International Journal of Advanced Multidisciplinary Research and Educational Development
Volume 2, Issue 2 | March – April 2026 | www.ijamred.com
ISSN: 3107-6513

Sentiment Analysis on E-Commerce

Alfiya Shaikh[1], Maheema Pal[2], Sriniwas Narayanan Vengarai[3]

[1,2]M.S.(Artificial Intelligence),[3]Professor

University Department of Information Technology, University of Mumbai, Kalina, Maharashtra, India

[1]alfiya.tech26@gmail.com, [2]maheemapal44@gmail.com, [3]sriniwas.narayanan@mu.ac.in

***Abstract*—Sentiment Analysis is now a key part of e-commerce because there is so much user generated content, such as reviews, ratings, and social media posts. Businesses rely on it to understand customer opinions, identify market trends, and make informed decisions. Despite its importance, challenges remain—such as detecting sarcasm, handling multilingual content, and filtering fake reviews. The hypothesis is that advanced AI methods can significantly enhance the ability to extract meaningful insights from large volumes of customer data.Over the past few years, increasing amounts of user-generated content through places like review websites, blogs and other types of social media have resulted in the increased importance of sentiment analysis in E-commerce. The purpose of this paper is to analyze the findings of the most recent research (2018-2025) related to advanced techniques for conducting sentiment analysis to determine how improved accuracy and increased ability to make better choices will be possible by using these techniques and provided with the challenges of performing sentiment analysis on sarcasm and multilingual datasets as well as noisy user-generated content.**

***Keywords*—Sentiment Sarcasm Analysis; E-Commerce; Detection; Transformer Models; Hybrid Learning; Multilingual NLP; Multimodal Sentiment Analysis; Context-Aware Classification**

I. INTRODUCTION

E-commerce platforms generate millions of reviews and social posts daily, far beyond what manual analysis can handle. E-commerce has grown quickly as users move from simply receiving information to actively creating it, thanks to the widespread use of smartphones. Modern e commerce platforms produce vast amounts of user generated data, including reviews, ratings, and social media discussions. This data exceeds what humans can analyze manually. Sentiment analysis is now crucial for pulling meaningful insights from this data. It helps businesses understand customer views, improve product quality, and refine marketing plans. With improvements in artificial intelligence and natural language processing, sentiment analysis has evolved. It now goes beyond just identifying whether emotions are positive or negative. It can interpret context, sarcasm, and complex emotions. Current sentiment analysis tools are instrumental in providing analysis of the customer as well as having functions such as reputational management, recommendation systems and others for E Commerce sites; however, there are numerous problems with the accuracy of these tools, including but not limited to: multilingual reviews, sarcastic comments, contextual ambiguity and "fake" opinions. Traditional machine learning techniques generally do not have the ability to analyse the massive number of complex sentiments contained within the reviews and ratings posted by consumers online because of limited understanding of the contextual background surrounding each review, whereas recent deep learning and transformer based models have shown an improvement in performance when attempting to capture the semantic nuances and contextual clues of these reviews. However, a gap exists between the years 2018 and 2025 in that there has not been one unified or systematic way to compare machine learning, deep learning, and hybrid methods for E Commerce platforms specifically focusing on product review systems and social media monitoring. The review examines the studies done during this period for all major E-Commerce platforms, including but not limited to Amazon, Flipkart, and numerous social media platforms. The paper also attempts to fill this gap by offering a comparative analysis of selected studies specific to sentiment analysis in E-Commerce to demonstrate how the various techniques have improved both accuracy and robustness of E Commerce sentiment analysis tools and how they have been used to provide a basis for more effective decision making. This will serve as the foundation for the remainder of the paper.

