International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Beyond Language: Multilingual Sentiment Analysis for Code-Mixed E-Commerce Reviews

Author Details:

Ms. Tejaswini S. Pawar[1], Ms. Pratiksha A. Bhamare[1], Ms. Lakshana H. Bachhav[1], Ms. Vaishnavi R. Patole[1], Ms. Asmita S. Pawar[1] [1] Information Technology Department, MVPS's KBTCOE, Nashik, Maharashtra, India
Corresponding Author Email: pawar.tejaswini@kbtcoe.org, bhamareprati@gmail.com, lakshanabachhav25@gmail.com, patolevaishnavi0209@gmail.com, asmitapawar141@gmail.com

https://doi.org/10.55041/ijsmt.v2i4.185

Cite this Article: Pawar, T. S., Bhamare, P. A., Bachhav, L. H., Patole, V. R. & Pawar, A. S. (2026). Beyond Language: Multilingual Sentiment Analysis for Code-Mixed E-Commerce Reviews. International Journal of Science, Strategic Management and Technology, *02*(04). https://doi.org/10.55041/ijsmt.v2i4.185

Abstract -

The online shopping sites have been registering phenomenal growth where colossal volumes of user-generated reviews are being created on a daily to a daily basis. These reviews are quite educative and offer great insight on customer perceptions, likes and experience on products. Such reviews are nevertheless typically written in multi- and code-mixed language, so they are a huge challenge particularly in the linguistically diverse regions such as India. This paper presents the design and implementation of a multilingual sentiment analysis system that is able to address these issues. This system is founded on hierarchical pipeline which includes preprocessing of data, transliteration and translating. It normalizes and eliminates the input data, e.g., noise, e.g., emojis, abbreviations, irregular text patterns. Language and translation of all texts to a standard format are identified with the help of transformer-based models to analyze them. A transformer-based model is used to classify the text into a few types of sentiments, once normalized. The system also has visualization and reporting functions that provide data as charts, graphs, and summaries to determine trends and patterns in customer feedback. Overall, this application demonstrates that the combination of preprocessing techniques and existing deep learning systems can be helpful in processing multilingual and code-mixed data.

Keywords- *multilingual sentiment analysis; code-mixed text; IndicBERT; mBART; transliteration; Hindi-English; Marathi-English; NLP*

I. INTRODUCTION

Consumer communication has been revolutionized by the booming growth of online retail. Millions of reviews on products are posted by shoppers each day on the e-commerce sites; these are texts that direct a potential buyer and give businesses a first hand review on the quality of products and services. There is a commercial and analytical relevance in systematically deriving actionable information out of this data [1][2].The linguistic environment in India is a very complex one. One review can contain Hindi, Marathi or other local languages in the Roman alphabet phonetically transcribed and mixed with standard English. This is often called code-mixing and it does not conform to the monolingual assumptions of traditional sentiment models and lead to catastrophic out-of-vocabulary issues that reduce classification accuracy [11][13].

The previous methods were based on lexicon look-up methods or classical supervised classifiers, like Naïve Bayes and Support Vector Machines. Though these may give a fair ground, they fail to reflect contextual meaning, idiomatic phrasing, and the ambiguity of mixed-language text innate to it. Transformer-based architectures enable new opportunities, but most pretrained models are optimised to use high-resource and single-language settings [10][14].

To cover these shortcomings, this paper presents a pipeline that is modular, integrating noise reduction, script normalisation, language-agnostic translation, and polarity scoring with deep-learning. The system has a visualisation layer that makes the insights available to business stakeholders and does not need NLP knowledge. The paper has the following structure: Section II is a review of previous works; Section III details the proposed system; Section IV presents findings; Section V-VI discuss benefits, limitations and future prospects; Section VII is the conclusion.

II. LITERATURE REVIEW

Sentiment analysis has developed beyond rule-based lexicons to become a full-fledged NLP sub-field. Prasetyo et al. [1] showed the effectiveness of domain-specific NLP pipes in enhancing consumer-insight extraction on e-commerce corpora, whereas Ismail et al. [2] validated that noise reduction by preprocessing special to informal digital text and improved precision of classifiers. Alsaedi et al. [3] conducted a survey of retail platform mining techniques, with the trade-off between complexity of the model and inference latency being noted.

