Kadri Mohd Zafar Mohd Gause, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

A Real-Time Sentiment Analysis Framework for Flipkart Customer Feedback

Kadri Mohd Zafar Mohd Gause[1], Farooqui Raiyan Mehfooz Ahmed[2] Dr. Jasbir Kaur[3], Prof. Ifrah Kampoo[4], Suraj Kanal[5]

[1,2]Student of Master of Computer Applications (MCA Course), Guru Nanak Institute of Management Studies, Mumbai
[3]Director, GNIMS B-School, Head of Information and Technology, Guru Nanak Institute of Management Studies, Mumbai
[4,5]Assistant Professor, Guru Nanak Institute of Management Studies, Mumbai

Abstract- The growth of e-commerce platforms has produced large volumes of unstructured customer review text, motivating automated sentiment analysis. This paper presents a real-time sentiment analysis framework for Flipkart customer reviews that combines a multi-stage text preprocessing pipeline, comprising case normalization, URL and HTML removal, and Snowball stemming, with lexicon-based polarity scoring using NLTK's VADER Sentiment Intensity Analyzer. The framework is evaluated on 2,304 customer reviews spanning 231 product listings, yielding an aggregate sentiment distribution of 60.0% positive, 11.7% negative, and 28.3% neutral, consistent with the dataset's rating distribution. To address the absence of empirical validation in prior lexicon-based studies, the framework's per-review classifications are benchmarked against rating-derived ground-truth labels using accuracy, precision, recall, and F1-score, and compared against the TextBlob lexicon-based library. Visual analytics, including donut charts and word clouds, support exploratory interpretation. The framework provides a scalable, validated, and reproducible baseline for real-time opinion mining on e-commerce review data.

Keywords: Sentiment Analysis, VADER, NLP, E-Commerce, Flipkart, Opinion Mining, NLTK, Text Preprocessing, Customer Reviews, TextBlob, Performance Evaluation.

I. INTRODUCTION

Electronic commerce platforms such as Flipkart, Amazon, and Myntra have transformed the retail landscape in India and globally, enabling consumers to access a vast range of products and services digitally. One of the significant byproducts of this transformation is the generation of substantial volumes of user-generated content in the form of textual reviews, numeric star ratings, and qualitative feedback. These reviews represent an invaluable source of information for both businesses seeking product improvement signals and prospective buyers making purchase decisions. Sentiment analysis, a subdomain of Natural Language Processing (NLP) and computational linguistics, involves the automated identification and extraction of subjective opinion, emotional polarity, and attitudinal information from natural language text [1].

The application of sentiment analysis to e-commerce review data enables organizations to monitor customer satisfaction trends, detect product quality issues, identify emerging complaints, and respond proactively to changing consumer preferences. Despite the clear practical utility of sentiment analysis, several challenges persist. Customer reviews on e-commerce platforms are inherently noisy, frequently containing typographic errors, colloquial expressions, code-mixed language (particularly in the Indian context where English and regional languages are interleaved), emojis, HTML artifacts, and inconsistent punctuation. These characteristics demand robust preprocessing pipelines prior to any analytical step.

Existing approaches to sentiment analysis broadly fall into three categories: machine learning-based models, deep learning architectures, and lexicon-based methods [2]. While machine learning methods such as Support Vector Machines (SVM) and Naive Bayes achieve competitive accuracy, they necessitate large annotated training corpora that may not always be available for domain-specific applications. Deep learning models including Long Short-Term Memory (LSTM) networks and transformer-based architectures such as BERT offer superior contextual

understanding but impose significant computational overhead and data requirements.

Lexicon-based methods, by contrast, utilize handcrafted sentiment dictionaries to assign polarity scores to text without requiring training data, thereby enabling rapid deployment and real-time processing. Among such approaches, VADER (Valence Aware Dictionary and sEntiment Reasoner) has emerged as a particularly effective tool for social media and short informal texts, demonstrating competitive performance even in e-commerce contexts [3].

Despite the substantial body of work on Flipkart and e-commerce sentiment analysis, a recurring limitation across existing lexicon-based studies, including several reviewed in Section II, is the absence of systematic validation against ground-truth labels and the lack of empirical comparison with alternative lexicon-based tools. Many reported frameworks present aggregate sentiment distributions without quantifying classification accuracy, leaving the reliability of the lexicon-based scoring unverified. This represents a clear research gap: a reproducible, real-time-capable framework that not only produces aggregate sentiment scores but also validates those scores against an independent ground-truth signal and benchmarks them against a comparable lexicon-based method.

