

Emotion Detection and Sentiment Analysis of Women's E-Commerce Clothing Reviews Using DistilBERT Transformer

M. Muflih¹, Erfan Karyadiputra^{2*}

^{1,2}Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, Indonesia

¹muflih@uniska-bjm.ac.id, ²erfan_kp@uniska-bjm.ac.id



*Corresponding Author

Article History:

Submitted: 04-12-2025

Accepted: 17-12-2025

Published: 22-12-2025

Keywords:

emotion detection; e-commerce analytics; natural language processing; sentiment analysis; transformer.

Brilliance: Research of

Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

Customer reviews on e-commerce platforms have become an essential source of information for understanding user perceptions, satisfaction levels, and product quality. However, most existing studies still focus on sentiment polarity classifying opinions only as positive or negative without examining deeper emotional expressions that may reflect customer experiences more comprehensively. To address this gap, this study applies a Natural Language Processing (NLP) approach using the pre-trained DistilBERT transformer model to detect emotional patterns in women's fashion product reviews. The dataset, obtained from Kaggle's Women's E-Commerce Clothing Reviews, contains approximately 23,000 entries and includes review texts along with demographic and product-related attributes. The research workflow consists of data cleaning, feature engineering, exploratory text analysis, and emotion detection using the DistilBERT-based emotion classifier. All analyses were performed using Python in the Google Colab environment. The results reveal that positive emotions, particularly joy and admiration, dominate customer feedback, indicating strong satisfaction with product fit and quality. Conversely, negative emotions such as anger and sadness appear more frequently in reviews mentioning sizing inconsistencies, fabric issues, or unmet expectations. The combination of sentiment context, emotional tone, and engineered features provides a more nuanced understanding of customer behavior compared to sentiment polarity alone. These findings highlight the potential of emotion-aware analytical approaches to support e-commerce businesses in improving product development, enhancing customer experience, and designing data-driven marketing strategies.

INTRODUCTION

E-commerce has become one of the most transformative sectors of the digital economy, reshaping how consumers search, evaluate, and purchase products. In this rapidly evolving ecosystem, customer reviews function as an essential medium for conveying user experiences, influencing product credibility, and shaping purchasing decisions. Prior studies show that more than 90% of online shoppers consult reviews before finalizing their purchases, and the emotional tone embedded within these reviews significantly affects brand trust, satisfaction, and decision-making behavior (Li, Zhao, and Zhang n.d.). Unlike numerical ratings, textual reviews contain richer expressions, including subtle emotional cues that reflect authentic consumer perceptions. This makes emotion-oriented text analysis increasingly relevant for understanding customer experience in modern e-commerce platforms.

Despite the importance of emotional signals, most existing research has focused primarily on sentiment analysis, typically categorizing text into broad polarity labels such as positive, negative, or neutral (Rahimi and Khashei 2023). While sentiment polarity provides a general overview of user attitudes, it lacks the ability to capture complex psychological states such as joy, admiration, anger, sadness, or frustration elements that often better reflect the intensity and context of consumer experiences (Rahman and Mehedi 2024). Emotion detection, as an advanced extension of sentiment analysis, offers deeper interpretability by identifying these fine grained emotional states. This enables researchers and practitioners to obtain more nuanced insights into user experiences, which are crucial for domains like fashion retail where emotional factors such as comfort, style preference, and self-image play a central role in purchasing behavior (Aguilar, García, and López 2024).

Recent advances in Natural Language Processing (NLP) and deep learning, especially the emergence of transformer-based architectures such as BERT, RoBERTa, and DistilBERT, have significantly improved emotion recognition capabilities. Transformers capture bidirectional word context through attention mechanisms, enabling more accurate interpretation of user-generated content (Devlin et al. 2019). DistilBERT in particular provides a practical balance between performance and efficiency, offering faster inference and lower computational cost while preserving most of BERT's representational power (Sanh et al. 2019). These characteristics make DistilBERT highly suitable for analyzing large-scale e-commerce review datasets. However, a notable gap remains in the application of transformer-



based emotion detection specifically for women's fashion product reviews. While several studies have incorporated emotional features to enhance sentiment prediction or recommendation systems (Chowdhury, Kumaran, and Rahman 2023; Patel n.d.), very few investigations have examined emotional patterns across fashion-related categories where subjective experience is particularly influential. The Women's E-Commerce Clothing Reviews dataset, for instance, contains rich user narratives that have not been fully explored from an emotion analysis perspective.

