

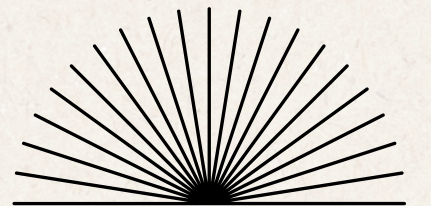
ASPECT-BASED SENTIMENT ANALYSIS FOR SINHALA ONLINE REVIEWS

PRESENTED BY:

Lakshani Galwatta - 258360N

SUPERVISED BY:

Dr. Nisansa de Silva



Agenda

02/38

01	Introduction
02	Research Problem
03	Research Objective
04	Literature Survey
05	Proposed Method

I AM HAPPY
GREAT SERVICES
EXCELLENT
FIVE STARS
HIGH QUALITY



INTRODUCTION

Background

- The rapid growth of digital ecosystems and e-commerce platforms has transformed how consumers share opinions and make purchasing decisions [1].
- Online marketplaces host millions of user-generated reviews, which have become a critical source of information for both consumers and businesses.
- To effectively analyze this large volume of textual data, organizations increasingly rely on automated Natural Language Processing techniques.
- These developments have also increased the importance of sentiment analysis in understanding customer feedback at scale [2, 3].

[1] N. de Silva, "Survey on publicly available Sinhala natural language processing tools and research," arXiv preprint arXiv:1906.02358, 2019. [Online]. Available: <https://arxiv.org/abs/1906.02358>

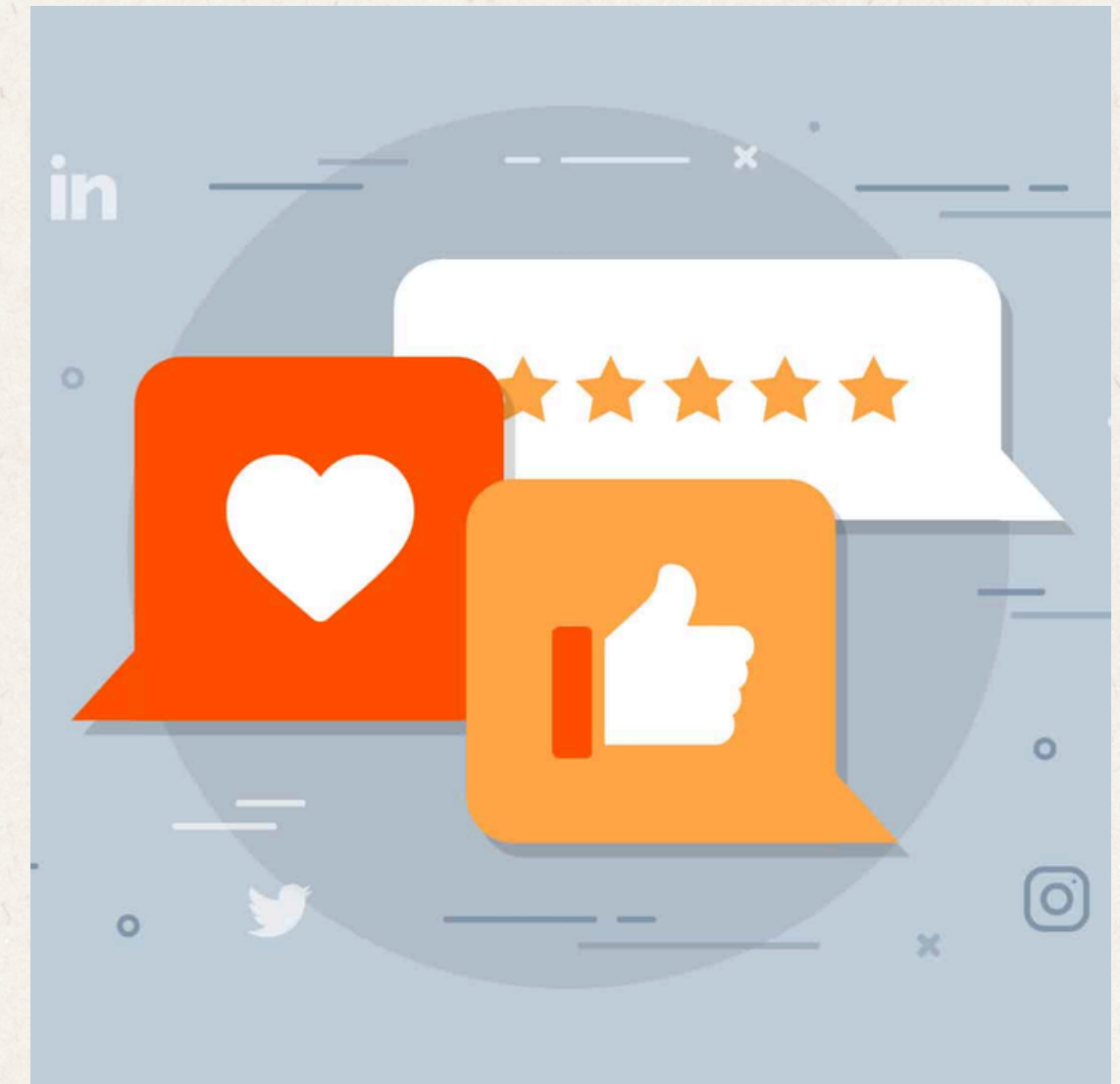
[2] K. P. 2025 Shyam Viththalani, "Aspect-based sentiment analysis advancements and applications in code-mixed text and gujarati language processing," in International Journal of Engineering Trends and Technology, vol. 73, 2025.

[3] K. Aziz, N. Ahmed, H. J. Hadi, A. Yusufu, M. A. Alshara, Y. Javed, and D. Ji, "Enhanced urduaspectnet: leveraging biaffine attention for superior aspect-based sentiment analysis," Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 9, p. 102221, 2024

Sentiment Analysis

05/38

- Sentiment Analysis is a subfield of Natural Language Processing that focuses on identifying subjective opinions in text [1].
- It typically classifies sentiment into **positive, negative, or neutral** categories [2, 3].
- Traditional approaches operate at the document or sentence level and assume that each text unit expresses a single sentiment.
- However, such methods fail to capture mixed or nuanced opinions commonly found in real-world reviews



[1] N. de Silva, "Survey on publicly available Sinhala natural language processing tools and research," arXiv preprint arXiv:1906.02358, 2019. [Online]. Available: <https://arxiv.org/abs/1906.02358>

[2] K. P. 2025 Shyam Viththalani, "Aspect-based sentiment analysis advancements and applications in code-mixed text and gujarati language processing," in International Journal of Engineering Trends and Technology, vol. 73, 2025.