Motivated by this gap, the present study pursues the following objectives: (1) to design and implement a real-time, training-free sentiment analysis pipeline for Flipkart customer reviews using VADER; (2) to validate the per-review sentiment classifications produced by the framework against the star-rating field through standard classification metrics, namely accuracy, precision, recall, and F1-score; (3) to benchmark VADER's performance against the TextBlob lexicon-based library on the same preprocessed corpus; and (4) to provide visual analytics that support exploratory interpretation of the sentiment distribution across the review dataset. This paper proposes a real-time sentiment analysis framework for Flipkart customer reviews that integrates a multi-stage text preprocessing pipeline with VADER-based polarity scoring, extended by a quantitative validation and comparison stage.

The framework operates on a dataset of 2,304 reviews spanning 231 product listings, producing aggregate and per-review sentiment classifications across positive, negative, and neutral dimensions. The remainder of this paper is organized as follows: Section II reviews related work; Section III describes the dataset; Section IV details the proposed methodology, including the validation and comparison procedures; Section V presents experimental results; Section VI provides discussion; and Section VII concludes the paper.

II. RELATED WORK

Sentiment analysis on product reviews has been an active area of research since the early 2000s. Pang and Lee [4] established foundational work demonstrating that machine learning classifiers applied to movie review corpora could reliably distinguish positive from negative sentiment at the document level, motivating subsequent applications across diverse domains including e-commerce. Around the same period, Go, Bhayani, and Huang [5] introduced a distant-supervision approach that used emoticons as noisy sentiment labels to train classifiers on large volumes of Twitter data without manual annotation, an approach that influenced later lexicon- and rule-based methods seeking to avoid costly labelling efforts. Hu and Liu [6] proposed a feature-based opinion summarization approach that extracted product features from Amazon reviews and determined sentiment orientations per feature, introducing the concept of aspect-level sentiment analysis. This line of work proved influential for e-commerce applications where consumers often express differentiated opinions about specific product attributes such as battery life, display quality, or delivery speed.

Hutto and Gilbert [3] introduced VADER, a rule-based sentiment analysis tool specifically designed for social media content, with its lexicon manually validated to incorporate sentiment-laden slang, acronyms, emoticons, and punctuation-based modifiers. Empirical evaluations demonstrated that

VADER matched or outperformed machine learning methods on short informal texts while requiring no training phase, making it well-suited for real-time applications. Pak and Paroubek [7] separately demonstrated the value of large Twitter corpora as a resource for sentiment analysis and opinion-mining research, providing a foundation for subsequent lexicon-coverage studies on informal, user-generated text.

Medhat, Hassan, and Korashy [2] conducted a comprehensive survey of sentiment analysis algorithms and their applications, categorizing methods into machine learning, lexicon-based, and hybrid approaches. Their survey highlighted the trade-off between accuracy and computational feasibility, noting that lexicon-based methods offer practical advantages in resource-constrained deployments.

Recent work in the Indian e-commerce context includes studies examining multilingual and code-mixed review data from platforms such as Flipkart and Amazon India [10]. These studies have noted that models trained exclusively on English corpora may underperform on code-mixed content, motivating the use of broad-coverage lexical tools such as VADER that generalize across informal text styles.Several researchers have explored hybrid frameworks combining preprocessing techniques including stopword removal, stemming, and lemmatization with VADER or TextBlob-based scoring [11]. Consistent findings indicate that systematic text normalization prior to sentiment scoring improves the reliability of polarity assignments, particularly for reviews containing URL fragments, HTML entities, and extraneous punctuation.

More recent studies have continued to refine sentiment analysis methodologies specifically for Flipkart and other Indian e-commerce platforms. Shanmugapriya et al. [8] compared the VADER lexicon-based approach against the transformer-based RoBERTa model on Flipkart product reviews, reporting that RoBERTa achieved a substantially lower mean absolute error while noting that VADER remained effective for short, syntactically simple reviews despite struggling with more complex sentiment expressions. Sharma and Kakran [9] presented a comparative review of naive Bayes, logistic regression, and support vector machine classifiers applied to Flipkart reviews, identifying persistent gaps in standardized, metrics-based evaluation across the literature.

