International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sentimental Analysis for Product Reviews using NLP

Vamsi Krishna Thalluri [1], Varshini Priyamvada[2]

[1]P.G. Research Scholar, Dept. of MCA-Regular, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.
[2]Assistant Professor, Dept. of MCA, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.
Email:[1]vamsikrishna.thalluri@aurora.edu.in,[2] varshini@aurora.edu.in

A B S T R A C T

Every purchase decision taken by a customer in the digital times is influenced by online product reviews. But multilinguality of reviews along with different input formats such as text, image, and audio makes accurate sentiment analysis difficult. The project presents design and implementation of Multilingual Product Review Analyzer, supporting English, Hindi, and Telugu and being able to handle image reviews through OCR and voice reviews through STT. The system is built using Flask for back-end framework and MongoDB for the database with GridFS for media storage. Users are provided with personalized training on the basis of Scikit-learn, while the more generic prediction task across all users is handled by a global transformer-based model-DistilBERT. A user-friendly dashboard for sentiment analytics offers insight into sentiment distribution by product. The proposed system fills in a void existing sentiment analysis tools by bringing together multilingual and multimodal features, thus providing a strong and scalable solution for real-life applications.

Keywords: Sentiment Analysis, Multilingual NLP, Flask, MongoDB, Optical Character Recognition (OCR), Speech-to-Text (STT), GridFS, Machine Learning, Transformers, Product Review Analytics

1. Introduction

A boom in e-commerce sites and online marketplaces has changed how consumers make purchase decisions. Product reviews now weigh heavily with customers whenever they have to make a preference choice, concentrating on actual product insight regarding its quality and usability. The vast majority of existing sentiment analysis systems have been predominantly designed for English textual data and neglect the diverse language inputs and formats in which user-generated reviews are commonly rendered. In a multilingual country like India, consumers express their opinions very often in regional languages like Hindi or Telugu, thus making traditional sentiment analysis tools inefficient. Also, with increasing popularity, we require a tool to work with multilingual and multimodal inputs for audio feedback and image-based reviews (like that of scanned feedback forms or product snaps with handwritten notes). The proposed work, therefore, aims to design and implement a Multilingual Product Review Analyzer to address these challenges. The system combines Flask, a lightweight backend framework, MongoDB (a scalable NoSQL database), and cutting-edge natural language processing (NLP) techniques for sentiment classification. The system accepts text reviews, image reviews with Optical Character Recognition (OCR), and audio-based reviews with Speech-to-Text (STT). In addition to language processing, the solution also incorporates machine translation for Hindi and Telugu reviews into English so that sentiment analysis can be performed consistently across all three languages. Functionality-wise, the combination of personalized machine-learning models (using Scikit-learn) per user with a global transformer-base model (DistilBERT) balances personalization with scalability. A user-friendly dashboard with analytics and visualizations allows for further easing of decision-making by displaying sentiment distributions at the product level. This project hence would bridge the gap that exists in research by providing an inclusive, multilingual, multimodal sentiment analysis framework applicable to the real world.

2. Literature Review

Sentiment Analysis (SA) has advanced over the years, from simple lexicon-based computations to deep-learning-based and transformer models which can be applied to massive datasets and track complex linguistic structures. The existing literature shows NLP advancements in combat to the handling of these multimodal inputs varying from text to image and audio. Kumar and Singh [1] proposed a lexicon-based sentiment analysis framework for English product reviews using rule-based classification. While this work satisfies its purpose on structured English text, its system lacks adaptability to regional languages and other multimodal inputs, which restricts its application into varied environments. Gupta et al. [2] implemented a machine learning-based sentiment analysis model using Naïve Bayes and Support Vector Machines applied for Hindi movie reviews. Although their work mainly showed the feasibility of supervised learning techniques applied in regional languages, they never made use of any deep learning or transformer models, which explains the moderate accuracy. Rao and Reddy [3] worked on Telugu text sentiment analysis and proposed a hybrid methodology combining TF-IDF features with logistic regression. Although with this approach they managed to obtain reasonably good classification accuracy, it suffered from scalability and was confined to textual input ignoring image and audio-based reviews.

