businesses have grown. However, it has grown to be very difficult for both buyers and sellers to manually analyze a large number of reviews and extract useful information. Many academics have been inspired by this to automate the analysis of reviews in order to find the hidden information that they contain. Due to Hindi's popularity in India, many customer reviews submitted by Indians contain both English and Hindi contents written in English script. These Hindi literatures frequently make use of everyday words with strong opinions that can be either favorable or bad, such as baahut acchha, baakbas, and peisa wassool. However, because previous studies on determining the polarity of opinions for product reviews were primarily focused on English-only texts, they largely ignored these texts. This study seeks to repair these inaccuracies and add Hindi term descriptions to the English texts in order to fill this vacuum in the literature. By gathering evaluations from well-known Indian e-commerce websites like Amazon.in and Flipkart.com, we were able to analyze the sentiment of product reviews that contain opinionated texts in both English and Hindi. The equivalent English terms were found after preprocessing the Hindi texts. The product review summary was computed using the Sent WordNet database. Section 2 presents the proposed system design and methodology, Section 3 presents our findings, and Section 4 brings the conclusion to a close.

II. PROPOSED WORK

We have gathered content from famous Indian e-commerce websites like flipkart.com and amazon.in that include product reviews. The collection contains numerous errors in the form of abbreviations with numerical characters, such as gr8 for wonderful, as well as joint terms like verygood. Additionally, the dataset includes Hindi words like "bakkbas," "bekkar," and "achchha" solely spelled in English script. Tables 2 and 3 provide a few representative typos that were collected from various Indian e-commerce websites. Words from Hindi texts transcribed in English are listed in Table 1 along with their English translations. Table 2 lists several common acronyms used online for review, messaging, etc. along with their English equivalents.

Table 3 displays a few joint words with blank spaces.

The main goal is to preprocess product reviews from Indian e-commerce websites in order to translate them into prose in English. The next stage is to use the Wordnets database to tag the text with Part of Speech (POS) after the reviews have been translated into English. The project then

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 9, September 2023

extracts adjectives, nouns, and adverbs from the text after POS tagging. Additionally, the adjectives used in the review are given numerical values using the SentiWordNet database. Figure 1 depicts the suggested system design.

Table1. List of Incorrect Hindi Words

Incorrect Text	Corrected Text
Ye	this
acchha	good
grt	great
Ni8	night
Som1	Some one

Table2. List of Incorrect English Words

Incorrect Text	Corrected Text
gud	good
excelent	excellent
bd	bad
goood	good
awsom	awesome

The algorithm we suggest is provided below using a pseudo code.

Suggested Algorithm

1. Tokenize in the first step using space.
2. **2.** Look up WorldNet
3. Go to the POS tagger if the word matches.
4. If not, correct the term and move to the POS Tagger.
5. **3.** The frequent feature database is used to store the noun, adverbs, and adjective.
6. **4.** Create the product summary in Step 4 using the SentiWordNet Lexical databases.
7. With the aid of the example provided here, the operation of our plan is analyzed. Yeh achha headphone h. eski sound quality bahut acchhi hai. Bass is too much. battery is bakkbas. design is verygud but not g8t.
8. The text is divided into multiple sentences using the [.], [?], and [!] symbols. These are examples of assessment sentences:
9. **Text1.** Yeh achha headphone h.
10. **Text2.** eski sound quality bahut achhi hai.
11. **Text3.** Bass is to much.
12. **Text4.** Battery is bakbas.
13. **Text5.** Design is verygd bt not g8t.
14. This is the English translation for the Hindi word Yeh. Achha is another Hindi term that translates to "good" in English. After correcting and converting every word to English, the entire review was written in English.
15. **Text1**. This headphone is good
16. **Text2.** The sound quality is very high.
17. **Text3.** Bass is too much.
18. **Text4.** Battery is bad.
19. **Text5.** Design is very good but not great.

