Journal of Computer Science

Research Article

KSWN (Kannada SentiWordNet): Developing a Sentiment Lexicon for Kannada using Translation and Word Embedding Techniques

[1]Rashmi Kariganuru Bheemarao, [1]Hassan Sadashiva Guruprasad and [2]Shambhavi Bangalore Ravi

[1]Department of Information Science and Engineering, B.M.S. College of Engineering, Bangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India
[2]Department of Computer Science and Engineering (DS), B.M.S. College of Engineering, Bangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India

Article history
Received: 13-12-2024
Revised: 11-04-2025
Accepted: 07-05-2025

Corresponding Author:
Rashmi Kariganuru Bheemarao
Department of Information Science and Engineering, B.M.S. College of Engineering, Bangalore, Visvesvaraya Technological University, Belagavi, Karnataka, India
Email: rashmikb.ise@bmsce.ac.in

Abstract: Opinion Mining has gained significant attention in recent years, especially due to the enormous growth of online content generation. However, finding the opinions expressed in comments and reviews is highly challenging in Indian regional languages due to the lack of annotated datasets. Opinion mining has predominantly been conducted in English, with recent efforts extending to Hindi and other languages. A primary resource in opinion mining is SentiWordNet, which aids in analyzing opinions by providing sentiment scores for words. Building a KSWN has been done to explore regional languages, as there is a notable absence of a comparable resource for Kannada. Thus, this study proposes creating a Kannada sentiment lexicon using a translation-based approach from various English sentiment lexicons. KSWN, a sentiment lexicon for Kannada developed using a translation approach, achieved an inter-annotator agreement, with Cohen's Kappa scores of 0.84 for positive words and 0.79 for negative words as verified by two Kannada annotators. The Kannada SentiWordNet, initially created, may not cover all sentiment-bearing words, word embeddings are employed to capture semantic similarity. As a seed lexicon can be the foundation for tagging a new corpus. Words in the new corpus are annotated by matching them with the seed list. New words with similar sentiment profiles are identified by applying similarity measures to the embedded word representations. These newly identified words are then added to the lexicon, further enriching it for sentiment analysis tasks.

Keywords: Natural Language Processing, SentiWordNet, Word Embeddings, Kannada Language, Sentiment Analysis, Lexicon Development

Introduction

In the age of digital transformation, the Internet is rapidly growing with a diverse range of multimodal content, including text, images, audio and video. Online communication, characterized by diverse language content, presents significant opportunities for gaining valuable insights through opinion mining, making it a critical task in the global context. Opinion mining is a technique used in marketing, customer service and social media monitoring to analyze large text volumes, identify patterns and categorize opinions for valuable insights.

Opinion Mining across English and other global languages has been the subject of much research, but regional languages are still understudied. Despite this, Kannada is frequently used on online forums and social media, especially because of the sizable immigrant populations worldwide. Kannada is a Dravidian language primarily spoken in the Indian state of Karnataka, with a rich literary tradition spanning over a thousand years. The volume of user-generated material in this language highlights the necessity for sentiment analysis tools specifically designed for Kannada. This lack of resources drives our creation of a Kannada sentiment lexical resource.

This resource supports real-world applications like social media sentiment tracking, customer feedback analysis and political discourse monitoring, while also aiding emotional cue detection for mental health insights

enhancing NLP for low-resource languages like Kannada.

Recent research has highlighted the challenges and opportunities in sentiment analysis for low-resource languages, which often lack sufficient annotated datasets and linguistic resources. Mohammed & Prasad (2023) presented a lexicon-based sentiment analysis approach for low-resource languages, using Hausa as a case study. It combines lexicon creation, data augmentation and model fine-tuning and is adaptable to other under-resourced languages. Aliyu *et al.* (2024) provided a comprehensive review of low-resource sentiment analysis approaches, emphasizing the importance of adaptation strategies and cross-lingual transfer to address data scarcity. The study highlights the growing use of transfer learning and transformer models in tackling multilingual sentiment tasks.

Developing sentiment lexicons for Kannada is challenging due to limited annotated data and the absence of existing resources. While translating from English lexicons offers a starting point, issues like loss of sentiment meaning, lack of direct equivalents and duplicates reduce accuracy and quality.