[3] K. Aziz, N. Ahmed, H. J. Hadi, A. Yusufu, M. A. Alshara, Y. Javed, and D. Ji, "Enhanced urduaspectnet: leveraging biaffine attention for superior aspect-based sentiment analysis," Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 9, p. 102221, 2024

Aspect-Based Sentiment Analysis (ABSA)

06/38

- Aspect-Based Sentiment Analysis extends traditional sentiment analysis by identifying specific aspects of a product or service and the sentiment associated with each aspect.
- Instead of assigning a single sentiment to an entire review, ABSA provides fine-grained aspect-level sentiment information [5].
- It involves tasks such as aspect extraction, opinion extraction, and aspect-sentiment classification.
- This detailed analysis enables more interpretable and actionable insights for business decision-making [7].



[5] X. Ding, J. Zhou, L. Dou, Q. Chen, Y. Wu, A. Chen, and L. He, "Boosting Large Language Models with Continual Learning for Aspect-based Sentiment Analysis," in Findings of the Association for Computational Linguistics: EMNLP 2024, pp. 4367–4377, 2024.

[7] A. Chakraborty, "End-to-end aspect based sentiment analysis using graph attention network," in Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, pp. 3663–3668, 2024.

Research Gap in Sinhala ABSA

07/38

- Despite significant progress in Aspect-Based Sentiment Analysis for high-resource languages, research on Sinhala ABSA remains limited.
- There is a lack of large-scale, publicly available annotated datasets for aspect-level sentiment analysis in Sinhala.
- The low-resource nature of the Sinhala language, along with linguistic challenges such as free word order and code-switching, further complicates model development [8-11].
- As a result, existing sentiment analysis approaches are insufficient for capturing fine-grained opinions in Sinhala online reviews.



[8] N. Sultana, R. Sultana, R. I. Rasel, and M. M. Hoque, "Aspect-based sentiment analysis of bangla comments on entertainment domain," in 2022 25th International Conference on Computer and Information Technology (ICCIT), pp. 953–958, IEEE, 2022.

[9] K. P. 2025 Shyam Viththalani, "Aspect-based sentiment analysis advancements and applications in code-mixed text and gujarati language processing," in International Journal of Engineering Trends and Technology, vol. 73, 2025.

[10] K. Aziz, N. Ahmed, H. J. Hadi, A. Yusufu, M. A. Alshara, Y. Javed, and D. Ji, "Enhanced urduaspectnet: leveraging biaffine attention for superior aspect-based sentiment analysis," Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 9, p. 102221, 2024.

[11] F. Ma, X. Hu, A. Liu, Y. Yang, S. Li, P. S. Yu, and L. Wen, "AMR-based network for aspect-based sentiment analysis," in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (A. Rogers, J. Boyd-Graber, and N. Okazaki, eds.), (Toronto, Canada), pp. 322–337, Association for Computational Linguistics, July 2023

RESEARCH PROBLEM

08/38



Research Problem

- Existing sentiment analysis studies in Sinhala mainly focus on document-level or sentence-level classification, which fails to capture sentiment toward individual aspects such as battery, camera, or delivery within a single review [1].
- Furthermore, there is a significant lack of publicly available Sinhala Aspect-Based Sentiment Analysis datasets and supporting linguistic resources, which hinders the development of fine-grained sentiment models for this low-resource language.
- This research addresses the challenge of performing effective aspect-based sentiment analysis on Sinhala online reviews under conditions of limited annotated data and linguistic resources.
- Addressing this problem is crucial for improving customer insight extraction, enabling better business intelligence for local e-commerce platforms, and advancing Sinhala Natural Language Processing research [2, 3].

[1] N. de Silva, "Survey on publicly available Sinhala natural language processing tools and research," arXiv preprint arXiv:1906.02358, 2019. [Online]. Available: <https://arxiv.org/abs/1906.02358>

[2] K. P. 2025 Shyam Viththalani, "Aspect-based sentiment analysis advancements and applications in code-mixed text and gujarati language processing," in International Journal of Engineering Trends and Technology, vol. 73, 2025.

[3] K. Aziz, N. Ahmed, H. J. Hadi, A. Yusufu, M. A. Alshara, Y. Javed, and D. Ji, "Enhanced urduaspectnet: leveraging biaffine attention for superior aspect-based sentiment analysis," Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 9, p. 102221, 2024