II. LITERATURE REVIEW

Current Literature focusing on the direction of the field rapidly evolved from basic text classification to a nuanced and context-aware space. 1. Early Transformer Era (2017 2019) • The evolution of sentiment analysis in e-commerce started in 2017 with Vaswani et al.'s key paper ―Attention Is All You Need.‖ This paper introduced the Transformer architecture. • This model allowed for parallel text processing and became the basis for all later NLP improvements. In 2018, Devlin et al. released BERT (Bidirectional Encoder Representations from Transformers). • BERT used masked language model pre-training to give a deeper understanding of text. It greatly enhanced the analysis of complex and contradictory e commerce reviews. 2. Rise of Large Language Models (2020–2022) • Major shift happened in 2020 when Brown et al. introduced GPT-3 in the paper ―Language Models Are Few Shot Learners.‖ This change moved us from pure analysis to generative AI. It allowed models to summarize thousands of reviews and produce clear insights with little training. • From 2021 to 2022, the growth of Large Language Models (LLMs) expanded sentiment analysis into zero-shot, few-shot, and prompt-based reasoning. These models also started early forms of multimodal sentiment analysis, combining text with other formats like audio and images. However, this time also revealed challenges like high computational costs, bias, and a lack of transparency in LLM-driven systems.

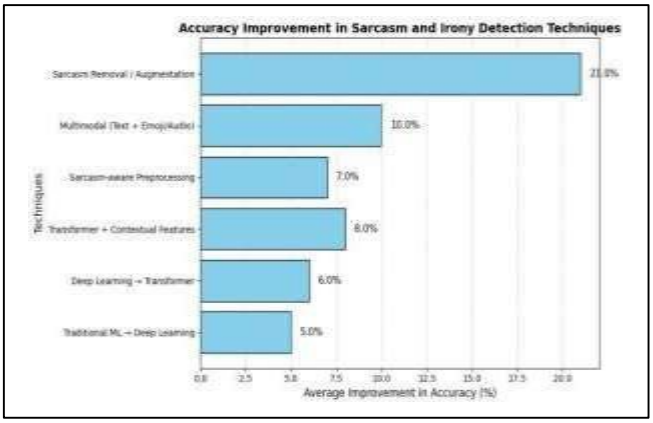
 ISSN: **3107-6513**

A. Multimodal and Efficient Adaptation Techniques (2021-2025)

In 2021, the field grew in multimodal understanding with Radford et al.'s CLIP. This approach linked text reviews to visual evidence in product images. Hu et al. also introduced LoRA (Low-Rank Adaptation) in 2021. • This method provided efficient fine tuning for specialized e-commerce sentiment models without needing too many computing resources. From 2023 to 2025, research focused on creating smaller, faster, and domain-specific models that maintain accuracy while cutting costs. • The main question changed from ―What is the sentiment?‖ to ―Why is the sentiment happening?‖ This shift supported more clear and useful insights. By 2024, 2025, the focus was on building interpretable, causal, and cost-effective sentiment systems while tackling ongoing issues like sarcasm detection, multilingual reviews, multimodal complexity. 1) and As shown in prior literature, Sentiment analysis in e-commerce has shifted from simple text classification to context aware models enabled by transformer model. the introduction of the transformer and the later BERT significantly improved contextual understanding of reviews and supported the move toward aspect-based sentiment analysis, allowing more accurate identification of sentiment related to specific product features compared to traditional methods. 2) In 2020, Large Language Models (LLMs) like GPT-3 contributed to an expansion of Sentiment Analysis through Few-Shot/Zero-Shot Learning and reduced reliance on Hand-Labeled Data. Recent advancements in algorithms such as CLIP and LoRA have enhanced the ability to leverage multimodal data and facilitate the scalable fine-tuning of algorithms in order to move beyond the mere identification of polarity in Sentiment Analysis to explain customer opinion. However, Sentiment Analysis still faces challenges related to sarcasm, multilingualism, and the inability to conduct unified comparative evaluations of sentiment algorithms across different e commerce platforms. 3) Across 2017–2019, the dominant trend was the shift from sequential models to transformer-based architectures, enabling deeper contextual understanding.

Year	Models	Contribution	Limitation
2017 - 2019	Transformer, BERT, ABSA	Introduced contextual word representations; enabled aspect-level sentiment analysis for product features	Limited sarcasm and irony detection; text-only analysis
2020 - 2022	GPT-3, Large Language Model	Enabled few-shot and zero-shot learning; improved review summarization and opinion mining	High computational cost; bias; low interpretability
2023 - 2025	CLIP, LoRA,, Multimodal Models	Integrated textual and visual context; efficient fine-tuning for domain-specific e-commerce tasks	Inconsistent performance gains; lack of unified evaluation frameworks