To address these challenges, this research makes the following contributions:

- Developing a KSWN: Creation of a sentiment lexicon specifically tailored for the Kannada language, enabling sentiment analysis for Kannada text and code-mixed Kannada-English text
- A novel dataset expansion approach: Introduction of an innovative method to expand the initial sentiment lexicon by leveraging corpus-driven techniques, including word embeddings and similarity measures, to identify and incorporate new sentiment words dynamically

Related Work

Opinion mining research initially focused on English but has expanded due to the growing prevalence of non-English content online. Two main approaches, Lexicon and Machine Learning, have been widely studied. While Machine Learning requires extensive annotated data, languages with limited resources often lack such corpora, making the lexicon approach a practical starting point. Methods for building sentiment lexicons include dictionary-based, WordNet-based and corpus-based approaches, which, despite providing out-of-context polarity scores, have proven effective as a reliable baseline.

Various opinion lexicons have been created for English. SentiWordNet (Sebastiani & Esuli, 2006) includes over 3 million words having positive, negative and objective scores. The Subjectivity Lexicon (Wilson *et al.*, 2005), part of Opinion Finder, contains words with POS tags, polarity and subjectivity categorized as strongly or weakly subjective. The Opinion Lexicon (Liu *et al.*, 2005) is derived from adjectives in annotated Twitter opinion sentences (Hu & Liu, 2004), while AFINN-111 (Nielsen, 2011) comprises manually valence-rated words. The Vader (Hutto & Gilbert, 2014) assigns sentiment scores.

Husnain *et al.* (2021) reviewed the contribution of SentiWordNet (SWN) in Opinion Mining (OM), categorizing its applications across lexical, sentence, document, thematic and conceptual levels. They noted that lexicon-based approaches are popular due to their independence from training data and versatility across domains, but highlighted two limitations: Finite word lists unsuitable for dynamic environments and fixed sentiment scores that ignore context. Despite these, SWN remains a reliable tool for detecting sentiment in blogs, reviews and social media, offering dynamic contextual representation through preassigned scores and polarities.

A comparative overview of various SentiWordNets is provided in Table (1). While prior studies have primarily focused on languages such as Hindi, Bengali and Tamil, the present work places a specific emphasis on the Kannada language. Moreover, unlike earlier approaches that predominantly rely on synset mapping techniques, this study employs a translation-based based by incorporating word embedding techniques, thereby enabling a more nuanced and data-driven framework for sentiment lexicon development in Kannada.

Ramanathan *et al.* (2019) introduced a new algorithm that incorporated informal words from Tamil tweets into Tamil SentiWordNet (TSWN), which previously had a limited number of adverbs and adjectives. TSWN was created using Tamil WordNet and English SentiWordNet. Google Translate was used in their approach. SentiPhraseNet for Telugu was built using a rule-based approach and validated with the ACTSA annotated corpus (Sunkapaka & Nageshwar, 2023). SentiPhraseNet addresses SentiWordNet's contextual limitations using sentiment phrases, dynamically learns unknown phrases and achieves 90.9% accuracy in Telugu sentiment analysis (Bharti *et al.*, 2021).

In the development of lexical tools, languages lacking resources typically rely on existing resources from other languages. This study utilized various English Sentiment lexicons for building KSWN.

The Kannada SentiWordNet developed in this study has broad practical relevance, enabling sentiment analysis in e-commerce, social media (including code-mixed content) and government policy feedback demonstrating its value beyond academic use. Kale *et al.* (2023) reviewed sentiment analysis in Indian regional languages, stressing the impact of dataset quality and preprocessing, especially for code-mixed data. They found that gold-standard datasets, ensemble methods and advanced feature extraction significantly improve model accuracy.

Girija *et al.* (2023) highlighted challenges in sentiment analysis for low-resource languages, including limited data, code-mixing and lack of NLP tools. They

suggested multilingual models, data augmentation and cross-lingual embeddings as effective solutions, emphasizing the role of quality data and collaboration in improving performance.

