Enhancing E-Commerce Recommender System Inputs Using Transformer-Based Aspect-Based Sentiment Analysis

Dr. Nazima Khanam, Dr Karen Robinson

Westcliff University, California, USA

ABSTRACT

Aspect-based sentiment analysis (ABSA) has become an important analytical technique in e-commerce research for understanding how customers perceive specific product features. Conventional sentiment analysis approaches typically treat a review as a single unit and assign an overall sentiment label, even though customers frequently express differing opinions about multiple product attributes within the same review. Such aggregation often obscures feature-level preferences and limits the usefulness of sentiment outputs for personalization and decision support. To address this limitation, this study proposes a transformer-based ABSA pipeline that integrates KeyBERT for unsupervised aspect extraction with Bidirectional Encoder Representations from Transformers (BERT) for contextual sentiment classification at the aspect level. The proposed approach is evaluated using a dataset of 10,000 Amazon product reviews obtained from publicly available open-source review data and is benchmarked against a widely used lexicon-based sentiment analysis method, Valence Aware Dictionary and Sentiment Reasoner (VADER). Model performance is assessed using precision, recall, and F1-score to capture both classification accuracy and balance across sentiment classes. Experimental results demonstrate that the transformer-based pipeline consistently outperforms the lexicon-based baseline, particularly in reviews containing mixed or contrasting sentiments across different product attributes. The findings show that contextual embeddings enable more accurate identification of sentiment polarity shifts and nuanced opinion expressions that are frequently missed by rule-based methods. Overall, the results indicate that transformer-based ABSA provides more reliable and interpretable sentiment representations, making it well suited for supporting personalized recommendations, feature-level analysis, and improved customer insight generation in e-commerce systems.

Keywords: Aspect-based sentiment analysis, transformer models, BERT, KeyBERT, e-commerce personalization, sentiment analysis

INTRODUCTION

Customers of e-commerce platforms encounter numerous products in the catalogs of e- commerce platforms. Recommender systems are integrated into e-commerce platforms to streamline this process. Current models incorporate structured data to provide recommendations through user ratings, purchase history, browsing history, among others. However, these methods do not seem to understand the user's sentiment regarding the product and its specific attributes, as they only capture user's general preferences [1], [2]. Similarly, these systems often recommend the same products to users if they have the same ratings, despite the products differing in attributes that users may find important.

Another data source that can be helpful in addressing the shortcomings of the structured data are customer reviews. Reviews are more likely to capture the user's sentiment through their detailed description of a user's experience. Reviews can also portray user sentiment through the varied attributes discussed in the review, such as quality, price, and durability, to name a few. Negative and positive sentiments can also be included in the same review. However, analyzing customer reviews not only through sentiment data, but through sentiment analysis methods that treat reviews as a singular unit by assigning distinct sentiment labels to the reviews, causes

important details to be lost. This could also result in losing mixed sentiments regarding the product, which could ultimately correlate with a user's preference. [3].

Aspect-based sentiment analysis (ABSA) was conceived with the goal of overcoming the limitations of traditional sentiment analysis, which is grossly inadequate as it treats a review in its entirety as a unit of analysis. By ascribing a sentiment to a review's different attributes, it becomes easier to unlock the complexity of customer feedback. Past work shows that aspect-level sentiment information is useful for personalization and for explaining the transparency of the recommendation system in a number of different application areas that employ multi-criteria evaluative techniques [4], [5].

More recently, the application of transformer-based models to sentiment analysis has been enabled by their state-of-the-art performance in a broad range of natural language processing tasks due to their ability to capture and represent the contextual meaning of a text at a high level. Bidirectional Encoder Representations from Transformers (BERT) is particularly illustrated as a strong candidate for sentiment classification because of its bidirectional left and right context representation in building its word vectors [6]. In sentiment analysis, as in many text processing tasks, keyword and phrase location is a quintessential means of aspect extraction. Unsupervised approaches such as KeyBERT are practical for large review datasets because they employ transformer embeddings to identify important keywords and phrases without prescriptive labeled training data [7].

Even though there are more advanced techniques, lexicon methods, such as Valence Aware Dictionary and Sentiment Reasoner (VADER), are still the most common due to their relative simplicity and low computational cost, and are thus appealing to practitioners developing large scale solutions [3]. VADER and other lexicon methods, on the other hand, handle the context, negation and contradictory views that are common in real-world e-commerce reviews, even with the simplicity of the methods. This brings to question whether other more complex techniques such as transformer-based ABSA are justifiably more complex and whether they are able to improve on the foundational sentiment understanding of the other methods on large datasets.