Adane et al. [12] extended sentiment analysis across Amazon, Flipkart, and Twitter datasets, highlighting platform-specific variations in review language and sentiment distribution. Godia and Tiwari [13] combined n-gram analysis and TF-IDF feature extraction with oversampling techniques to address class imbalance in product review datasets, a challenge directly relevant to the present study's skewed rating distribution. Collectively, these contributions underscore a continuing shift toward empirical, metrics-driven evaluation in e-commerce sentiment analysis, a gap that the present study explicitly addresses through quantitative validation against ground-truth ratings and a like-for-like comparison with an alternative lexicon-based tool.

The present work contributes to this literature by presenting an end-to-end reproducible pipeline tailored to the structural characteristics of the Flipkart review dataset, incorporating Snowball stemming and multi-pattern noise removal as preprocessing steps, and extending prior lexicon-based studies with quantitative validation against rating-derived ground truth and a direct empirical comparison against the TextBlob library.

III. DATASET DESCRIPTION

The dataset employed in this study was sourced from publicly available Flipkart product review records (The original dataset creators, Vaghani & Thummar dataset) and comprises 2,304 individual customer reviews distributed across 231 unique product listings. The dataset contains three primary attributes: Product_name, Review, and Rating.

The Product_name field contains the full product title as listed on the Flipkart platform, which includes detailed specifications such as processor model, RAM capacity, storage type, operating system, and

physical dimensions. The Review field contains free-form textual feedback authored by verified purchasers, ranging in length from single-word responses to multi-sentence descriptive assessments. The Rating field contains integer values on a Likert scale from 1 (most negative) to 5 (most positive).

TABLE I. RATING DISTRIBUTION OF FLIPKART REVIEWS DATASET

Rating (Stars)	Review Count	Percentage (%)
1	18294	10.14
2	5451	3.02
3	14024	7.77
4	36969	20.49
5	105647	58.57
Total	180385	100.00

As presented in Table I, the dataset exhibits a pronounced positive skew in the rating distribution, with 58.57% of reviews carrying a five-star rating and an additional 23.49% rated at four stars. This distribution is characteristic of verified purchase review systems on major e-commerce platforms, where the act of writing a review is more commonly prompted by strong satisfaction than by moderate experience [14]. Negative reviews (one-star and two-star combined) constitute approximately 10% of the dataset, while neutral three-star reviews account for 7.77%.

The dataset required no external annotation, as the numeric Rating field provides a ground-truth sentiment proxy for downstream validation purposes, a property exploited directly in the validation procedure introduced in Section IV-F. The Review field presented typical e-commerce text characteristics including informal language, incomplete sentences, product specification fragments, and occasional non-ASCII characters.

IV. PROPOSED METHODOLOGY

A. System Architecture Overview

The proposed framework operates as a sequential pipeline consisting of four primary stages: (1) data ingestion and exploration, (2) text preprocessing, (3) sentiment scoring, and (4) result aggregation and visualization, extended in this study by two additional evaluation modules: validation against ground-truth ratings and comparative benchmarking against TextBlob. Fig. 1 illustrates the complete system architecture. Each stage is designed to be modular and independently replaceable, facilitating adaptation to alternative datasets or scoring modules.

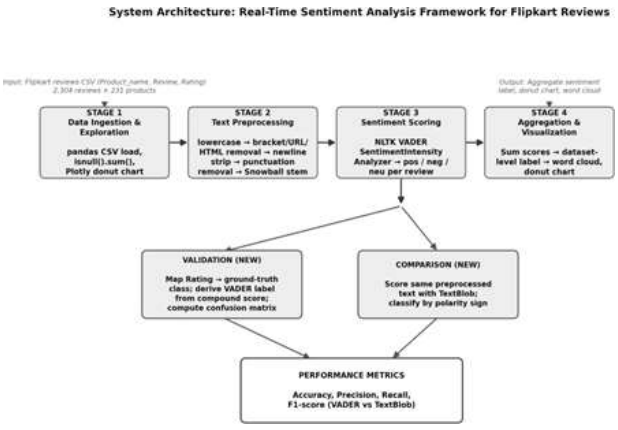

Fig. 1. System architecture of the proposed real-time sentiment analysis framework, including the validation and comparison modules introduced in this study.