Table 1: SentiWordNets of different languages

Authors	Language	Approach Adopted	Dataset Used	Remarks
Mohanty *et al*. (2017)	Odia	The procedure creates a source lexicon, transfers polarities via synset IDs and evaluates it through manual annotation with agreement scoring	Bengali, Tamil, Telugu and Odia WordNets were utilized. Bengali, Tamil and Telugu SentiWordNets were used	Effectively leverages existing SentiWordNets and WordNets for building sentiment lexicons in low-resource languages through synset alignment
Asghar *et al*. (2019)	Urdu Lexicon	word-level translation scheme	A set of English opinion words, SentiWordNet, an English–Urdu dictionary and Urdu modifiers	Demonstrates a practical approach to lexicon development for Urdu using word-level translation and modifier handling for improved sentiment coverage
Kannan *et al*. (2016)	Tamil	Translation approach	The creation of Tamil SentiWordNet involved using the English SentiWordNet 3.0, Subjectivity Lexicon, AFINN-111 and Opinion Lexicon	Utilizes multiple English sentiment resources for comprehensive lexicon translation, contributing to Tamil sentiment analysis in a resource-constrained setting
Das & Bandyopadhyay (2010)	Bengali, Hindi and Telugu	An online, interactive game was developed to create and validate the SentiWordNet(s) by engaging the internet population	SentiWordNet and the Subjectivity Word List for English	Introduces an innovative crowdsourcing approach through an interactive game for constructing and validating SentiWordNets in Indian languages
Shelke *et al*. (2023)	Marathi	Linguists labeled Marathi news and sentiment analysis was done using machine learning models	Marathi newspapers and channel websites	Combines expert linguistic annotation with machine learning to build a sentiment analysis system grounded in real-world Marathi news data
Ranjitha & Bhanu (2021)	Kannada	The sentiment analysis was enhanced using an optimized data dictionary and the Decision Tree algorithm	Randomly chosen online reviews	Applies machine learning with an optimized sentiment dictionary to improve classification performance on Kannada online reviews
Bakay *et al*. (2019)	Turkish	Enlarge SentiTurkNet in terms of synset number by using a different Turkish WordNet	WordNets in Turkish, such as KeNet and TR-WordNet from BalkaNet	Effectively expands Turkish SentiWordNet by integrating multiple Turkish WordNets, enhancing synset coverage and resource richness
Gohil & Patel (2019)	Gujarati	Polarity scores from Hindi SeniWordNet are mapped to IndoWordNet(IWN) synsets and synonym relationships within IWN are used to generate G-SWN	Hindi SentiWordNet and IndoWordNet	Leverages cross-lingual polarity mapping and IndoWordNet synset relationships to systematically construct a sentiment lexicon for Gujarati
Garg & Lobiyal (2020)	Hindi	Assign emotional affinity to words in IndoWordNet	IndoWordNet	Enhances Hindi sentiment resources by assigning emotional affinity to words in IndoWordNet, contributing to emotion-aware sentiment analysis
Joshi *et al*.(2010)	Hindi	The H-SWN algorithm maps sentiment scores from SentiWordNet synsets to corresponding synsets in Hindi WordNet	English SentWordNet, Hindi WordNet	Pioneers' sentiment score mapping from English SentiWordNet to Hindi WordNet, laying the groundwork for sentiment lexicon development in Hindi
Bakliwal *et al*. (2012)	Hindi	A graph-based WordNet expansion method utilizes synonym and antonym relations	English SentiWordNet and Google Translate	Employs a graph-based approach leveraging synonym-antonym relations for expanding Hindi SentiWordNet, enhancing lexical coverage with translated resources
Das & Bandyopadhyay (2010)	Bangla	Prepared using an English-Bengali bilingual dictionary and SentiWordNet	SentiWordNet and Subjectivity Word List and English-Bengali bilingual dictionary	Constructs a Bangla sentiment lexicon using bilingual resources, effectively bridging linguistic gaps with English-based sentiment tools

Yashaswini and Padma 2015 demonstrated the use of sentiment analysis on Kannada product reviews, highlighting its potential to enhance customer insights and product evaluations in the e-commerce sector. Hande *et al.* (2020) focused on analyzing sentiment and offensive content in Kannada-English code-mixed social media comments, illustrating the importance of sentiment analysis in monitoring public opinion and user behavior on social platforms. Ijeri and Patil (2024) demonstrated the use of sentiment analysis in Kannada to assess public opinions on social media. This underscores the lexicon's utility in gauging sentiments on social issues within the Kannada-speaking community. Shetty *et al.* (2022) explored the sentiment analysis of Twitter posts in Kannada, among other languages, focusing on e-commerce platforms. Their work illustrated how sentiment analysis can be applied to understand customer feedback, enhancing service delivery and customer satisfaction in the e-commerce sector.