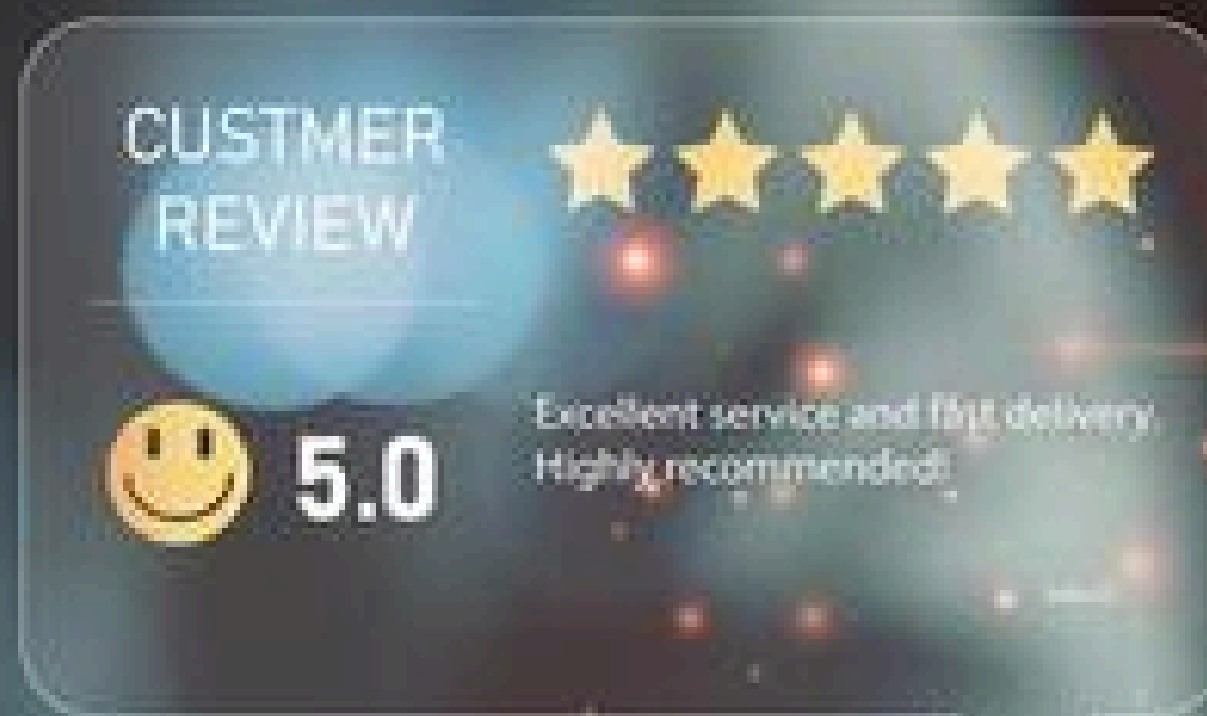
RESEARCH OBJECTIVES



Research Objectives

The primary objectives of this research are as follows:

1. Curate a Sinhala aspect-based sentiment analysis dataset by collecting and annotating online product reviews.
 2. Explore state-of-the-art approaches to improve aspect-level sentiment understanding by modeling syntactic and semantic relationships.
 3. Develop an ABSA model fine-tuned for Sinhala to perform aspect-level sentiment classification.
 4. Evaluate model performance using standard ABSA metrics.
-



LITERATURE SURVEY

Generative & Unified Frameworks for ABSA

- Traditional ABSA treats aspect extraction, opinion extraction, and sentiment classification as separate tasks, limiting their ability to capture semantic interdependencies between subtasks.
- Recent approaches reformulate ABSA as unified sequence generation, enabling end-to-end modeling of aspects, opinions, and polarities within a single framework [4–10].
- These generative frameworks primarily employ **paraphrase-based generation**, **pointer-based sequence generation**, or **structured output generation** to jointly model multiple sentiment elements.
- Paraphrase-based methods generate natural language representations of sentiment structures, while pointer-based approaches predict token-level spans with sentiment labels, both reducing error propagation common in pipeline architectures [4–7].

[4] W. Zhang, Y. Deng, X. Li, Y. Yuan, L. Bing, and W. Lam, “Aspect sentiment quad prediction as paraphrase generation,” arXiv preprint arXiv:2110.00796, 2021.

[5] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “Towards generative aspect-based sentiment analysis,” Association for Computational Linguistics, 2021.

[6] H. Yan, J. Dai, T. Ji, X. Qiu, and Z. Zhang, “A unified generative framework for aspect-based sentiment analysis,” in Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers), 2021, pp. 2416–2429.

[7] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang et al., “Sentiprompt: Sentiment knowledge enhanced prom

Generative & Unified Frameworks for ABSA (Cnt.)

- More structured generative approaches further extend this idea by producing schema-constrained outputs such as ordered quadruples or opinion trees, which capture both semantic and syntactic dependencies and improve interpretability [8, 9].
- Such structured representations also enable explicit modeling of hierarchical relationships between aspects and opinions, facilitating more accurate extraction of complex and implicit sentiment expressions [8, 9].
- Overall, unified generative ABSA frameworks achieve higher extraction accuracy, better handling of implicit sentiments, and improved adaptability to low-resource settings, representing a shift from task-specific classifiers to integrated sentiment modeling architectures [4–10].

[8] J. J. Peper and L. Wang, “Generative aspect-based sentiment analysis with contrastive learning and expressive structure,” arXiv preprint arXiv:2211.07743, 2022.

[9] X. Bao, Z. Wang, X. Jiang, R. Xiao, and S. Li, “Aspect-based sentiment analysis with opinion tree generation.” in IJCAI, vol. 2022, 2022, pp. 4044–4050.

[10] H. Wan, Y. Yang, J. Du, Y. Liu, K. Qi, and J. Z. Pan, “Target-aspect-sentiment joint detection for aspect-based sentiment analysis,” in Proceedings of the AAAI conference on artificial intelligence, vol. 34, no. 05, 2020, pp. 9122–9129.

Graph-Based Approaches & Dependency Parsing

- Graph-based models explicitly encode syntactic relationships through dependency trees, capturing long-range dependencies and complex structures that sequential models miss [11-13]
- Early studies demonstrated that integrating dependency graphs with neural architectures significantly improves aspect-sentiment linking performance [11].
- Subsequent advances combined contextual embeddings from BERT with graph convolutional networks to jointly model syntactic structure and contextual semantics, leading to more robust aspect-level representations [11, 12].
- Enhanced graph mechanisms introduced aspect-specific gating, word importance weighting, and syntax-guided regularization to focus on sentiment-relevant words [12].
- Syntax-aware Graph Neural Networks effectively align linguistic structure with contextual representations for robust ABSA [11-13]

[11] Z. Xiao, J. Wu, Q. Chen, and C. Deng, "Bert4gcn: Using bert intermediate layers to augment gcn for aspect-based sentiment classification," arXiv preprint arXiv:2110.00171, 2021. 28

[12] A. P. B. Veyseh, N. Nouri, F. Deroncourt, Q. H. Tran, D. Dou, and T. H. Nguyen, "Improving aspect-based sentiment analysis with gated graph convolutional networks and syntax-based regulation," in Findings of the association for computational linguistics: EMNLP 2020, 2020, pp. 4543–4548.

[13] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, "Dual graph convolutional networks for aspect-based sentiment analysis," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 6319–6329.

Graph-Based Approaches & Dependency Parsing (Cnt.)

- Dual-graph architectures jointly model syntactic and semantic relationships, outperforming single approaches in complex and informal text [13]
- Abstract Meaning Representation-based semantic graphs capture meaning-level relations less sensitive to surface syntax variations [14]
- Graph-based ABSA has also been adapted to specialized domains such as legal opinion analysis, with entity-centric dependencies and long structured sentences [15].
- Rule-based and hybrid graph systems exploit dependency parsing to produce interpretable and domain-aligned sentiment predictions, highlighting the continued relevance of linguistic structure [16].

[13] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, "Dual graph convolutional networks for aspect-based sentiment analysis," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 6319–6329.