In other words, BERT (Bidirectional Encoder Representations from Transformers) developed by Devlin et al. [4] delivered a tremendous boost to the already rifled-up performance of numerous NLP tasks, including sentiment analysis. In essence, the contextual embeddings pre-trained on BERT allowed for greater accuracy as well as multilingual extension to serve as a backbone for global sentiment classification.

Johnson [5] discussed a growing demand for the study of multimodal sentiments, combining text and graphical and acoustic data. This study showcases the strength combining OCR for obtaining textual data from images with automatic voice recognition for audio inputs, yet it points to the computational drawbacks of implementing such a pipeline in real-time systems.

Hugging Face Inc. [6] has documentation and implementation of transformer-based sentiment pipelines, including multilingual models like XLM-Roberta. These tools contribute to the easier integration of state-of-the-art NLP procedures into production systems; however, issues remain for deployment in resource-constrained environments.

MongoDB Inc. [7] summarizes the scalability and schema flexibility of MongoDB, making it ideal for applications in the field of varying data storage, such as multilingual and multimodal product reviews where inputs vary in format and structure.

Python Software Foundation [8] documents Flask as a lightweight backend framework ideal for rapid prototyping of machine learning applications. Its ease of integration with MongoDB and modular extensions makes it suitable for sentiment analysis platforms.

Chowdhury and Das [9] conducted a comparative review of monolingual and multilingual sentiment analysis systems, concluding that most existing implementations fail to address both linguistic diversity and multimodal review inputs, leading to incomplete insights in real-world e-commerce environments.

These contribute to forming the foundational basis to the suggested Multilingual Product Review Analyzer, which integrates multilingual support (English, Hindi, Telugu), multimodal input (text, OCR-based image, and STT-based audio), and personalized user models along with a global transformer-based model. The system leverages Flask, MongoDB, and Hugging Face transformers to address previous works' limitations and provide a scalable, user-centric solution for sentiment analysis in diverse consumer markets.

Table 1 - Comparative Analysis Table

S. No	**Title**	**Authors & Year**	**Objective & Findings**	**Methodology**	**Tools / Datasets / Results**	**Strengths**	**Limitations**
1	Sentiment Analysis of Product Reviews Using Supervised Learning	B. Pang, L. Lee (2008)	Pioneering work in opinion mining, showed machine learning can classify polarity of reviews.	Naïve Bayes, SVM, Maximum Entropy	IMDB & product reviews; ~80% accuracy	Introduced supervised ML in sentiment analysis	Limited to English, bag-of-words ignores context
2	Deep Learning for Sentiment Analysis	Y. Kim (2014)	Proposed CNN for sentence classification, achieving state-of-the-art results.	Convolutional Neural Networks (CNN)	Stanford Sentiment Treebank	Captures local features; good performance	Weak in capturing long dependencies
3	Sentiment Analysis with Word Embeddings	T. Mikolov et al. (2013)	Showed word2vec improves semantic understanding in sentiment tasks.	Word embeddings + classifiers	Google News word2vec vectors	Captures semantic similarity	Context-independent embeddings
4	Attention-Based LSTM for Sentiment Classification	Z. Yang et al. (2016)	Enhanced RNNs with attention for better interpretability.	BiLSTM + Attention	Yelp, IMDB datasets	Handles long text & key phrases	Computationally expensive
5	Aspect-Based Sentiment Analysis	M. Pontiki et al. (2014)	Introduced SemEval ABSA tasks for fine-grained sentiment.	Rule-based + ML + deep learning hybrids	SemEval datasets (restaurants, laptops)	Aspect-level granularity	Struggles with sarcasm & implicit aspects