Eshwarappa and Shivasubramanyan (2024) proposed a K-BERT-PC classifier combining a modified BERT model with probability clustering, achieving 91% accuracy on a translated SemEval-based Kannada dataset. Their work demonstrates the effectiveness of deep learning and clustering for Kannada sentiment analysis with limited labeled data. Dhiman and Toshniwal (2020) proposed an enhanced text classification framework to analyze public engagement with health-related government policies on Twitter. Though not Kannada-specific, their methods can be adapted for Kannada social media analysis to gauge public sentiment toward governmental initiatives.

Data Acquisition

The acquisition of the source lexicon involved applying various filtering techniques to prominent English sentiment analysis resources, including the SentiWordNet 3.0 (Baccianella *et al.*, 2010), Opinion Lexicon (Liu *et al.*, 2005), AFINN 111 (Nielsen, 2011) and VADER (Hutto & Gilbert, 2014).

The proposed Kannada SentiWordNet (K-SWN) is derived from English SentiWordNet, where each synset has positive, negative and objective scores summing to 1. SentiWordNet is an enhanced lexical resource for opinion mining and sentiment classification, based on WordNet with added subjective data. The latest version, SentiWordNet 3.0, includes about 2 million entries, each annotated with positive, negative and objective scores. Synsets are identified by their Princeton WordNet IDs and tagged with part-of-speech information. AFINN-111 is a sentiment lexicon of 2,477 English words and phrases rated from -5 to +5, designed for social media analysis and focused solely on valence for simplified labeling. The Opinion Lexicon, introduced in 2004, contains 6,800 words (2,006 positive and 4,794 negative) and is widely used for social media analysis and refining ambiguous entries in sentiment lexicons. The Valence Aware Dictionary and Sentiment Reasoner (VADER) tool is a rule-based sentiment analysis model designed for short, informal texts like tweets and reviews. It assigns sentiment scores (positive, negative, neutral and compound) based on a lexicon and specific rules for modifiers and punctuation.

Materials and Methods

The study was carried out using Python in the Jupyter Notebook environment, with the Anaconda distribution providing a consistent setup. Key libraries included NLTK for WordNet integration, Pandas and NumPy for data processing, Matplotlib for visualization and Google Translate for translation. FastText word embeddings trained on Kannada were used to compute semantic similarity, leveraging subword information suitable for morphologically rich languages. Four English sentiment lexicons were used as source resources: SentiWordNet (SWN), the Opinion Lexicon, AFINN-111 and VADER. A bilingual English–Kannada dictionary and a cleaned Kannada word corpus were also utilized in the development and expansion of the Target Lexicon.

From SentiWordNet's 117,659 entries, words with positive or negative scores greater than 0.4 were selected, yielding 3,525 positive words and 2,044 negative words. Similarly, the Opinion Lexicon contributed 2,008 positive words and 4,647 negative words after removing duplicates and untranslated entries. The AFINN 111 dataset provided 879 positive words and 1,600 negative words. Additionally, 2248 positive words and 3118 negative words were finalized from the Vader dataset based on their mean values. This systematic approach ensured the creation of a comprehensive and balanced Kannada sentiment lexicon for sentiment analysis. A sentiment lexicon provides context-independent polarity scores for terms, making it a valuable resource for low-resource languages. A detailed selection of words is mentioned in Table (2).

