Analyzing Customer Sentiment in Online Product Reviews Using Machine Learning Techniques

Jaliba Sherin K.J.[1], Dr. Pramod K [2]

1(MCA, Nehru College of Engineering and Research Centre, Thrissur, Kerala, India
Email: Sherinjaliba48@gmail.com)
2(1(MCA, Nehru College of Engineering and Research Centre, Thrissur, Kerala, India
Email: pramod.mca@ncerc.ac.in)

Abstract:

Online product reviews have become an invaluable resource for consumers seeking information to make informed purchasing decisions. Sentiment analysis, a branch of Natural Language Processing (NLP), plays a crucial role in automatically extracting sentiments or opinions from these reviews. This research paper presents a comprehensive study on sentiment analysis in online product reviews, specifically focusing on Amazon product reviews. The study employs advanced machine learning techniques, including feature engineering and deep learning models such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Bidirectional Encoder Representations from Transformers (BERT), to accurately classify sentiment expressed in customer reviews. The methodology encompasses data collection from Amazon reviews, comprehensive preprocessing techniques including tokenization, stop-word removal, and feature extraction, followed by model training and evaluation. The BERT model demonstrates superior performance with 89% accuracy, 88% precision, 89% recall, and 88% F1-score, significantly outperforming traditional machine learning approaches such as Logistic Regression (83.1% accuracy) and Decision Trees (75% accuracy). This research provides valuable insights for e-commerce businesses to understand customer opinions better and make data-driven decisions regarding product improvements, marketing strategies, and customer satisfaction enhancement.

Keywords ---- : *Sentiment Analysis, Machine Learning, Deep Learning, BERT, Amazon Reviews, Natural Language Processing, Customer Opinion Mining, Product Reviews*

I. INTRODUCTION

The explosive growth of e-commerce platforms in recent years has fundamentally transformed how consumers make purchasing decisions. Online marketplaces like Amazon have become dominant platforms where millions of product reviews are posted daily, creating a rich source of customer feedback and opinion data[1]. These reviews not only assist prospective customers in making educated selections but also provide businesses

with valuable insights into product performance, customer preferences, and market trends[2].

Sentiment analysis, also known as opinion mining, is a computational technique that examines people's opinions, evaluations, sentiments, attitudes, and emotions toward various entities such as products, organizations, services, and events[3]. It represents a specialized branch of Natural Language Processing (NLP) that focuses on automatically identifying, extracting, and categorizing opinions expressed in textual data. The fundamental objective is to determine and characterize the polarity of text, classifying it as positive, negative, or neutral[4].

The importance of sentiment analysis in the e-commerce domain cannot be overstated. With the abundance of user-generated content available on online platforms, traditional manual analysis methods have become impractical and time-consuming[5]. According to recent studies, approximately 90% of consumers consult online reviews before making purchase decisions, highlighting the critical role these reviews play in consumer behavior[6]. However, the sheer volume and diversity of reviews present significant challenges for both consumers trying to make informed decisions and businesses attempting to understand customer sentiment at scale.

field of sentiment analysis has witnessed significant advancements in recent years, particularly with the emergence of sophisticated machine learning and deep learning techniques[7]. Traditional approaches relied primarily on lexicon-based methods and simple machine learning algorithms, but these methods often struggled to capture the complex, contextual nature of human language, including challenges such as sarcasm, polarity shifts, and domain-specific sentiment expressions[8]. Modern approaches leverage advanced neural network architectures that can learn hierarchical representations of text and capture long-term dependencies, resulting in more accurate sentiment classification[9].

This research addresses these challenges by developing and evaluating a comprehensive sentiment analysis framework specifically designed for e-commerce product reviews. The study focuses on Amazon product reviews across multiple categories including computers, tablets, smartphones, televisions, and video surveillance equipment. The primary contributions of this research include: (1) implementation of advanced preprocessing techniques to handle noisy and diverse review data, (2) development of sophisticated feature engineering methods combining lexical, syntactic, and semantic features, (3) application of state-of-the-art deep learning models including BERT for sentiment classification, and (4) comprehensive performance evaluation demonstrating significant improvements over traditional approaches.

The remainder of this paper is organized as follows: Section II presents a comprehensive literature review of existing sentiment analysis approaches and techniques. Section III details the methodology employed in this study, including data collection, preprocessing, feature engineering, and model development. Section IV presents the experimental results and comparative analysis. Section V discusses the implications and applications of the findings. Finally, Section VI concludes the paper and outlines directions for future research.

II. LITERATURE REVIEW

The field of sentiment analysis has evolved significantly over the past two decades, with researchers exploring various approaches ranging from lexicon-based methods to sophisticated deep learning architectures. This section provides a

comprehensive overview of related work in sentiment analysis, with particular emphasis on e-commerce product reviews.

A. Fundamental Approaches to Sentiment Analysis

Sentiment analysis research can be broadly categorized into three primary approaches: lexicon-based methods, machine learning-based methods, and hybrid approaches that combine elements of both[10]. Lexicon-based methods rely on sentiment lexicons or dictionaries containing pre-defined sentiment information about words. These lexicons assign specific weights or scores to words based on their polarity, indicating whether they convey positive, negative, or neutral sentiment[11]. While computationally efficient and interpretable, lexicon-based methods often struggle with contextual interpretation, particularly in cases involving sarcasm, irony, or ambiguous language[12].

Maks (2012)[13] introduced a sophisticated lexicon model specifically designed for sentiment analysis and opinion mining applications. The model provides detailed descriptions of verbs, nouns, and adjectives, focusing on subjectivity relations between actors in sentences. The research demonstrated that human annotators could reliably identify subtle subjectivity relations, validating the effectiveness of the lexicon approach. The model's emphasis on capturing perspectives of different actors, including the author, contributes to more comprehensive sentiment analysis.

Machine learning-based methods, in contrast, rely on supervised or unsupervised learning algorithms to automatically classify sentiment from labeled training data[14]. These approaches employ algorithms such as Support Vector Machines (SVM), Naive Bayes, and Random Forests, which can generalize patterns learned from training data to classify sentiment in new, unseen text[15]. Saberi

(2017)[16] provided a comprehensive survey of sentiment analysis approaches, highlighting both machine learning and lexical-based methods, along with their application areas and associated challenges.