The advent of large pretrained transformers was a turning point. Rana et al. [10] presented BERT-BiGRU-Senti-GCN, a hybrid encoder, which combines bidirectional encoding with graph-based reasoning, and obtained good results on English review benchmarks. In the case of Indian languages, particularly, Thara andPoornachandran [11][14]examined Malayalam-English code-mixed social media and demonstrated that when language identification is done with dedication, prior to classification, it is seen that the misclassification rate is significantly reduced.

Shanmugavadivel et al. [13] carried this line of research into several pairs of Indic languages, with results that multilingual models trained jointly, rather than per-language fine-tuning, perform better in cases where annotated data is limited. The native-script and Romanized identification of Indic languages, presented by Madhani et al. [16] as BhashaAbhijnaanam, is a direct inspiration of what is now known as the transliteration stage of the current system.

Transliterated Hindi and Marathi review sentiment were specifically addressed by Sutar and Desai [18] and Khan and Sawarkar [19], which showed that lexicon-enhanced transformer models perform better than vanilla fine-tuning in the situation of sparse domain vocabulary. This advancement notwithstanding, there are few standardised five-class benchmarks of code-mixed Indian retail reviews, which this work fills to a certain degree.

III. METHODOLOGY

The suggested pipeline works with raw reviews and undergoes five consecutive, modular steps. The successive stages generate cleaner, more semantically homogenous representations, which feed the subsequent ones, ensuring the data integrity of the successive stages.

Figure: 1

A. *Input and User Interaction*

The user interaction module enables the reception of the raw e-commerce reviews into the system. The administrator has control over the data stream so that appropriate datasets of reviews (which, in many cases, have a multifaceted mixture of English and regional languages) are properly placed in line to be processed. This module will act as an access point to the system and it will utilize initial data points to enable the smooth integration along the pipeline.

Figure: 2

B. *Data Preprocessing*

Noise removal is performed systematically on raw review text. HTML tags, special characters and too much punctuation are removed. Emojis are substituted by text descriptions or they are eliminated based on the relevance of the sentiment. Unofficial shortenings and internet slang (e.g., omg, def) are translated into official equivalents in a curated slang dictionary. This step significantly minimises the errors of tokenisation in subsequent steps.

Figure: 3

C. *Transliteration*

Phonetically spelled the Latin letters are often used in Hindi or Marathi; the phonetic tokens are handled as out-of-vocabulary English words, with disastrous out-of-vocabulary effects. The linguistic identity of each token is recreated by transliterating Romanized strings with a special transliteration module, which transliterates the Romanized strings into Devanagari script [16], reducing the sparseness of the data.

Figure: 4

D. *Language Detection and Translation*

mBART [17], a multilingual sequence-to-sequence transformer trained on 25 languages, is a post-transliteration model that identifies which language the text is written in

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

and translates non-English sections into English. The volume of reviews can be directed to one target language, which allows the downstream processing to be uniform without language-conditional model branches.

Figure: 5

E. Sentiment Classification

A fine-tuned IndicBERT model classifies the text which is translated and normalised. IndicBERT is a small version of BERT that is pretrained on 12 large Indian languages and English with a shared subword set [20], and is naturally sensitive to Indic language patterns. The dependent variable uses five labels of polarity on a scale, Very Bad, Bad, Average, Good, and Very Good to offer a finer scale than binary positive/negative paradigms.

Figure: 6

F. Visualisation and Reporting

The outputs of the classification are summarized in pie charts, bar graphs and temporal trend lines and displayed in a dashboard, which is interactive, to provide the entire picture of the data. Export reports to PDF and PNG to make sure that non-technical stakeholders are able to monitor the sentiment trends and respond to new trends without being required to access the model directly. The centralized visualization layer is used to mediate the interface between the complex processing of algorithms and the strategic process of making decisions by offering real-time filtering and drill-down capabilities. Moreover, automated warning systems can be integrated that notifies the management of the sudden shift in the customer perception in order to be responsive and adaptable to the shift in the market.

Figure: 7

G. Business Insights Generation

The final phase involves the synthesis of actionable business insights. The system can trace a long-term consumer behavior and product-specific trends by consolidating investigated data. These understandings enable organizations to make decisions that are data-driven to maximize product offerings and improve customer engagement strategies in the competitive digital market.