The architecture of building the Kannada SentiWordNet is depicted in Figure (1). It illustrates a systematic procedure for constructing and expanding the KSWN. This translation-based methodology is segmented into three key stages: Firstly, gathering the Source Lexicon, which involves utilizing existing sentiment resources in the source language, exemplified here with four English sentiment resources, secondly, translating to the Target Lexicon, where the acquired source lexicon is translated into the language using Google translation tool and a dictionary with careful attention to avoid ambiguity and maintain contextual accuracy. Finally, the Target Lexicon is evaluated in this study through manual assessment by domain-specific linguistic annotators, with annotator agreement scores reported to gauge reliability. The annotations were carried out by two native Kannada speakers, both of whom are professors with academic backgrounds in Kannada language and literature. Their expertise ensured

accurate interpretation and labeling of nuanced linguistic expressions. Annotation was conducted using shared sheets, supported by predefined guidelines that were followed throughout the process to maintain consistency.

Table 2: Filtering the source sentiment lexicons

Source sentiment lexicons	Total words	After filtering		After the removal of duplicates and not translated words	
		Positive	Negative	Positive	Negative
SWN	117659	3525	2044	2683	1204
Opinion lexicon	6800	2008	4647	1760	4082
AFINN 111	2477	879	1600	798	1408
Vader	7520	2248	3118	1151	2117
		After merging and removing duplicates from all source lexicons		3831	4726

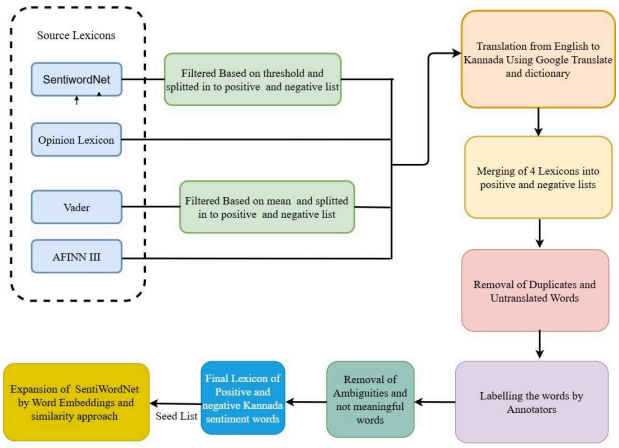

Fig. 1: Architecture of building the Kannada SentiWordNet

The KSWN extended through corpus-based methods to identify Kannada sentiment terms. This step includes using a developed target lexicon as a seed list to expand the SentiWordNet. Kannada word corpus required preprocessing, including the removal of English words. These unlabeled words were assigned the labels using Kannada word embeddings from fastText and cosine similarity techniques. The similarity is measured with a matching frequency of 0.6. After the expansion, 450 words are added to the positive list and 632 words are added to the negative list. Google Translate was used to convert the final set of words from English to Kannada, with Kannada annotators cross-checking for errors, multi-word entries and ambiguous translations. However, Multi-Word Expressions (MWEs) often resulted in inaccurate or contextually incorrect translations and such entries were excluded to maintain the lexicon's reliability and subjectivity.

Word embeddings, such as Fast Text, were used due to their ability to capture semantic similarity, even in morphologically rich languages like Kannada. Fast Text handles subword information, making it suitable for agglutinative languages. This data-driven approach helps expand the lexicon effectively beyond manual annotations. It offers better coverage and contextual relevance compared to rule-based or translation methods.

Results and Discussion

Cohen's Kappa is a statistical measure that evaluates the agreement between two annotators, accounting for chance agreement, making it more reliable than simple percent agreement.

The formula for Cohen's Kappa (κ) is:

$$\kappa = Po - Pe / 1 - Pe$$

where:

- *Po*: Observed agreement (the proportion of times the raters agree)
- *Pe*: Expected agreement (the proportion of agreement expected by chance)

Interpretation of Cohen's Kappa, the value of κ ranges from -1 to 1:

- 1: Perfect agreement
- 0: Agreement is no better than chance
- Negative values: Agreement is worse than chance

Table (3) presents the evaluation of the Kannada SentiWordNet, including the interpretation of Cohen's Kappa scores. The lexicon achieved a kappa score of 0.84 (95% CI: 0.823–0.857, $p<0.001$) for positive words and 0.79 (95% CI: 0.773–0.807, $p<0.001$) for negative words. These values indicate strong inter-annotator agreement and confirm that the sentiment assignments are statistically significant and consistent with human interpretation. Since the study is not measuring the degree of subjectivity, in this case tagging the corpus as either positive or negative comes from each annotator's predefined cognitive knowledge. Clearly stated guidelines are therefore necessary to ensure that the process remains unambiguous. In this case, the annotators are from a literature background. These annotators tagged the entire corpus independently. If the tags contradict each other about the exact ratings to be awarded, the token is removed.

