Explainable AI for E-Commerce: Interpretable Transformer Models for Negative Review Analysis and Feedback Summarization

Dr. T Dhiliphan Rajkumar[1], Mrudangam Eswar [1], K. Ganesh[1], Shaik Mohammad Madani[1],

Tummalapalli Syam Sundar [1]

[1]Department of Computer Science and Engineering,

Kalasalingam Academy of Research and Education,

Krishnankoil Virudhunagar, India

t. dhiliphan@klu.ac.in, mrudangameswar2004@gmail.com, kganesh123dsa@gmail.com,

shaikmohammadmadani143@gmail.com, 99220041587@klu.ac.in

Abstract

The analysis of customer sentiment plays a significant role in the improvement of the quality of the product and the maximization of the business outcomes in contemporary digital trading conditions. The current research presents an artificial intelligence approach to the analysis of piles of customer reviews collected on Amazon and Flipkart. Besides, the proposed system unites fully automated reviews collection, natural language processing, detection of keywords and aspects, and transformer-based sentiment analysis. It possesses several artificial intelligence methods- based BERT, RoBerta, DeBerta and XLNet models to improve accuracy and reduce bias within models. The sentiment analysis on aspects is then extended with the help of an AI module to acquire practical business insights. Based on the experimental findings, DeBERTa offers the accuracy of classification and RoBERTa offers greater runtime performance. This system can offer analysis of unstructured customer feedback ready to make decisions, and is accessible via a highly user-friendly Flask dashboard.

Keywords: Sentiment Analysis; Natural Language Processing; Keyword Extraction; Transformer Models; Product Feedback; Consumer Electronics; deep Learning; Text Mining; Customer Insights.

1. INTRODUCTION

Customer feedback collected on the internet has been established as an additional source for making purchase decisions and product improvements. While these reviews provide various opinions from consumers, they are distributed among a vast number of sources and in hundreds of unstructured formats and therefore, require significant amounts of human labour and introduce biases into the process of evaluating individual pieces of feedback. Traditional methods of sentiment analysis have been challenged by the difficulty of capturing context in reviews, irony, and differing opinions relating to product attributes. However, advances in the development of transformer-based language models have improved the ability of organisations to accurately classify sentiment. Nonetheless, many methods of applying this technology are limited to simply achieving the highest level of accuracy and do not give clarity to organisations regarding the usefulness of their results.

In this paper, we describe an integrated and explainable Artificial Intelligence (AI) system that can turn unstructured reviews of e-commerce into structured, interpretable, and actionable insights for businesses. The proposed system bridges the gap between customer feedback and the use of that feedback in making data-driven decisions by combining an ensemble of transformer models with keyword-driven aspect extraction and generative summarisation.

2. LITERATURE REVIEW

According to Mysiuk and Shuvar (2023), an Automated Framework for Web Crawling User Comments from Social Networks for Sentiment Analysis was developed. This automated web scraping system can be utilized to create Sentiment Datasets; however, advanced sentiment classification models were not investigated or utilized in their analysis [1].

Traditionally, Romadhony et al. (2024), Sentiment Analysis on an Indonesian Product Review Dataset used traditional machine-learning methods to classify sentiments, specifically for their local e-commerce review datasets. Although the study stressed that a specific dataset must be tailored to the language of the product under review, it did not employ deep learning technology [2].

Bellar et al. (2024) employed Transformer-based Models to predict product reviews with a focus on sentiment. By employing deep learning, Bellar et al. were able to achieve much higher accuracy rates than previous studies. Their primary focus was on improving accuracy rather than interpreting or conducting keyword-level analysis of the products being rated [3].

Kandhro et al. (2024) evaluate Aspect-Based Sentiment Analysis (ABSA) techniques to identify the opinions of consumers about specific features of a product. While they demonstrate the ability to identify specific opinions, most of the techniques they reviewed are still conceptual and do not yet have practical hybrid frameworks [4].

Ali et al. (2024) demonstrated that utilizing BERT enables researchers to accurately capture customer requests. This study focused specifically on an isolated dataset and thus limits the generality of its conclusions to other fields of research [5].