B. Deep Learning Approaches for Sentiment Analysis

Recent advances in deep learning have revolutionized sentiment analysis, enabling models to capture complex linguistic patterns and contextual information more effectively[17]. Convolutional Neural Networks (CNNs) have proven particularly effective in extracting local features from text by utilizing filters that scan small segments of review texts, making them especially useful for identifying sentiment cues within shorter segments such as individual sentences or phrases[18].

Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), excel at capturing sequential dependencies and contextual information crucial for understanding sentiment in longer, context-dependent reviews[19]. Kumar and Saviour Devaraj (2021)[20] introduced a semantic-based feature selection approach utilizing SentiWordNet, a lexical resource in the WordNet database, combined with classifiers such as Naive Bayes, FLR, and AdaBoost, demonstrating improved performance through thoughtful feature selection.

The emergence of transformer-based models, particularly BERT (Bidirectional Encoder Representations from Transformers), represents a significant breakthrough in NLP and sentiment analysis[21]. BERT's ability to capture bidirectional context and understand nuanced language patterns has led to state-of-the-art performance across various sentiment analysis tasks. The model's pre-training on massive text corpora enables it to learn

rich linguistic representations that can be fine-tuned for specific sentiment classification tasks[22].

C. Sentiment Analysis in E-commerce and Social Media

The application of sentiment analysis to e-commerce platforms has received considerable attention from researchers due to its practical implications for businesses and consumers. Ghous (2021)[23] explored methods for classifying Amazon reviews as positive, negative, or neutral, applying data engineering and data science techniques to analyze consumer sentiment effectively. The study highlighted the importance of appropriate preprocessing and feature extraction techniques for achieving high classification accuracy.

Messaoudi, Guessoum, and Ben Romdhane (2022)[24] presented a comprehensive survey addressing the rise of social networks and their impact on opinion detection. The research identified key challenges in sentiment analysis, including natural language processing ambiguity, detecting spam opinions, identifying sarcasm, and handling abbreviations. The survey provided detailed exploration of well-known classical and recent approaches in opinion mining, offering valuable insights into the evolution of the field.

Abayomi-Alli et al. (2022)[25] conducted a study analyzing public opinions from social media microblogs, demonstrating the application of sentiment analysis techniques including VADER (Valence Aware Dictionary for Sentiment Reasoning) and the Liu-Hu method. The research employed topic modeling techniques such as Latent Dirichlet Allocation (LDA) and Latent Semantic Indexing (LSI) algorithms, providing insights into the main themes discussed in the dataset. The study confirmed the efficacy of VADER in analyzing unstructured social media data containing non-

English slang, conjunctions, emoticons, and other informal elements.

D. Feature Engineering and Text Preprocessing

Effective feature engineering is crucial for successful sentiment analysis. Various techniques have been developed to extract meaningful features from text data, including n-grams, part-of-speech tags, sentiment lexicons, and word embeddings[26]. Tuz Zohra Anny and Islam (2022)[27] proposed a comprehensive approach to sentiment polarity classification, encompassing both sentence-level classification and review-level categorization. The study emphasized the importance of capturing sentiment at multiple granularities to provide nuanced insights into the opinions expressed in text.

Lin et al. (2022)[28] conducted a systematic literature review of opinion mining in software engineering, categorizing opinion mining-related activities and examining available approaches. The review identified relevant datasets for performance evaluation and tool customization, highlighting concerns and limitations that researchers should consider when applying sentiment analysis techniques. The findings offered valuable references for selecting appropriate opinion mining tools and advancing techniques within specific domains.

E. Performance Evaluation and Comparative Studies

Comparative analysis of different sentiment analysis approaches has been essential for understanding their relative strengths and limitations. Umarani and Julian Deepa (2021)[29] explored various machine learning and deep learning techniques for sentiment analysis, comparing the performance of different methods in extracting insights from textual data. The research demonstrated that deep learning approaches

generally outperform traditional machine learning methods, particularly when dealing with large-scale datasets and complex linguistic patterns.

Kabir et al. (2021)[30] examined various machine learning techniques for sentiment analysis, evaluating and comparing the effectiveness of different models in classifying text sentiment. The study provided empirical evidence supporting the superiority of ensemble methods and deep learning architectures over single-algorithm approaches. He, Zhou, and Zhao (2022)[31] investigated e-commerce product experience based on a fusion sentiment analysis method, combining multiple sentiment analysis techniques to gain deeper insights into consumer feedback and product experience.

The literature review reveals that while significant progress has been made in sentiment analysis, several challenges remain. These include handling domain-specific language, managing class imbalance in sentiment datasets, detecting subtle emotional nuances, and improving model interpretability. The current research addresses these challenges by employing advanced preprocessing techniques, sophisticated feature engineering methods, and state-of-the-art deep learning models, particularly BERT, to achieve superior performance in sentiment classification of Amazon product reviews.

III METHODOLOGY

This research employs a comprehensive methodology for sentiment analysis of Amazon product reviews, incorporating advanced machine learning and deep learning techniques. The methodology encompasses several critical stages: data collection, data preprocessing, feature engineering, model training, and performance evaluation. This section provides detailed descriptions of each stage, ensuring reproducibility and clarity of the research process.

A. Research Framework

The sentiment analysis system is designed to leverage advanced methods for extracting customer opinions from online product reviews. The framework combines traditional feature engineering techniques with state-of-the-art deep learning models to achieve accurate sentiment classification. The overall research framework follows a systematic pipeline that transforms raw review text into meaningful sentiment predictions.

The system's effectiveness lies in its multi-faceted approach to feature extraction and model training. By capturing both local and contextual information through carefully designed features and sophisticated neural network architectures, the framework enables comprehensive understanding of sentiment expressions within customer reviews.

B. Data Collection

The dataset used in this research was collected from Amazon.com, one of the largest and most popular e-commerce platforms worldwide. The collection process focused on gathering comprehensive product reviews across multiple categories to ensure diversity and representativeness of the dataset. The dataset comprises over 28,000 customer reviews covering more than 60 different product types, including computers, tablets, smartphones, televisions, video surveillance systems, and various consumer electronics.

Reviews were collected in JSON (JavaScript Object Notation) format, which provides structured representation of review data including key attributes such as review text, rating scores, product identification, user information, review date, and helpfulness votes. The JSON format facilitates efficient data parsing and preprocessing operations.

