Sentiment analysis in Amazon reviews for binary classification of user comments

Stephanie Batista[1], Aranzazu Berbey-Alvarez[1], Gustavo Quintero[1], Arkadiusz Kampczyk[2]
[1]*Universidad Tecnológica de Panamá, Panama,* Stephanie.batista@utp.ac.pa, Aranzazu.berbey@utp.ac.pa, Gustavo.quintero@utp.ac.pa
[2] AGH University of Krakow, Poland, kampczyk@agh.edu.pl

***Abstract*—This paper presents a comprehensive sentiment analysis study applied to Amazon product reviews, using Natural Language Processing (NLP) techniques and supervised machine learning. The dataset comes from the Amazon Product Reviews dataset available on Kaggle. The workflow included text cleaning, normalization, and vectorization using Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) methods. TF-IDF outperformed BoW by emphasizing the relative weight of less-frequent but discriminative terms. Two supervised classifiers — Multinomial Naive Bayes and Logistic Regression — were trained and compared based on accuracy, precision, recall, and F1-score. Results show that Logistic Regression achieved superior performance, with an accuracy of 87% and an F1-score of 0.85, outperforming Naive Bayes (0.77 and 0.67 respectively). These findings highlight the effectiveness of classical models in opinion mining for e-commerce environments and underline their relevance compared to more complex architecture such as BERT and deep neural networks.**

Index terms*— *Amazon, e-commerce reviews, feature extraction, logistic regression, machine learning, Naive Bayes, NLP, opinion mining, sentiment analysis, text classification

I. Introduction

The growth of e-commerce has profoundly transformed consumer dynamics and how users express their opinions about products and services. Platforms like Amazon generate millions of reviews every day, constituting an invaluable source of information on customer perception and satisfaction[1][2]. When properly analyzed, these reviews allow for the detection of behavioral patterns, the evaluation of the quality of a product or service[3], rating prediction performance [4], opinion mining [5][6] and the improvement of data-driven marketing strategies[7][8][9], [10], [11], [12].

Sentiment analysis (SA) also known as opinion mining, represents one of the most relevant areas within Natural Language Processing (NLP), as it focuses on determining the polarity of opinions expressed in text, generally classifying them as positive, negative, or neutral[13][14]. Its application in the context of e-commerce makes it possible to identify customer satisfaction trends and design automated recommendation and response systems [15][16][6].

In the digital era, online reviews play a crucial role in influencing consumer decisions and shaping the reputation of products and services. Among e-commerce platforms, Amazon stands out as one of the largest sources of user-generated feedback, where millions of customers share their experiences through ratings and written comments. However, the enormous volume of reviews makes it difficult for users and companies to efficiently identify the most relevant and informative opinions. Sentiment analysis has emerged as an important technique in the field of Natural Language Processing (NLP) to automatically detect and classify emotions expressed in textual data. By analyzing emotional patterns in user reviews, it is possible to better understand customer satisfaction, preferences, and concerns. Furthermore, incorporating emotional information into review evaluation systems can improve the automatic ranking of user comments, highlighting reviews that provide more meaningful and trustworthy insights.

While deep learning models achieve higher accuracy, their computational demands limit deployment in resource-constrained academic settings. This work shows preliminary results as an academic exercise of the subject called *Proyecto integrador III* (Integrative Project III), which is a course of engineering inside a research project in a master's program of the Faculty of Industrial Engineering.

This work aims to compare the performance of two widely used supervised learning algorithms for text classification: Multinomial Naive Bayes and Logistic Regression. Through a reproducible experimental approach, the paper seek to determine which of these models performs better in classifying Amazon reviews, considering the accuracy and stability of the results. In particular, we evaluate the effectiveness of classical models against more recent deep learning methods, analyzing their advantages in terms of simplicity, interpretability, and computational efficiency. This study contributes to the understanding of how traditional methods can remain highly competitive in large-scale opinion mining scenarios.

The first section of this article is the introduction, which presents the problem, the motivation, and the objective of the comparison between classifiers. The second section reviews the scientific literature on sentiment analysis and related works. The third section describes the research methodology. This methodology combines the PRISMA inquiry framework with mathematical analysis methods such as Naive Bayes and Logistic Regression, using the Amazon Product Reviews dataset obtained from the Kaggle platform. The fourth section presents the preliminary results, namely, the comparative performance of the Naive Bayes and Logistic Regression models in sentiment analysis, and the interpretation of the confusion matrix for both models. The fifth section discusses the preliminary results and the limitations of this study. Finally, the conclusions and potential future research are presented.