Figure: 8

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

IV. RESULTS AND DISCUSSION

The model was evaluated using stratified k-fold cross-validation on a labelled set of Indian e-commerce reviews with five sentiment classes. The pipeline achieved an overall accuracy of 68.32 % (σ = ±0.0477) and a macro F1-score of 0.7009 (σ = ±0.0455). The emphasis was made on macro averaging because each class was treated equally although there was a moderate imbalance in the datasets.

Table I: Class-wise Performance Metrics

Class	Precisio n	Recall	F1-Scor e
Average	0.7819	0.8729	0.8211
Bad	0.7620	0.7473	0.7478
Good	0.5911	0.3456	0.4271
Very Bad	0.7910	0.9167	0.8457
Very Good	0.5944	0.7591	0.6626

The highest F1-scores (0.846 and 0.821 respectively) are obtained with the Very Bad and Average classes, as indicated in Table I, presumably due to the use of more salient vocabulary in strongly polarised opinions. The lowest F1-score (0.427) of the Good class can be explained by semantic overlap with Very Good and contextual ambiguity in mildly positive language, which are also observed in other related multilingual sentiment work [10][13].

The confusion matrix analysis proves that misclassifications are concentrated around the contiguous polarity boundaries (Good↔Very and Bad↔Average), which points to the fact that the five-point scale presents itself as the source of subjectivity. Future work may use ordinal regression loss or soft-label training to learn this polarity continuum better.

Figure: 9

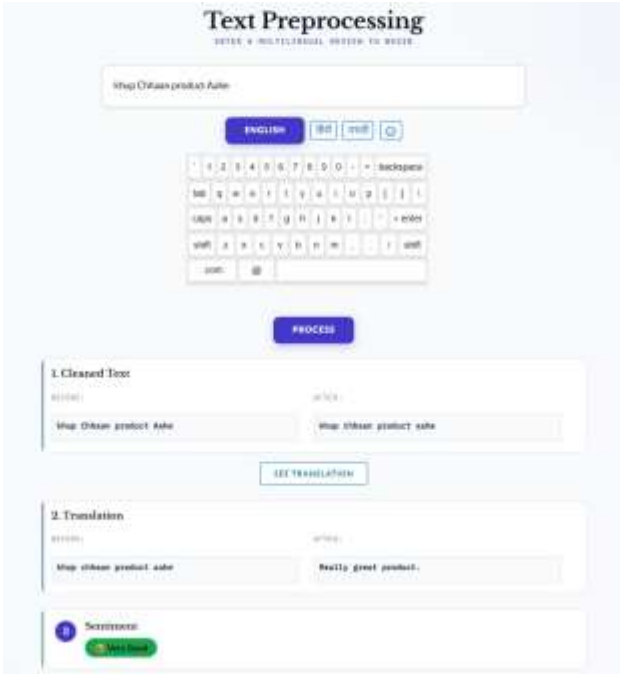
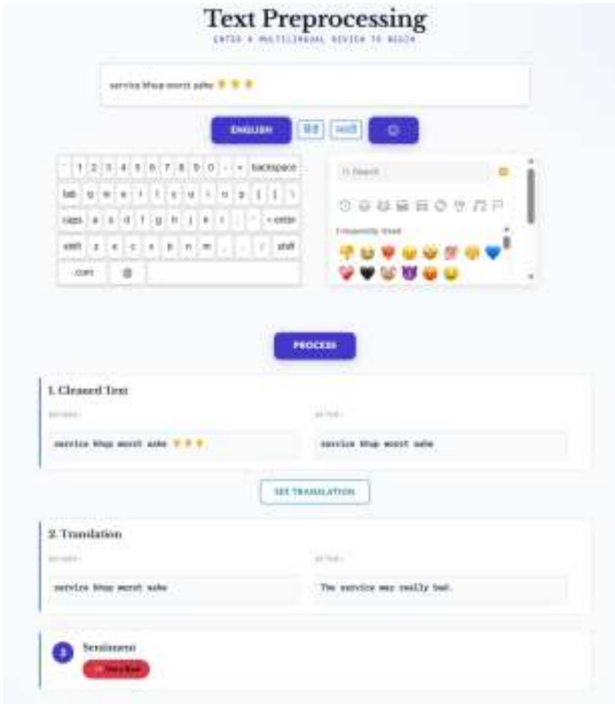