This are the issues in sentiment ABSA methodologies that the current study seeks to understand through the empirical comparison of ABSA methodologies. To this end, a transformer-based aspect sentiment modeling pipeline using 10,000 product reviews on Amazon obtained from the publicly available Amazon open-source review dataset is analyzed. This review explains the use of KeyBERT for aspect extraction and BERT for sentiment classification. This pipeline's VADER benchmark and benchmarked against VADER on the basis of precision, recall and F1 score. The aim of this work is not to build a complete recommender system, but to assess the underlying aspect-level sentiment data to provide personalization and interpretability to e-commerce systems.

This work has a few notable contributions. It first evaluates the performance of a transformer-based ABSA model on a massive data set of e-commerce reviews. It places second on the comparison of contextual transformer and lexicon-based baseline models on aspect-level sentiment. It finally helps determine how suitable transformer-based sentiment models would be for supporting personalization in e-commerce.

RELATED WORK

The present work is concerned with sentiment analysis in e-commerce applications; aspect-based sentiment analysis; and sentiment analysis using transformer models. These three branches of sentiment analysis research comprise the interdisciplinary field of sentiment analysis in e-commerce. Therefore, it is necessary to consider the work done in these branches before clearly describing the research gap that this work intends to fill.

Sentiment Analysis in E-Commerce

In e-commerce, the sentiment expressed by customers in reviews and feedback comments is analyzed to understand customer views. Early work in this field analyzed user sentiment at the document or sentence level, where an entire review or a single sentence was labeled as positive, negative, or neutral [9]. These techniques

have been relevant to tasks such as product ranking, reputation analysis, and customer satisfaction assessment. However, document-level sentiment analysis, while effective in revealing overall sentiment trends, provides limited insight into user preferences related to specific product features.

In recent years, some research has focused on combining sentiment scores with collaborative filtering and content-based recommendation approaches [10], [11]. These systems perform better than models built on ratings alone; however, they still operate using aggregated sentiment scores. As a result, they have difficulty capturing the nuanced and complex sentiments expressed in textual reviews. Such approaches often overlook positive or negative opinions related to specific product attributes when products share similar overall sentiment scores.

Aspect-Based Sentiment Analysis

Aspect-based sentiment analysis extends traditional sentiment analysis methods by evaluating individual attributes mentioned in text and associating sentiment with each attribute. This approach has been applied in domains such as hospitality, consumer electronics, and online retailing [12], [10]. Due to the finer-grained nature of ABSA systems, sentiment associated with each attribute is explicitly separated, making ABSA a strong candidate for supporting personalization and decision-making processes.

Prior research shows that aspect-level sentiment analysis helps identify which specific attributes trigger positive or negative sentiment and improves the interpretability of recommendation outputs [11]. In e-commerce settings, ABSA techniques have also been used to summarize customer reviews and determine overall positive and negative opinions about products, supporting feature-level comparison between competing items [10]. However, many early ABSA approaches rely on lexicon-based rules or supervised learning models that require large labeled datasets, which limits their scalability.

Transformer-Based Models for Sentiment Analysis

Recent developments in deep learning have encouraged the use of transformer-based models for sentiment analysis. Bidirectional Encoder Representations from Transformers (BERT)–based models have demonstrated strong performance across a range of sentiment analysis tasks due to their ability to capture contextual and semantic relationships within text [6]. Transformer-based approaches have been shown to outperform traditional machine learning and lexicon-based methods in multiple sentiment classification benchmarks.

To reduce the dependence on annotated datasets, unsupervised and weakly supervised techniques for aspect extraction have also gained attention. Methods such as KeyBERT use transformer embeddings to identify important keywords and phrases that represent potential aspects in text [7]. These techniques are particularly suitable for large review datasets, as they avoid the need for labor-intensive manual annotation. Despite their potential, gaps remain in the application of end-to-end transformer-based ABSA pipelines within large-scale e-commerce contexts.

Research Gap

Previous studies demonstrate the potential benefits of sentiment analysis and ABSA for e-commerce applications. However, several gaps remain in the existing literature. First, many studies rely on limited datasets, which restricts their ability to represent real-world scenarios. Second, while individual transformer models have shown strong performance, fewer studies evaluate complete ABSA pipelines in comparison with widely used lexicon-based methods. Finally, there is limited empirical evidence demonstrating whether the increased complexity of transformer-based ABSA systems leads to meaningful improvements over simpler, lexicon-based approaches when applied to large-scale datasets.

RESEARCH METHODOLOGY

This study employs an experimental research methodology to examine whether transformer-based aspect-level sentiment modeling offers measurable advantages over traditional lexicon-based sentiment analysis in the context of e-commerce reviews. The methodological framework is designed to ensure reproducibility, scalability, and fair comparison between models with fundamentally different underlying assumptions. By combining unsupervised aspect extraction with contextual sentiment classification, the proposed approach seeks to address limitations commonly observed in document-level sentiment analysis and to assess whether these improvements justify the additional computational complexity.