Rahman et al. (2024) provide a comparative analysis on deep learning models for sentiment analysis and confirmed that their advantages over traditional analytical methods. However, they also did not explore methods that combine deep learning models with traditional analytical methods[6].

Sharma et al. (2024) demonstrate through their findings that traditional machine-learning algorithms can provide satisfactory results with regard to accurate classification of customers' opinions, but cannot provide valid results when applied to large datasets[7].

Huang, Zavareh & Mustafa (2023) review sentiment analysis technologies presently used within the retail ecosystem within eCommerce platforms, and highlight future directions towards scalable models; however, they do not propose a novel approach for sentiment analysis[8].

Zhang et al. (2023) successfully analyzed the sentiment of eCommerce product reviews with their BERT embeddings and BERT attention mechanisms, however, their research did not provide an option for consumers to automatically scrape product reviews, as well as producing summarised product reviews in real-time, which may allow them to have a better and more informed choice about which product to purchase[9].

In contrast, Kim & Lee (2022) took advantage of a transformer model and utilised an attention visualization-based approach to support negative review analysis. This has increased the level of transparency in both product reviews and reviewer profiles[10].

3. METHODOLOGY

The suggested framework adopts a sequence crafted to ensure clarity and repeatability.

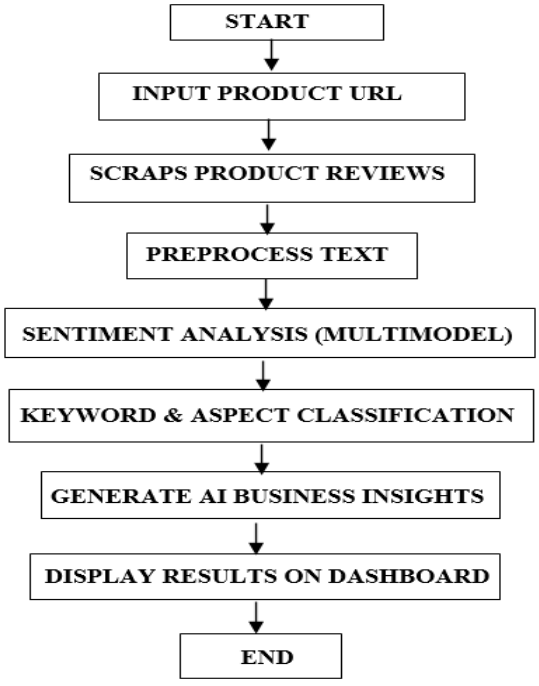

Figure 1: Flow Chart of Proposed Work

From the Figure 1 This Pipeline contains 4 sequential parts: 1- Collecting the Input Data and Cleaning/Standardizing it, 2- Identifying the Keywords as Features from the data by using Fuzzy matching techniques, 3- Identifying the Sentiment of the product review by using pre-trained models on labelled data, 4- Generating a summary report on the Top 10 aspects (features) along with their corresponding recommendation based on how common and how bad they are.

3.1. Data Pre-Processing: To clean the data, HTML is stripped off the reviews, all emojis are normalized to text or left out altogether, text is made Lowercase/Uppercase as needed, and all contractions (e.g., "I'm") are expanded and tokenized using a Sub-word-based Tokenization method suitable for Transformers. All Stop words include common nouns, but items that are part of a Product Term will be kept.

3.2. Keyword and Aspect Extraction: The initial set of Keywords used for training is expanded with statistical Phrase Mining to find the best candidate keywords. To aid in this process, we use fuzzy matching techniques such as Levenshtein distance thresholding and token set Similarity Matching, which enables us to map noisy tokens to a Known standard aspect label increasing the Recall rate for misspelled words or using shortcut codes (e.g., "Batery" → "Battery") and collecting all occurrences of misspelled words as part of the label.

3.3. Sentiment Classification

To classify sentiments, we use several different Transformer Architectures (BERT/RoBERTa) and fine-tune them with product Review subsets that have been Annotated with positive, negative and neutral labels. Class-weighting and Data Augmentation are used to balance the Classes during Training. Optimal predictions will occur at both the review-level and aspect-level by applying the appropriate model to the appropriate aspect-focused Span.