II. LITERATURE REVIEW

According IA[17], Sentiment analysis is "*defined as the process of identifying and analyzing subjective information in user-written text, enabling the automatic prediction of sentiment as "positive," "negative," or "neutral." It is widely applied in various business contexts, such as movie and product reviews, to assess opinions and influence revenue*". In Tonkin [18], Sentiment analysis is defined as "*a specific subtask within the broad area of opinion mining; in short, the classification of texts according to the emotion that the text appears to convey. Sentiment analysis typically classifies texts according to positive, negative and neutral classification*"
Sentiment analysis has evolved considerably over the past two decades, driven by advances in Natural language processing (NLP) and machine learning. These disciplines have enabled the automation of human language comprehension and the conversion of large volumes of text into structured and useful information [1][2]. One of the foundations of this progress has been the development of vector representations such as Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF), which transform text into numerical matrices suitable for statistical models [19][20]. These techniques, although simple, have demonstrated remarkable effectiveness in sentiment classification tasks and remain widely used due to their interpretability and computational efficiency[13][21].

In terms of classification algorithms, Logistic Regression has established itself as one of the most robust methods for text categorization, offering better generalization capabilities compared to probabilistic models like Naive Bayes, especially with high-dimensionality datasets or large vocabularies [14][22][23]. However, Naive Bayes remains relevant due to its simplicity, training speed, and low resource consumption, qualities that make it ideal for real-time applications and computationally limited environments [24] [8].
In recent years, the rise of deep learning has introduced more advanced architectures, such as Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT), which have significantly improved accuracy in sentiment analysis tasks [25][26]. These models are capable of capturing complex contextual dependencies, detecting semantic nuances, and recognizing figurative or sarcastic expressions, something that classical models do not achieve with the same effectiveness [22][27]. However, their implementation requires considerable computational resources and technical expertise, which limits their applicability in academic projects or resource-constrained environments[1][2].

Another critical aspect highlighted in the literature is text preprocessing, an essential stage for ensuring the quality of the final model. Processes such as noise removal, normalization, tokenization, and stop word removal are crucial for the success of classifiers [28][29]. Likewise, the appropriate selection of features and word representation directly influences the model's stability and its ability to generalize results [24][14].

In this context, the present study adopts an approach based on best practices in text mining and supervised learning, with the aim of comparing two classic algorithms—Naive Bayes and Logistic Regression—within a reproducible experimental framework. This review situates the research within the current landscape of sentiment analysis, combining established techniques and robust methodologies that prioritize accuracy, interpretability, and efficiency in language processing.

III. METODOLOGY

This section describes the dataset (Kaggle), the preprocessing (NLTK), the vector representations (BoW and TF-IDF), the 80/20 partitioning, the cross-validation training, and the evaluation with standard metrics. Seven specific publications databases- google scholar, ResearchGate, IEEE, Science Direct, Elsevier, Web of Sciences and Scopus, were used to identify the research gap of this manuscript (See table 1).

TABLE I. Description of databases for source

Data base	Justification
Scopus (1)	Scopus Largest indexed database in engineering and CS; allows advanced searches by TITLE-ABS-KEY. The paper itself used most extensively.
Web of Science (WoS) (2)	Web of Science (WoS) High impact factor; preferred for citation metrics (h-index, JIF).
IEEE Xplore (3)	IEEE Xplore essential for NLP and ML in the IEEE context; direct access to conference proceedings.
ScienceDirect / Elsevier (4)	ScienceDirect / Elsevier Excellent coverage of Elsevier journals; relevant for Computers & Electrical Engineering papers.
Google Scholar (5)	Google Scholar Largest overall coverage but google scholar don't have the quality control like Scopus, IEEE and Elsevier. However, it is useful for grey literature and theses.
ResearchGate (6)	ResearchGate Academic is a popular known social network for academics, scientists and researchers; useful for preprints but not recommended as unique formal source for systematic reviews in comparison with Scopus, IEEE and science direct.

The eligibility criteria for the articles used in the review are as follows:

- The article must be in English in open access.
- The articles must be published in a reliable journal with impact factors.
- The articles must focus on analyzing emotions in Amazon reviews for automatic ranking of user texts.
- Also, the research considered related studies, report, thesis, technical documents and EN guidelines about this topic.
- The article must have been published in the past 7 years (2019–2026).