The methodology consists of dataset construction and preprocessing, aspect extraction, sentiment classification, baseline comparison, and evaluation using both categorical and probabilistic performance measures. Each stage of the pipeline is described in detail in the following subsections.

Research Question and Hypotheses

The research methodology is guided by a clearly defined research question that reflects the central objective of this study:

RQ:

How effectively does transformer-based aspect-level sentiment analysis improve sentiment classification outcomes compared to lexicon-based methods in the analysis of e-commerce product reviews?

This research question is motivated by the growing adoption of deep learning–based language models in sentiment analysis and the need to evaluate whether such models provide tangible benefits beyond established, low-cost lexicon-based techniques. While prior studies have demonstrated the effectiveness of transformer architectures in general sentiment classification tasks, fewer studies have rigorously compared complete ABSA pipelines against traditional baselines in realistic e-commerce settings.

Based on this research question, the following hypotheses are formulated:

Null Hypothesis (H_0):

There is no statistically significant difference in sentiment classification performance, measured by accuracy, precision, recall, and F1-score, between transformer-based aspect-level sentiment analysis and lexicon-based sentiment analysis methods when applied to e-commerce product reviews.

Alternative Hypothesis (H_1):

Transformer-based aspect-level sentiment analysis achieves significantly higher sentiment classification performance, measured by accuracy, precision, recall, and F1-score, than lexicon-based sentiment analysis methods when applied to e-commerce product reviews.

These hypotheses are evaluated through quantitative performance metrics, error pattern analysis, and qualitative examination of aspect-level sentiment outputs.

Dataset Description

The empirical evaluation was conducted using a dataset of **10,000 Amazon product reviews**, sourced from a publicly available open-source review corpus. Amazon reviews are widely used in sentiment analysis research due to their diversity, scale, and detailed user feedback. Reviews in the dataset span multiple product categories, including consumer electronics, apparel, and household goods, thereby providing a heterogeneous testing environment.

Each review consists of free-text feedback and an associated star rating. Consistent with prior sentiment analysis literature, sentiment labels were derived from the rating values to create a three-class sentiment framework [1], [2]. Reviews with ratings of 1 or 2 were labeled as *negative*, reviews with a rating of 3 were labeled as *neutral*, and reviews with ratings of 4 or 5 were labeled as *positive*. This labeling scheme reflects commonly adopted practices in e-commerce sentiment modeling and supports comparison with prior work.

The selected dataset size balances computational feasibility with sufficient sample diversity to observe meaningful performance differences between models.

Aspect Extraction Using KeyBERT

Aspect extraction constitutes a critical stage of the proposed methodology, as it determines which product attributes are subsequently evaluated for sentiment. In this study, aspect extraction was performed using KeyBERT, an unsupervised keyword extraction technique that leverages transformer-based embeddings to identify semantically salient terms and phrases [7].

KeyBERT operates by generating document-level embeddings and comparing them with candidate n-gram embeddings to identify terms that best represent the content of a review. This approach enables the extraction of meaningful aspects without requiring manually annotated training data. The use of an unsupervised method was intentionally selected to improve scalability and to avoid domain-specific labeling constraints that often limit supervised ABSA approaches.

Extracted aspects typically correspond to product attributes frequently discussed by customers, such as quality, price, shipping, durability, and customer service. These aspects serve as the basis for subsequent sentiment classification.

Sentiment Classification Using BERT

Sentiment classification was performed using Bidirectional Encoder Representations from Transformers (BERT), a transformer-based language model that has demonstrated state-of-the-art performance across a wide range of natural language processing tasks [6]. BERT's bidirectional attention mechanism allows it to capture contextual relationships between words, enabling more accurate interpretation of sentiment expressions that depend on surrounding context.

For each extracted aspect, contextual embeddings were generated by feeding the corresponding review text into the BERT model. Sentiment polarity was then assigned at the aspect level, allowing the model to capture opposing sentiments associated with different attributes within the same review. This aspect-level classification framework addresses a key limitation of document-level sentiment analysis, which assigns a single sentiment label to an entire review.

The ability of BERT to handle negation, concessive clauses, and sentiment shifts is particularly important in e-commerce reviews, where users often express both satisfaction and dissatisfaction in a single piece of feedback.

Baseline Method: VADER

To evaluate the effectiveness of the transformer-based pipeline, results were benchmarked against Valence Aware Dictionary and Sentiment Reasoner (VADER), a widely used lexicon-based sentiment analysis tool. VADER assigns sentiment scores based on a predefined dictionary of sentiment-laden words and a set of heuristic rules designed to handle punctuation, capitalization, degree modifiers, and negation [8].