Table3. Combination of Word

Joint Text	Corrected Text
Verygood	Very good
Verybad	Very bad
Bahutbekar	Bahut bekar

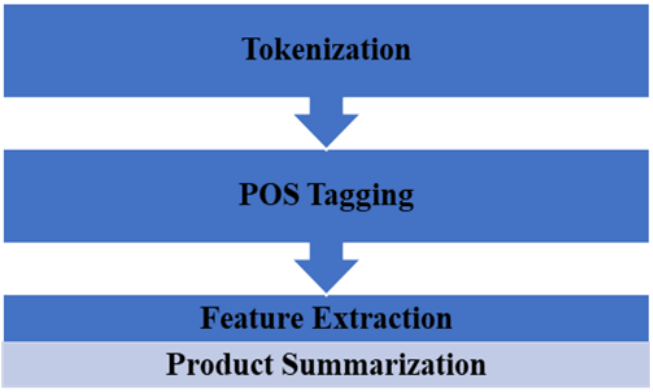


Figure1. Proposed Architecture

The POS labeling for a single sentence is implemented as shown below. Part of Speech Tagging was done using the Penn Treebank tagset.

This======= [DT]
Good======= [Good JJ]
Headphone===== [headset NN]
I s========= [Is VBZ] .
==========[. UH

The SentiWordNet database will next be consulted to determine the adjective's priority and the sentiment value of the sentence. The SentiWordNet lexical database provides the score for each adverb and adjective. Some of them from Table 4 are included here; they will be used in the aforementioned example. Where neutral stands for a word's orientation that is neither positive nor negative. Additionally, negation stands for a multiplier factor with a value of -1.

In the aforementioned illustration, sentence s1's sentiment score for the term good is +0.76. The sentence "text1" has an emotion score of +0.76 since there are no adverbs or negations in it. Due to the presence of the adverb (very) and the adjective (good) in the second sentence (text2), the sentiment score of text2 is 1.27, which is the result of adding the good (0.77) and very (0.5) values. Table 5 displays the sentiment score for each sentence.

Table 4. Score List of Words

Word	Word Orientation	Score
Good	positive	.76
Great	positive	.88
Awesome	positive	.88

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 9, September 2023

Word	Word Orientation	Score
Excellent	positive	1
Well	positive	.76
Average	positive	.38
Enough	neutral	.88
Bad	negative	.66
Very	neutral	.5
Not	negation	-1

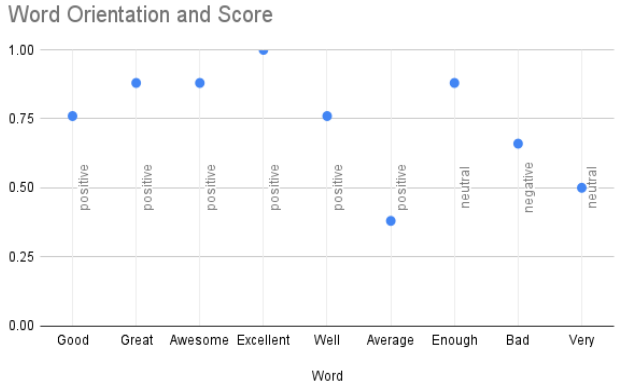
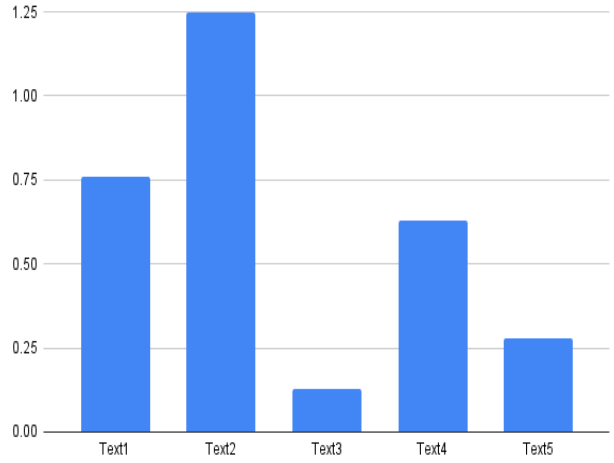