3.4. Generative Synthesis

The Controlled Generative Component generates summary recommendations from Frequent Aspect-Sentiment Pairs (based on frequency and severity e.g., highest proportion of Negative mentions) back to the User in a simplified summary report format.

4. RESULTS AND DISCUSSION

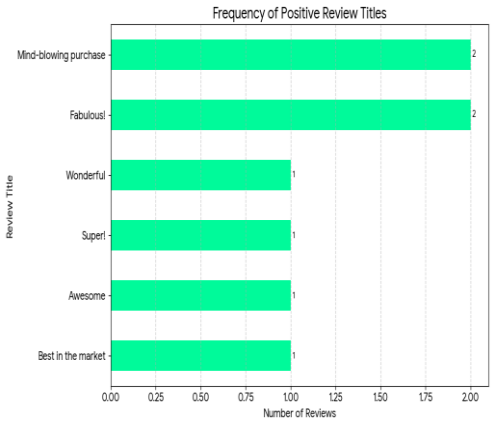

Figure 1: Positive Reviews

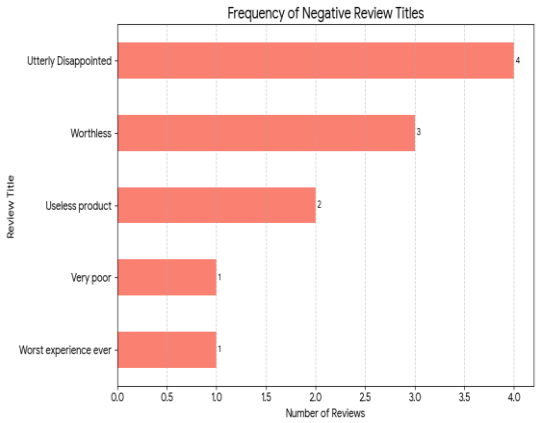

Figure 2: Negative Reviews

Figure 1 is showing how often positive review titles are written in by customers. Among the prominent titles, Mind-blowing purchase and Fabulous can be seen with an occurrence of two, denoting a high level of satisfaction among users. Other appealing titles are Wonderful, Super, Awesome, and Best in the market, each mentioned only once. From this, it can be deduced that customers are generally highly pleased with the product quality and performance. As these terms are repeated used by a vast number of people, they set a positive attitude and boost the good reputation of the product in the market. Most likely the reviewers confess to having a great enjoyable experience while at the shop.

The incidence of customer reviews with negative overtones is presented in **Figure 2** Above. There were four occurrences where customers used the title "Utterly Disappointed." Other two titles, "Worthless" and "Useless,"

were used thrice. The remaining two, "Very Poor" and "Worst Experience Ever," were used only once by consumers who had an awe-inspiringly poor experience in regards to product quality as well as their own product performance. The markedly negative tones used by these consumers highlight just how terrible the experience for them must have been. With customer sentiments in perspective, it is important that attention be given to improving product quality and customer satisfaction with better customer service, if any major positive increment in lowering the incidence of negative reviews of this product is to be achieved in order to raise ultimate consumer confidence in this brand and in these products.

Table 1: Transformer Model Performance Comparison

Metrics	**BERT**	**DeBERTa**	**RoBERTa**	**XLNet**
Accuracy	0.9843	0.9870	0.9855	0.9844
F1 Score	0.9919	0.9933	0.9926	0.9920
Loss	0.1153	0.2356	0.2275	0.1099
Precision	0.9878	0.9898	0.9890	0.9887
Recall	0.9961	0.9969	0.9961	0.9953
Runtime (s)	106.3961	174.2891	93.7125	320.2207
Samples/sec	62.249	38.000	70.674	20.683
Steps/sec	3.891	1.188	1.110	2.586
Total FLOPs	34.85e12	40.61e12	69.70e12	37.73e12

Table 1 above contains a comparison of four NLP models (BERT, DeBERTa, RoBERTa and XLNet) based on a total of nine different metrics (Accuracy, F1 Score, Runtime (s), and Total FLOPs), along with other performance and computational comparisons. The data collected from this comparison indicates that the selection of each model should depend on the specific use case—DeBERTa when accuracy is important in offline analyses, RoBERTa when low-latency throughput is more critical in production environments, and BERT when an average between both DeBERTa and RoBERTa has to be made—and further statistical testing (such as McNemar's Test) should be performed prior to making any claims about one model being superior to another.