VADER was selected as the baseline due to its popularity in applied sentiment analysis and its low computational requirements. Its rule-based design makes it particularly attractive for large-scale applications where computational resources are limited. However, the absence of contextual modeling limits VADER's ability to

capture nuanced sentiment expressions, making it an appropriate comparator for evaluating the benefits of transformer-based approaches.

Evaluation Metrics

Model performance was evaluated using precision, recall, and F1-score under both macro-averaged and weighted configurations. Macro-averaged metrics treat each sentiment class equally and are useful for assessing performance on minority classes, while weighted metrics account for class distribution and reflect overall classification effectiveness.

In addition to these aggregate metrics, confusion matrices were examined to analyze class-specific error patterns and misclassification tendencies. For the transformer-based model, probability-based evaluation metrics were also computed, including Kullback–Leibler (KL) divergence and cosine similarity. These measures provide insight into how closely predicted probability distributions align with ground truth labels and offer a more nuanced understanding of model confidence and calibration.

Algorithm 1: Transformer-Based Aspect Sentiment Analysis Pipeline

Input:

- Review dataset R={r1,r2,…,rn}

- Pre-trained BERT model

- KeyBERT model

- VADER sentiment lexicon

Output:

- Aspect–sentiment pairs for each review

- Performance metrics (Precision, Recall, F1-score)

Steps:

1. **Data Preparation**

   Load the Amazon open-source review dataset and select review text fields.

   Remove empty or extremely short reviews.

2. **Text Preprocessing**

   For each review:

   - Convert text to lowercase

   - Remove punctuation and non-alphanumeric characters

   - Remove stop words

   - Tokenize the cleaned text

3. **Aspect Extraction (KeyBERT)**

For each preprocessed review:

- Apply KeyBERT to extract a set of candidate aspects

- Store extracted keywords and key phrases as aspect candidates

4. **Aspect-Level Sentiment Classification (BERT)**

For each extracted aspect in review :

- Identify contextual text surrounding

- Input the context–aspect pair into the BERT sentiment classifier

- Assign a sentiment polarity label (positive, negative, or neutral)

5. **Baseline Sentiment Analysis (VADER)**

For each review:

- Apply VADER to compute a document-level sentiment score

- Assign sentiment polarity based on VADER thresholds

6. **Performance Evaluation**

- Compare transformer-based aspect-level sentiment predictions with VADER outputs

- Compute precision, recall, and F1-score for both approaches

7. **Result Aggregation**

- Aggregate aspect–sentiment pairs across all reviews

- Analyze performance differences between transformer-based and lexicon-based methods

End

Experimental Results and Analysis

This section presents the experimental evaluation of the proposed transformer ABSA approach and compares its performance with the lexicon VADER. The evaluation was conducted using 10,000 Amazon product reviews, and the results are analyzed using both quantitative performance metrics and qualitative aspect-level interpretations.

Dataset Characteristics

The distribution of sentiment classes in the dataset is summarized in Table 1. Among the 10,000 reviews, 4,076 reviews (40.8%) are labeled as negative, 3,972 reviews (39.7%) as positive, and 1,952 reviews (19.5%) as neutral. This distribution indicates that the dataset is relatively balanced between positive and negative sentiment, with a smaller but meaningful proportion of neutral reviews.

Table 1. Distribution of Reviews by Sentiment Class

Sentiment Class	Frequency	Percentage
Negative	4,076	40.8%
Positive	3,972	39.7%
Neutral	1,952	19.5%

The same distribution is visually illustrated in Figure 1, which confirms that no single sentiment class dominates the dataset. This balance supports the use of macro-averaged evaluation metrics and reduces bias toward any one class.

Figure 1. Sentiment class distribution of the Amazon review dataset (N = 10,000)
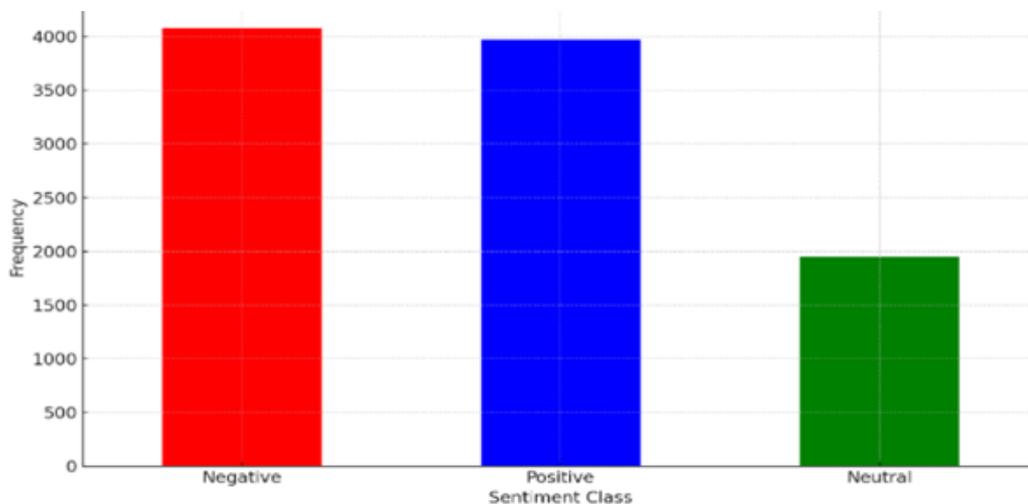
In addition to sentiment balance, the dataset was analyzed for review length variability. Table 2 reports descriptive statistics of review length in terms of word count. Reviews range from very short comments to longer, detailed narratives, with a mean length of approximately 75 words and a standard deviation of 43 words. This variability increases the likelihood that individual reviews contain multiple aspects and mixed sentiments.