Table 3: Evaluation details of Kannada SentiWordNet

Positive Labels		Negative Labels	
Total Words Considered	3831	Total Words Considered	4726
Neutral words	245	Neutral words	866
Negative words	887	Positive words	1042
Inter Annotator Agreement (Cohen's Kappa)	0.84 (0.823, 0.857)	Inter Annotator Agreement (Cohen's Kappa)	0.79 (0.773, 0.807)
95% Confidence Interval p-value	<0.001	95% Confidence Interval p-value	<0.001
Interpretation: Strong agreement			

Table (4) presents a subset of manually annotated Kannada words along with sentiment labels assigned by

two independent annotators. The table highlights the level of agreement between annotators, which was used as a criterion for inclusion in the final sentiment lexicon. Words with consistent annotations were marked as "Considered," while those with disagreement or deemed irrelevant were excluded. Additionally, English glosses are provided to aid comprehension for non-Kannada readers. This manual validation step was crucial in ensuring the accuracy, consistency and linguistic relevance of the annotated resource.

During the annotation process, several challenges were encountered that impacted the accuracy and consistency of the sentiment lexicon. Ambiguous sentiment profiles were a frequent issue; for instance, the word ಶರಣಾಗು ("surrender") could be seen as positive in devotional contexts but neutral or negative in others. Similarly, polysemous words like ಬಿಸಿ ("hot") varied in sentiment depending on usage, such as food (positive), weather (neutral/negative), or emotions (negative). A notable number of words were also duplicated across forms—inflections, spelling variants, or redundant entries—which required manual de-duplication to ensure dataset integrity. Translation posed another significant hurdle. Some words did not translate properly from English to Kannada or produced outputs that lacked grammatical or contextual relevance, leading to confusion or meaningless results. For example, certain crowd-sourced or machine-translated terms, even after conversion, failed to convey a valid sentiment in Kannada and had to be discarded. Additionally, Romanized Kannada and code-mixed expressions introduced spelling, phonetics and sentiment interpretation inconsistencies, complicating annotation further. To address these issues, words with unclear sentiment, translation errors, or annotator disagreement were excluded, ensuring only high-quality and contextually valid entries were retained in the final lexicon.

Table 4: Sample of annotated Kannada words with inter-annotator labels and final decision

Word	Annotator 1 Label	Annotator 2 Label	Agreement	Final Remark	English Gloss
ಅಸಹ್ಯ	Negative	Negative	Yes	Considered	Disgusting
ನಿರಾಶೆ	Negative	Negative	Yes	Considered	Disappointment
ಉತ್ಸಾಹ	Positive	Positive	Yes	Considered	Enthusiasm
ಉಂಟುಮಾಡುತ್ತವೆ	Neutral	Irrelevant	No	Deleted	They create/cause
ಶರಣಾಗು	Positive	Neutral	No	Deleted	Surrender

Conclusion

The KSWN developed in this study serves as a baseline for sentiment analysis and future enhancements. A translation approach and bilingual dictionaries supported the lexicon's construction. Additionally, a corpus-based approach was employed to capture language-specific words, using the KSWN as a seed list.

Currently, the lexicon is classified into two sentiment categories (positive and negative). Future work could expand this to a multi-point sentiment scale and incorporate subjectivity scores using annotated Kannada corpora. The lexicon's accuracy can be evaluated by applying it to Kannada social media data and comparing automated sentiment annotations with manual ones. Addressing challenges such as capturing the sentiment of multi-word expressions will be essential for refining and broadening the practical applications of the Kannada SentiWordNet. The expansion of a sentiment lexicon is directly influenced by the corpus being analyzed; a corpus with richer and more diverse sentiment expressions leads to greater opportunities for lexicon expansion.