The search did correspond to the years 2019 to June 2026 in Scopus database(n=22). The chain´s search was:

(TITLE-ABS-KEY (sentiment analysis) AND TITLE-ABS-KEY (opinion mining) OR TITLE-ABS-KEY (text classification) AND TITLE-ABS-KEY (amazon data)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (EXACTKEYWORD , "Sentiment Analysis")) AND (LIMIT-TO (OA , "publisherfullgold")) AND PUBYEAR > 2018 AND PUBYEAR < 2026

The table II presents a review about sentiment analysis of a wide of amazon product reviews [30] [31] [32] [33] [34] [35] [36] 46]like: restaurants[37], mobile[38] [39] [40] [41], air pods, fine foods [42], DVD[43], Electronics[43] [44], and Kitchenware[43], as watches[40], mobile[40], shoes[40], clothes[40], and perfumes[40], cuisine[45], IMDb, TripAdvisor47], and cricket[46].

TABLE II. Brief of sentiment analysis review

This study was conducted using the Amazon Product Reviews dataset, obtained from the Kaggle platform, which contains thousands of user-written product reviews, accompanied by ratings from one to five stars, titles, and relevant metadata [52]. This dataset provided a solid foundation for sentiment analysis, as it included both positive and negative comments across a wide range of product categories. This study excluded of 3-star reviews because they are neutral reviews. The decision to exclude 3-star reviews is justified because neutral ratings often contain mixed or ambiguous sentiments, making sentiment classification less reliable. Prior research has shown that reviews with extreme ratings provide clearer opinion signals and reduce labeling ambiguity, thereby improving the validity and interpretability of sentiment analysis results according Pang & Lee[13], Liu[2] and Maas et al.,[53].

The data preparation and preprocessing process was crucial to ensuring the quality of the analysis. First, null values were removed, ensuring that only unique reviews were considered. Subsequently, reviews with a three-star rating were excluded, as these typically represent neutral opinions, thus reducing computational effort. The review text was converted to lowercase, and punctuation, numbers, and stop words were removed using tools from the natural language toolkit (NLTK) package [28]. This process allowed for the normalization of the text corpus and the reduction of semantic noise.

1: BERT 2: Ontology 3: Logistic regression(Machine learning) 4: Latent Dirichlet Allocation (LDA) technique 5: Tabu Search (TS) 6: Stream Evolution Dynamics (SED) 7: NRC-Emotion Lexicon approach 8: TF-IDF 9: Long Short-Term Memory (LSTM) 10: Convolutional Neural Networks (CNN) 11: review 12: NLP 13: Random Forest (RF), 14: Support Vector Machine (SVM), 15: Naïve Bayes (NB) 16: Gradient Boosting Machine (GBM), 17: deep learning (dl) 18: ROBERTa 19: PCA 20: GSK 21: Recurrent neural networks (RNN) 22: Possibilistic Fuzzy C-Means (PFCM)

Authors	Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Context
Gueddes y Mahjoub[37]	2026	x	x																					Restaurant benchmark, product reviews (Amazon) and microblogs (Twitter)
Daraghmi and Zyadeh[38]	2026			x						x	x			x	x	x	x	x	x					Food Reviews and Unlocked Mobile Reviews
Lakshmi et al., [30]	2025							x														x		Amazon product reviews
Firza et al., [42]	2025								x															Consumer Electronics Evaluation (Amazon's AirPods reviews)
Nabil et al., [47]	2025									x	x													IMDB, Amazon, and Amazon Fine Foods datasets
Malik and Bilad[39]	2024											x												Fine Food Reviews and Unlocked Mobile Reviews
Pradeepa et al., [31]	2024			x	x																			Amazon review data set
Rajagopal et al.,[32]	2024					x	x																	Big data
Dhamayanth and Lavanya[43]	2024																			x				benchmark datasets from Amazon, including DVD, Electronics, and Kitchenware
Kumar et al.,[33]	2024																				x			Amazon website
Krosuri and Aravapalli[34]	2023												x											Amazon and Twitter
Boumhid et al., [35]	2022																							Twitter, Facebook, Amazon.
Marwat et al., [40]	2022								x				x											Amazon store and Flipkart web application covering distant products, such as watches, mobile, shoes, clothes, and perfumes
Fadel and Behadili[45]	2022			x					x				x		x	x								Twitter data about OK cuisine reviews obtained from the Amazon website
Adewole et al., [36]	2021													x	X	x								Amazon, Yelp, IMDB and Kaggle
Guha and Mohan[44]	2021									x								x						MR, IMDB, SST and Amazon datasets indicated
Geetha and Renuka[48]	2021	x								x			x		x	x		x						Amazon products Electronics reviews
Benlahbib and Nfaqui[49]	2020	x											x											IMDb, TripAdvisor and Amazon websites show
Muhammad et al., [50]	2020							x																iPhone 5s reviews gathered from Amazon, Facebook, and Reevoo blog
Kumaravel and Bizu[51]	2019				x						x													Dataset: amazon customer review
Kauffman[41]	2019												x											Big data extracted from Amazon online reviews (specifically cell phones)
Bodini[46]	2019																							Bangla language, restaurant reviews, f comments and posts related to the sport of cricket