Table 2. Descriptive statistics of review length

Statistic	Value
Mean	75 words
Std. Dev	43 words
Minimum	8 words

25th %ile	39 words
Median	67 words
75th %ile	105 words
Maximum	198 words

Aspect Extraction Results

Aspect extraction was performed using KeyBERT, and representative examples are presented in Table 3. The extracted aspects correspond to semantically meaningful product attributes such as *battery*, *screen*, *delivery*, *price*, and *quality*. These results demonstrate that the unsupervised extraction approach successfully identifies relevant attributes without requiring annotated training data.

Table 3. Examples of aspect extraction using KeyBERT

Original Review	Cleaned Review	Extracted Aspects
"The screen and camera is bright but the battery life is too short"	"screen camera bright battery life short"	[screen, camera, battery]
"Fair looking dress design, but the shipping took too long"	"fair dress design, long shipping"	[design, shipping]
"Price is reasonable and quality is excellent"	"price reasonable, quality excellent"	[price, quality]

The extracted aspects form the basis for aspect-level sentiment classification and enable the model to analyze opinions associated with individual product features rather than treating the review as a single sentiment unit.

Aspect-Level Sentiment Classification

Aspect-level sentiment outputs generated using the BERT model are illustrated in Table 4. These examples show how sentiment polarity is assigned independently to each extracted aspect. In reviews where users express both positive and negative opinions, the transformer-based model correctly identifies opposing sentiments associated with different attributes.

Table 4. Examples of aspect-level sentiment classification using BERT

Cleaned Review	Extracted Aspects	Aspect Sentiments	Why BERT is Superior
"screen camera bright battery life short"	[screen, camera, battery]	{screen: Positive, camera: Positive battery: Negative}	VADER only gives Neutral; BERT distinguishes both positive and negative aspects.
"fair dress design, long shipping"	[design, shipping]	{design: Positive, shipping: Negative}	Lexicon-based VADER misses the contrast introduced by "but"; BERT captures opposing sentiments.
"price reasonable, quality excellent"	[price, quality]	{price: Positive, quality: Positive}	VADER interprets "reasonable price" weakly; BERT assigns positive polarity to both attributes.
"dress beautiful though stitching loose"	[dress, stitching]	{dress: Positive, stitching: Negative}	VADER is swayed by "beautiful" and ignores stitching flaw; BERT captures mixed sentiment.
"necklace shiny but clasp broke"	[necklace, clasp]	{necklace: Positive, clasp: Negative} {service: Negative, delivery: Positive}	Both return Neutral overall, but only BERT exposes the flaw in clasp at aspect-level.
"customer service rude yet delivery fast"	[service, delivery]		VADER focuses on "rude" and misses concession in "yet"; BERT separates

These results highlight a key advantage of ABSA: the ability to preserve sentiment granularity. In contrast to document-level sentiment analysis, which collapses multiple opinions into a single label, aspect-level classification provides a more faithful representation of customer feedback.

Quantitative Performance Comparison

The quantitative performance of the transformer-based ABSA model and the VADER baseline is summarized in Table 5, which reports precision, recall, and F1-score using both macro-averaged and weighted metrics.

Table 5. Performance comparison of VADER and BERT-based ABSA

| Model	Precision(Macro)	Recall (Macro)	F1 (Macro)	Precision (Weighted)	Recall (Weighted)	F1 (Weighted)
VADER	0.455	0.437	0.394	0.515	0.532	0.471
BERT	0.570	0.502	0.491	0.638	0.512	0.531

Across all reported metrics, the transformer-based model outperforms the VADER baseline. The improvement is particularly notable in F1-score, indicating a better balance between precision and recall. While VADER performs reasonably well in detecting strongly polarized sentiment, its overall performance is lower due to its limited ability to handle contextual and mixed sentiment expressions.

Confusion Matrix Analysis

To further examine classification behavior, confusion matrices were generated for both models.