This study addresses the research questions by demonstrating the feasibility of building a Kannada sentiment lexicon using a corpus-driven approach, starting from a curated seed list and expanding it using language-specific sentiment-bearing expressions. The effectiveness of Kannada SentiWordNet is evaluated through its application in sentiment classification tasks, particularly in handling code-mixed Kannada-English data. The lexicon has practical applicability in real-world contexts such as sentiment analysis of Kannada social media content, regional customer feedback systems and public opinion mining. Future work could involve integrating the lexicon into deep learning architectures, including transformer-based models, to support token-level sentiment annotation and enhance the handling of contextual nuances such as sarcasm, idiomatic phrases and code-mixed constructs.

Acknowledgment

We thank the domain-specific linguistic annotators for their invaluable contributions in manually assessing the Target Lexicon. Their expertise and meticulous evaluations were instrumental to this study. We also appreciate the editorial team's efforts in reviewing and refining our work.

Funding Information

This research received no specific grant from any funding agency.

Author’s Contributions

All the authors confirm whole responsibility for the following: Study conception and design, data collection, analysis and interpretation of results and manuscript preparation.

Ethics

The authors confirm that this article has not been published in any other journal. The corresponding author confirms that all the authors have read and approved the manuscript. Additionally, no ethical issues are involved in the manuscript or the dataset and no conflicts of interest are involved

References

Aliyu, Y., Sarlan, A., Usman Danyaro, K., Rahman, A. S. B. A., & Abdullahi, M. (2024). Sentiment Analysis in Low-Resource Settings: A Comprehensive Review of Approaches, Languages, and Data Sources. *IEEE Access*, *12*, 66883-66909. https://doi.org/10.1109/access.2024.3398635

Asghar, M. Z., Sattar, A., Khan, A., Ali, A., Masud Kundi, F., & Ahmad, S. (2019). Creating sentiment lexicon for sentiment analysis in Urdu: The case of a resource-poor language. *Expert Systems*, *36*(3), e12397. https://doi.org/10.1111/exsy.12397

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. *Lrec*, 2200-2204.

Bakay, O., Ergelen, O., & Yildiz, O. T. (2019). Integrating Turkish Wordnet KeNet to Princeton WordNet: The Case of One-to-Many Correspondences. *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1-5. https://doi.org/10.1109/asyu48272.2019.8946386

Bakliwal, A., Arora, P., & Varma, V. (2012). Hindi subjective lexicon: A lexical resource for hindi polarity classification. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC)*, 1189-1196.

Bharti, S. K., Naidu, R., & Babu, K. S. (2021). Dynamic SentiPhraseNet to Support Sentiment Analysis in Telugu. *Mathematical Modeling, Computational Intelligence Techniques and Renewable Energy*, 183-191. https://doi.org/10.1007/978-981-15-9953-8_16

Das, A., & Bandyopadhyay, S. (2010). SentiWordNet for Indian languages. *Proceedings of the 8th Workshop on Asian Language Resources*, 56-63.

Dhiman, A., & Toshniwal, D. (2020). An enhanced text classification to explore health based indian government policy tweets. *ArXiv:2007.06511v2*. https://doi.org/10.48550/arXiv.2007.06511

Eshwarappa, S. M., & Shivasubramanyan, V. (2024). Enhancing sentiment analysis in Kannada texts by feature selection. *International Journal of Electrical and Computer Engineering (IJECE)*, *14*(6), 6572. https://doi.org/10.11591/ijece.v14i6.pp6572-6582

Garg, K., & Lobiyal, D. K. (2020). Hindi EmotionNet. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *19*(4), 1-35. https://doi.org/10.1145/3383330

Girija, V. R., Sudha, T., & Cheriyan, R. (2023). Analysis of Sentiments in Low Resource Languages: Challenges and Solutions. *2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*, 1-6. https://doi.org/10.1109/rasse60029.2023.10363469

Gohil, L., & Patel, D. (2019). A Sentiment Analysis of Gujarati Text using Gujarati Senti word Net. *International Journal of Innovative Technology and Exploring Engineering*, *8*(9), 2290-2292. https://doi.org/10.35940/ijitee.i8443.078919

