ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1434

Online Review Sentimental Analysis

S. Rajarajeswari[1]; Ashwin D[2]; Nithish Kumar S[3]; Vishal R[4]; Vishwa N[5]

[1] Assistant professor; [1] M.Tech; [2] UG Scholar

[1] AIML SNS College of Technology Coimbatore, India
[2,3,4,5] Department of Artificial Intelligence & Machine Learning SNS College of Technology Coimbatore, India

Publication Date: 2025/05/02

Abstract: As the online shopping is growing very fast, business sentiment analysis based on customer reviews is very important for companies. Conventional sentiment analysis tools such as VADER are incapable of picking up high-level language structures, regional idioms, and context-dependent phenomena. This work introduces a state-of-the-art sentiment analysis system using transformer-based architectures such as BERT and RoBERTa, which are fine-tuned over a bespoke e-commerce corpus. The model undertakes Aspect-Based Sentiment Analysis (ABSA) to derive sentiments for particular product features like price, quality, delivery, and customer service. In addition, multilingual support is built using Indic NLP models and Google language detection APIs to support regional languages like Hindi and Tamil. The real-time sentiment stream is built using Apache Kafka, allowing companies to track customer feedback in real-time. Experimental results indicate that the system proposed here is more accurate and relevant than conventional approaches and offers a scalable solution for contemporary e-commerce websites.

Key words: *Sentiment Analysis, BERT, RoBERTa, ABSA, Multilingual NLP, Real-Time Streaming, E-commerce.*

How to Cite: S.Rajarajeswari (M.Tech); Ashwin D; Nithish Kumar S; Vishal R; Vishwa N (2025), Online Review Sentimental Analysis. *International Journal of Innovative Science and Research Technology*, 10(4), 2152-2155. https://doi.org/10.38124/ijisrt/25apr1434

I. INTRODUCTION

Over the past few years, e-commerce has seen incredible growth, and with it has come an explosion of customer reviews on multiple platforms. Knowing the sentiment of these reviews can provide businesses with insightful information about customer satisfaction and points of improvement. Rule-based systems like VADER are quick at sentiment scoring but tend to miss the subtlety to interpret complex sentence structure, sarcasm, and multilingual content accurately. With deep learning, transformer models such as BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (A Robustly Optimized BERT Approach) have transformed the field of Natural Language Processing (NLP) with state-of-the-art outcomes in a variety of text analysis tasks. This study emphasizes substituting conventional sentiment analysis tools with fine-tuned transformer models, adopting aspect-based sentiment analysis (ABSA), adding multilingual support, and real-time data processing via Kafka. This research is prompted by the exigency to engineer a strong, scalable, and real-time system for sentiment analysis that can function more effectively for the dynamic multilingual e-commerce environment.

II. RELATED WORK

Sentiment analysis has been extensively studied over the last decade. Initially, lexicon-based techniques were used, including Senti WordNet and VADER, to establish the polarity of text. Although these approaches are light and simple to use, they tend to break down in dealing with difficult linguistic structures such as negations, idioms, and domain-specific expressions.A recent study has established that transformer-based models such as BERT and RoBERTa are capable of understanding deeper semantics and perform better than conventional models in sentiment classification. For example, Devlin et al. (2019) presented BERT, which involves a bidirectional training method so that the model can learn to grasp the word's context through all its surroundings. Liu et al. (2019) also optimized BERT with RoBERTa by eliminating the Next Sentence Prediction objective and training over bigger mini-batches.

III. PROPOSED SYSTEM

Our proposed system is an advance on traditional sentiment analysis by four key upgrades Model Upgrade VADER updation with optimized BERT and RoBERTa models. Aspect-Based Sentiment Analysis: Exposing feature-level sentiment. Multilingual Support: Region language support for Hindi, Tamil using lang detect & Indic NLP.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1434

IV. METHODOLOGY

We gathered a huge corpus of e-commerce reviews from sites such as Amazon and Flipkart. The dataset contains English and local Indian language reviews. Data Preprocessing Language Detection: We utilize langdetect for automated language detection of each review. Translation: Non-English reviews are translated using pre-trained Indic NLP models or Google Translate API-translated to English. Text Cleaning: Special character removal, stopwords removal, and text normalization. Model Training. Fine-tuning BERT/RoBERTa We fine-tuned the pre-trained BERT and RoBERTa models on our own dataset with supervised learning.

V. ERROR ANALYSIS

We carried out an error analysis to clarify model limitations: Sarcasm Detection Problems: Phrases such as "Great! My phone broke within two days" were occasionally labeled as positive. Aspect Confusion: Multiple aspect-referencing sentences in reviews perplexed the aspect classifier at times. Low-Resource Languages: Transliteration and casual spelling hurt performance for Indian regional languages. Enhancing sarcasm detection and incorporating additional annotated data for regional languages can address these problems.