Figure 2 presents the confusion matrix for VADER. The matrix reveals a strong tendency to classify reviews as positive, including a substantial number of negative and neutral reviews incorrectly predicted as positive. This bias reflects the limitations of lexicon-based approaches when applied to complex, real-world review text.

Figure 2. Confusion matrix for VADER sentiment classification
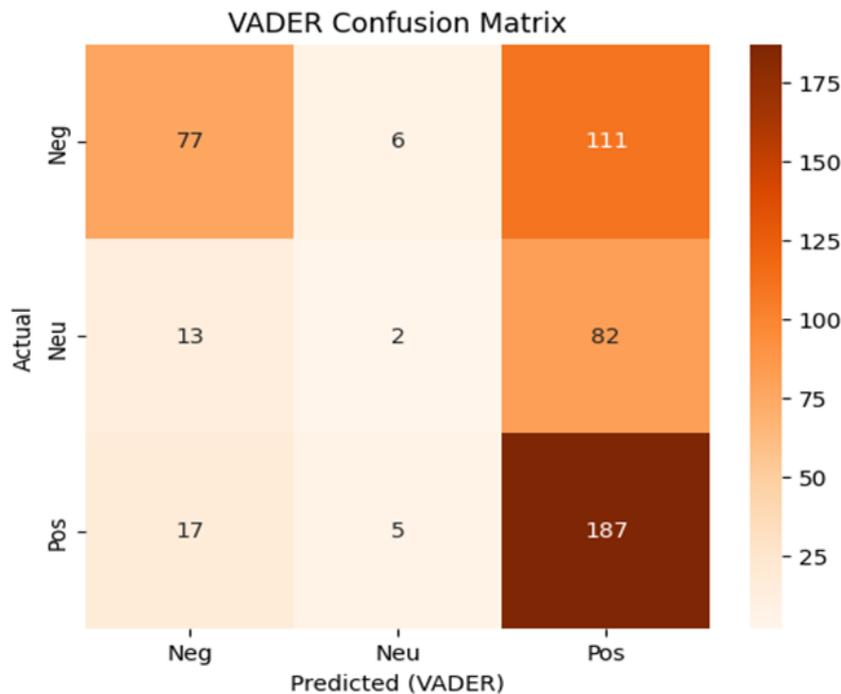
In contrast, Figure 3 shows the confusion matrix for the BERT-based sentiment classifier. The transformer-based model demonstrates improved discrimination across all three sentiment classes, particularly for neutral reviews, and exhibits a more balanced distribution of classification errors.

Figure 3. Confusion matrix for BERT-based sentiment classification
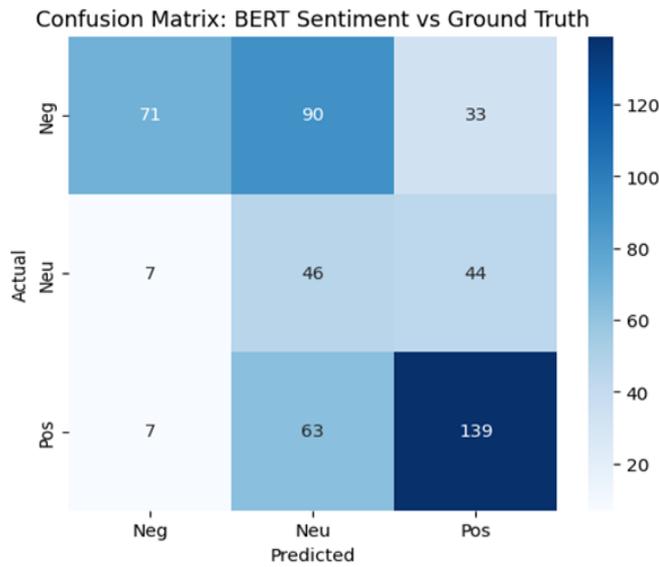
Distributional Evaluation of Transformer Predictions

Beyond categorical accuracy metrics, distributional evaluation metrics were computed to assess the quality and stability of the BERT model's probabilistic predictions.

Figure 4 illustrates the distribution of Kullback–Leibler (KL) divergence values between predicted sentiment probability distributions and ground-truth labels. Most values are concentrated at lower divergence levels, with a median value of approximately 0.75, indicating reasonable alignment between predictions and true labels.