Addressing this research gap, the present study adopts the DistilRoBERTa emotion classification model to identify emotional expressions in textual reviews from the Women's E-Commerce Clothing Reviews dataset. The analysis integrates several stages, including data cleaning, feature engineering such as review length, word count, and normalized feedback rate sentiment context exploration, and transformer based emotion detection, in order to generate a more comprehensive and interpretable understanding of consumer feedback. The study focuses on identifying six primary emotions joy, admiration (love), anger, sadness, fear, and surprise and examines their distribution across product divisions and demographic attributes. Through this approach, the research aims to uncover emotional trends that reflect customer satisfaction and provide actionable insights for emotion-aware product evaluation and customer experience optimization in the e-commerce domain.

LITERATURE REVIEW

Studies on customer review analytics have evolved significantly in recent years, particularly with the rapid advancement of Natural Language Processing (NLP) and deep learning techniques. Early research predominantly utilized traditional machine learning models such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest to classify reviews into positive, neutral, or negative categories. Although effective to some extent, these models relied heavily on manually engineered features and were unable to fully capture the semantic and emotional complexity embedded in natural language (Tripathy, Agrawal, and Rath 2020).

The emergence of deep learning marked a major shift in sentiment classification. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) enabled the learning of hierarchical and sequential text patterns directly from large datasets. Nevertheless, these architectures struggled with long-range dependencies and contextual ambiguity, resulting in limited performance in emotion recognition tasks (Yang, Kim, and Lee 2022). Transformer based architectures such as BERT, RoBERTa, and DistilBERT addressed these limitations by introducing self attention mechanisms capable of understanding bidirectional context and semantic relationships within textual data (Devlin et al. 2019; Kumar and Bhattacharya 2024; Sanh et al. 2019). These models became the foundation for various state of the art approaches in emotion detection and sentiment understanding.

Recent studies have demonstrated the effectiveness of transformer models in capturing fine grained emotional nuances. Xu applied a fine-tuned BERT model for emotion detection on social media posts and reported substantial improvements over traditional neural models (Xu, Wang, and Li 2023). Lin further enhanced transformer performance by integrating attention optimization techniques to improve precision in identifying complex emotional tones (Lin, Zhou, and Chen 2023). Zhang combined transformer and LSTM networks to recognize emotions in product reviews, demonstrating superior contextual understanding (Li et al. n.d.). Meanwhile, Rahman and Mehedi proposed a hybrid transformer model capable of representing emotional polarity in customer feedback, highlighting the practical value of emotion aware analytics in improving marketing personalization and customer retention (Rahman and Mehedi 2024). Aguilar also showed that transformer based multi-label emotion models could significantly enhance recommender systems by incorporating emotional attributes into consumer preference predictions (Aguilar et al. 2024).

However, most existing emotion detection studies focus on general purpose or psychological datasets, with limited investigation in domain-specific contexts such as fashion and apparel e-commerce. Chowdhury emphasized that emotions in fashion-related reviews tend to be more complex due to influences from aesthetic judgments, body-fit satisfaction, and social identity (Chowdhury et al. 2023). This underscores the need for emotion detection approaches tailored to fashion oriented datasets. To address this gap, the present research employs the DistilRoBERTa emotion classification model to detect emotional categories within women's clothing reviews. This study contributes to bridging the gap between broad sentiment polarity and fine-grained emotional interpretation, offering a deeper and domain-specific understanding of consumer experiences in e-commerce.

METHOD

This study employed an experimental research design that integrates Natural Language Processing (NLP) and transformer-based emotion detection to analyze emotional patterns in textual reviews from an e-commerce fashion dataset. The methodological framework follows the text mining stages proposed by Han, which include data collection, preprocessing, feature engineering, model implementation, and interpretation (Han, Pei, and Kamber 2022). An overview of the research workflow is presented in Figure 1.