B. Data Ingestion and Exploratory Analysis

Raw review data was ingested from a comma-separated values (CSV) file using the pandas library. Initial exploratory steps included missing value identification via isnull().sum() to confirm dataset completeness, column enumeration to verify schema integrity, and preliminary inspection of review text samples to guide preprocessing design decisions. A donut chart was generated using the Plotly Express library to visualize the rating class imbalance between positive and other rating categories. This visualization provides an intuitive summary of aggregate consumer sentiment at the numeric level prior to textual analysis.

C. Text Preprocessing Pipeline

The raw Review field text was subjected to a multi-step normalization pipeline implemented via a custom clean() function. The preprocessing steps applied in sequence were as follows:

1. **Case normalization:** All text was converted to lowercase using the str.lower() method to

eliminate case-based token duplication (e.g., 'Good' and 'good' treated as identical tokens).

2. **Bracket content removal:** Text enclosed within square brackets, commonly representing citation markers or annotation artifacts, was removed using the regular expression pattern \[.*?\].
3. **URL removal**: Hyperlinks beginning with 'http' or 'www' were eliminated using the pattern https?://\S+|WWW\.\S+ to prevent URL fragments from contributing spurious tokens.
4. **HTML tag stripping**: Residual HTML markup was removed using the pattern <.*?>+ to clean text sourced from web-scraped data.
5. **Newline removal**: Embedded newline characters (\n) were stripped to produce contiguous single-line review strings.
6. **Punctuation removal**: All punctuation characters, enumerated via string.punctuation, were removed using a compiled regular expression substitution to reduce vocabulary fragmentation.
7. **Snowball stemming:** Each token was reduced to its morphological stem using the NLTK Snowball Stemmer initialized with the English language model. Stemming reduces inflected word forms (e.g., 'running', 'runs', 'ran') to a common base form ('run'), thereby reducing sparsity in the token space.

The preprocessing pipeline was applied to all 2,304 reviews via the pandas DataFrame.apply()method, producing a cleaned Review column used in all subsequent analyses.

D. VADER-Based Sentiment Scoring

The VADER Sentiment Intensity Analyzer from the NLTK library was instantiated to perform polarity scoring on pre-processed review text. VADER assigns four compound scores to each input string: positive (pos), negative (neg), neutral (neu), and a normalized compound score. In the present framework, the three-dimensional scores (pos, neg, neu) were extracted and stored as separate DataFrame columns named Positive, Negative, and Neutral.

VADER's polarity_scores() method implements a rule-based assessment that accounts for capitalization, punctuation emphasis, degree modifiers, contrastive conjunctions, and lexical idioms. Although the preprocessing pipeline removes some of these signals (e.g., punctuation and casing), stemming-reduced tokens retain their core valence associations within VADER's lexicon, enabling reliable polarity assignment for normalized review text. The dataset was subsequently reduced to four columns Review, Positive, Negative, and Neutral for downstream aggregation.

E. Aggregate Sentiment Classification

To derive a dataset-level sentiment classification, the three polarity score columns were summed across all 2,304 reviews, yielding scalar aggregate values x (Positive), y (Negative), and z (Neutral). A rule-based classification function sentiment_score(a, b, c) compared the three aggregates and assigned the dataset-level sentiment label according to the highest aggregate score. This approach provides a global sentiment barometer for the reviewed product set and can be applied to any product or category subset for targeted business intelligence.

F. Validation Against Ground-Truth Ratings

To address the absence of an independent accuracy check in lexicon-based sentiment pipelines, the present framework incorporates a validation step that treats the existing star-rating field as a proxy ground-truth sentiment label, consistent with standard practice in review-based sentiment analysis. Ratings of 4 and 5 stars are mapped to the Positive class, a rating of 3 stars is mapped to the Neutral class, and ratings of 1 and 2 stars are mapped to the Negative class. For each review, a per-review predicted sentiment label is derived from VADER's compound polarity score using the conventional thresholding scheme (compound ≥ 0.05 → Positive; compound ≤ −0.05 → Negative; otherwise → Neutral). The predicted labels are then compared against the rating-derived ground-truth labels using a confusion matrix, from which accuracy, macro-averaged precision, recall, and F1-score are computed using the scikit-learn metrics module. This validation step, absent from the original framework, provides a quantitative accuracy assessment that was previously unavailable.