Once the texts were cleaned, vectorization techniques were applied to convert the reviews into numerical representations. Two approaches were implemented: Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF), both of which have proven highly effective in text classification tasks by preserving the relevance of terms within the corpus[19][20]. In this study, TF-IDF performed better by highlighting the relative importance of less frequent words, thus improving the model's accuracy.

The dataset was divided into 80% for training and 20% for testing, aiming to maintain a balanced distribution between positive and negative reviews. Two supervised learning algorithms were trained and compared: Multinomial Naive Bayes and Logistic Regression [13] [22][23]. Both models were fitted using cross-validation techniques and evaluated with standard metrics, including accuracy, precision, recall, and F1-score, which allowed measuring the overall performance and generalization capacity of the classifier [14][24].

Processing and normalization of the data set led to a cleaned corpus of 90,955 Amazon reviews after the mandatory removal of neutral 3-star ratings. Following a rigorous cross-validation and evaluation methodology, the data are detected divided by a stratified proportion of 80% for training (72,764 records) and 20% for testing, resulting in a strict test sample volume of exactly 18,191 comments which constitutes the representativeness framework for all confusion matrices and associated metrics.

Finally, the evaluation phase included generating word clouds to visualize the most frequent terms in positive and negative reviews, as well as confusion matrices that allowed for a detailed analysis of classification successes and errors. These visualizations, along with the calculation of performance metrics, provided a comprehensive understanding of each model's behavior and supported the comparison between Naive Bayes and Logistic Regression [1][2].

IV. Results

Regarding the scope of the preliminary results, the authors consider it important to point out that sentiment analysis is limited to measuring the general polarity of a text or comment, simply classifying it as positive, negative, or neutral. Emotion analysis, on the other hand, identifies specific psychological states (such as joy, sadness, anger, fear, or disgust) [17] [18].

Sentiment analysis focuses almost exclusively on text data (such as reviews or social media posts). But emotion analysis typically incorporates multimodal data, analyzing not only text but also tone of voice and even facial expressions.The preliminary experimental results obtained from the Amazon review dataset showed clear differences between the evaluated models. Table 1 summarized the performance metrics calculated for the Naive Bayes and Logistic Regression algorithms.

TABLE III Comparative performance of the Naive Bayes and Logistic Regression models in sentiment analysis

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.773	0.75	0.77	0.675
Logistic regression	0.870	0.85	0.87	0.850

The comparative evaluation of traditional models was complemented through the Area Under the ROC Curve (AUC-ROC) metric, providing a robust evaluation that overcomes the limitation of the rigid classification threshold of 0.5 typical of standard confusion matrices.

The Logistic Regression model reached an outstanding AUC-ROC of 0.93, which corroborates its excellent probabilistic discrimination capacity in high vector dimensionality environments (TF-IDF), significantly exceeding the value of 0.81 obtained by the Multinomial Naive Bayes classifier.

Logistic regression performed better across all metrics, particularly in the F1 score, indicating a better balance between accuracy and completeness. These results was consistent with previous research showing that linear models, when incorporating regularization, achieve more effective generalization in contexts with high text dimensionality [14][21][22].

On the other hand, the inferior performance of *Naive Bayes* can be attributed to its assumption of conditional independence between words, which reduced its ability to capture contextual relationships in natural texts [13][2]. However, studies such as those by Archak *et al*.,[8] and Moraes et al. [16] highlighted its usefulness in applications where computational efficiency was prioritized over maximum accuracy (See figure 1).