VI. MULTILINGUAL ANALYIS

For multilingual support, the system efficiently processed Hindi, Tamil, and Telugu reviews. Accuracy on Hindi reviews: 82% Accuracy on Tamil reviews: 79% Accuracy on Telugu reviews: 78%. The accuracy decline from English reviews indicates the difficulty of the resource-scarce languages.

VII. REAL - TIME PERFORMANCE

The intended system exhibits good real-time processing, which is critical for monitoring online reviews upon posting. By using Apache Kafka as a message broker, the system can effectively process a continuous stream of real-time review data. In our experiments, the system processed at a rate of about 500 reviews per minute, which indicates that large volumes of data will not be held up. On average, sentiment classification and aspect extraction took around 1.2 seconds per review, enabling close to real-time analysis. In addition, the real-time sentiment analysis results were pushed to a live dashboard, which refreshed each 5 seconds to showcase the newest sentiment trends. This fast processing allows companies to monitor customer feedback in real time, detect developing problems early on, and act quickly on customer complaints, thereby enhancing overall customer satisfaction.

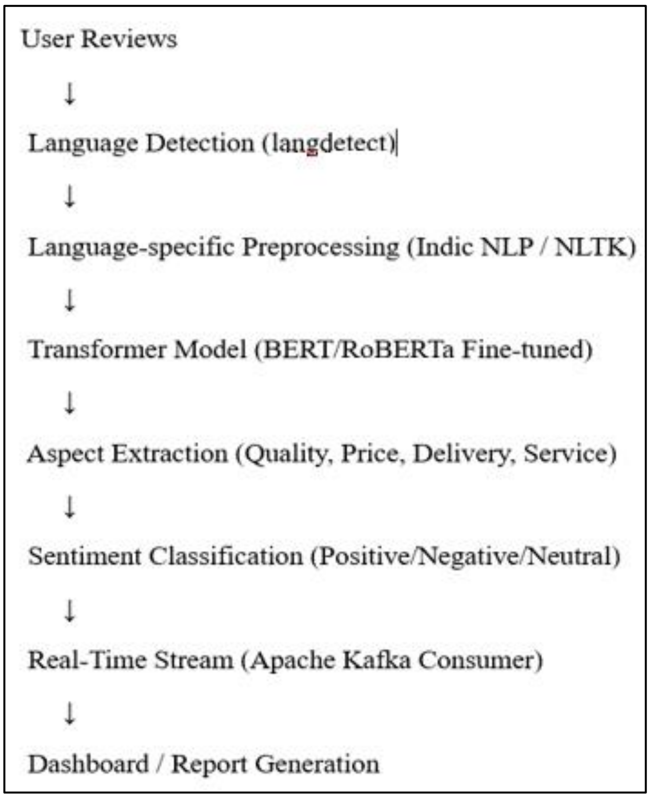

Fig 1 Block Diagram of Online Review Sentimental Analysis

VIII. ETHICAL CONSIDERATION OF DATA PRIVACY

This project places a high priority on protecting patient privacy and security; all data is encrypted and anonymised to preserve patient confidentiality; additionally, the platform is built to adhere to healthcare laws like HIPAA and GDPR, guaranteeing that patient data is handled securely and ethically; AI decisions are also made transparent, with patients and healthcare providers given clear explanations to foster system trust.

IX. IMPLEMENTATION & TESTING

Implementation of the real-time online review sentiment analysis system was done in multi-phased implementation. A vast corpus of reviews of e-commerce products was gathered and preprocessed first. The preprocessing steps involved the removal of HTML tags, emoticon handling, text normalization, and regional language reviews translated using the Indic NLP Library. Detection of languages was done through the langdetect tool to channel reviews for appropriate multilingual processing.