[14] F. Ma, X. Hu, A. Liu, Y. Yang, S. Li, P. S. Yu, and L. Wen, "AMR-based network for aspect-based sentiment analysis," in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 322–337. [Online]. Available: <https://aclanthology.org/2023.acl-long.19/>

[15] I. Rajapaksha, C. R. Mudalige, D. Karunaratna, N. de Silva, A. S. Perera, and G. Rathnayaka, "Sigmalaw pbsa-a deep learning model for aspect-based sentiment analysis for the legal domain," in International Conference on Database and Expert Systems Applications. Springer, 2021, pp. 125–137.

[16] I. Rajapaksha, C. R. Mudalige, D. Karunaratna, N. de Silva, G. Rathnayaka, and A. S. Perera, "Rule-based approach for party-based sentiment analysis in legal opinion texts," in 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer). IEEE, 2020, pp. 284–285.

Instruction Tuning & Prompt-Based ABSA

- Instruction tuning and prompt-based learning enable a single model to perform multiple ABSA tasks by following natural language instructions instead of using task-specific architectures [7].
- These approaches unify aspect extraction, opinion extraction, sentiment classification, and triplet/quadruple generation, reducing complexity and improving zero-shot and few-shot performance.
- Prompt-based methods use either manual templates or learnable soft prompts, with learnable prompts providing more stable and effective sentiment representations [7].
- Instruction-tuned models trained on diverse task instructions can generalize across different ABSA subtasks and domains without extensive retraining [17].
- Frameworks such as InstructABSA and its hybrid extensions demonstrate strong performance, cross-domain adaptability, and suitability for low-resource languages like Sinhala [17, 18].

[7] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang et al., “Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis,” arXiv preprint arXiv:2109.08306, 2021.

[17]. Scaria, H. Gupta, S. Goyal, S. Sawant, S. Mishra, and C. Baral, “Instructabsa: Instruction learning for aspect based sentiment analysis,” in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), 2024, pp. 720–736.

[18] D. Jayakody, A. Malkith, K. Isuranda, V. Thenuwara, N. de Silva, S. R. Ponnampereuma, G. Sandamali, and K. Sudheera, “Instruct-deberta: A hybrid approach for aspect-based sentiment analysis on textual reviews,” arXiv preprint arXiv:2408.13202, 2024

ABSA for Low-Resource Languages - Overview

Key Challenges in Low-Resource ABSA

- Critical shortage of annotated corpora and limited availability of linguistic tools for processing [19, 20]
- Morphological complexity and richness that complicates tokenization and feature extraction [20, 25]
- Code-mixing between native scripts and English, along with romanized script variations [2, 25]
- Domain-specific vocabulary gaps that make general models less effective [24]
- Inconsistent orthography in user-generated content from social media and reviews [22, 2]

Common Solutions Across Languages:

- Creation of high-quality datasets through rigorous native speaker annotation processes [19, 23]
- Deployment of multilingual transformer models such as XLM-R, IndicBERT, and mBERT [20, 2]
- Hybrid modeling approaches combining sentiment lexicons with deep learning architectures [2]
- Data augmentation using techniques like T5-based paraphrasing and back-translation [20]

[7] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang et al., “Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis,” arXiv preprint arXiv:2109.08306, 2021.

[17]. Scaria, H. Gupta, S. Goyal, S. Sawant, S. Mishra, and C. Baral, “Instructabsa: Instruction learning for aspect based sentiment analysis,” in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), 2024, pp. 720–736.

[18] D. Jayakody, A. Malkith, K. Isuranda, V. Thenuwara, N. de Silva, S. R. Ponnampereuma, G. Sandamali, and K. Sudheera, “Instruct-deberta: A hybrid approach for aspect-based sentiment analysis on textual reviews,” arXiv preprint arXiv:2408.13202, 2024

Notable Recent Low-Resource ABSA Implementations

Language	Dataset Size	Key Features & Achievements
Telugu [19]	5,027 reviews	Movie domain with six categories, LSTM baselines established
Odia [20]	5,780 reviews, 9,400+ aspects	Seven product/service domains with T5 augmentation and XLM-R fine-tuning
Uzbek [23]	3,500+ reviews, 6,100+ sentences	Multi-task approach using KNN methodology
Vietnamese [24]	16,227 reviews	E-commerce beauty products using BiGRU+Conv1D
Urdu [25]	84,743 reviews	USA-BERT model with script normalization
Bangla[22]	4,012 comments	Entertainment domain (YouTube). Machine-learning classifiers like SVC, Random Forest or Linear Regression were used

[7] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang et al., “Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis,” arXiv preprint arXiv:2109.08306, 2021.

[17]. Scaria, H. Gupta, S. Goyal, S. Sawant, S. Mishra, and C. Baral, “Instructabsa: Instruction learning for aspect based sentiment analysis,” in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), 2024, pp. 720–736.

[18] D. Jayakody, A. Malkith, K. Isuranda, V. Thenuwara, N. de Silva, S. R. Ponnampereuma, G. Sandamali, and K. Sudheera, “Instruct-deberta: A hybrid approach for aspect-based sentiment analysis on textual reviews,” arXiv preprint arXiv:2408.13202, 2024

Implications for Sinhala ABSA

Similar Linguistic and Resource Challenges:

- Morphological complexity requiring specialized tokenization approaches (parallels with Telugu and Odia) [19, 20]
- Script variation where users alternate between native Sinhala and romanized Singlish (similar to Urdu and Gujarati) [2, 25]
- Extensive code-mixing with English in social media and informal contexts [2]
- Limited availability of high-quality annotated resources and NLP tools [20, 23]

Proven Methodologies Applicable to Sinhala

Preprocessing [2]:

- Language detection for code-switched segments
- Transliteration between native script and romanized forms
- Normalization for informal user-generated content

Modeling Approaches:

- Unsupervised: LISA model for settings without annotated datasets [21]
- Lightweight: KNN-based methods for establishing initial baselines [23]
- Advanced: Multilingual transformers like IndicBERT and XLM-R [20, 25]
- Hybrid: Lexicon-transformer systems combining dictionaries with BERT [2]

[7] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang et al., “Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis,” arXiv preprint arXiv:2109.08306, 2021.

[17]. Scaria, H. Gupta, S. Goyal, S. Sawant, S. Mishra, and C. Baral, “Instructabsa: Instruction learning for aspect based sentiment analysis,” in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), 2024, pp. 720–736.