Hande, A., Priyadharshini, R., & Chakravarthi, B. R. (2020). KanCMD: Kannada CodeMixed dataset for sentiment analysis and offensive language detection. *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, 54-63.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177. https://doi.org/10.1145/1014052.1014073

Husnain, M., Missen, M. M. S., Akhtar, N., Coustaty, M., Mumtaz, S., & Prasath, V. B. S. (2021). A systematic study on the role of SentiWordNet in opinion mining. *Frontiers of Computer Science*, *15*(4), 154614. https://doi.org/10.1007/s11704-019-9094-0

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550

Joshi, A., Balamurali, A. R., & Bhattacharyya, P. (2010). A fall-back strategy for sentiment analysis in hindi: a case study. *Proceedings of ICON 2010: 8th International Conference on Natural Language Processing Macmillan Publishers*, 1-6.

Kale, S. D., Prasad, R., Potdar, G. P., Mahalle, P. N., Mane, D. T., & Upadhye, G. D. (2023). A Comprehensive Review of Sentiment Analysis on Indian Regional Languages: Techniques, Challenges, and Trends. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(9s), 93-110. https://doi.org/10.17762/ijritcc.v11i9s.7401

Kannan, A., Mohanty, G., & Mamidi, R. (2016). Towards building a SentiWordNet for Tamil. *Proceedings of the 13th International Conference on Natural Language Processing*, 30-35.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the Web. *Proceedings of the 14th International Conference on World Wide Web*, 342-351. https://doi.org/10.1145/1060745.1060797

Mohammed, I., & Prasad, R. (2023). Building lexicon-based sentiment analysis model for low-resource languages. *MethodsX*, *11*, 102460. https://doi.org/10.1016/j.mex.2023.102460

Mohanty, G., Kannan, A., & Mamidi, R. (2017). Building a SentiWordNet for Odia. *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 143-148. https://doi.org/10.18653/v1/w17-5219

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *ArXiv:1103.2903*. https://doi.org/10.48550/arXiv.1103.2903

Ramanathan, V., Meyyappan, T., & Thamarai, S. M. (2019). Predicting Tamil Movies Sentimental Reviews Using Tamil Tweets. *Journal of Computer Science*, *15*(11), 1638-1647. https://doi.org/10.3844/jcssp.2019.1638.1647

Ranjitha, P., & Bhanu, K. N. (2021). Improved Sentiment Analysis for Dravidian Language-Kannada Using Dicision Tree Algorithm with Efficient Data Dictionary. *IOP Conference Series: Materials Science and Engineering*, 012039. https://doi.org/10.1088/1757-899x/1123/1/012039

Sebastiani, F., & Esuli, A. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of the 5th International Conference on Language Resources and Evaluation*, 417-422.

Shelke, M. B., Alsubari, S. N., Panchal, D. S., & Deshmukh, S. N. (2023). Lexical Resource Creation and Evaluation: Sentiment Analysis in Marathi. *Smart Trends in Computing and Communications*, 187-195. https://doi.org/10.1007/978-981-16-9967-2_19

Shetty, S., Hegde, S., Shetty, S., Shetty, D., Sowmya, M. R., Shetty, R., Rao, S., & Shetty, Y. (2022). Sentiment Analysis of Twitter Posts in English, Kannada and Hindi languages. *Recent Advances in Artificial Intelligence and Data Engineering*, 361-375. https://doi.org/10.1007/978-981-16-3342-3_29

Sunkapaka, S., & Nageshwar, V. (2023). Sentiment Analysis Using Telugu SentiWordNet. *2023 2nd International Conference on Futuristic Technologies (INCOFT)*, 1-5. https://doi.org/10.1109/incoft60753.2023.10425658

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). OpinionFinder: a system for subjectivity analysis. *Proceedings of HLT/EMNLP on Interactive Demonstrations*, 34-35. https://doi.org/10.3115/1225733.1225751

Xiao, W. (2024). A Comparative Review of Advanced Techniques for Financial Sentiment Analysis. *Proceedings of the International Conference on Modeling, Natural Language Processing and Machine Learning*. CMNM 2024: International Conference on Modeling, Natural Language Processing and Machine Learning, Xi'an China. https://doi.org/10.1145/3677779.3677791