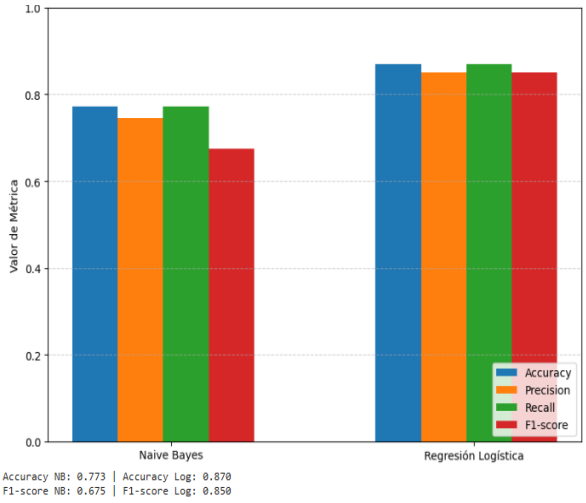


Fig. 1 Metrics: Naives bayes vs Logistic regression

To complement the interpretation, figure 2 presents the confusion matrix of both models, showing that Naive Bayes tends to misclassify negative reviews as positive, while Logistic Regression achieves a better balance between both classes of categories [24][1].

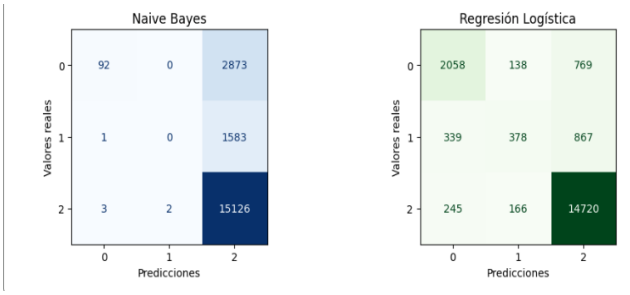


Fig. 2. Visual comparison of Naive Bayes and Logistic Regression confusion matrices Standardized test sample, N = 18,191 reviews)

.

Figure 2 showed the confusion matrices of both models. "The Naive Bayes model exhibited a strong bias towards the majority (positive) class, misclassifying a critical volume of negative comments as positive. This behavior is evident in the high number of false positives (2,873 negative comments classified as positive), which explains the drop in its F1 measure to 0.675 and its lower discriminative stability compared to Logistic Regression.

In contrast, Logistic Regression showed a more balanced distribution of correct classifications across classes, with a higher number of true positives (14,720) and a significant reduction in false negatives. This behavior alignedd with previous studies highlighting Logistic Regression's ability to handle unbalanced and highly dimensional datasets [21][24][22]. These differences demonstrated that, while Naive Bayes is more computationally efficient, its performance degrades when relationships between features was not independent. Logistic Regression, on the other hand, achieved more stable generalization, consistent with the literature on opinion mining and natural language processing[1][2].

V. Discussion

A. Analysis of the results

The preliminary results showed that the Logistic regression demonstrated more balanced performance, consistent with findings from previous studies [1] [14]. Its ability to handle interdependencies between terms improves its performance compared to the conditional independence assumed by Naive Bayes [21][22]. The results also highlight the importance of preprocessing and vectorization, crucial factors in the model's success [19][24].

The Naive Bayes model, while more computationally efficient, is sensitive to class imbalances and large vocabularies [2][13]. In contrast, Logistic Regression offers a better balance between simplicity, interpretability, and accuracy, making it ideal for real-time applications [22][27] [23].

Finally, more recent models such as BERT and GPT have demonstrated improvements in the detection of semantic nuances [25][26]. However, their use may not be viable in environments with processing limitations or a need for interpretability [1].

The results show that Logistic Regression, with an accuracy of 87% and an F1-score of 0.85, significantly outperforms the Naive Bayes method, which achieved an accuracy of 77.3% and an F1-score of 0.67. This confirms that linear models with regularization offer better generalization capabilities in text analysis, as they do not depend on the conditional independence of words.

Although BERT achieves state-of-the-art accuracy on Amazon review datasets [ref], the practical deployment advantages of Logistic Regression — including interpretability, training speed, and minimal hardware requirements — make it a competitive choice for resource-constrained academic environments. LR with 87% accuracy can be considered competitive in the current context, where BERT achieves 90–96% on similar datasets because the difference of ~3–9 accuracy points can be acceptable given LR's 10–100× lower computational cost.

B. Limitations of the study

(a) The dataset comes from a single source (Kaggle), which may introduce selection bias;

(b) The analysis is binary (positive/negative), which oversimplifies the reality of the emotional space;

(c) The exclusion of neutral reviews reduces the generalizability of the model; and

(d) Classical models can suffer from slang, sarcasm, and irony, phenomena widely documented in Amazon reviews, so it is recommended that a filter layer be used in future research.