[18] D. Jayakody, A. Malkith, K. Isuranda, V. Thenuwara, N. de Silva, S. R. Ponnampereuma, G. Sandamali, and K. Sudheera, “Instruct-deberta: A hybrid approach for aspect-based sentiment analysis on textual reviews,” arXiv preprint arXiv:2408.13202, 2024

Synthetic Data Generation for ABSA

The Core Challenge:

- Training ABSA models requires large annotated datasets, which low-resource languages like Sinhala lack [27]
- Traditional data collection is time-consuming, expensive, and requires linguistic expertise

LLM-Based Synthetic Generation [27]:

- Large Language Models generate synthetic training samples with minimal seed data (as few as 25 examples)
- Creates aspect-rich sentences across diverse sentiment classes and linguistic structures
- Improves both ATE and APC performance in low-resource settings
- Provides scalable, language-agnostic pipeline

Augmentation Strategies [20]:

- Ensemble augmentation: back-translation + T5-based paraphrasing
- Semantic similarity filtering using Universal Sentence Encoder for quality control
- Successfully enhanced XLM-R and IndicBERT performance for morphologically rich languages

[20] L. Dewangan, Z. A. Sayeed, and C. Maurya, “Benchmark creation for aspect based sentiment analysis in low-resource odia language and evaluation through fine-tuning of multilingual models,” in Proceedings of the 31st International Conference on Computational Linguistics, 2025, pp. 5863–5869.

[27] N. C. Hellwig, J. Fehle, and C. Wolff, “Exploring large language models for the generation of synthetic training samples for aspect-based sentiment analysis in low resource settings,” Expert Systems with Applications, vol. 261, p. 125514, 2025.

Multilingual and Sinhala / Code-Mixed Studies

Sinhala Banking Domain Study [28, 29]:

Dataset & Domain:

- 15,000 banking reviews: 5,000 Sinhala and code-mixed samples [29]
- Covers loans, online banking, transactions, and customer service aspects [28]

Technical Approach:

- Domain-specific vocabulary construction using embedding-based keyword identification [28]
- Hybrid modeling: XLM-RoBERTa combined with Sinhala sentiment lexicons [29]
- Preprocessing pipeline: transliteration and normalization for non-standard orthography [29]
- Explainability using SHAP and LIME for interpretability [29]

Results:

- Hybrid Sinhala model achieved >88% accuracy [29]
- Outperformed classical models and zero-shot GPT baselines [29]

Key Insight: Domain lexicon development is essential for handling Sinhala code-mixing [28]

[28] F. Rizvi, T. Navojith, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, “Keyword extraction, and aspect classification in sinhala, english, and code-mixed content,” arXiv preprint arXiv:2504.10679, 2025.

[29] A. Rizvi, N. Thamindu, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, “Enhancing multilingual sentiment analysis with explainability for sinhala, english, and code-mixed content,” arXiv preprint arXiv:2504.13545, 2025.

Deep Learning Foundations and Architectures

1. Context Denoising Model [30]

- **Problem:** Contextual noise and irrelevant information in user-generated text mislead sentiment detection
- **Solution:** Diffusion-based denoising mechanism with syntactically-informed noise injection
- **Approach:** Uses dependency parsing to target irrelevant words, applied on BERT and LLaMA
- **Results:** Superior performance on LAP14 and REST14-REST16 benchmarks
- **Sinhala Relevance:** Handles colloquial expressions, free word order, and irrelevant information

2. Contextual Syntax-Aware Extraction (CSAE) Model [31]

- **Problem:** Modern embeddings struggle with multi-word aspect detection (e.g., "quality of food")
- **Solution:** Integrates POS tags, dependency embeddings, and syntactic relative distance metrics
- **Mechanism:** Self-attention emphasizes syntactically significant context for boundary detection
- **Results:** Significant improvements on SemEval-2014 restaurant and laptop datasets
- **Sinhala Relevance:** Critical for handling flexible word order, compound expressions, and rich morphology

[30] Y. Tian, C. Liu, Y. Song, F. Xia, and Y. Zhang, "Aspect-based sentiment analysis with context denoising," in Findings of the Association for Computational Linguistics: NAACL 2024, 2024, pp. 3083–3095.

[31] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in Proceedings of the 58th annual meeting of the association for computational linguistics, 2020, pp. 3211–3220.

Deep Learning Foundations and Architectures (Cnt.)

3. Aspect-to-Scope Multi-View Contrastive Learning (A2SMvCL) [32]

- **Problem:** Multi-aspect interference where sentiment words for one aspect incorrectly influence another
- **Innovation:** Defines "aspect-specific scope" as precise sentiment-relevant regions per aspect
- **Architecture:** Dual graph views (semantic attention-based + syntactic dependency-based)
- **Mechanism:** Intra-view and inter-view contrastive learning with self-gating fusion
- **Results:** Consistently outperforms existing methods across six benchmark datasets
- **Sinhala Relevance:** Handles sentiment encoded implicitly across syntactically dispersed elements

[32] H. Chai, Z. Yao, S. Tang, Y. Wang, L. Nie, B. Fang, and Q. Liao, "Aspect-to-scope oriented multi-view contrastive learning for aspect-based sentiment analysis," in Findings of the Association for Computational Linguistics: EMNLP 2023, 2023, pp. 10 902–10 913.

Deep Learning Foundations and Architectures (Cnt.)

4. Contrastive Learning Framework [33]

- **Supervised Variant:** Uses sentiment labels to bring same-polarity samples closer in embedding space
- **Unsupervised Variant:** Generates positive samples via back-translation and synonym replacement
- **Results:** More compact sentiment clusters and improved polarity separation on REST14, LAP14, Twitter
- **Sinhala Relevance:** Improves robustness in noisy, small-scale datasets with domain augmentation

5. Frequency-Based Domain Lexicon Framework [34]

- Lightweight, interpretable approach using contextual frequency and opinion-aspect pairs
- Language-agnostic foundation complementing deep learning architectures

[33] L. Xu and W. Wang, "Improving aspect-based sentiment analysis with contrastive learning," Natural Language Processing Journal, vol. 3, p. 100009, 2023.

[34] T. M. Zayet, M. A. Ismail, and K. D. Varathan, "Aspect extraction in domain Lexicon generation: A new frequency-based approach," IEEE Access, 2024.

Hybrid and Multi-Component Architectures

Why Hybrid Architectures?