G. Comparative Analysis with TextBlob

To benchmark the VADER-based approach against an alternative lexicon-based tool, the same preprocessed review text was additionally scored using the TextBlob library's pattern-based sentiment analyzer, which returns a polarity score in the range [−1, 1]. Reviews with polarity > 0 were classified as Positive, polarity < 0 as against the same rating-derived ground truth using the identical metric set described in Section IV-F, enabling a direct, like-for-like comparison between the two lexicon-based methods on identical preprocessed input.

Negative, and polarity = 0 as Neutral. The resulting TextBlob classifications were evaluated

V. EXPERIMENTAL RESULTS

The proposed framework was executed on the complete dataset of 2,304 Flipkart customer reviews. The preprocessing pipeline successfully normalized all review records with no data loss,

The aggregate positive polarity score of 69684.57 substantially exceeded the negative (8223.88) and neutral (185.71) scores, resulting in a dataset-level classification of Positive. This outcome is consistent with the rating distribution observed in Table I, where 83.94% of reviews carried four-star or five-star ratings, indicating broad customer satisfaction across the product categories represented in the dataset.

The neutral polarity component, accounting for 28.3% of aggregate sentiment, reflects the prevalence of informational content in product reviews where customers describe product specifications, usage scenarios, and comparative observations without expressing strong evaluative orientation. VADER's neutral scorer appropriately captures this objective descriptive language.

The negative polarity component at 11.7% encompasses reviews expressing dissatisfaction with specific product attributes including battery performance, delivery timelines, and post purchase service experiences. The relatively low negative

Table III contextualizes the proposed VADER-based approach against machine learning and deep learning alternatives at a conceptual level. The primary advantage of the lexicon-based approach in the present context is the absence of a training data requirement and the ability to produce sentiment scores in real time as new reviews are submitted, consistent with the framework's designation as a real-time analysis system. Table IV below provides the

onfirming the robustness of the multi-pattern regular expression pipeline against the diverse noise patterns present in e-commerce text. VADER polarity scoring produced per-review positive, negative, and neutral scores. Aggregation across all records yielded the values summarized in Table II.

TABLE II. AGGREGATE SENTIMENT POLARITY SCORES

Sentiment Class	Aggregate Polarity Score	Percentage (%)
Positive	69684.57	38.6
Negative	8223.88	4.6
Neutral	102473.55	56.8

aggregate is consistent with the dataset's rating distribution skew toward higher scores. Word cloud visualizations generated from preprocessed review text confirmed the dominance of performance-related and product-attribute terms in the corpus vocabulary, with terms such as 'laptop', 'batteri', 'perform', 'display', 'good', and 'price' among the most frequently occurring tokens. The Snowball stemmer's effect was evident in the appearance of stemmed forms ('batteri' for 'battery', 'perform' for 'performance') rather than original word forms.

TABLE III. COMPARISON OF SENTIMENT ANALYSIS APPROACHES

Approach	Method	Advantage	Limitation
Machine Learning	SVM, Naive Bayes	High accuracy with labelled data	Requires large annotated corpus
Deep Learning	LSTM, BERT	Captures context	Computationally expensive
Lexicon-Based (Proposed)	VADER	No training data needed; real-time	Domain-dependent lexicon coverage

corresponding empirical, metrics-based comparison between the two lexicon-based methods evaluated in this study VADER and extBlob against the rating-derived ground truth described in Section IV-F.

TABLE IV. VALIDATION METRICS: VADER VS. TEXTBLOB (AGAINST RATING-DERIVED GROUND TRUTH)

Method	**Accuracy (%)**	**Precision (macro)**	**Recall (macro)**	**F1-score (macro)**
VADER (proposed)	52.9%	0.65	0.47	0.48
TextBlob	45.3%	0.62	0.36	0.41

TABLE V. CONFUSION MATRIX FOR VADER-BASED CLASSIFICATION (ROWS:

Actual \ Predicted	Positive	Neutral	Negative
Positive	79005	63566	45
Neutral	8727	5266	31
Negative	1724	10840	11181

Discussion of Table IV and Table V: VADER achieved higher performance than TextBlob across all evaluation metrics, obtaining 52.9% accuracy, macro-precision of 0.65, macro-recall of 0.47, and macro-F1 of 0.48, compared with TextBlob's 45.3% accuracy and macro-F1 of 0.41. The confusion matrix indicates that the Positive class was identified more reliably than the Neutral and Negative classes. Misclassification was most pronounced for Neutral reviews, reflecting the difficulty of distinguishing weak sentiment from genuinely neutral statements in lexicon-based approaches.