Each review in the dataset contains both textual content (the actual review text written by customers) and numerical ratings (typically on a scale from 1 to 5 stars).

The dataset selection criteria ensured representation across diverse product domains and included reviews with a wide range of sentiment expressions. This diversity is crucial for training robust sentiment classification models that can generalize well to different product categories and review styles. The collection process utilized web scraping techniques and API access methods to gather reviews systematically while respecting platform usage policies and rate limiting requirements.

To ensure data quality, reviews with insufficient text content (single-word reviews) were excluded from the dataset. Additionally, reviews were filtered to remove duplicates and spam content. The final dataset maintains a balanced representation of different sentiment polarities, although natural imbalances exist due to the tendency of satisfied customers to provide positive feedback more frequently than negative feedback.

C. Data Preprocessing

Data preprocessing constitutes a critical phase in the sentiment analysis pipeline, as raw text data contains noise, inconsistencies, and irrelevant information that can negatively impact model performance. The preprocessing stage transforms unstructured review text into a clean, standardized format suitable for feature extraction and model training. This research implements a comprehensive preprocessing pipeline consisting of multiple sequential operations.

1) Lowercasing: The first preprocessing step converts all text characters to lowercase. This normalization ensures consistency in word representation and reduces vocabulary size by treating words with different capitalizations (e.g., "Great," "great," "GREAT") as identical tokens. Lowercasing is implemented using Python's built-in string methods, transforming all uppercase and mixed-case text to lowercase while preserving numerical and special characters.

2) HTML Tag and URL Removal: Many online reviews contain HTML formatting tags, URLs, and other web-specific content that do not contribute meaningful sentiment information. Regular expressions are employed to identify and remove HTML tags (e.g., <div>, <p>, <br>), complete URLs (e.g., http://example.com), and partial web addresses. This cleaning operation ensures that the text analysis focuses on actual review content rather than markup or reference information.

3) Punctuation Removal: Punctuation marks are systematically removed from the review text, as they generally do not carry sentiment polarity information in the context of bag-of-words and similar feature representations. The punctuation removal process eliminates characters such as periods, commas, semicolons, colons, exclamation marks, question marks, quotation marks, and other special characters. However, apostrophes in contractions are initially preserved for the contraction expansion phase.

4) Contraction Expansion: Contractions are common in informal written language, particularly in customer reviews. This preprocessing step expands contractions to their full forms (e.g., "don't" → "do not," "can't" → "cannot," "I'm" → "I am"). Contraction expansion is implemented using a predefined dictionary mapping contracted forms to their expanded equivalents. This normalization improves consistency in text representation and ensures that sentiment-bearing words are properly captured.

5) Stop Word Removal: Stop words are frequently occurring words that typically do not contribute significant semantic content or sentiment information. Examples include articles ("a," "an," "the"), prepositions ("in," "on," "at"), conjunctions ("and," "but," "or"), and common verbs ("is," "are," "was"). The NLTK (Natural Language Toolkit) library provides comprehensive stop word lists for English and other languages. Stop word removal reduces data dimensionality and focuses the analysis on content-bearing words that express sentiment.

6) Tokenization: Tokenization segments continuous text into discrete units called tokens, typically individual words or subword units. This research employs word-level tokenization, where review texts are split into sequences of words based on whitespace and punctuation boundaries. The TextBlob library's pretrained tokenizer is utilized for this purpose, providing robust tokenization that handles various edge cases and linguistic patterns. Tokenization produces a list of words for each review, enabling subsequent operations such as part-of-speech tagging and feature extraction.

7) Part-of-Speech (POS) Tagging: Part-of-speech tagging assigns grammatical categories to each word in the tokenized text. POS tags include classifications such as noun (NN), verb (VB), adjective (JJ), adverb (RB), and others following standard linguistic taxonomies. POS information captures syntactic structure and can be valuable for sentiment analysis, as certain grammatical categories (particularly adjectives and adverbs) often carry strong sentiment polarity. The NLTK library's POS tagger is employed to annotate each token with its grammatical category.

8) Stemming and Lemmatization: Stemming and lemmatization reduce words to their root or base forms, decreasing vocabulary size and grouping morphological variants. Stemming employs rule-based approaches to remove suffixes (e.g., "running," "runs," "ran" → "run"), while lemmatization uses vocabulary and morphological analysis to return dictionary base forms. This research applies the Porter Stemmer algorithm from NLTK, which provides a balanced approach between computational efficiency and stemming accuracy.

D. Feature Engineering

Feature engineering transforms preprocessed text into numerical representations that machine learning algorithms can process. This research implements multiple feature extraction techniques to capture different aspects of sentiment expression in review text.

1) Bag-of-Words (BoW) Representation: The Bag-of-Words model represents text as vectors of word frequencies, disregarding grammar and word order but maintaining information about word occurrence counts. Each unique word in the corpus constitutes a dimension in the feature space, and each review is represented as a vector indicating the frequency of each word. The BoW approach provides a straightforward and effective baseline representation for sentiment analysis tasks.

The BoW implementation begins by constructing a vocabulary from the training corpus, identifying all unique words after preprocessing. Each review is then transformed into a fixed-length vector, where each dimension corresponds to a vocabulary word and the value represents the word's frequency in that review. While BoW does not capture semantic relationships or word order, it effectively quantifies lexical content and enables efficient processing of large text collections.

2) N-gram Features: N-grams extend the bag-of-words representation by capturing sequences of

consecutive words rather than individual words alone. Unigrams represent single words, bigrams represent two-word sequences, and trigrams represent three-word sequences. N-gram features are particularly valuable for sentiment analysis because they can capture common sentiment-bearing phrases (e.g., "not good," "very satisfied," "highly recommend") that convey meaning through word combinations rather than individual words.

This research extracts both unigram and bigram features to balance the trade-off between feature dimensionality and contextual information. Higher-order n-grams (trigrams and beyond) can capture more specific patterns but also increase feature space dimensionality exponentially and may suffer from data sparsity issues.

3) Sentiment Lexicon Scoring: Sentiment lexicons provide predefined sentiment scores for individual words based on their inherent polarity. This research employs an opinion lexicon containing approximately 5,000 positive words and 4,500 negative words, each associated with polarity values. The sentiment score for a review is calculated by matching words in the review text against the lexicon and aggregating their individual sentiment scores.