Figure: 10



V. ADVANTAGES AND LIMITATIONS

A. Advantages

- Multilingual Flexibility: The system supports Hinglish, Marathi-English, and other code-mixed inputs, in contrast to English-only analysers, through the mBART and IndicBERT models.
- Less Sparsity in Data: Transliteration allows Romanized tokens to be viewed as known, which significantly boosts the amount of signal that can be used.
- Granular Classification: The five-point sentiment scale provides the businesses with an excellent ability to understand customer satisfaction in a more detailed way as compared to binary models.
- Actionable Reporting: The dashboard will transform high-dimensional outputs into exportable PDF/PNG reports that can be used by non-technical decision-makers.
- Scalable Architecture: The decoupled modular architecture enables single stages to be upgraded or replaced without a complete redesign of the entire pipeline.

B. Limitations

- Sarcasm Detection: Transformer models continue to fail when users use positive wording to express highly negative experience, resulting in a linguistic paradox.
- Class Imbalance: The amount of training data included more positive and neutral reviews than negative ones, causing classification bias to signals of satisfied customers.
- Computational Cost: mBART and IndicBERT are intensive in terms of GPU,which can be limiting in real-time usage in smaller organisations.
- Domain Specificity: The system is specific to the e-commerce domain; retraining to other domains, like the healthcare domain, or political domain, is necessary.
- Noisy Data Sensitivity: Despite the preprocessing phase, extreme spelling difference, over-use of special characters or repetitive strings may still interfere with tokenisation.

VI. FUTURE SCOPE

A number of directions are to be extended to the current system. By adding Aspect-Based Sentiment Analysis (ABSA) the pipeline would be able to assign a certain amount of polarity to particular product characteristics like delivery speed, battery life, or the quality of the packaging, instead of a single review level score. Another near-term priority is enhanced contrastive or counterfactual training goals on the detection of sarcasm and irony [23].

Multimodal Sentiment Integration is also considered: the system might cross-check negative complaints with photographic proof of a defect by combining textual review signals with concomitant product photos through computer-vision models. Increasing language coverage to Tamil, Telugu and Bengali will increase its applicability throughout the Indian subcontinent.

Replacement of batch processing with real-time inference through streaming APIs would enable businesses to respond to negative sentiment as soon as they arise instead of hours. Lastly, by incorporating Explainable AI (XAI) with attention-map visualisations, administrators would be able to determine which specific keywords or regional slang words lead to a certain sentiment score, establishing trust inautomated results among the stakeholders [24][25].

VII. CONCLUSION

This paper introduced a multilingual sentiment analysis pipeline that is intended to analyze code-mixed Indian e-commerce reviews. The system fuses systematic noise elimination, phonetic transliteration, mBART-based neural translation, and IndicBERT sentiment classification, to attain competitive accuracy (68.32 0.7009, macro F1) on a tasking five-class benchmark, without language-specific model variants per regional language pair.

The visualisation dashboard, which is built-in, can convert raw model predictions into charts and exportable reports, and hence provide insights to business users at a glance. The modular design will make sure that every step can be upgraded individually with the advancement of NLP technology.

Subsequent releases will include ABSA, sarcasm detection, multimodal fusion and expanded coverage of Indic languages, gradually evolving the platform into a proactive real-time intelligence platform to operators of global e-commerce.

ACKNOWLEDGMENT

The authors would like to express their gratefulness to the faculty members of the Information Technology Department, KBTCOE of MVPS, Nashik, who have guided and given constructive feedback to the project all along. The institution is appreciatively recognized to have offered the computational facilities needed to train and test models. The researchers also wish to extend their gratitude to the creators of Hugging Face Transformers, IndicBERT, and mBART, who have made their tools available to the research community.

REFERENCES

[1] A. Prasetyo, R. D. Lestari, and R. Hartono, "Using NLP to Enhance Customer Sentiment Analysis in Ecommerce," 2024. doi: 10.62951/ijies.v1i2.90

[2] W. S. Ismail, M. M. Ghareeb, and H. Youssry, "Enhancing Customer Experience through Sentiment Analysis and NLP in E-commerce," J. Wireless Mobile Netw., vol. 15, no. 3, pp. 60-72, 2024.