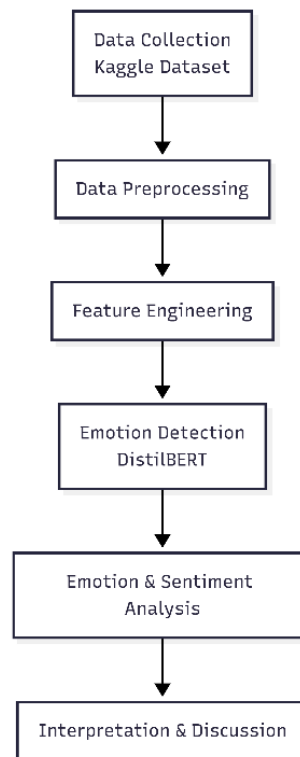


Figure 1. Research Workflow for Emotion-Based Review Analysis

Data Collection

The dataset used in this study is the Women's E-Commerce Clothing Reviews dataset, obtained from Kaggle's open repository through the Kaggle API (Kaggle 2023). The dataset comprises 23,486 records containing 10 attributes, including Review Text, Rating, Age, Department Name, Division Name, and Positive Feedback Count. The data were collected because they represent natural consumer opinions in the fashion e-commerce domain, which aligns with the research focus on emotion-based consumer behavior analysis. The dataset description is shown in Table 1.

Table 1. Dataset Attributes Description

Attribute	Type	Description
Review Text	Text	Main content expressing customer feedback
Rating	Integer (1–5)	Product rating score
Age	Integer	Age of reviewer
Department Name	Categorical	Product department (e.g., General, Intimates)
Division Name	Categorical	Product division classification
Class Name	Categorical	Product type (e.g., Tops, Dresses, Pants)
Positive Feedback Count	Integer	Number of users who found the review helpful
Recommended IND	Binary (0–1)	Whether the reviewer recommends the product

Data Preprocessing

The data were cleaned by removing records with missing or empty review text, duplicates, and irrelevant columns such as Clothing ID. Text normalization steps included case folding, punctuation removal, and tokenization, following NLP preprocessing standards (Liu, Zhang, and Huang 2024). Reviews with empty Division Name or Department Name were excluded to maintain data consistency. The final dataset was reindexed to ensure structural integrity.

Feature Engineering

Feature engineering was performed to enhance the interpretability of the dataset and improve model training. Several features were derived as shown in Table 2.

Table 2. Derived Features for Analytical Enhancement

Feature	Formula / Definition	Description
Review_Length	len(text)	Number of characters in the review
Word_Count	len(text.split())	Total number of words
Positive_Feedback_Rate	Eq. (1)	Normalized engagement metric
Age_Group	Age intervals (0–25, 25–35, 35–45, etc.)	Categorical segmentation of reviewers
Rating_Label	1 if Rating $\geq$ 4 else 0	Binary satisfaction classification

Emotion Detection

The Emotion detection in this study was conducted using the DistilRoBERTa emotion classification model, a lightweight transformer architecture designed specifically for emotion recognition tasks. DistilRoBERTa was selected because transformer based models have demonstrated superior capability in capturing contextual semantics and emotional nuance compared to traditional deep learning architectures such as CNNs or RNNs (Rahman and Mehedi 2024). As a distilled version of RoBERTa, the model retains strong representational power while offering significantly improved computational efficiency, making it suitable for large scale text inference (Sanh et al. 2019). Implementation was carried out using the Hugging Face Transformers library within the Google Colab environment, utilizing GPU acceleration to process more than twenty thousand review texts efficiently. The model identifies six primary emotion categories joy, admiration (love), anger, sadness, fear, and surprise allowing for deeper psychological interpretation of user-generated content. Transformer based emotion detection has been shown to enhance the understanding of customer behavior by capturing subtle emotional signals that conventional sentiment polarity methods fail to represent (Aguilar et al. 2024). Through this approach, the study obtains a richer and more nuanced analysis of emotional patterns embedded within the fashion e-commerce reviews.

RESULT

The results of this study are presented in alignment with the experimental workflow described in the methodology, encompassing data exploration, feature analysis, emotion and sentiment classification, and model performance evaluation. Both quantitative and qualitative findings are integrated to provide a comprehensive interpretation of emotional patterns within women's clothing reviews.