The scoring process follows a simple but effective algorithm: for each word in the preprocessed review, the system looks up the word in the sentiment lexicon. If the word is found, its associated sentiment score is added to a cumulative sentiment total for the review. After processing all words, the cumulative score is normalized by the review length to produce a final sentiment score. Reviews with positive cumulative scores are classified as positive, reviews with negative cumulative scores as negative, and reviews with scores near zero as neutral.

4) Word Embeddings: Word embeddings represent words as dense, low-dimensional vectors in a continuous vector space, where semantically similar words are positioned closer together. Unlike sparse representations such as bag-of-words, word embeddings capture semantic relationships and contextual similarities between words. This research explores multiple word embedding approaches including Word2Vec, GloVe (Global Vectors for Word Representation), and contextual embeddings from BERT.

Word2Vec employs neural network architectures to learn word representations by predicting words from their context (Skip-gram model) or predicting context from words (Continuous Bag-of-Words model). GloVe learns embeddings through matrix factorization of word co-occurrence statistics. BERT's contextual embeddings provide dynamic word representations that vary based on surrounding context, enabling more nuanced understanding of word meaning in different contexts.

E. Deep Learning Models

This research implements multiple deep learning architectures for sentiment classification, leveraging their capabilities to learn hierarchical representations and capture complex patterns in text data.

1) Convolutional Neural Networks (CNNs): Convolutional Neural Networks, originally developed for computer vision tasks, have proven highly effective for text classification. CNNs apply convolutional filters to text sequences, extracting local n-gram features and identifying relevant patterns regardless of their position in the text. The CNN architecture for sentiment analysis typically consists of an embedding layer, one or more convolutional layers with various filter sizes,

pooling layers for dimensionality reduction, and fully connected layers for final classification.

The convolutional operation applies filters of different sizes (e.g., 3, 4, 5 words) to capture n-gram patterns of varying lengths. Each filter produces a feature map highlighting the presence of specific patterns throughout the text. Max-pooling operations select the most salient features from each feature map, creating fixed-length representations regardless of review length. This architecture enables the model to identify important sentiment-bearing phrases and patterns while maintaining computational efficiency.

2) Recurrent Neural Networks (RNNs):
Recurrent Neural Networks are specifically designed to process sequential data by maintaining internal hidden states that capture information from previous time steps. For sentiment analysis, RNNs process review text word by word, updating their hidden state at each step to accumulate contextual information. This architecture enables modeling of long-range dependencies and understanding of how sentiment builds throughout a review.

Standard RNN architectures suffer from vanishing gradient problems that limit their ability to capture long-term dependencies. This research employs advanced RNN variants, specifically Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), which address these limitations through gating mechanisms. LSTM networks utilize input gates, forget gates, and output gates to control information flow, enabling selective retention and updating of relevant information over long sequences. GRUs provide a simplified gating mechanism with comparable performance and reduced computational complexity.

3) Bidirectional Encoder Representations from Transformers (BERT): BERT represents a

paradigm shift in NLP, employing transformer architecture with attention mechanisms to capture bidirectional context. Unlike traditional left-to-right or right-to-left language models, BERT processes text bidirectionally, considering both preceding and following context for each word. This bidirectional understanding enables more nuanced comprehension of language and improved performance on various NLP tasks including sentiment analysis.

The BERT model used in this research consists of multiple transformer layers, each containing multi-head self-attention mechanisms and feed-forward networks. The embedding layer combines word embeddings, position embeddings, and segment embeddings to create rich input representations. The self-attention mechanism computes attention weights between all pairs of words in the input, enabling the model to identify relevant relationships and dependencies regardless of distance.

For sentiment classification, the research employs a fine-tuning approach: a pre-trained BERT model is adapted to the specific task of sentiment analysis on Amazon reviews. The BERT encoder processes the review text and produces contextualized representations for each token. The representation of the special [CLS] token, which aggregates sequence-level information, is fed through additional fully connected layers to produce sentiment predictions. The model is trained using cross-entropy loss, optimizing the parameters to minimize classification error on the labeled training data.

4) Hybrid CNN-RNN Architecture: To leverage the complementary strengths of CNNs and RNNs, this research also explores a hybrid architecture that combines both approaches. The CNN component extracts local n-gram features from the text, while the RNN component models sequential

dependencies and long-term context. The hybrid model processes review text first through convolutional layers to extract salient local features, then feeds these feature representations through recurrent layers to capture temporal dependencies and sequential patterns. This architecture enables the model to benefit from CNN's efficient local pattern recognition while also capturing the sequential nature of language through RNNs. The combined approach has demonstrated improved performance on various text classification tasks, providing a robust framework for sentiment analysis.

F. Model Training and Optimization

The dataset is partitioned into training (70%), validation (10%), and testing (20%) subsets using stratified sampling to maintain class distribution across splits. The training set is used to optimize model parameters through backpropagation and gradient descent. The validation set guides hyperparameter tuning and early stopping to prevent overfitting. The test set provides unbiased evaluation of final model performance.

Model training employs the PyTorch framework, a flexible and efficient deep learning library. The BERT model is fine-tuned using the AdamW optimizer with learning rate scheduling and gradient clipping. Training hyperparameters including learning rate, batch size, number of epochs, and dropout rates are tuned through systematic grid search and validation performance monitoring. Cross-entropy loss serves as the optimization objective for multi-class sentiment classification. The loss function measures the discrepancy between predicted probability distributions and true labels, guiding the model to produce accurate sentiment predictions. Regularization techniques including dropout and weight decay prevent overfitting and improve generalization to unseen data.

G. Performance Evaluation Metrics

Model performance is evaluated using standard classification metrics that provide comprehensive assessment of sentiment classification accuracy and reliability.

1) Accuracy: Accuracy measures the proportion of correctly classified instances among all instances. It is calculated as:

$$\text{Accuracy} = \frac{+}{+\ +\ +}$$

where TP represents true positives, TN true negatives, FP false positives, and FN false negatives. While accuracy provides an overall performance measure, it may be misleading for imbalanced datasets where one class dominates.

2) Precision: Precision quantifies the proportion of true positive predictions among all positive predictions:

$$\text{Precision} = \frac{}{+}$$

High precision indicates that when the model predicts a positive sentiment, it is likely to be correct. Precision is particularly important when false positives are costly.