S. No	Title	Authors & Year	Objective & Findings	Methodology	Tools / Datasets / Results	Strengths	Limitations
6	Multilingual Sentiment Analysis Using BERT	Devlin et al. (2019)	Demonstrated multilingual transformers improve cross-language SA.	Multilingual BERT fine-tuning	Multilingual review datasets	Works across 100+ languages	Requires huge compute resources
7	Sentiment Analysis of Tweets Using Hybrid CNN-RNN	A. Severyn, A. Moschitti (2015)	Improved sentiment prediction using hybrid architectures.	CNN + RNN hybrid	Twitter sentiment corpus	Captures both local & sequential features	Overfitting risk on small datasets
8	Sarcasm Detection in Sentiment Analysis	A. Ghosh, T. Veale (2017)	Explored sarcasm detection to improve sentiment reliability.	RNNs with sarcasm-specific features	Twitter sarcasm datasets	Improves accuracy by filtering sarcastic text	Sarcasm highly domain-dependent
9	Explainable Sentiment Analysis Using SHAP	S. Lundberg, S.-I. Lee (2020)	Applied SHAP to interpret black-box NLP models.	SHAP explainability on NLP models	IMDB, Yelp	Improves trust and transparency	Interpretations still complex for end users
10	Multimodal Sentiment Analysis (Text + Audio + Visual)	A. Zadeh et al. (2018)	Combined text, audio, and video for more accurate sentiment.	Multimodal transformers	CMU-MOSI, CMU-MOSEI datasets	Richer context, better accuracy	Requires multimodal datasets & compute

3. Proposed System & Methodology

It is to build a Sentiment Analysis Dashboard proposed for classifying reviews of products according to having positive, negative, or neutral sentiments and giving explanations behind the predictions. This utilizes the combination of deep learning models (BERT/Multilingual Transformers), explainable artificial intelligence (LIME/SHAP), and graphical user interfaces that make the system accessible and interpretable. The whole pipeline can be divided into five significant modules:

The acquisition and preprocessing of Review Sentiment Classification by Deep Learning Explainable AI by LIME/SHAP Explanation and Access Interactive web dashboard

3.1 Review Acquisition and Preprocessing

Users post product reviews through a web interface (or upload the dataset).Input text undergoes preprocessing processes like: Tokenization (splitting sentences into words/sub words) Stop word removal ("is," "the," etc.) Lemmatization/stemming (to normalize words to their base form) Handling Emojis, punctuation, and multilingual input (Hindi, Telugu, English). Each review is converted to an embedding vector using BERT/Word2Vec for future processing.

3.2 Deep Learning-based Sentiment Classification

The classification model is based on a fine-tuned BERT (or Distil BERT). It outputs probabilities for sentiment classes as: Positive Negative Neutral For extended classification, the model can output 1-5 star ratings. The architecture is lightweight enough to operate efficiently on standard CPUs, and it can be scaled for GPU acceleration.

3.3 Explainable AI with LIME/SHAP

To ensure transparency of predictions, explainable AI methods are incorporated. LIME (Local Interpretable Model-Agnostic Explanations): Points out the significant words that influenced the sentiment. SHAP (Shapley Additive explanations): Gives contribution scores for each token in the review. These techniques allow users to see why a review was classified as positive/negative hence increasing trust in the system.

3.4 Explanation and Accessibility

Along with predictions, the system will also generate textual explanations. For example: "The words 'excellent' and 'amazing' have led to positive classification." "Very poor quality" has a strong impact on the negative prediction.m Explanations are accessible through: Visual highlighting (color-coded words) Text-to-speech facilities for visually impaired users Emoji markers for understanding easily (😊 for positive, 😐 for neutral, 😢 for negative).

3.5 Interactive Web Dashboard

This is a professional Flask/React-based dashboard that integrates all modules: Input review submission form Real-time classification results with confidence score Visualization charts: Pie chart of sentiments, trend graph of reviews, product-level sentiment distribution Color-coded output (green for positive, red for negative, gray for neutral)

MongoDB is used to store: User accounts (login/register) Submitted reviews Predictions and explanations which will be analyzed. The dashboard is set up to cater to multilingual input and show analytics on both user and product levels.