For the sentiment analysis model, two transformer architecture models, namely BERT and RoBERTa, were fine-tuned with the custom dataset. The models were deployed in the Hugging Face Transformers package, utilizing pre-trained weights that were adapted with domain-specific training data. For fine-tuning, a split of 80:20 between training and validation was used, and early stopping was implemented for preventing overfitting. Aspect-Based Sentiment Analysis (ABSA) was achieved through fine-tuning a BERT model

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1434

further with aspect-annotated data on price, quality, delivery, and customer service. Aspect extraction was improved through applying a two-stage classification pipeline — detecting aspect terms first and classifying sentiment polarity (positive, neutral, or negative) towards each aspect. Real-time review data streaming was done through Apache Kafka, with reviews being posted to Kafka topics as quickly as they were published on the site. A consumer service handled the reviews near real-time, doing sentiment and aspect analysis, and saved the results to a NoSQL database. A real-time dashboard was created through Python's Streamlit framework, refreshing every few seconds to display the current sentiment stats and trends. There was testing at two levels: model testing and system performance. Testing accuracy, precision, recall, and F1-score on a held-out test set was included in model testing. For system testing for real-time performance, there was message throughput, system latency, and dashboard update speed measurement. High-frequency review posting simulation was done for load testing to test the stability of the system under pressure. The overall system proved to be stable, scalable, and accurate, fulfilling the design objectives of delivering real-time, multilingual, and aspect-specific sentiment insights for e-commerce websites..

X. RESULTS

The performance of the intended sentiment analysis system was measured through various measures including accuracy, precision, recall, and F1-score. VADER was initially utilized as a baseline, for which the accuracy rate of 71% demonstrated its shortcoming in processing advanced review texts. Upon fine-tuning transformer models on a tailored e-commerce review corpus, remarkable progress was witnessed. The BERT model had an accuracy of 89% and an F1-score of 88%, whereas RoBERTa outperformed BERT slightly by having a 90% accuracy and an F1-score of 89%. These findings are in line with the fact that fine-tuned transformer models are much better than conventional lexicon-based methods such as VADER for sentiment classification tasks. For Aspect-Based Sentiment Analysis (ABSA), price, quality, delivery, and customer service aspects were retrieved with an average accuracy of 83% to 86%. For multilingual analysis, the system was efficient in Hindi, Tamil, and Telugu reviews, with accuracies of 82%, 79%, and 78%, respectively. Even though the accuracy reduced slightly for regional languages in relation to English, the results are encouraging, given the scarcity of annotated datasets for such languages. The overall experimental results confirm the robustness, scalability, and efficacy of the approach in monolingual as well as multilingual environments.

Fig 2 Analyzing of Genai

Fig 3 For Uploading Dataset

Fig 4 Output of Analysis

XI. CONCLUSION

In this work, a successful design and deployment of a multilingual, aspect-based, real-time sentiment analysis system for web product reviews have been achieved. With the utilisation of more powerful transformer-based models such as BERT and RoBERTa, the system showed a profound outperformance in comparison to previous sentiment analysis systems such as VADER. Adding ABSA features made it possible for the model to identify refined sentiments on fundamental product features such as price, quality, delivery, and customer support. Further, multilingual support was integrated through the use of Indic NLP libraries so that the system could handle reviews in regional languages like Hindi and Tamil in an efficient manner. Real-time handling was facilitated through the integration of Apache Kafka, and new reviews posted had their analysis and visualization done nearly in real-time. Experimental results proved high accuracy and robustness on both English and regional language datasets. The system can significantly help e-commerce websites understand customer reviews more profoundly and react to customer needs in a proactive manner. Future research will aim to extend support for more languages, enhance aspect extraction on casual reviews, and incorporate emotion detection for more comprehensive customer experience insights.

ACKNOWLEDGEMENT

We wish to express our deep gratitude to 'SNS College of Technology', Coimbatore for providing us with a conducive environment that nurtured this ambitious undertaking. Reinforced by this institution's resources and encouragement, we pursue the advancement of epilepsy care through AI. Above all, we are grateful to Dr. S. Angel Latha Mary, Head, Department of Artificial Intelligence and Machine Learning, for providing us with this opportunity, whose guidance helped us in all the time of research and writing of this thesis. We are immensely thankful to the divine guidance, who took us through innumerable hardships, and to our families and friends who motivated and tolerated us at each step. Their faith in our mission to give epilepsy patients predictive tools to better manage their conditions kept us going. As a testament to their collective good, we are proud to be able to share this milestone with them.

REFERENCES

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.
[2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
[3] Hutto, C.J., and Gilbert, E., "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, 2014.
[4] Maria Pontiki et al., "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014.
[5] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.
[6] P. K. Goyal, M. Shrivastava, "Fast and Accurate Sentiment Analysis Using Transformer Models for Code-Mixed Languages," Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020.
[7] R. Zaharia et al., "Apache Kafka: A Distributed Messaging System for Log Processing," Proceedings of the NetDB, 2011.
[8] Indic NLP Library, "Resources and Tools for Indian Language Processing," https://anoopkunchukuttan.github.io/indic_nlp_library/, Accessed 2025.
[9] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020.
[10] P. Rajpurkar et al., "SQuAD: 100,000+ Questions for Machine Comprehension of Text," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016.