VI. Conclusions

This study focuses on sentimental analysis in Amazon reviews to develop an automatic ranking approach for user comments. The objective was to explore how emotional content influences the perceived usefulness and relevance of reviews, while applying machine learning and text mining techniques to enhance the organization and prioritization of customer feedback. These findings highlight the effectiveness of classical models in opinion mining for e-commerce environments and underline their relevance compared to more complex architecture such as BERT and deep neural networks. The results of this research may contribute to improving recommendation systems, customer experience, and decision-making processes in e-commerce environments. Future works should explore these issues:

(1) the incorporation of fine-grained emotion detection using lexicon-based methods (NRC Lexicon, LIWC) combined with LR for richer sentiment characterization;
(2) the fine-tuning of BERT or RoBERTa on domain-specific Amazon review corporal.
(3) the extension of the framework to multilingual reviews to assess cross-cultural sentiment patterns.

ACKNOWLEDGMENT

The authors express their gratitude to Universidad Tecnológica de Panama and AGH University of Krakow (Poland) for the support provided in carrying out this research.

REFERENCES

[1] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," in *Knowledge-based approaches to concept-level sentiment analysis*, IEEE Computer Society, 2013, pp. 15–21. doi: 10.1109/MIS.2013.30.

[2] B. Liu, *Sentiment analysis and opinion mining*, 1st ed. Springer Cham, 2012. doi: 10.1007/978-3-031-02145-9.

[3] S. Zhou and Y. Liu, "How do cultural ecosystem services affect visitor emotions ? Evidence from social media data in Shanghai urban parks," no. January, 2026, doi: 10.3389/fpsyg.2025.1711498.

[4] J. Yoon, H. Lim, S. Woo, B. Lee, and J. Kim, "Analyzing Influential Factors in Review-Based Restaurant Recommender Systems : The Role of Review Length , Aspect , and Emotion," *Electronics*, vol. 15, no. 1821, pp. 1–20, 2026, doi: 10.3390/electronics15091821.

[5] S. Biswas and G. Poornalatha, "A Human-Psychology-Informed Multi-Task Learning Framework for Robust Opinion Mining A Human-Psychology-Informed Multi-Task Learning Framework for Robust," *Appl. Artif. Intell.*, vol. 40, no. 1, 2026, doi: 10.1080/08839514.2026.2663635.

[6] K. Balasaranya and P. Ezhumalai, "Text Optimized XLNet Based Sentimental Analysis for Opinion Mining Towards Products From Twitter Data," *Int. J. Comput. Intell. Syst. Res.*, vol. 19, no. 165, pp. 1–41, 2026, doi: 10.1007/s44196-026-01265-4.

[7] B. Hu, M., & Liu, "Mining and summarizing customer reviews," in *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD'04, Ed., Seattle WA USA: IEEE, 2004, pp. 168–177. doi: 10.1145/1014052.1014073.

[8] P. G. Archak, N., Ghose, A., & Ipeirotis, "Deriving the pricing power of product features by mining consumer reviews," *Manage. Sci.*, vol. 57, no. 8, pp. 1485–1509, 2011, doi: 10.1287/mnsc.1110.1370.

[9] L. Teixeira, Y. Alencar, L. Bastos, P. Rodrigues, R. Pignatelli da Silva, and A. Lopes Damian, "Emotions and Experiences on the Road : Unveiling UX in Automotive Infotainment Through YouTube Comments," in *Proceedings ofthe 27th International Conference on Enterprise Information Systems (ICEIS 2025)*, (ICEIS 2025, Ed., 2025, pp. 436–447. doi: 10.5220/0013295700003929.

[10] S. He and H. Zou, "A bert-based algorithm for fusion sentiment analysis of user comments," in *2025 International Conference on Big Data and Informatization Education (ICBDIE 2025), February 2123, 2025, Suzhou, China*, ACM, Ed., Suzhou, China: Association for Computing Machinery, 2025, pp. 491–498. doi: 10.1145/3729605.3729691.

[11] G. Wang, Z. Zhang, Y. Gao, Y. Cui, H. Cheng, and Z. Zhang, "Research on User Experience Optimization and Functional Design of Cat Snacks Based on the Kano Model," in *2025 International Conference on Digital Economy and Information Systems (DEIS 2025), April 2527, 2025, Guangzhou, China*, ACM, Ed., Guangzhou, China.: Association for Computing Machinery, 2025, pp. 567–574. doi: 10.1145/3745133.3745220.

[12] W. Yang, "Research on Emotion Tendency Analysis Model Based on Deep Learning," in *2025 2nd International Conference on Modeling, Natural Language Processing and Machine Learning (CMNM 2025), May 1618, 2025, Fuzhou, China*, CMNM 2025, Ed., Fuzhou, China: Association for Computing Machinery, 2025, pp. 347–350. doi: 10.1145/3757110.3757169.