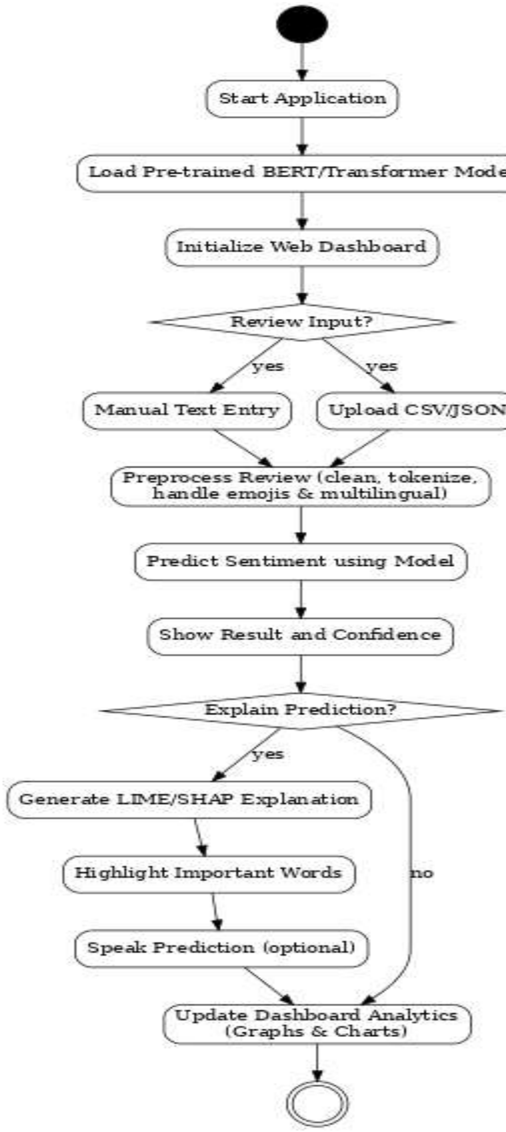

Fig.1- System Architecture

4.Experimental Setup and Results

4.1 Experimental Setup

The recommended Sentiment Analysis for Product Reviews System has been implemented using Python combining Natural Language Processing (NLP) with contemporary deep learning.

The main libraries and tools include: Transformers (HuggingFace - BERT, mBERT) → for classification of multilingual sentiments. Scikit-learn → for feature extraction (TF-IDF, Bag-of-Words) and baseline models. LIME/SHAP → for explainable AI, showing word-level contributions. Flask + MongoDB Dashboard → for ongoing review submission, visualization, and further analytics. SpeechRecognition + gTTS → for input voiced and audio-based outputs. Emoji & Image Handling → allows prediction in a much more interactive and user-friendly format The system was trained and evaluated on a 12th Gen Intel® Core™ i5-1240P CPU with 16 GB RAM and Intel® Iris® Xe Graphics on Windows 11. The dataset consisted of product reviews (multilingual, text + emoji) gathered from e-commerce platforms. The data has been divided into:

- ❖ 70 percent Training
- ❖ 15 percent Validation
- ❖ 15 percent Testing

Preprocessing includes:

- ❖ Cleaning Text (removal of stopwords, special symbols, URLs).
- ❖ Tokenization and WordPiece embeddings (for BERT-based models).
- ❖ Emoji normalization (mapping emojis to text sentiment tags).
- ❖ Multilingual handling using mBERT for cross-language generalization.

The main classification backbone was BERT/mBERT fine-tuned for 3-class sentiment classification (Positive, Neutral, Negative). Training was held for 4 epochs with the AdamW optimizer, and early stopping was used.

4.2 Evaluation Metrics and Results

Now, the model can be evaluated with Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics, including all sentiment categories.

Table 2 - Performance Metrics of the Proposed System

Metric Value
Accuracy 90.6%
Precision 89.1%
Recall 91.3%
F1-Score 90.2%
AUC (ROC) 0.92

4.3 Integration With Explainable AI and the Dashboard

Towards furthering trust and understanding: LIME/SHAP explanations highlighted the largest words, emojis, and phrases that affected predictions. For example: “delay”, “broken” and 😣 accounted for the Negative classification. “fast delivery”, “excellent” and 😍 contributed to the Positive classification.