Table5. Sentence-level grade

Text Number	Score	Polarity Type
Text1	0.76	Positive
Text2	1.25	Positive
Text3	0.13	Positive
Text4	0.63	Negative
Text5	0.28	Positive



To determine the sentiment of a product review, we analyze the polarity of each feature mentioned across all reviews and calculate a feature-weighted average of the overall polarity.

III. RESULT

At the time this paper was being written, we had gathered 16000 reviews of three well-known headphones brands from flipkart.com and amazon.in. The findings demonstrate that when it comes to headphones, consumers are most concerned with qualities like sound quality, battery life, design, and value for money. Of these features, battery life, sound quality, and design are discovered to be more prevalent across all headphones. Table 6 presents the outcomes. No comment in either the positive or negative rows indicates that no one has expressed an opinion regarding that product feature.

IV. CONCLUSION

We created a method for sentiment analysis that can look at both Hindi and English text reviews to determine how customers feel about a given product. To repair incorrect terms and replace Hindi text with their English equivalents, we used a dictionary-based strategy. In order to repair incorrect words and Hindi terms, we also wish to augment this system with a machine learning technique. A weighted average of all the aspects of the product under evaluation makes up the final opinion score. In order to determine the final opinion score, we are also attempting to pinpoint a product's most salient characteristics.

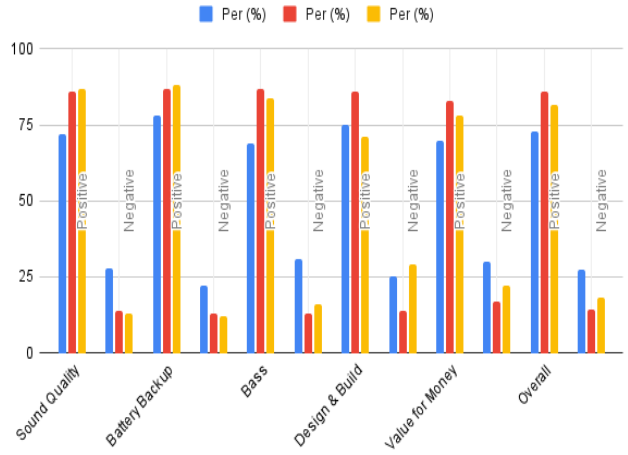
Table 6. A Brief Summary of the Product Review Boult Audio Gearpods, boAt Airdopes 161 and boAt Rockerz 550

Headphone Feature		Boult Audio Gearpods	boAt Airdopes 161	boAt Rockerz 560
Feature Name	Score Type	Per (%)	Per (%)	Per (%)
Sound Quality	Positive	72	86	87
	Negative	28	14	13
Battery Backup	Positive	78	87	88
	Negative	22	13	12
Bass	Positive	69	87	84
	Negative	31	13	16
Design & Build	Positive	75	86	71
	Negative	25	14	29
Value for Money	Positive	70	83	78
	Negative	30	17	22
Overall	Positive	72.8	86	81.6
	Negative	27.2	14.2	18.4

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 9, September 2023



REFERENCES

[1] Saini Jacob Soman1, P. Swaminathan, R. Anandan, K. Kalaivani: A comparative review of the challenges encountered in sentiment analysis of Indian regional language tweets vs English language tweets (IJET) 319-322(2018).

[2] G. Vinodini, R. M. Chandrasekaran" Sentiment Analysis, and Opinion Mining: A Survey" (IJARCSSE) 282-292(2012).