- Overcome limitations of purely supervised or transformer-only approaches [29]
- Integrate linguistic knowledge, statistical reasoning, and model interpretability [35]
- Particularly effective for low-resource languages with limited annotated data [21, 29]

Key Hybrid Architecture Examples:

1. Explainable Sinhala-English Banking System [29]

- **Components:** XLM-R + BERT-base + Sinhala financial lexicons + SHAP/LIME visualization
- **Capabilities:** Handles code-mixed text, corrects Sinhala-specific linguistic errors, provides interpretable outputs
- **Result:** Demonstrates how symbolic resources combined with transformers improve under-resourced contexts

[21] M. Shams, N. Khoshavi, and A. Baraani-Dastjerdi, “Lisa: language-independent method for aspect-based sentiment analysis,” IEEE Access, vol. 8, pp. 31 034–31 044, 2020.

[28] F. Rizvi, T. Navojith, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, “Keyword extraction, and aspect classification in sinhala, english, and code-mixed content,” arXiv preprint arXiv:2504.10679, 2025.

[29] A. Rizvi, N. Thamindu, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, “Enhancing multilingual sentiment analysis with explainability for sinhala, english, and code-mixed content,” arXiv preprint arXiv:2504.13545, 2025.

[35] A. Maroof, S. Wasi, S. I. Jami, and M. S. Siddiqui, “Aspect based sentiment analysis for service industry,” IEEE Access, 2024.

Hybrid and Multi-Component Architectures (Cnt.)

2. Sinhala Keyword Extraction & Aspect Classification [28]

- **Pipeline:** Embedding-based keyword extraction → vocabulary expansion → transliteration → aspect mapping
- **Key Insight:** Domain-specific preprocessing and component integration enhance accuracy

3. LISA Unsupervised Model [21]

- **Components:** Lexicon translation + co-occurrence statistics + EM-based topic modeling
- **Advantage:** Minimal supervised data requirements with language flexibility

4. Service Industry Two-Stage Pipeline [35]

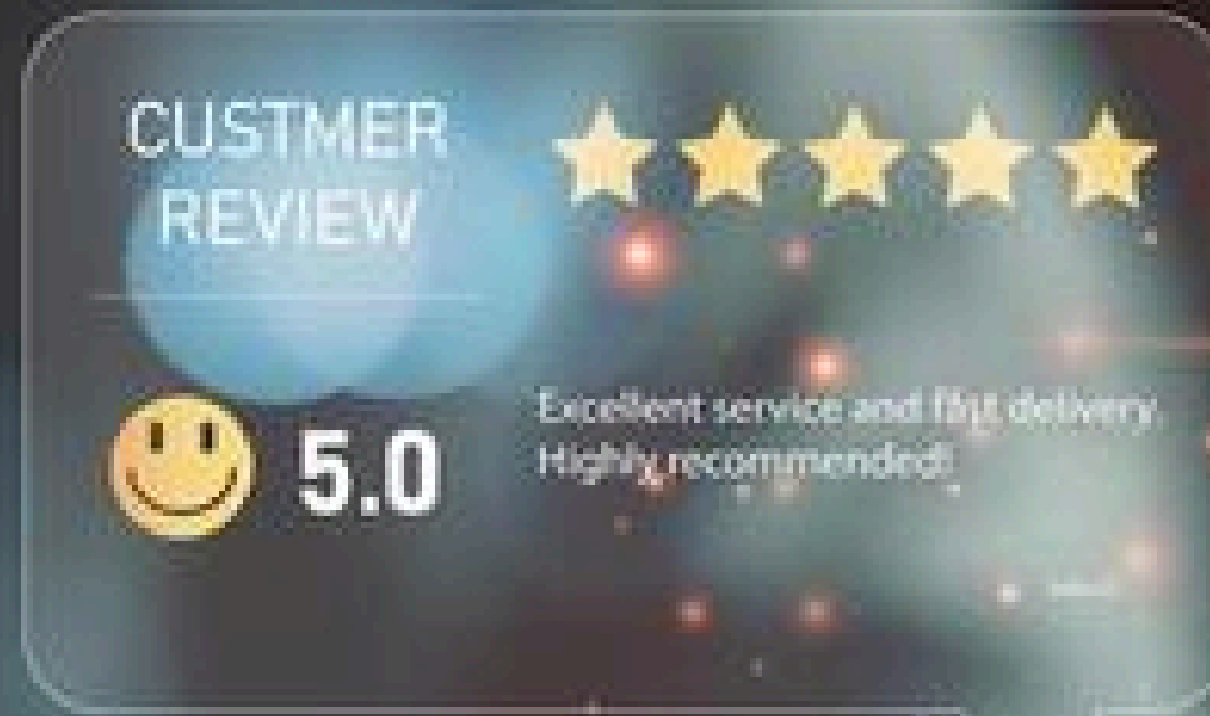
- **Approach:** Linguistic rules (POS, dependency parsing, lexicons) + ML/DL models (SVM, LSTM, BERT)
- **Strength:** Captures implicit aspects, multi-word expressions, and contextual sentiment shifts

[21] M. Shams, N. Khoshavi, and A. Baraani-Dastjerdi, "Lisa: language-independent method for aspect-based sentiment analysis," IEEE Access, vol. 8, pp. 31 034–31 044, 2020.

[28] F. Rizvi, T. Navojith, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, "Keyword extraction, and aspect classification in sinhala, english, and code-mixed content," arXiv preprint arXiv:2504.10679, 2025.

[29] A. Rizvi, N. Thamindu, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, "Enhancing multilingual sentiment analysis with explainability for sinhala, english, and code-mixed content," arXiv preprint arXiv:2504.13545, 2025.

[35] A. Maroof, S. Wasi, S. I. Jami, and M. S. Siddiqui, "Aspect based sentiment analysis for service industry," IEEE Access, 2024.