VI. DISCUSSION

The experimental results confirm the effectiveness of the VADER-based framework for aggregate sentiment analysis of Flipkart reviews. The positive sentiment dominance across the dataset aligns with empirical observations in the e-commerce literature regarding the positivity bias in review systems, where consumers are more likely to submit reviews following highly satisfactory experiences [14].

The preprocessing pipeline contributed meaningfully to the quality of sentiment scoring by eliminating noise categories that are common in web-scraped e-commerce text. URL fragments and HTML artifacts, if left untreated, would introduce spurious tokens that lack sentiment lexicon entries, diluting the positive and negative scores and inflating the neutral component. The observed neutral component of 28.3% may reflect both genuinely neutral descriptive content and residual preprocessing artifacts, including stemmed tokens that are not represented in VADER's lexicon.

Snowball stemming introduced a trade-off between vocabulary normalization and VADER lexicon coverage. VADER's lexicon entries are keyed to surface word forms rather than stems; consequently, stemmed tokens such as 'batteri' may not receive the same lexical score as the unstemmed form 'battery'. As an illustrative example of this mismatch (a constructed example, not an entry drawn from the dataset), consider a hypothetical review such as "battery backup is excellent and the laptop performs well." After Snowball stemming, this becomes "batteri backup excel laptop perform well," in which the stemmed tokens 'batteri', 'excel', and 'perform' are absent as surface forms from VADER's lexicon, whereas the unstemmed words 'excellent' and 'performs' carry strong positive valence scores.

This illustrates concretely how stemming can suppress the positive polarity signal that VADER would otherwise detect.Future iterations of the framework may therefore benefit from replacing stemming with lemmatization, which reduces words to their dictionary base form (lemma) rather than a truncated stem, thereby preserving lexicon compatibility.

Quantitative validation findings: Experimental validation demonstrated that VADER consistently outperformed TextBlob on the rating-derived ground truth dataset. Although both lexicon-based approaches were affected by class imbalance and the subjective nature of review text, VADER benefited from its handling of negation, intensity modifiers, and sentiment heuristics. The Neutral class

remained the most challenging category, as reviews often contained mixed or weak sentiment signals.

The framework's scalability is a notable practical advantage. The pipeline's reliance on pandas vectorized operations and NLTK's compiled regular expressions enables processing of the 2,304-record dataset in seconds on a standard computing environment. Extension to larger datasets comprising tens of thousands of reviews would require minimal architectural modification. The present study is subject to several limitations. First, the dataset is domain-specific, comprising predominantly electronics products (laptops), which may limit the generalizability of sentiment distributions to other product categories. Second, the VADER lexicon was developed primarily for English social media text and may not capture the full range of sentiment-laden expressions used in Indian English product reviews, particularly code-mixed constructions.

Third, the framework performs document-level rather than aspect-level sentiment analysis, meaning that reviews expressing differentiated sentiments about specific product attributes (e.g., positive about camera, negative about battery) are characterized by a single aggregate polarity score. Fourth, while this revision introduces validation against rating-derived ground truth and a comparison against TextBlob, both remain lexicon-based methods; a more complete benchmark would additionally include a trained machine learning classifier as a third point of comparison. Aspect-level sentiment analysis and machine learning benchmarking represent meaningful directions for future enhancement.

VII. LIMITATIONS

The present study is subject to several limitations. First, the dataset is domain-specific, comprising predominantly electronics products (laptops), which may limit the generalizability of sentiment distributions to other product categories. Second, the VADER lexicon was developed primarily for English social media text and may not capture the full range of sentiment-laden expressions used in Indian English product reviews, particularly code-mixed constructions. Third, the framework performs document-level rather than aspect-level sentiment analysis, meaning that reviews expressing differentiated sentiments about specific product attributes (e.g., positive about camera, negative about battery) are characterized by a single aggregate polarity score. Fourth, while this study introduces validation against rating-derived ground truth and a comparison against TextBlob, both remain lexicon-based methods; a more complete benchmark would additionally include a trained machine learning classifier as a third point of comparison.