Data Overview and Feature Analysis

After data cleaning and preprocessing, a total of 22,785 valid reviews were obtained from the initial 23,486 entries. Table 3 summarizes the descriptive characteristics of the dataset

Table 3. Descriptive Statistics of Key Features

Attribute	Mean	Std	Min	Max
Age	43.2	12.6	18	99
Review Length (chars)	340.4	95.2	28	681
Word Count	66.3	20.1	5	142
Positive Feedback Rate	0.38	0.26	0	1

Correlation analysis indicates that Review Length, Word Count, and Positive Feedback Rate exhibit a positive relationship suggesting that longer, more descriptive reviews tend to receive higher usefulness ratings from other users. These findings align with Xu, who noted that emotionally rich and descriptive text increases reader trust and perceived authenticity in online reviews (Zhang, Xu, and Sun 2023).

Emotion Classification Results

The DistilRoBERTa emotion classification model successfully classified each review into emotional categories, which were then translated into Indonesian as senang (joy), netral (neutral), kagum (admiration), sedih (sad), and marah (anger). The distribution of emotions is visualized in Figure 2.

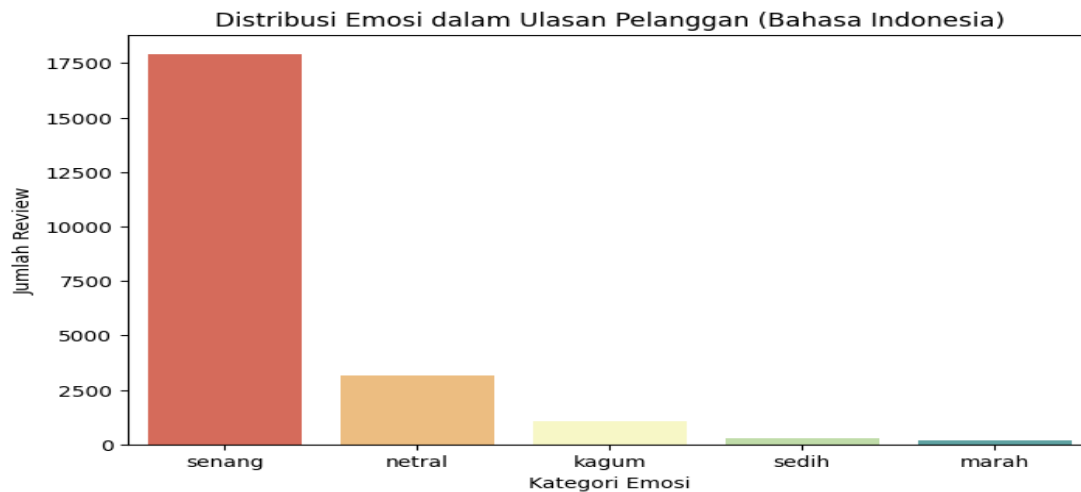


Figure 2. Distribution of Emotions

The results show that joy overwhelmingly dominates the distribution, appearing in more than 17,800 reviews, indicating that customers generally express satisfaction and positive experiences with the products. The second most frequent category is neutral, found in approximately 3,200 reviews, which reflects reviews containing descriptive or factual statements without strong emotional tone. The admiration category appears in around 1,000 reviews and represents a deeper sense of appreciation toward product quality or design. Meanwhile, sadness and anger appear in significantly smaller proportions (each below 500 reviews), typically associated with issues related to product size, material inconsistency, or unmet expectations. These findings align with Rahman and Mehedi, who noted that fashion-related customer reviews tend to exhibit predominantly positive emotional expression, especially when customers feel satisfied with style, comfort, and product fit (Rahman and Mehedi 2024).

Sentiment Polarity and Departmental Trends

Sentiment polarity analysis revealed that positive sentiment is dominant across all product departments. Categories such as Tops, Dresses, and Intimate consistently show more than 90% positive reviews, suggesting high customer satisfaction in product segments with standardized sizing and stable quality. By contrast, departments like Trend and Jackets exhibit slightly more variation in sentiment, including higher proportions of neutral and negative reviews. These mixed responses often involve concerns about fit, design variation, or return experiences. The overall pattern highlights that product characteristics strongly influence sentiment distribution. This observation is consistent with the findings of Chowdhury, who reported that apparel categories with unpredictable sizing or fashion-driven characteristics tend to generate more emotionally diverse feedback due to higher uncertainty in product fit and customer expectations (Chowdhury et al. 2023).