PROPOSED METHOD

Dataset Creation and Augmentation

Data Collection:

- Sources: Daraz.lk and Sinhala social media platforms
- Diverse product categories: electronics, household, fashion, personal care
- Ethical web scraping of publicly accessible content only

Data Preprocessing Pipeline:

- Remove HTML tags, URLs, duplicates, and full English reviews
- Translate Romanized Sinhala (Singlish) to proper Sinhala script
- Normalize Unicode characters and correct spelling variations
- Handle code-mixed text: retain sentiment-bearing English words
- Sinhala-compatible tokenization and sentence segmentation

Data Augmentation [22, 33]:

- Back-translation via English/Tamil
- Synonym replacement using embeddings
- Transformer-based paraphrasing

[22] N. Sultana, R. Sultana, R. I. Rasel, and M. M. Hoque, "Aspect-based sentiment analysis of bangla comments on entertainment domain," in 2022 25th International Conference on Computer and Information Technology (ICCIT). IEEE, 2022, pp. 953–958

[33] L. Xu and W. Wang, "Improving aspect-based sentiment analysis with contrastive learning," Natural Language Processing Journal, vol. 3, p. 100009, 2023.

Dataset Creation and Augmentation

Annotation Strategy:

- Aspect Terms: Explicit features (price, quality, delivery)
- Aspect Categories: Higher-level groupings (Product Quality, Service)
- Sentiment Polarity: Positive, negative, or neutral per aspect
- Multiple native speakers, Cohen's Kappa for reliability



Negative



Neutral



Positive

Model Development and Evaluation

Multi-Task Problem Formulation:

- **Task 1:** Aspect Extraction
- **Task 2:** Sentiment Classification (polarity per aspect)

Modeling Approaches:

1. **Transformer-Based Models** : mBERT, XLM-RoBERTa, IndicBERT fine-tuned on Sinhala dataset
2. **Joint/Multi-Task Learning** : Shared encoder with task-specific output layers
3. **Graph-Based Approaches** : GCNs/GATs on dependency-parsed syntactic graphs

Evaluation:

- **Metrics:** Accuracy, Precision, Recall, F1-score
 - **Comparisons:** Models, architectures, with/without augmentation
 - **Error Analysis:** Implicit aspects, sarcasm, negation, code-mixing
-

References

- [1] J. Wang, B. Xu, and Y. Zu, “Deep learning for aspect-based sentiment analysis,” in 2021 International conference on machine learning and intelligent systems engineering (MLISE). IEEE, 2021, pp. 267–271.
- [2] Y. Tian, C. Liu, Y. Song, F. Xia, and Y. Zhang, “Aspect-based sentiment analysis with context denoising,” in Findings of the Association for Computational Linguistics: NAACL 2024, 2024, pp. 3083–3095.
- [3] M. H. Phan and P. O. Ogunbona, “Modelling context and syntactical features for aspect-based sentiment analysis,” in Proceedings of the 58th annual meeting of the association for computational linguistics, 2020, pp. 3211–3220.
- [4] N. de Silva, “Survey on Publicly Available Sinhala Natural Language Processing Tools and Research,” arXiv preprint arXiv:1906.02358v26, 2026.
- [5] X. Ding, J. Zhou, L. Dou, Q. Chen, Y. Wu, A. Chen, and L. He, “Boosting Large Language Models with Continual Learning for Aspect-based Sentiment Analysis,” in Findings of the Association for Computational Linguistics: EMNLP 2024, 2024, pp. 4367–4377.
- [6] Y. Tian, G. Chen, and Y. Song, “Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble,” in Proceedings of the 2021 conference of the North American chapter of the association for computational Linguistics: human language technologies, 2021, pp. 2910–2922.
- [7] A. Chakraborty, “End-to-end aspect based sentiment analysis using graph attention network,” in Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, 2024, pp. 3663–3668.
- [8] Z. Xiao, J. Wu, Q. Chen, and C. Deng, “Bert4gcn: Using bert intermediate layers to augment gcn for aspect-based sentiment classification,” arXiv preprint arXiv:2110.00171, 2021.
- [9] L. Dewangan, Z. A. Sayeed, and C. Maurya, “Benchmark creation for aspect-based sentiment analysis in low-resource odia language and evaluation through fine-tuning of multilingual models,” in Proceedings of the 31st International Conference on Computational Linguistics, 2025, pp. 5863–5869.
- [10] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “Towards generative aspect-based sentiment analysis.” Association for Computational Linguistics, 2021.

References

- [11] H. Yan, J. Dai, T. Ji, X. Qiu, and Z. Zhang, “A unified generative framework for aspect-based sentiment analysis,” in Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers), 2021, pp.2416–2429.
- [12] J. B. Disanayake, Say it in Sinhala. Lake House Investments Limited, 1985.
- [13] K. P. 2025 Shyam Viththalani, “Aspect-based sentiment analysis advancements and applications in code-mixed text and gujarati language processing,” in International Journal of Engineering Trends and Technology, vol. 73, 2025.
- [14] K. Aziz, N. Ahmed, H. J. Hadi, A. Yusufu, M. A. Alshara, Y. Javed, and D. Ji, “Enhanced urduaspectnet: leveraging biaffine attention for superior aspect-based sentiment analysis,” Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 9, p. 102221, 2024.
- [15] W. Zhang, Y. Deng, X. Li, Y. Yuan, L. Bing, and W. Lam, “Aspect sentiment quad prediction as paraphrase generation,” arXiv preprint arXiv:2110.00796, 2021.
- [16] C. Li, F. Gao, J. Bu, L. Xu, X. Chen, Y. Gu, Z. Shao, Q. Zheng, N. Zhang, Y. Wang et al., “Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis,” arXiv preprint arXiv:2109.08306, 2021.
- [17] J. J. Peper and L. Wang, “Generative aspect-based sentiment analysis with contrastive learning and expressive structure,” arXiv preprint arXiv:2211.07743, 2022.
- [18] X. Bao, Z. Wang, X. Jiang, R. Xiao, and S. Li, “Aspect-based sentiment analysis with opinion tree generation.” in IJCAI, vol. 2022, 2022, pp. 4044–4050.
- [19] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, “Semeval-2015 task 12: Aspect based sentiment analysis,” in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). Denver, Colorado: Association for Computational Linguistics, 2015, pp. 486–495. [Online]. Available: <https://aclanthology.org/S15-2082/>
- [20] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq et al., “Semeval-2016 task 5: Aspect based sentiment analysis,” in Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), 2016, pp. 19–30.