3) Recall: Recall (also called sensitivity or true positive rate) measures the proportion of actual positive instances correctly identified:

$$\text{Recall} = \frac{}{+}$$

High recall indicates that the model successfully identifies most positive instances. Recall is crucial when missing positive cases (false negatives) has significant consequences.

4) F1-Score: The F1-score provides a harmonic mean of precision and recall, balancing both metrics:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-score is particularly useful for evaluating model performance on imbalanced datasets, as it accounts for both false positives and false negatives.

IV. RESULTS AND DISCUSSION

This section presents comprehensive experimental results and analysis of the sentiment analysis models developed in this research. The evaluation encompasses multiple aspects including exploratory data analysis, model performance metrics, comparative analysis with baseline approaches, and visualization of results.

A. Exploratory Data Analysis

Exploratory data analysis provides insights into the characteristics and distribution of the Amazon product review dataset. The dataset comprises 28,000 reviews across more than 60 product categories, with ratings ranging from 1 to 5 stars. Analysis of sentiment score distribution reveals that positive reviews (ratings 4 and 5) constitute approximately 65% of the dataset, negative reviews (ratings 1 and 2) represent about 20%, and neutral reviews (rating 3) account for approximately 15%.

Word cloud visualization techniques were employed to identify the most frequently occurring terms in reviews categorized by sentiment polarity. For positive reviews, prominent terms include "excellent," "great," "perfect," "love," "amazing," "best," "recommend," and "satisfied," reflecting favorable customer experiences. Negative reviews frequently contain terms such as "terrible," "poor," "worst," "disappointed," "broken," "useless," "waste," and "defective," indicating product dissatisfaction and quality issues. Neutral reviews display more balanced language with terms related to specific product features and factual descriptions.

Product category analysis demonstrates variation in sentiment distribution across different product types. Electronics categories such as smartphones and tablets receive generally positive sentiment with occasional strong negative reviews related to defects or performance issues. Video surveillance equipment shows more neutral and technical language in reviews, focusing on specifications and functionality rather than emotional expressions. This category-specific variation highlights the importance of domain-adaptive sentiment analysis approaches.

B. BERT Model Performance

The BERT model demonstrates exceptional performance in sentiment classification of Amazon product reviews, achieving state-of-the-art results across all evaluation metrics. The model attains 89% accuracy, indicating that nearly 9 out of 10 reviews are correctly classified. Precision reaches 88%, demonstrating high reliability in positive sentiment predictions. Recall of 89% indicates strong capability in identifying actual positive instances. The F1-score of 88% reflects balanced performance considering both precision and recall.

Analysis of the training process reveals steady improvement in both training and validation accuracy over five epochs. The training accuracy curve shows consistent increase from approximately 0.870 to 0.890, demonstrating effective learning. The validation accuracy exhibits some fluctuation but maintains similar levels to training accuracy, indicating good generalization without significant overfitting. The loss curves display corresponding decrease over epochs, with training loss declining from approximately 0.38 to

0.26 and validation loss showing similar downward trends.

The confusion matrix analysis provides detailed insights into classification performance across sentiment categories. For negative sentiment, the model achieves high true negative rate with relatively few false positives and false negatives, demonstrating strong capability in identifying negative reviews. Positive sentiment classification also shows excellent performance with high true positive rate and low error rates. However, neutral sentiment presents greater challenges, with higher rates of misclassification where neutral reviews are often categorized as positive or negative. This difficulty with neutral sentiment is common in sentiment analysis due to the ambiguous nature of neutral expressions and the relative scarcity of truly neutral reviews compared to strongly polarized opinions.

The BERT model's superior performance can be attributed to several factors. First, the bidirectional context modeling enables comprehensive understanding of word meanings based on surrounding text. Second, the attention mechanisms identify relevant words and relationships that express sentiment, even when they are distant in the text. Third, the pre-training on massive text corpora provides rich linguistic knowledge that transfers effectively to the sentiment classification task. Fourth, the fine-tuning process adapts the general language understanding to the specific characteristics of Amazon product reviews.

C. Comparative Analysis with Baseline Models

To demonstrate the effectiveness of the BERT approach, comprehensive comparative analysis was conducted against traditional machine learning baselines including Logistic Regression and Decision Trees. The results clearly establish the superiority of deep learning approaches for sentiment analysis tasks.

1) Logistic Regression Performance: Logistic Regression, a widely used linear classification algorithm, achieves 83.1% accuracy with 87.7% precision, 83.1% recall, and 84.9% F1-score. While these results represent respectable performance, they lag significantly behind the BERT model. The linear nature of Logistic Regression limits its ability to capture complex, non-linear relationships in text data. However, the model's interpretability and computational efficiency make it a reasonable baseline for comparison.

2) Decision Tree Performance: Decision Trees demonstrate the weakest performance among the compared approaches, achieving 75% accuracy, 79% precision, 73% recall, and 73% F1-score. The relatively poor performance can be attributed to several factors. Decision Trees' tendency to overfit training data limits generalization to unseen reviews. The discrete, rule-based splitting decisions may not capture the continuous, nuanced nature of sentiment expression. Additionally, Decision Trees struggle to effectively represent the high-dimensional, sparse nature of text data.

3) Performance Comparison Summary: The performance comparison clearly demonstrates the advantages of deep learning approaches, particularly BERT, for sentiment analysis. BERT outperforms Logistic Regression by 5.9 percentage points in accuracy and surpasses Decision Trees by 14 percentage points. Similar improvements are observed across precision, recall, and F1-score metrics. These substantial performance gains justify the increased computational requirements of deep learning models for practical sentiment analysis applications.

Model	Accuracy	Precision	Recall	F1-Score
BERT	89.0%	88.0%	89.0%	88.0%
Logistic Regression	83.1%	87.7%	83.1%	84.9%
Decision Tree	75.0%	79.0%	73.0%	73.0%

Table 1: Performance comparison of sentiment analysis models

D. Sentiment Analysis by Product Category

Analysis of sentiment distribution across product categories reveals interesting patterns and insights. Computers and tablets receive predominantly positive reviews (approximately 70% positive), with common praise for performance, design, and functionality. Negative reviews typically cite issues such as battery life, software problems, or hardware defects. Smartphones demonstrate similar positive sentiment trends (approximately 68% positive), with customers frequently praising camera quality, display, and user experience. Negative feedback often relates to durability concerns, software updates, and customer service experiences. The smartphone category shows higher review volume and engagement compared to other categories, reflecting the product's importance in consumers' lives.