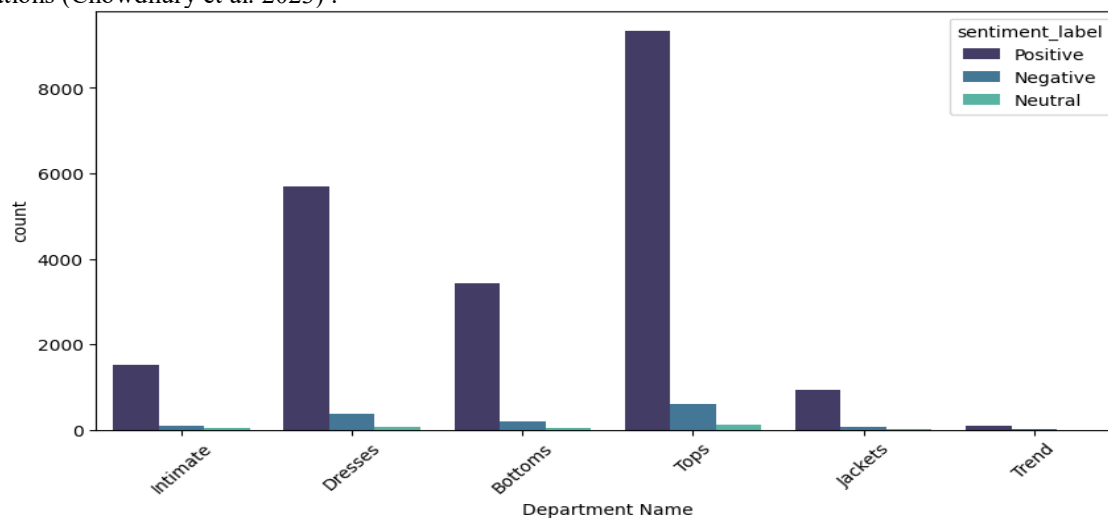


Figure 3. Sentiment Polarity by Department

Average Rating and Recommendation Analysis

The rating analysis presented in Figure 3 shows notable variation across product departments. The Intimate and Jackets categories receive the highest average ratings, both slightly above 4.2, indicating strong customer satisfaction toward product quality, comfort, and fit. The Bottoms and Tops departments follow closely with average ratings around 4.1–4.2, reflecting consistent positive user experience. Meanwhile, Dresses remain stable within the positive range, though slightly lower compared to Intimate.

The Trend department records the lowest average rating, approximately 3.8, suggesting that customers may experience more variability in product suitability such as fashion style, size availability, or material expectations. Despite these differences, all categories maintain average scores above 3.5, demonstrating that overall customer perception remains positive. This pattern supports previous findings by Lin, who reported that strongly positive emotional tone contributes to higher product appraisal and rating behavior in online fashion markets (Lin et al. 2023). The alignment between emotional expression and numeric rating suggests that emotional satisfaction plays a central role in shaping product evaluation and future purchase intentions.