References

- [21] H. Wan, Y. Yang, J. Du, Y. Liu, K. Qi, and J. Z. Pan, “Target-aspect-sentiment joint detection for aspect-based sentiment analysis,” in Proceedings of the AAAI conference on artificial intelligence, vol. 34, no. 05, 2020, pp. 9122–9129.
- [22] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, “Semeval-2014 task 4: Aspect based sentiment analysis,” in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: Association for Computational Linguistics, 2014, pp. 27–35.
- [23] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, “A challenge dataset and effective models for aspect-based sentiment analysis,” in Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), 2019, pp. 6280–6285.
- [24] A. P. B. Veyseh, N. Nouri, F. Dernoncourt, Q. H. Tran, D. Dou, and T. H. Nguyen, “Improving aspect-based sentiment analysis with gated graph convolutional networks and syntax-based regulation,” in Findings of the association for computational linguistics: EMNLP 2020, 2020, pp. 4543–4548.
- [25] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, “Dual graph convolutional networks for aspect-based sentiment analysis,” in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 6319–6329.
- [26] F. Ma, X. Hu, A. Liu, Y. Yang, S. Li, P. S. Yu, and L. Wen, “AMR-based network for aspect-based sentiment analysis,” in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 322–337. [Online]. Available: <https://aclanthology.org/2023.acl-long.19/>
- [27] I. Rajapaksha, C. R. Mudalige, D. Karunarathna, N. de Silva, A. S. Perera, and G. Ratnayaka, “Sigmalaw pbsa-a deep learning model for aspect-based sentiment analysis for the legal domain,” in International Conference on Database and Expert Systems Applications. Springer, 2021, pp. 125–137.
- [28] I. Rajapaksha, C. R. Mudalige, D. Karunarathna, N. de Silva, G. Rathnayaka, and A. S. Perera, “Rule-based approach for party-based sentiment analysis in legal opinion texts,” in 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer). IEEE, 2020, pp. 284–285.
- [29] K. Scaria, H. Gupta, S. Goyal, S. Sawant, S. Mishra, and C. Baral, “Instructabsa: Instruction learning for aspect based sentiment analysis,” in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), 2024, pp. 720–736.

References

- [30] D. Jayakody, A. V. A. Malkith, K. Isuranda, V. Thenuwara, N. de Silva, S. R. Ponnampereuma, G. G. N. Sandamali, and K. L. K. Sudheera, "Instruct-DeBERTa: A Hybrid Approach for Aspect-based Sentiment Analysis on Textual Reviews," *International Journal on Advances in ICT for Emerging Regions (ICTer)*, vol. 18, no. 2, 2025.
- [31] Y. R. Regatte, R. R. R. Gangula, and R. Mamidi, "Dataset creation and evaluation of aspect based sentiment analysis in telugu, a low resource language," in *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 5017–5024.
- [32] M. Shams, N. Khoshavi, and A. Baraani-Dastjerdi, "Lisa: language-independent method for aspect-based sentiment analysis," *IEEE Access*, vol. 8, pp. 31 034–31 044, 2020.
- [33] N. Sultana, R. Sultana, R. I. Rasel, and M. M. Hoque, "Aspect-based sentiment analysis of bangla comments on entertainment domain," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 953–958.
- [34] S. G. Matlatipov, J. Rajabov, E. Kuriyozov, and M. Aripov, "Uzabsa: Aspect-based sentiment analysis for the uzbek language," in *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages@LREC-COLING 2024*, 2024, pp. 394–403.
- [35] Q.-L. Tran, P. T. D. Le, and T.-H. Do, "Aspect-based sentiment analysis for vietnamese reviews about beauty product on e-commerce websites," in *Proceedings of the 36th Pacific Asia conference on language, information and computation*, 2022, pp. 767–776.
- [36] M. R. Ashraf, Y. Jana, Q. Umer, M. A. Jaffar, S. Chung, and W. Y. Ramay, "Bert-based sentiment analysis for low-resourced languages: A case study of urdu language," *IEEE Access*, vol. 11, pp. 110 245–110 259, 2023.
- [37] Z. Wang, Q. Xie, and R. Xia, "A simple yet effective framework for few-shot aspect-based sentiment analysis," in *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, 2023, pp. 1765–1770.
- [38] N. C. Hellwig, J. Fehle, and C. Wolff, "Exploring large language models for the generation of synthetic training samples for aspect-based sentiment analysis in low resource settings," *Expert Systems with Applications*, vol. 261, p. 125514, 2025.

References

- [39] F. Rizvi, T. Navojith, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, “Keyword extraction, and aspect classification in sinhala, english, and code-mixed content,” arXiv preprint arXiv:2504.10679, 2025.
- [40] A. Rizvi, N. Thamindu, A. Adhikari, W. Senevirathna, D. Kasthurirathna, and L. Abeywardhana, “Enhancing multilingual sentiment analysis with explainability for sinhala, english, and code-mixed content,” arXiv preprint arXiv:2504.13545, 2025.
- [41] H. Chai, Z. Yao, S. Tang, Y. Wang, L. Nie, B. Fang, and Q. Liao, “Aspect-to-scope oriented multi-view contrastive learning for aspect-based sentiment analysis,” in Findings of the Association for Computational Linguistics: EMNLP 2023, 2023, pp. 10 902–10 913.
- [42] S. Salawu, J. Lumsden, and Y. He, “A large-scale English multi-label Twitter dataset for cyberbullying and online abuse detection,” in Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021), A. Mostafazadeh Davani, D. Kiela, M. Lambert, B. Vidgen, V. Prabhakaran, and Z. Waseem, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 146–156. [Online]. Available: <https://aclanthology.org/2021.woah-1.16/>
- [43] L. Xu and W. Wang, “Improving aspect-based sentiment analysis with contrastive learning,” Natural Language Processing Journal, vol. 3, p. 100009, 2023.
- [44] T. M. Zayet, M. A. Ismail, and K. D. Varathan, “Aspect extraction in domain lexicon generation: A new frequency-based approach,” IEEE Access, 2024.
- [45] A. Maroof, S. Wasi, S. I. Jami, and M. S. Siddiqui, “Aspect based sentiment analysis for service industry,” IEEE Access, 2024.
- [46] D. Jayakody, K. Isuranda, A. Malkith, N. De Silva, S. R. Ponnampereuma, G. Sandamali, K. Sudheera, and K. G. Sarathchandra, “Enhanced aspect-based sentiment analysis with integrated category extraction for instruct-deberta,” in Proceedings of the 38th Pacific Asia Conference on Language, Information and Computation, 2024, pp. 665–674.
- [47] C. Chen, Z. Teng, Z. Wang, and Y. Zhang, “Discrete opinion tree induction for aspect-based sentiment analysis,” in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 2051–2064.

Thank you