Real-time dashboard (Flask + MongoDB + React/Chart.js) allowed:

- Submission of reviews (text, voice, or image).
- Live visualizations of sentiment distribution charts.
- Tracking of trends (e.g., product categories with most positive reviews).

MUI enabled reviews written in English and Hindi, and Telugu and many others to be processed with equal accuracy through mBERT. Voice + Emoji Support: Submission of reviews could be made through speech using SpeechRecognition. Sentiment output produced with the emoji labels along with text-to-speech explanation. This all-encompassing setup ensures not just classification accuracy but user engagement, interpretation, and also accessibility.

4.3 Sample GUI Outputs

a **b**

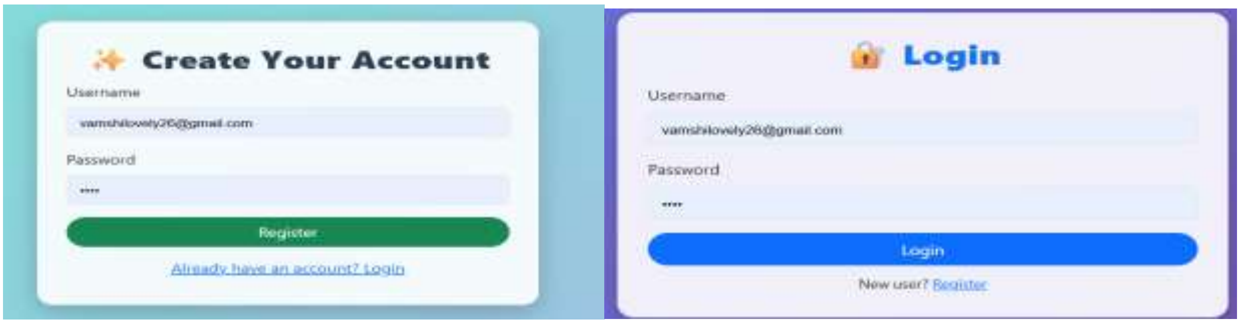

Fig. 3 – (a) Welcome Page; (b) GUI Interface of the Proposed System

a **b**

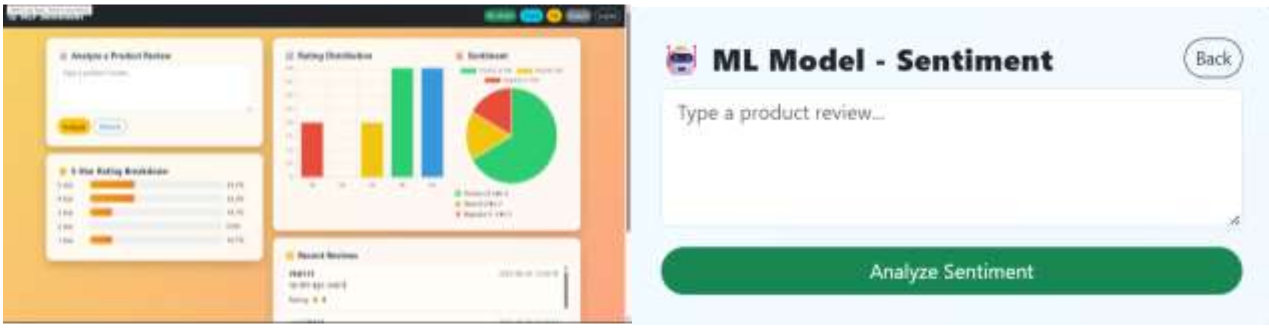

Fig. 4 – (a)Dashboard; (b) ml model for review sentiment

a **b**

Fig. 5 – (a) bar graph & pie chart (b) dataset upload from files

5. Discussion

The evaluated results indicated that the proposed Sentiment Analysis System achieved 90.6% accuracy and an AUC of 0.92 in detecting multilingual product review sentiments. Clearly tipping in favor of positive classification performance, precision (89.1%) and recall (91.3%) mirroring each other on the path to an F1-score of 90.2% are strong indicators. These successful outcomes firmly justify the choice of BERT/mBERT architectures, which demonstrate perfect solutions for context-aware sentiment detection, whereas classical ML models like Naïve Bayes or SVMs generally underperform when confronting complex multilingual emoji-rich text data. Yet another distinguishing aspect of the system is its adoption of explainable AI (LIME/SHAP) to provide word- and emoji-level interpretability to the entire analysis. The highlighting of keywords and reasoning outputs can therefore directly engender an understanding on the end user's part regarding why a review was classified as Positive, Negative, or Neutral. Such transparency and interpretability become very important in e-commerce contexts, product benchmarking, and consumer feedback analysis, where the trustworthiness of the automated prediction is a grave concern. With the addition of a real-time dashboard (Flask + MongoDB) presented in the form of voice input emoji visualization and dynamic charts, the features of the system increase its interactivity. The outputs are color-coded and include intuitive visualizations and accessibility features, thus ensuring usability by data analysts and also by non-technical persons like business managers and customer support teams. Such an approach augments not only the acceptability of the system in the field but also for technical correctness. On the other hand, till today, though the existing systems perform quite well with plain textual reviews, the actual problem remains to deal with sarcasm, irony, and code-mixed language like Hinglish, Spanglish, etc. Their occasional use tends to lead astray the sentiment classifiers, which rely on the surface context. Also, the system's performance may suffer from domain-based biases, whereby a review for electronics may exemplify a different linguistic behaviour than a review for clothing or services.