[13] L. Pang, B., & Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 1, no. 1–2, p. 94, 2008, doi: 10.1561/1500000011.

[14] H. Medhat, W., Hassan, A., & Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[15] D. Chen, H., & Zimbra, "AI and opinion mining," *IEEE Intell. Syst.*, vol. 25, pp. 47–80, 2010, doi: 10.1109/MIS.2010.75.

[16] W. P. Moraes, R., Valiati, J. F., & Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, 2013, doi: 10.1016/j.eswa.2012.07.059.

[17] AI, "Sentiment Analysis," Computational Intelligence Applications for Text and Sentiment Data Analysis. [Online]. Available: https://www.sciencedirect.com/topics/social-sciences/sentiment-analysis

[18] E. L. Tonkin, *A Day at Work (with Text): A Brief Introduction*. Elsevier Ltd., 2016. doi: 10.1016/B978-1-84334-749-1.00002-0.

[19] H. Manning, C. D., Raghavan, P., & Schütze, *Introduction to Information Retrieval*, First. Cambridge, England: Cambridge University Press, 2008. [Online]. Available: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

[20] J. Leskovec and J. D. Rajaraman, A., & Ullman, *Mining of Massive Datasets*. Cambridge University Press, 2011. [Online]. Available: http://infolab.stanford.edu/~ullman/mmds/book.pdf

[21] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surv.*, vol. 34, no. 1, pp. 1–47, 2002, doi: 10.48550/arXiv.cs/0110053.

[22] B. Zhang, L., Wang, S., & Liu, "Pretrained models for sentiment analysis: A comprehensive review," 2022.

[23] J. Zhang, X., & Zhao, "Research on text classification algorithm based on improved logistic regression," 2019, doi: 10.1016/j.knosys.2023.110761.

[24] E. A. He, H., & Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

[25] K. Devlin, J., Chang, M. W., Lee, K., & Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol.*, vol. 1, 2019, doi: 10.18653/v1/N19-1423.

[26] E. Young, T., Hazarika, D., Poria, S., & Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," *Comput. Lang.*, p. 32, 2018, doi: 10.48550/arXiv.1708.02709.

[27] T. Tang, D., Qin, B., & Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Ed., Lisbon, Portugal: Association for Computational Linguistics, 2015, pp. 1422–1432. doi: 10.18653/v1/D15-1167.

[28] E. Bird, S., Klein, E., & Loper, *Natural Language Processing with Python*, First Edit. O'Reilly Media, Inc, 2009. [Online]. Available: https://tjzhifei.github.io/resources/NLTK.pdf

[29] S. Loria, "TextBlob Documentation," TextBlob: Simplified Text Processing. [Online]. Available: https://textblob.readthedocs.io/en/dev/

[30] L. Lakshmi *et al.*, "Opinion mining in e-commerce: Evaluating machine learning approaches for sentiment analysis," *Results Control Optim.*, vol. 19, 2025, doi: 10.1016/j.rico.2025.100575.

[31] S. Pradeepa *et al.*, "HGATT_LR: transforming review text classification with hypergraphs attention layer and logistic regression," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-70565-6.

[32] S. Rajagopal, S. Jayaprakash, and K. Subburathinam, "Sentiment Analysis on Big Data: A Hybrid SED-TABU Feature Selection Method," *Teh. Vjesn.*, vol. 31, no. 6, pp. 2079–2086, 2024, doi: 10.17559/TV-20231004000989.

[33] L. K. Kumar *et al.*, "Analyzing Public Sentiment on the Amazon Website: A GSK-Based Double Path Transformer Network Approach for Sentiment Analysis," *IEEE Access*, vol. 12, pp. 28972–28987, 2024, doi: 10.1109/ACCESS.2024.3368441.

[34] L. R. Krosuri and R. S. Aravapalli, "Novel heuristic-based hybrid ResNeXt with recurrent neural network to handle multi class classification of sentiment analysis," *Mach. Learn. Sci. Technol.*, vol. 4, no. 1, 2023, doi: 10.1088/2632-2153/acc0d5.

[35] A. Boumhidi, A. Benlahbib, and E. H. Nfaoui, "Cross-Platform Reputation Generation System Based on Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 10, pp. 2515–2531, 2022, doi: 10.1109/ACCESS.2021.3139956.