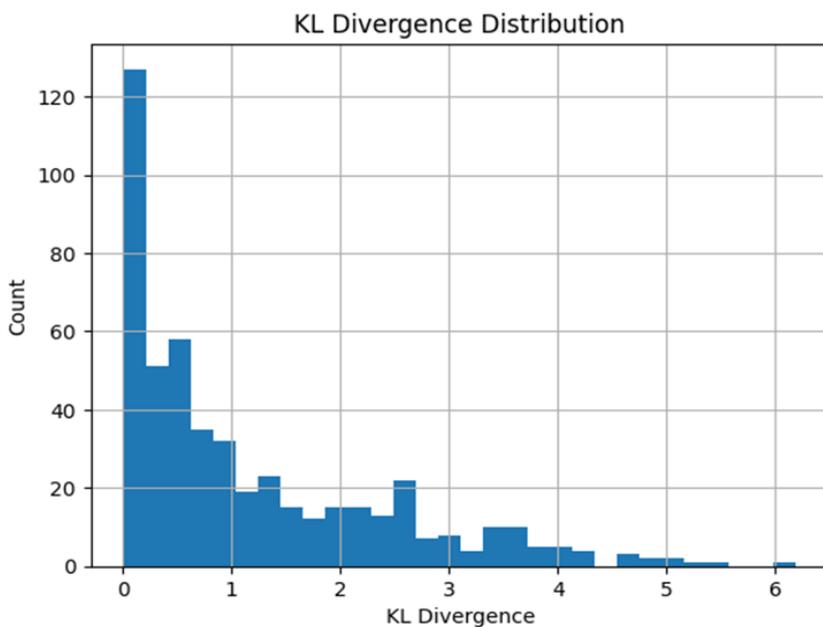**Figure 4. KL divergence distribution for BERT predictions**

Figure 5 presents the cosine similarity distribution between predicted and true sentiment vectors. The concentration of similarity values close to 1.0 indicates strong directional agreement and suggests that the model produces stable and well-calibrated probability outputs.

Figure 5. Cosine similarity distribution for BERT predictions

Cosine Similarity Distribution
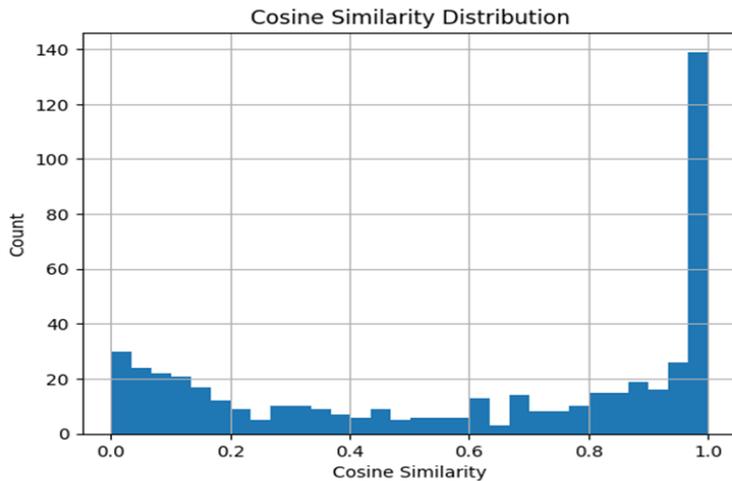
Summary of Results

Overall, the experimental results demonstrate that the transformer-based ABSA approach consistently outperforms the lexicon-based VADER baseline across quantitative metrics, error analysis, and qualitative interpretation. The findings show that contextual modeling and aspect-level sentiment classification provide more accurate and informative representations of customer sentiment, particularly in reviews containing multiple or conflicting opinions.

Interpretation and Implications

This section interprets the experimental findings and discusses their practical relevance, and broader implications for sentiment analysis and personalization in e-commerce systems

Practical Implications

From a practical standpoint, the results demonstrate that lexicon-based sentiment analysis tools such as VADER may oversimplify complex customer feedback. When reviews contain multiple opinions about different product features, VADER assigns a single aggregated sentiment label, which can distort representations of customer perceptions and obscure actionable insights.

In contrast, the transformer-based ABSA pipeline disaggregates sentiment at the feature level, enabling more precise and actionable interpretations. This capability supports data-driven decision-making in product development, targeted marketing, and recommendation personalization. By reducing reliance on coarse-grained sentiment scores, organizations can better align product offerings with specific customer preferences and sources of dissatisfaction.

Aspect-Level Utility: Illustrative Use Cases

The practical utility of aspect-based sentiment analysis is illustrated through representative examples drawn from multiple product categories:

- **Electronics (Smartphone):**

     Review: *"The camera takes stunning photos, but the battery drains within hours."*

     Aspect-Level Output: Camera = Positive; Battery = Negative

Implication: The product can be recommended to photography-oriented users, while battery performance can be flagged for engineering improvement.

- **Clothing (Dress):**

  Review: *"The fabric feels premium and elegant, although the stitching came loose after one wash."*

  Aspect-Level Output: Fabric = Positive; Stitching = Negative

  Implication: Retailers can emphasize fabric quality while addressing stitching issues through supplier quality control.