Television products receive mixed sentiment with approximately 60% positive reviews. Positive feedback emphasizes picture quality, smart features, and value for money. Negative reviews commonly mention shipping damage, setup difficulties, and remote control issues. The television category demonstrates higher proportion of neutral reviews

compared to other categories, possibly reflecting the more objective, technical nature of performance assessments.

Video surveillance equipment shows unique sentiment patterns with higher proportion of neutral and technical reviews (approximately 50% positive, 30% neutral). Reviews often focus on installation processes, compatibility issues, and technical specifications. The lower emotional content in this category suggests that customers approach these products with more practical, functional considerations rather than emotional attachment.

E. Sentiment Polarity Distribution Analysis

Detailed analysis of sentiment polarity scores reveals the distribution of sentiment intensity across reviews. The majority of reviews cluster at the extremes, with strong positive sentiment scores (above 0.7) and strong negative sentiment scores (below -0.5) being more common than moderate sentiment. This polarization reflects the tendency of customers to write reviews when they have particularly strong opinions, either very satisfied or very dissatisfied.

The distribution of sentiment scores also reveals interesting patterns in reviewer behavior. Customers who assign 5-star ratings typically write reviews with consistently high positive sentiment scores, often exceeding 0.8 on the normalized polarity scale. In contrast, 1-star ratings correspond to strongly negative sentiment scores, frequently below -0.6. The correspondence between numerical ratings and sentiment polarity validates the effectiveness of the sentiment analysis approach.

Reviews with moderate ratings (3 stars) display more diverse sentiment scores, ranging from slightly negative to slightly positive. This diversity reflects the ambiguous nature of moderate opinions where customers may have mixed feelings about

products, appreciating certain features while finding others disappointing. The challenge in classifying these neutral sentiments accurately explains the confusion matrix patterns observed in the BERT model performance analysis.

F. Feature Importance and Model Interpretation

Analysis of feature importance in the BERT model reveals which words and patterns contribute most significantly to sentiment predictions. Attention weight visualization demonstrates that the model focuses heavily on adjectives and adverbs that express sentiment intensity, such as "excellent," "terrible," "amazing," and "awful." Additionally, the model identifies negation patterns (e.g., "not good," "don't recommend") as important signals that reverse sentiment polarity.

Comparative analysis of different feature engineering approaches demonstrates the value of combining multiple feature types. Bag-of-words representations alone achieve moderate performance, while the addition of n-gram features improves classification by capturing sentiment-bearing phrases. Sentiment lexicon scores provide valuable baseline features, particularly for reviews containing many explicit sentiment words. Word embeddings offer the richest representations by encoding semantic relationships and contextual information.

The superior performance of BERT's contextual embeddings compared to static word embeddings (Word2Vec, GloVe) highlights the importance of context-dependent representations. Words like "cheap" can express positive sentiment ("cheap price") or negative sentiment ("cheap quality"), and BERT's ability to distinguish these contexts based on surrounding words contributes significantly to classification accuracy.

G. Error Analysis and Model Limitations

Despite impressive overall performance, the BERT model exhibits certain error patterns and limitations that warrant discussion. As noted in the confusion matrix analysis, neutral sentiment classification presents the greatest challenge. Many misclassifications involve neutral reviews being categorized as slightly positive or slightly negative, suggesting difficulty in identifying truly ambiguous or balanced opinions.

Sarcasm and irony represent another source of classification errors. Reviews containing phrases like "great, another broken product" or "just what I needed, more problems" may be misclassified as positive due to the presence of apparently positive words. While BERT's contextual understanding helps mitigate these issues compared to simpler approaches, perfect sarcasm detection remains challenging.

Domain-specific terminology and technical language can also lead to errors. Highly technical reviews discussing specifications, compatibility, and performance metrics may lack explicit sentiment expressions, making classification more difficult. Reviews written in non-standard English, including heavy use of slang, abbreviations, or grammatical errors, occasionally cause classification difficulties.

The model's performance varies across different review lengths. Very short reviews (one or two sentences) may lack sufficient context for accurate classification, while extremely long reviews may contain mixed sentiments that average out to moderate scores. The optimal review length for accurate classification appears to be 50-200 words, where sufficient context exists without excessive complexity.

H. Practical Applications and Business Implications

The sentiment analysis framework developed in this research offers numerous practical applications for e-commerce businesses. Real-time sentiment monitoring enables companies to track customer satisfaction continuously, identifying emerging issues before they escalate. Automated routing of negative reviews to customer service teams facilitates rapid response to dissatisfied customers, potentially resolving problems and improving retention.

Product development teams can leverage sentiment analysis insights to identify features that customers love and aspects requiring improvement. Analysis of sentiment trends over time reveals the impact of product updates, design changes, and quality initiatives. Competitive analysis becomes more data-driven through sentiment comparison across brands and product categories.

Marketing strategies can be optimized based on sentiment insights. Positive reviews and testimonials can be automatically identified and highlighted in promotional materials. Understanding which product features generate the most positive sentiment enables targeted marketing messaging. Sentiment analysis of competitor products identifies market opportunities and differentiation strategies.

Inventory management and demand forecasting benefit from sentiment analysis through early detection of quality issues or emerging trends. Sudden shifts in sentiment patterns may signal supply chain problems, manufacturing defects, or changing customer preferences. Proactive adjustment of inventory levels based on sentiment trends reduces costs and improves customer satisfaction.

Customer segmentation based on sentiment patterns enables personalized marketing and product recommendations. Customers who consistently write positive reviews represent brand advocates worthy of special recognition and engagement. Customers expressing negative sentiment require different communication strategies focused on problem resolution and relationship repair.

II. CONCLUSION

This research has presented a comprehensive study on sentiment analysis in online product reviews using advanced machine learning techniques, with specific focus on Amazon product reviews across multiple categories. The study successfully developed and evaluated a robust sentiment analysis framework that combines sophisticated preprocessing techniques, comprehensive feature engineering, and state-of-the-art deep learning models to achieve accurate sentiment classification.