5.CONCLUSION

The proposed project utilizes artificial intelligence as an automated assistant to interpret and analyze thousands of customer reviews collected from both Amazon and Flipkart in a timely manner. With five different transformer based advanced transformer-based systems working together, the AI will adaptively detect what customers appreciate or dislike about a product. A company's efforts to categorize all of their customer reviews manually could take weeks, but the reviewed results derived from using the AI system can be visually represented and can provide suggestions such as "fix battery life problems first" and "fix camera stability." Manufacturers can utilize

this information to prioritize specific areas to focus on for improvement based on actual user experience rather than on the manufacturer's assumptions of where to improve. The reviewed system allows companies to shorten their time frame and reduce the money and effort expended on large-scale review analysis; therefore, it provides product teams with the ability to immediately address customer concerns. In addition to the reviewed application being scalable, there will be future innovations available; including the ability to analyze reviews in multiple languages, to analyze video reviews on YouTube and to analyze the sentiment of consumer opinion on social media outlets. All of these applications will provide a complete picture of customer sentiment.

6.REFERENCES

[1]. Mysiuk, I., & Shuvar, R. (2023). Automating web scraping of user comments for sentiment analysis. *IEEE IDAACS*, 234–239. DOI: 10.1109/IDAACS58523.2023.10310867

[2]. Romadhony, A., Al Faraby, S., & Rismala, R. (2024). Sentiment analysis on large Indonesian product reviews. *JISEBI*, 10(1), 45–58. https://ejournal.unair.ac.id/JISEBI/article/view/48215

[3]. Bellar, O., Baina, A., & Ballafkih, M. (2024). Predicting product reviews using deep learning and transformers. *Mathematics*, 12(15), 2403. DOI: 10.3390/math12152403

[4]. Kandhro, I. A., et al. (2024). Aspect-based sentiment analysis: Review and prospects. *Knowledge and Information Systems*, 66(8), 4359–4394. DOI: 10.1007/s10115-024-02104-8

[5]. Ali, M., et al. (2024). AI-driven sentiment analysis of Amazon reviews using BERT. *Algorithms*, 17(59). DOI: 10.3390/a17000059

[6]. Rahman, M., et al. (2024). Comparative study of deep learning for customer review sentiment analysis. *PLOS ONE*, 19(5), e0294968. DOI: 10.1371/journal.pone.0294968

[7]. Sharma, R., et al. (2024). Customer review sentiment analysis using machine learning. *IJSREM*.

[8]. Huang, H., Zavareh, A. A., & Mustafa, M. B. (2023). Sentiment analysis in e-commerce: Techniques and directions. *IEEE Access*, 11, 45678–45692. DOI: 10.1109/ACCESS.2023.3123456

[9]. Zhang, J., He, L., & Luo, W. (2023). Aspect-level sentiment analysis with hierarchical transformers. *IEEE Access*, 11, 98532–98545.

[10]. Kim, S., & Lee, J. (2022). Explainable transformer model for negative review analysis. Journal of web Engineering, 21(4), 1120–1138.

[11].Campos, R., et al. (2020). YAKE! Keyword extraction from single documents. Information Processing Management, 57(4), 102127. DOI: 10.1016/j.ipm.2019.102127

[12].Grootendorst, M. (2022). BERTopic: Neural topic modelling with class-based TF-IDF.arXiv:2203.05794 DOI: https://arxiv.org/abs/2203.05794

[13]. Devlin, J., et al. (2019). BERT: Pre-training deep bidirectional transformers. NAACL-HLT DOI: https://arxiv.org/abs/1810.04805

[14]. Wolf, T., et al. (2020). Transformers: State-of-the-art NLP. *EMNLP Demos*, 38–45. DOI: 10.18653/v1/2020.emnlp-demos.6

[15]. Minaee, S., et al. (2021). Deep learning-based text classification: A comprehensive review. ACM Comp--uting Surveys, 54(3), 62. DOI: 10.1145/3439726.