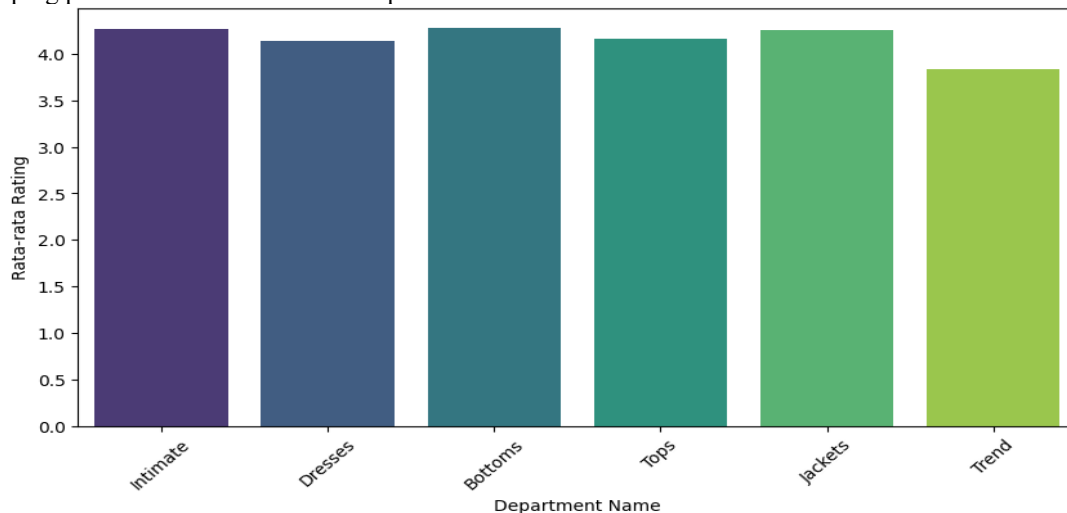


Figure 4. Average Rating Across Departments

Sentiment Rating Relationship

The relationship between sentiment polarity and numeric rating demonstrates strong consistency. Reviews classified with positive sentiment correspond predominantly to ratings of 4–5, confirming alignment between emotional tone and quantitative assessment. Conversely, reviews expressing negative sentiment are primarily associated with ratings of 1–2, typically reflecting dissatisfaction related to sizing problems, material quality, or delivery issues. Reviews with neutral sentiment generally correspond to mid-range ratings around 3, indicating descriptive feedback without strong emotional engagement. This consistency validates the reliability of the emotion classification model and reflects a coherent pattern between textual sentiment and star ratings. These findings further support the work of Lin, who showed that emotional intensity in review text is strongly aligned with rating behavior (Lin et al. 2023).

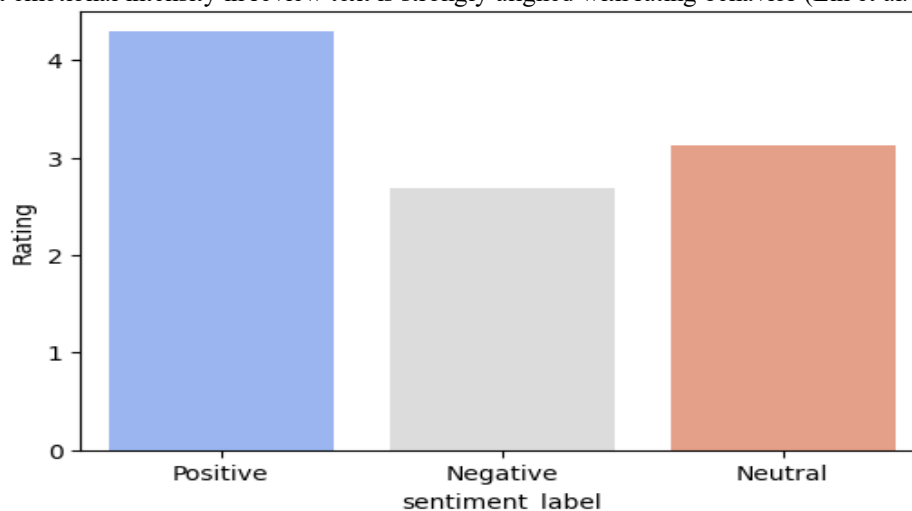


Figure 5. Relationship Between Sentiment and Rating

DISCUSSION

The results of this study reveal several essential insights into emotion-based review analysis. Overall, the findings indicate a dominance of positive emotions, where most customers express happiness and admiration, reflecting a high level of satisfaction toward product quality and service experience. Emotional tone also varies across departments; categories with standardized sizing such as Tops and Intimate demonstrate higher emotional positivity, while fashion-trend items show more polarized reactions due to subjective style preferences. Moreover, a strong coherence between emotional expression and numerical ratings was observed, confirming that the detected emotions align with actual customer satisfaction levels and validating the interpretability of the DistilBERT model. The model's robustness and computational efficiency further demonstrate its suitability for real-time e-commerce applications, offering accuracy comparable to larger transformer architectures with reduced resource requirements. These results are consistent with Aguilar, who emphasize that emotion-aware transformer models can enhance personalized marketing analytics and recommendation systems (Aguilar et al. 2024).

Despite these encouraging findings, several limitations should be acknowledged. The dataset focuses solely on women's apparel reviews, which may constrain generalizability to other product domains. In addition, the language scope is limited to English, and translating emotional categories into Bahasa Indonesia could introduce minor semantic deviations. Cultural diversity in emotional expression may also influence interpretation accuracy, while the pre-trained DistilBERT corpus lacks deep adaptation to fashion-specific terminology. Future research should therefore explore multilingual fine-tuning, cross-category dataset expansion, and multimodal emotion recognition that integrates both textual and visual information to achieve a more comprehensive understanding of consumer emotions in diverse contexts.