[36] K. S. Adewole *et al.*, "Hybrid feature selection framework for sentiment analysis on large corpora," *Jordanian J. Comput. Inf. Technol.*, vol. 7, no. 2, pp. 130–151, 2021, doi: 10.5455/jjcit.71-1609858713.

[37] A. Gueddes and M. A. Mahjoub, "BERT-OntoSent : combining BERT language model with sentiment ontology for enhanced sentiment analysis on social media BERT-OntoSent : combining BERT language model," *J. Inf. Telecommun.*, vol. 10, no. 1, pp. 1–23, 2026, doi: 10.1080/24751839.2025.2528363.

[38] E. Daraghmi and N. Zyadeh, "Sentiment Classification of Amazon Product Reviews Based on Machine and Deep Learning Techniques: A Comparative Study," *Futur. Internet*, vol. 18, no. 3, 2026, doi: 10.3390/fi18030138.

[39] N. Malik and M. Bilal, "Natural language processing for analyzing online customer reviews: a survey, taxonomy, and open research challenges," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/PEERJ-CS.2203.

[40] M. I. Marwat *et al.*, "Sentiment Analysis of Product Reviews to Identify Deceptive Rating Information in Social Media: A SentiDeceptive Approach," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 3, pp. 830–860, 2022, doi: 10.3837/tiis.2022.03.005.

[41] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining," *Sustain.*, vol. 11, no. 15, 2019, doi: 10.3390/su11154235.

[42] N. Firza, A. Bakiu, and A. Monaco, "Machine Learning for Quality Diagnostics: Insights into Consumer Electronics Evaluation," *Electron.*, vol. 14, no. 5, 2025, doi: 10.3390/electronics14050939.

[43] N. Dhamayanthi and B. Lavanya, "A Novel Framework for Sentiment Analysis: Dimensionality Reduction for Machine Learning (DRML)," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 6, pp. 777–794, 2024, doi: 10.14569/IJACSA.2024.0150678.

[44] T. Guha and K. G. Mohan, "A Hybrid Capsule Network with Attention and BiLSTM for Opinion Mining in Text," *Webology*, vol. 18, no. Special Issue, pp. 426–441, 2021, doi: 10.14704/WEB/V18SI04/WEB18138.

[45] F. H. Fadel and S. F. Behadili, "A Comparative Study for Supervised Learning Algorithms to Analyze Sentiment Tweets," *Iraqi J. Sci.*, vol. 63, no. 6, pp. 2712–2724, 2022, doi: 10.24996/ijs.2022.63.6.36.

[46] M. Bodini, "Aspect extraction from bangla reviews through stacked auto-encoders," *Data*, vol. 4, no. 3, 2019, doi: 10.3390/data4030121.

[47] S. Nabil, J. El Bouhdidi, and M. Y. Chkouri, "Enhancing Recommender Systems Using Sentiment Analysis: Addressing Cold Start Issues and Improving Recommendation Quality," *J. Adv. Inf. Technol.*, vol. 16, no. 4, pp. 556–567, 2025, doi: 10.12720/jait.16.4.556-567.

[48] M. P. Geetha and D. Karthika Renuka, "Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model," *Int. J. Intell. Networks*, vol. 2, pp. 64–69, 2021, doi: 10.1016/j.ijin.2021.06.005.

[49] A. Benlahbib and E. H. Nfaoui, "Aggregating customer review attributes for online reputation generation," *IEEE Access*, vol. 8, pp. 96550–96564, 2020, doi: 10.1109/ACCESS.2020.2996805.

[50] W. Muhammad, M. Mushtaq, K. N. Junejo, M. Y. Khan, S. Louis, and U. States, "Sentiment Analysis of Product Reviews in the Absence of Labelled Data Using Supervised Learning Approaches," *Malaysian J. Comput. Sci.*, vol. 33, no. 2, pp. 118–132, 2020.

[51] T. Kumaravel and B. Bizu, "Convolutional neural network for customer's opinion on Amazon products," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 6634–6643, 2019, doi: 10.35940/ijrte.C5670.098319.

[52] J. McAuley, J., & Leskovec, *Hidden factors and hidden topics: Understanding rating dimensions with review text*. Cambridge: Cambridge University Press, 2013. doi: 10.1145/2507157.2507163.

[53] C. Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, "Learning Word Vectors for Sentiment Analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, D. Lin, Y. Matsumoto, and R. Mihalcea, Eds., Portland, Oregon, USA: Association for Computational Linguistics, 2011, pp. 142–150. [Online]. Available: https://aclanthology.org/P11-1015.pdf