- **Jewelry (Watch):**

  Review: *"The design is gorgeous and looks expensive, but the strap feels cheap."*

  Aspect-Level Output: Design = Positive; Strap = Negative

  Implication: Products can be positioned toward design-conscious customers while improving component durability.

- **Appliances (Vacuum Cleaner):**

  Review: *"It cleans carpets thoroughly, but the noise is unbearable."*

  Aspect-Level Output: Cleaning Performance = Positive; Noise = Negative

  Implication: The product can be marketed to users prioritizing cleaning efficiency over noise sensitivity.

These examples demonstrate how ABSA enables feature-driven recommendations rather than generalized sentiment-based selection.

Future Implications

The findings suggest that transformer-based ABSA pipelines can be effectively integrated into recommender systems, product feedback loops, and customer support triage processes. Prior studies emphasize that aspect-level sentiment analysis improves not only classification accuracy but also interpretability, which is increasingly important for the adoption of AI-driven systems in business environments [9].

The improved performance observed in this study supports broader application of transformer-based sentiment models in domains requiring nuanced decision-making and transparent analytical outputs.

Recommendations for Future Research

Future research should focus on cross-domain validation of ABSA pipelines, extension to multilingual datasets, and exploration of hybrid models that combine lexicon-based features with transformer embeddings. Additionally, integrating ABSA-driven insights into live recommender systems and examining human–AI collaboration in decision-making contexts represent promising avenues for further investigation.

CONCLUSION

This study demonstrates that transformer-based ABSA approaches, specifically the integration of KeyBERT and BERT, provide clear performance and interpretability advantages over lexicon-based sentiment analysis methods. The findings confirm that the additional model complexity yields meaningful improvements in

sentiment classification for multi-opinion e-commerce reviews and directly supports feature-level personalization strategies.

REFERENCES

1. Babaali, M., Fatemi, A., & Nematbakhsh, M. A. (2024). Aspect extraction with enriching word representation and post-processing rules. Expert Systems with Applications, 240, Article 120304. https://doi.org/10.1016/j.eswa.2024.124174

2. Bellar, O., Baina, A., & Ballafkih, M. (2024). Sentiment analysis: Predicting product reviews for e-commerce recommendations using deep learning and transformers. Mathematics, 12(15), Article 2403. https://doi.org/10.3390/math12152403

3. Cai, H., Xie, Q., Zhao, Q., & Li, K. (2023). Memd-absa: A multi-element multi-domain dataset for aspect-based sentiment analysis. Language Resources and Evaluation, 59(3), 2501–2529. https://doi.org/10.1007/s10579-025-09820-9

4. Chauhan, G. S., Nahta, R., & Meena, Y. K. (2023). Aspect-based sentiment analysis using deep learning approaches: A survey. Computer Science Review, 48, Article 100505. https://doi.org/10.1016/j.cosrev.2023.100576

5. Cui, Y., Zhou, P., Yu, H., Sun, P., Cao, H., & Yang, P. (2024). ASKAT: Aspect sentiment knowledge graph attention network for recommendation. Electronics, 13(1), Article 216. https://doi.org/10.3390/electronics13010216

6. Darraz, N., Karabila, I., El-Ansari, A., Alami, N., & El Mallahi, M. (2025). Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems. Expert Systems with Applications, 261, Article 125533. https://doi.org/10.1016/j.eswa.2024.125533

7. Davoodi, L., Mezei, J., & Heikkilä, M. (2025). Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms. Electronic Commerce Research, 1–43. https://doi.org/10.1007/s10660-025-09948-4

8. Dogra, V. (2024, July). Aspect-based approaches for measuring customer feedback in the e-commerce industry. In Proceedings of the 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS) (pp. 479–484). IEEE. https://doi.org/10.1109/ICSCSS60660.2024.10625381

9. Elzeheiry, S., Gab-Allah, W. A., & Mekky, N. (2023). Sentiment analysis for e-commerce product reviews: Current trends and future directions. Preprints. https://doi.org/10.20944/preprints202305.1649.v1

10. Haq, B., Daudpota, S. M., Imran, A. S., Kastrati, Z., & Noor, W. (2023). A semi-supervised approach for aspect category detection and aspect term extraction from opinionated text. Computers, Materials & Continua, 77(1), 115–137. https://doi.org/10.32604/cmc.2023.040638

11. Xu, Y., & Ibrahim, N. F. (2022). Cross-domain aspect-based sentiment analysis for enhancing customer experience in electronic commerce. Advances in Artificial Intelligence and Machine Learning, 2(1), 112–127. https://doi.org/10.54364/aaiml.2024.43151

12. Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., & Li, Q. (2024). Recommender systems in the era of large language models. IEEE Transactions on Data Engineering, 36, 6889-6907. https://doi.org/10.1109/tkde.2024.3392335