CONCLUSION

This study presents an emotion-based review analysis using the DistilRoBERTa emotion classification model to classify emotional expressions in Women's E-Commerce Clothing Reviews. The findings demonstrate that positive emotions especially joy and admiration dominate customer feedback, while negative emotions such as anger and sadness appear in relatively small proportions and are mostly triggered by issues related to product size and fabric inconsistency. Emotion patterns were also found to be consistent with sentiment polarity and numerical rating scores, indicating that the model successfully captured emotional signals that reflect real customer satisfaction.

Based on these findings, the research contributes to two important outcomes. First, it answers the research question by proving that transformer-based emotion classification provides deeper insight compared to binary sentiment analysis. Second, this work demonstrates that DistilBERT can analyze large-scale textual reviews efficiently, making it suitable for real-time e-commerce applications such as dynamic recommendation engines, product quality monitoring, and marketing personalization.

However, this study is limited to reviews written in English and specifically within the women's apparel domain, which may reduce generalizability across product categories and languages. Future research should involve multilingual datasets, cross-category review comparisons, and multimodal emotional reasoning that merges text, rating, images, and user interaction behavior. Expanding these directions will further enhance the applicability of transformer-based emotional analytics in a broader e-commerce environment.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin (UNISKA) for providing research facilities, academic guidance, and institutional support throughout this study. Special appreciation is extended to the Faculty of Information Technology for encouraging research in artificial intelligence and data analytics. The authors also acknowledge the open-access dataset providers on Kaggle for making the women's e-commerce review data publicly available, which was essential for conducting this analysis. Finally, the authors thank all colleagues and reviewers whose valuable insights and feedback contributed significantly to the improvement of this manuscript.

REFERENCES

- Aguilar, J., C. García, and M. López. 2024. "Emotion-Aware Transformer Models for Recommendation in e-Commerce Platforms." *Expert Systems with Applications* 235:120112.
- Chowdhury, S., A. Kumaran, and M. Rahman. 2023. "Emotional Drivers of Fashion Product Satisfaction: An NLP Perspective." *Journal of Retail Analytics* 12(4):55–72.
- Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers."
- Han, J., J. Pei, and M. Kamber. 2022. *Data Mining: Concepts and Techniques (4th Edition)*. Morgan Kaufmann.
- Kaggle. 2023. "Women's E-Commerce Clothing Reviews Dataset."
- Kumar, S., and A. Bhattacharya. 2024. "Emotion Detection Using Lightweight Transformer Models."
- Li, H., Q. Zhao, and L. Zhang. n.d. "A Hybrid Machine Learning Model for Customer Churn and Purchase Propensity



- Prediction.” *Expert Systems with Applications* 214.
- Lin, Y., S. Zhou, and J. Chen. 2023. “Transformer-Based Sentiment and Emotion Analysis for Retail Reviews.” *Neural Computing and Applications* 35.
- Liu, W., K. Zhang, and Y. Huang. 2024. “Standard Text Preprocessing Pipeline for Sentiment Mining.” *Information Processing & Management* 61(2).
- Patel, R. n.d. “Machine Learning Approaches for Consumer Decision-Making Prediction.” *Decision Analytics Journal* 7:100239.
- Rahimi, M., and M. Khashei. 2023. “Text Emotion Recognition with Deep Contextual Embedding.” *Knowledge-Based Systems* 260:111955.
- Rahman, M., and A. Mehedi. 2024. “Hybrid Transformer Architecture for Emotion Detection.” *Applied Intelligence* 54(6):11325–41.
- Sanh, V., L. Debut, J. Chaumond, and T. Wolf. 2019. “DistilBERT: Smaller, Faster & Lighter Transformer Architecture.”
- Tripathy, A., A. Agrawal, and S. Rath. 2020. “Sentiment Classification Using CNN and RNN.” *Expert Systems with Applications* 171.
- Xu, M., F. Wang, and H. Li. 2023. “Emotion Detection Using Enhanced Transformer Architecture.” *IEEE Access* 11:44512–26.
- Yang, S., J. Kim, and H. Lee. 2022. “Improving Sentiment Classification with BiLSTM Attention Model.” *International Journal of Data Science* 7(2):88–99.
- Zhang, L., C. Xu, and H. Sun. 2023. “Hybrid Transformer-LSTM for e-Commerce Emotion Analysis.” *ACM Transactions on Intelligent Systems* 41(3).