[3] Akhtar MS, Ekbal A, Bhattacharyya P (2016a) Aspect-based sentiment analysis: category detection and sentiment classification for Hindi. In: 17th International Conference on intelligent text processing and computational linguistics, pp 1–12.

[4] Akhtar MS, Ekbal, Bhattacharyya P (2016b) Aspect based sentiment analysis in Hindi: resource creation and evaluation. In: Proceedings of the 10th international conference on language resources and evaluation, pp 1–7.

[5] Arora P (2013) Sentiment analysis for Hindi language. MS by Research in Computer Science.

[6] Asghar MZ, Khan A, Zahra SR, Ahmad S, Kundi FM (2017) Aspect-based opinion mining framework using heuristic patterns. Clustering Computing https://doi.org/10.1007/s10586-017-1096-9.

[7] Asghar MZ, Khan A, Khan F, Kundi FM (2018a) Rift: a rule induction framework for Twitter sentiment analysis. Arab Journal of Science & Engg. 43(2):857–877.

[8] Asghar M.Z., Kundi F.M, Ahmad S, Khan A, Khan F(2018b) T-saf: Twitter sentiment analysis framework using a hybrid classification scheme. Expert System 35(1):1–19.

[9] Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Proc Lang Resource Evaluation 10:2200–2204.

[10] Bakliwal A, Arora P, Verma V (2012) Hindi subjective lexicon: a lexical resource for Hindi polarity classification. In: Proceedings of the eight international conference on language resources and evaluation, pp 1189–1196.

[11] Balamurali A, Joshi A, Bhattacharyya P (2012) Cross-lingual sentiment analysis for Indian languages using linked Wordnets. In: Proceedings of 24th international conference on computational linguistics: posters, pp 73–82.

[12] Bansal N, Ahmed UZ, Mukherjee A (2013) Sentiment analysis in Hindi. Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India, pp 1–10.

[13] Bhattacharyya P (2017) Indowordnet In: The WordNet in Indian languages. Springer, pp 1–18.

[14] Chand S (2016) Indian languages: classification of Indian languages. http://www.yourarticlelibrary.com/language/indian-languages-classification-of-indian-languages/19813/.

[15] Sharma R, Bhattacharyya P (2014) A sentiment analyzer for Hindi using Hindi SentiWordNet Lexicon. In: 11th International Conference on natural language processing, pp 1–6.

[16] Sharma R, Nigam S, Jain R (2014) Polarity detection movie reviews in Hindi language, pp1–9 : ArXivpreprint: 1409.3942.

[17] Sharma Y, Mangat V, Kaur M (2015) A practical approach to sentiment analysis of Hindi tweets. In: 1st International conference on next generation computing technologies. IEEE, pp 677–680.

[18] Das A, Bandopadhya S (2010a) Phrase-level polarity identification for Bangla. Int J Comput Linguist Appl 1(1–2):169–182.

[19] Das A, Bandyopadhyay S (2010c) Sentiword net for Indian languages. In: Asian federation for natural language processing, pp 56–63.

[20] Venugopalan M, Gupta D (2015) Sentiment classification for Hindi tweets in a constrained environment augmented using tweet specific features. In: International conference on mining intelligence and knowledge exploration. Springer, pp 664–670.

[21] Esuli A, Sebastiani F (2007) SentiWordNet: a high-coverage lexical resource for opinion mining. In: International conference on language resources and evaluation, pp 1–26.

[22] Govindan R, Haroon RP (2016) A survey on sentiment and emotion classification in Indo-Dravidian languages:1040–1042.

[23] Vandana Yadav, Parul Verma & Vinodini Katiyar Long short term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in Hindi languages International Journal of Information Technology volume 15, pages759–772 (2023).

[24] Vandana Yadav, Parul Verma & Vinodini Katiyar E-Commerce Product Reviews Using Aspect Based Hindi Sentiment Analysis **DOI:** 10.1109/ICCCI50826.2021.9402365,ISSN:2329-7190.