[3] T. Alsaedi et al., "Sentiment Mining in E-Commerce," Int. J. Electr. Comput. Eng. Syst., vol. 15, no. 8, pp. 641-650, 2024.

[4] J. M. Jose and P. Narayanan, "Sentiment Analysis With NLP," in Advances in Computational Intelligence and Robotics, 2024, pp. 211-256.

[5] D. G. Takale, "NLP-Powered Virtual Shopping Assistants and Sentiment Analysis in E-commerce," 2024. doi: 10.48001/jofsn.2024.221-5

[6] B. M. Vamsi, "Sentiment Analysis on Online E-commerce Product Reviews using NLP," Int. J. Res. Appl. Sci. Eng. Technol., 2024.

[7] E. H. Aydogan and F. Y. Okay, "Sentiment Analysis of Reviews for E-Commerce Applications," pp. 1-6, 2024.

[8] R. Lourdusamy, P. Thangavel, and S. Johnbosco, "Sentiments Unleashed: Pioneering the Frontier of Sentiment Analysis," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 10, no. 5, pp. 205-220, 2024.

[9] H. Sharma and V. Kakran, "Sentiment Analysis: Analyzing Flipkart Product Reviews using NLP and Machine Learning," 2024.

[10] M. R. R. Rana et al., "BERT-BiGRU-Senti-GCN: an Advanced NLP Framework for Customer Sentiments in E-Commerce," 2025. doi: 10.1007/s44196-025-00747-1

[11] S. Thara and P. Poornachandran, "Social media text analytics of Malayalam-English code-mixed data using deep learning," J. Big Data, vol. 9, p. 45, 2022.

[12] M. Marreddy et al., "NLP Tasks in Telugu: Data Sets, Embeddings, and Models," ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 22, 2022.

[13] K. Shanmugavadivel et al., "Deep learning-based sentiment analysis on multilingual code-mixed data," Sci. Rep., vol. 12, p. 21557, 2022.

[14] S. Thara and P. Poornachandran, "Transformer Based Language Identification for Malayalam-English Code-Mixed Text," IEEE Access, vol. 9, pp. 118837-118850, 2021.

[15] K. Shalini et al., "Sentiment Analysis for Code-Mixed Indian Social Media Text," in Proc. ICACCI, 2018, pp. 1126-1131.

[16] Y. Madhani, M. M. Khapra, and A. Kunchukuttan, "BhashaAbhijnaanam: Language Identification for 22 Indic Languages," arXiv:2305.15814, 2023.

[17] Y. Liu et al., "Multilingual Denoising Pre-training for Neural Machine Translation," Trans. Assoc. Comput. Linguist., vol. 8, pp. 726-742, 2020.

[18] R. Sutar and K. Desai, "Sentiment Analysis of Transliterated Hindi and Marathi Using Lexicon-Enriched Transformers," Int. J. Environ. Sci., vol. 11, pp. 1228-1238, 2025.

[19] Z. F. Khan and S. D. Sawarkar, "Enhancing Sentiment Analysis of Marathi-English Code-Mixed Texts," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 18s, pp. 741-750, 2024.

[20] A. Kunchukuttan et al., "IndicGLUE: NLU Benchmark for Indian Languages," in Proc. EMNLP, 2020, pp. 2110-2146.

[21] R. Joshi, "L3Cube-MahaSent: A Large-scale Marathi Sentiment Analysis Dataset," in Proc. AACL-IJCNLP, 2022.

[22] R. Dhadhal, A. Kunchukuttan, and P. Kumar, "IndicBART: Multilingual Pre-trained Model for Indic NLG," arXiv:2109.02903, 2021.

[23] M. S. Akhtar et al., "Sarcasm Detection in Hindi-English Code-Mixed Social Media Text," Knowl.-Based Syst., vol. 194, 2020.

[24] A. Kumar and V. Jaiswal, "A Deep Learning Approach for Multilingual Sentiment Analysis in E-Commerce Reviews," Int. J. Cogn. Comput. Eng., vol. 3, pp. 115-125, 2022.

[25] P. Sharma and R. Gupta, "Building Scalable Real-time Sentiment Analysis Pipelines for E-Commerce Platforms," IEEE Trans. Serv. Comput., vol. 17, no. 1, pp. 88-102, 2024.