The BERT model demonstrated exceptional performance, achieving 89% accuracy, 88% precision, 89% recall, and 88% F1-score, significantly outperforming traditional machine learning approaches such as Logistic Regression (83.1% accuracy) and Decision Trees (75% accuracy). These results validate the effectiveness of transformer-based architectures and contextual embeddings for capturing the nuanced nature of sentiment expression in customer reviews. The superior performance of BERT stems from its bidirectional context modeling, attention mechanisms, and ability to capture long-term dependencies in text.

The comprehensive methodology developed in this research, encompassing data collection, preprocessing, feature engineering, and model training, provides a systematic framework that can be applied to sentiment analysis tasks in various domains beyond e-commerce. The preprocessing pipeline effectively handles noisy text data, informal language, and diverse linguistic patterns common in online reviews. The feature engineering

approach successfully combines multiple representation techniques to capture lexical, syntactic, and semantic aspects of sentiment expression.

The practical implications of this research extend to multiple stakeholders in the e-commerce ecosystem. Businesses can leverage automated sentiment analysis to gain real-time insights into customer opinions, enabling data-driven decisions regarding product improvements, marketing strategies, and customer service enhancements. The ability to process thousands of reviews efficiently and accurately provides scalability that manual analysis cannot match. Consumers benefit from sentiment analysis through improved product recommendations and summarization of collective opinions.

The research also contributes to the broader field of natural language processing by demonstrating effective techniques for handling class imbalance, neutral sentiment classification, and domain-specific language patterns. The comparative analysis of different models and feature representations provides valuable insights for researchers and practitioners developing sentiment analysis systems.

Despite the impressive results, the study acknowledges certain limitations and challenges. Neutral sentiment classification remains difficult, with higher error rates compared to positive and negative categories. The model occasionally struggles with sarcasm, irony, and highly technical language. Very short or extremely long reviews present classification challenges due to insufficient or excessive contextual information. Computational requirements of the BERT model, while justified by performance gains, may limit deployment in resource-constrained environments.

The findings of this research have significant implications for the future of sentiment analysis in e-commerce and beyond. The demonstrated effectiveness of deep learning approaches, particularly BERT, suggests that continued advancement in transformer architectures and pre-training techniques will drive further improvements. The framework developed in this study provides a foundation for extending sentiment analysis to multilingual contexts, multimodal data (incorporating images and videos), and fine-grained aspect-based sentiment analysis.

In conclusion, this research successfully achieved its objectives of developing an accurate and robust sentiment analysis system for online product reviews, demonstrating the superiority of advanced deep learning techniques, and providing practical insights for e-commerce applications. The comprehensive methodology, rigorous evaluation, and detailed analysis contribute valuable knowledge to both the academic understanding of sentiment analysis and practical implementation of customer opinion mining systems.

VI. CHALLENGES AND FUTURE WORK

While this research has achieved significant success in sentiment analysis of online product reviews, several challenges remain that present opportunities for future investigation and improvement.

A. Current Challenges

E-commerce sentiment analysis faces several persistent challenges that affect model performance and practical deployment. Noisy data, including spelling errors, slang, emoticons, and non-standard grammar, complicates accurate text processing and feature extraction. The informal nature of online reviews requires robust preprocessing and error-correction mechanisms. Class imbalance, where positive reviews substantially outnumber negative

and neutral reviews, poses challenges for training models that perform equally well across all sentiment categories.

The massive volume of reviews generated daily on e-commerce platforms creates scalability challenges for real-time sentiment analysis. Efficient processing pipelines and model optimization are necessary to handle high-throughput requirements. Multilingual reviews present additional complexity, requiring language-specific models or robust multilingual approaches. Domain-specific terminology and technical language vary significantly across product categories, necessitating domain adaptation techniques.

Subtle emotional expressions, including sarcasm, irony, and mixed sentiments, remain difficult to detect accurately. While BERT's contextual understanding helps address these issues, perfect detection requires more sophisticated modeling of pragmatics and discourse structure. The lack of model interpretability in deep learning approaches limits trust and understanding of prediction rationale, particularly important for business decision-making contexts.

B. Future Research Directions

Several promising directions exist for extending and improving this research. Aspect-based sentiment analysis represents a natural progression, enabling identification of specific product features and aspects (e.g., battery life, camera quality, price) and associated sentiments. This fine-grained analysis provides more actionable insights for product development and marketing.

Multimodal sentiment analysis incorporating text, images, and videos offers richer understanding of customer opinions. Product reviews increasingly include photos and videos, which provide additional context and information about customer experiences. Developing fusion models that effectively combine textual and visual information could significantly enhance sentiment analysis accuracy.

Temporal sentiment analysis tracking sentiment evolution over time enables detection of trends, seasonal patterns, and the impact of events such as product updates or competitive launches. Time-series analysis of sentiment data could provide early warning of emerging issues and opportunities.

Cross-domain transfer learning addresses the challenge of limited labeled data in specific product categories by leveraging knowledge learned from abundant-data domains. Explainable AI techniques could enhance model interpretability, providing clear rationales for sentiment predictions. Attention visualization, saliency mapping, and feature importance analysis enable users to understand which words and patterns drive classification decisions.

Integration of external knowledge bases and common-sense reasoning could improve handling of complex linguistic phenomena. Few-shot and zero-shot learning approaches could enable sentiment analysis for new product categories or languages with minimal labeled training data. Meta-learning techniques train models to quickly adapt to new tasks with limited examples.

Adversarial robustness testing and improvement ensures models remain accurate when confronted with intentionally misleading or manipulated reviews. Active learning strategies could optimize the labeling process by intelligently selecting the most informative reviews for human annotation. This approach reduces labeling costs while maintaining model performance. Fairness and bias mitigation techniques ensure sentiment analysis systems perform equitably across different demographic groups and product categories.

C. Practical Implementation Considerations

Future work should address practical deployment considerations including model compression and efficiency optimization for real-time applications. Techniques such as knowledge distillation, quantization, and pruning can reduce model size and inference time while maintaining acceptable accuracy. Edge deployment on mobile devices or embedded systems requires further optimization and specialized implementations.

Integration with existing business systems and workflows necessitates development of APIs, dashboards, and visualization tools that make sentiment analysis insights accessible to non-technical stakeholders. User-friendly interfaces and automated reporting facilitate adoption and actionable use of sentiment analysis results.