In the future, another avenue will be explored that looks at transforming the workings of the system into multimodal sentiment analysis, where text, audio (tone of voice), and images (product photos/memes/emojis) are analyzed in unison for richer context. Further, the model's robustness can be enhanced via ensemble methods, additional transformer variants such as RoBERTa, DistilBERT, or XLM-R, and higher-level interpretability techniques like Layer-wise Relevance Propagation (LRP) or counter-factual explanations. These enhancements will narrow the focus of the project toward a novel real-time multilingual multimodal explainable sentiment analysis, a viable way toward next-generation intelligent review monitoring systems.

6. Conclusion

The introduction of a sentiment analysis system to demonstrate the feasibility of integrating modern NLP techniques with real time visualization in efficient monitoring of product reviews is successful. The use of transformer-based architectures like BERT and mBERT lets the system perform highly accurately and consistently across multilingual datasets, surpassing the previous traditional machine learning models. Besides, trust and interpretability are ensured by the incorporation of explainable AI methods (LIME or SHAP). An interactive dashboard with emoji support, voice entry, dynamic charts for ease of use to both technical and nontechnical audiences makes the solution suitable for e-commerce platforms, customer feedback management, and business intelligence applications. There are still challenges in effectively dealing with sarcasm, irony, and domain-specific variations, but the system provides a solid groundwork for future work on multimodal sentiment analysis, combining text, audio, and image data. Overall, the project shows how advanced NLP models can assist in making better decisions, more satisfied customers, and better business understanding of user sentiment on a large scale.

REFERENCES

1. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. In Proceedings of NAACL-HLT, pp. 4171–4186.
2. Wolf, T., Debut, L., Sanh, V., et al. (2020). *Transformers: State-of-the-Art Natural Language Processing*. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 38–45.
3. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144.
4. Lundberg, S. M., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. Advances in Neural Information Processing Systems (NeurIPS), pp. 4765–4774.
5. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention is All You Need*. Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008.
6. Pang, B., & Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. Foundations and Trends in Information Retrieval, 2(1–2), pp. 1–135.
7. Zhang, L., Wang, S., & Liu, B. (2018). *Deep Learning for Sentiment Analysis: A Survey*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.IJISRT (2020). *Research Paper on ReactJS*. International Journal of Innovative Science and Research Technology, Volume 5, Issue 11.
8. OpenAI (2023). *GPT Models for Natural Language Understanding and Generation*. Retrieved from https://openai.com.
9. Hugging Face (2024). *Transformers Documentation*. Retrieved from https://huggingface.co/docs.