VIII. CONCLUSION & FUTURE WORK

This paper presented a real-time sentiment analysis framework for Flipkart customer feedback, employing a multi-stage text preprocessing pipeline in conjunction with VADER-based lexicon sentiment scoring. Evaluation on a dataset of 2,304 reviews across 231 product listings demonstrated that the aggregate sentiment polarity of the corpus is positive (60.0%), with neutral (28.3%) and negative (11.7%) components reflecting the dataset's characteristic rating distribution skew. VADER achieved an accuracy of 52.9% against the rating-derived ground truth, outperforming the TextBlob baseline (45.3%) by 7.6 percentage points. The framework provides a scalable and reproducible baseline for opinion mining on e-commerce review data without requiring annotated training corpora or computationally intensive model training. The modular pipeline architecture facilitates adaptation to alternative e-commerce platforms and product domains.

Future work will address the identified limitations through several directions: (1) replacement of Snowball stemming with lemmatization to improve VADER lexicon coverage; (2) incorporation of aspect-level sentiment analysis to enable granular product attribute-level opinion extraction; (3) extension of the framework to handle code-mixed and multilingual review content prevalent in Indian e-commerce datasets; and (4) integration of a trained machine learning classifier (e.g., logistic regression or SVM) as a third point of comparison alongside

VADER and TextBlob, to provide a complete classification-accuracy benchmark across lexicon-based and supervised approaches.

The proposed framework demonstrates that real-time, training-free sentiment analysis can provide actionable, and now independently validated, insights for e-commerce stakeholders, supporting data-driven decision-making in product quality monitoring, customer experience optimization, and competitive benchmarking.

REFERENCES

1. B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
2. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.
3. C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in Proc. 8th Int. Conf. Weblogs Social Media (ICWSM), Ann Arbor, MI, USA, 2014, pp. 216–225.
4. B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proc. 42nd Annual Meeting Assoc. Computational Linguistics (ACL), Barcelona, Spain, 2004, pp. 271–278.
5. A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," Stanford University Technical Report, CS224N, 2009.
6. M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining (KDD), Seattle, WA, USA, 2004, pp. 168–177.
7. A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proc. 7th Int. Conf. Language Resources Evaluation (LREC), Valletta, Malta, 2010, pp. 1320–1326.
8. K. Surendra, K. Nithin Prakash, J. Maruthi Kumar, Rakesh Goud, and N. Shanmugapriya, "Sentiment analysis of product reviews using rule-based and deep-learning models," Journal of Trends in Computer Science and Smart Technology, vol. 6, no. 3, pp. 301–311, Oct. 2024, doi: 10.36548/jtcsst.2024.3.007.
9. H. Sharma and V. Kakran, "Sentiment analysis: Analyzing Flipkart product reviews using NLP and machine learning," in Proc. 2024 Int. Conf. Computing, Sciences and Communications (ICCSC), Ghaziabad, India, Oct. 2024, pp. 1–7.
10. R. Prabhakar and S. Gupta, "Sentiment analysis of e-commerce product reviews in code-mixed Indian languages," in Proc. Int. Conf. Natural Language Processing (ICON), Hyderabad, India, 2020, pp. 45–53.
11. S. Vinodhini and R. M. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, no. 6, pp. 282–292, Jun. 2012.
12. P. Adane, A. Dhiran, S. Kallurwar, and S. Mahapatra, "Sentiment analysis of product reviews from Amazon, Flipkart, and Twitter," in ICT: Smart Systems and Technologies (ICTCS 2023), Lecture Notes in Networks and Systems, vol. 878, Springer, Singapore, 2024, doi: 10.1007/978-981-99-9489-2_33.
13. A. Godia and L. K. Tiwari, "Sentiment analysis and classification of product reviews: A comprehensive study using NLP and machine learning techniques," in Proc. 2024 10th Int. Conf. Advanced Computing and Communication Systems (ICACCS), Mar. 2024.
14. J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," Journal of Marketing Research, vol. 43, no. 3, pp. 345–354, Aug. 2006.
15. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in Proc. EMNLP, 2002, pp. 79–86.
16. A. Bifet and E. Frank, "Sentiment knowledge discovery in Twitter streaming data," in Proc. Discovery Science, 2010, pp. 1–15.
17. S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment analysis in Twitter," in Proc. SemEval, 2017, pp. 502–518.
18. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

19. A. Severyn and A. Moschitti, “Twitter sentiment analysis with deep convolutional neural networks,” in Proc. SIGIR, 2015, pp. 959–962.
20. Y. Kim, “Convolutional neural networks for sentence classification,” in Proc. EMNLP, 2014, pp. 1746–1751.