Privacy and ethical considerations require careful handling of user data, compliance with regulations, and transparent communication about data usage. Continuous learning and model updating mechanisms enable sentiment analysis systems to adapt to evolving language patterns, emerging products, and changing customer preferences. Online learning approaches update models incrementally as new data becomes available, maintaining accuracy without complete retraining. Feedback loops incorporating human corrections and verifications improve model performance over time.

ACKNOWLEDGMENT

REFERENCES

[1] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," in 2018 IEEE International Conference on Innovative Research and Development (ICIRD), pp. 1-6, 2018.

[2] X. Fang and J. Zhan, "Sentiment analysis using product review data," Journal of Big Data, vol. 2, no. 1, pp. 1-14, 2015.

[3] R. S. Kumar and R. Saviour Devaraj, "Sentiment analysis using semantic-based feature selection approach utilizing SentiWordNet," International Journal of Advanced Computer Science and Applications, vol. 12, no. 9, pp. 123-134, 2021.

[4] B. Kaur and N. Kumari, "A hybrid approach to sentiment analysis of technical article reviews," International Journal of Education and Management Engineering, vol. 6, no. 6, pp. 1-9, 2016.

[5] C. Messaoudi, Z. Guessoum, and L. Ben Romdhane, "Opinion mining and sentiment analysis: A survey," Expert Systems with Applications, vol. 185, pp. 115-129, 2022.

[6] Q. T. Ghous, "Sentiment analysis of Amazon customer product reviews: A review," International Journal of Scientific Research Engineering & Development, vol. 4, pp. 564-595, 2021.

[7] V. Umarani, A. Julian, and J. Deepa, "Sentiment analysis using various machine learning and deep learning techniques," Journal of the Nigerian Society of Physical Sciences, vol. 3, pp. 385-394, 2021.

[8] B. Saberi, "Opinion mining or sentiment analysis: A survey," International Journal of Advanced Research in Computer Science, vol. 8, no. 5, pp. 1234-1241, 2017.

[9] S. Tammina and S. Annareddy, "Sentiment analysis on customer reviews using a convolutional neural network," in 2020 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-6, 2020.

[10] M. Kabir, M. M. J. Kabir, S. Xu, and B. Badhon, "An empirical research on sentiment analysis using machine

learning approaches," *International Journal of Computer Applications, vol. 182, no. 14, pp. 15-19, 2021.*

[11] E. Maks, "A lexicon model for deep sentiment analysis and opinion mining applications," *Decision Support Systems, vol. 53, no. 4, pp. 680-688, 2012.*

[12] M. Wahyudi and D. A. Kristiyanti, "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization," *Journal of Theoretical and Applied Information Technology, vol. 91, no. 1, pp. 189-201, 2016.*

[13] E. Maks, "A lexicon model for deep sentiment analysis and opinion mining applications," *Computational Linguistics, vol. 38, no. 2, pp. 245-267, 2012.*

[14] N. Shrestha and F. Nasoz, "Deep learning sentiment analysis of Amazon.com reviews and ratings," *International Journal of Soft Computing and Artificial Intelligence, vol. 8, no. 1, pp. 01-09, 2019.*

[15] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," *in Advances in Intelligent Systems and Computing, vol. 810, pp. 639-647, 2019.*

[16] B. Saberi, "Opinion mining or sentiment analysis: A survey of techniques," *International Journal of Computer Applications, vol. 168, no. 9, pp. 23-29, 2017.*

[17] Q. Yu, H. Zhao, and Z. Wang, "Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis," *in ACM International Conference Proceeding Series, pp. 48-52, 2019.*

[18] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Information Sciences, vol. 488, pp. 190-204, 2019.*

[19] W. Tan, X. Wang, and X. Xu, "Sentiment analysis for Amazon reviews using RNN-GRU models," *Research Report, pp. 1-9, March 2019.*

[20] R. S. Kumar and R. Saviour Devaraj, "Feature selection for sentiment analysis using SentiWordNet," *IEEE Access, vol. 9, pp. 45123-45138, 2021.*

[21] S. Alaparthi and M. Mishra, "BERT: A sentiment analysis odyssey," *Journal of Marketing Analytics, vol. 9, pp. 118-126, 2021.*

[22] C. Sun, X. Qiu, Y. Xu, and X. Huang, "How to fine-tune BERT for text classification," *in Lecture Notes in Computer Science, vol. 11856, pp. 194-206, 2019.*

[23] Q. T. Ghous, "Sentiment analysis of Amazon customer product reviews," *International Journal of Scientific Research Engineering & Development, vol. 4, no. 6, pp. 564-595, 2021.*

[24] C. Messaoudi, Z. Guessoum, and L. Ben Romdhane, "Opinion mining and sentiment analysis challenges: A comprehensive survey," *Expert Systems with Applications, vol. 203, pp. 117-135, 2022.*

[25] A. Abayomi-Alli, O. Abayomi-Alli, S. Misra, and L. Fernandez-Sanz, "Sentiment analysis of social media data using machine learning," *Computer Networks, vol. 182, pp. 107-121, 2022.*

[26] F. Tuz Zohra Anny and O. Islam, "Sentiment polarity classification: A comprehensive approach," *International Journal of Computer Applications, vol. 184, no. 25, pp. 15-21, 2022.*

[27] F. Tuz Zohra Anny and O. Islam, "Sentiment analysis at sentence and review level," *IEEE Transactions on Computational Social Systems, vol. 9, no. 3, pp. 765-778, 2022.*

[28] B. Lin, N. Cassee, A. Serebrenik, G. Bavota, N. Novielli, and M. Lanza, "Opinion mining for software engineering: A systematic literature review," *ACM Transactions on Software Engineering and Methodology, vol. 31, no. 3, pp. 1-58, 2022.*

[29] V. Umarani, A. Julian, and J. Deepa, "Sentiment analysis using various machine learning and deep learning techniques," *Journal of the Nigerian Society of Physical Sciences, vol. 3, no. 4, pp. 385-394, 2021.*

[30] M. Kabir, M. M. J. Kabir, S. Xu, and B. Badhon, "Sentiment analysis using machine learning approaches: A comprehensive study," *International Journal of Computer Applications, vol. 182, no. 14, pp. 15-19, 2021.*

[31] H. He, G. Zhou, and S. Zhao, "Exploring e-commerce product experience based on fusion sentiment analysis method," *IEEE Access, vol. 10, pp. 110248-110260, 2022.*