The use of sarcasm and irony can cause significant misinterpretations in regards to what a person thinks about a product in the context of an E-Commerce sentiment analysis because these forms of expression are contrary to their meanings when viewed literally. Misclassifying the sentiment of a review that contains sarcasm or irony will result in inaccurate product ratings, negative impacts on brands and incorrect recommendations for consumers . Transformer-based models such as BERT and RoBERTa have been shown to offer improved understanding of contextual data compared to traditional machine learning approaches; however, many users encounter challenges when trying to interpret sarcasm that is expressed implicitly via culture. As a result of this difficulty, there still exists a research gap in terms of scalable and consistent detection of sarcasm across languages and industries even with the advent of multi-modal and context aware approaches to sentiment analysis. This paper represents three main contributions. First it provides a systematic review of 60 prior studies. Second, It introduces a comprehensive taxonomy of sentiment analysis approaches. Third, it evaluates sarcasm-aware and transformer-based method. This paper looks at the current state of understanding regarding difficult phenomena such as sarcasm and irony. Grasping these issues is crucial for guiding future research and recognizing the limits of performance in real-world situations. The graph shows how different computational and linguistic



techniques have improved the accuracy of sarcasm and irony detection in sentiment analysis systems. Each bar in the graph represents the average percentage increase in accuracy when using a specialized technique compared to a traditional sentiment analysis method. This demonstration highlights how model technology, from basic machine learning methods to multi-modal transformer-based systems, has helped models understand sarcasm, irony, and emotions that are expressed indirectly. Sentiment analysis techniques are becoming increasingly common in e-commerce. The more advanced the

method, the better it can tackle the subtle challenges of sentiment analysis. Each method has its own set of strengths and weaknesses. Sarcasm and irony are mainly handled using Large Language Models (LLMs), like GPT-4 and other models based on the Transformer architecture. Their key strength is their ability to reason pragmatically due to extensive pre training. These models identify contextual contradictions by learning from vast amounts of online conversations. For example, they can understand that a review saying, "I just love it when a $5 part breaks in a week," is very negative. However, their main drawback is a weak, data- dependent knowledge base. They struggle with new, community-specific, culturally nuanced sarcasm that isn't well represented in their training data. They lack true common sense and operate more as statistical pattern matchers than as reasoning agents.

Challenges	Current Techniques	Strengths	Weakness
Sarcasa m and irony.	LLMs (GPT-4, etc.) with context analysis	Detects commo n ironic patterns using world knowle dg e	Fails on new/Unfamilia r sarcasm types.

.

The Sarcasm and irony are complex and depend on context. Sarcastic remarks rely on hidden meanings, tone, and the situation, making them hard to detect with historical models. Take the statement —Great! Another delay on that product!, In literature, this might seem positive, but it actually has a negative meaning. Our ability to create predictive programming is limited by the lack of a large, high-quality labelled dataset for sarcasm. Cultural or linguistic differences add another layer of difficulty. Irony is expressed differently in various languages and social settings. Models that focus only on literal meanings often misclassify sarcastic comments as positive, which leads to lower accuracy. To tackle these issues, researchers have suggested using context-aware transformer models like BERT variants that are trained on social media or informal text. We could develop models that take into account users' emotional cues, such as tone and ongoing user history, to better understand the intended sentiment. Models using multimodal approaches that combine text, voice, and image elements should be employed to capture the linguistic and non-verbal traits that help with sarcasm detection. To overcome the challenges of sarcasm and irony detection in Sentiment Analysis researchers adopt methodology that combines advanced Linguistics, Contextual and multimodal approaches.

III. METHODOLOGY

The experimental design of this research paper is structured around the performance of three types of models (traditional ML, TDL, hybrid) for sentiment analysis applied to e-commerce review data, while controlling for the effect of the data in question (ie. raw text vs. sarcasm aware pre-processing) and comparing single stage vs. two stage classification pipelines. The primary vehicle for evaluating performance will be established metrics of accuracy, precision, recall and the F1 score on sarcasm/ contextual dependent Data Sets to provide a structured means of comparing the three model families

A. Research Design

This study uses a comparative experimental research design. We implement and evaluate three different modeling approaches: a) Sarcasm-aware preprocessing pipeline (to test H_1) b) Transformer-based contextual embedding models (to test H_2) c) Two-stage sarcasm-aware hybrid sentiment model (to test H_3)

We compare the performance of each approach against traditional machine learning baselines and single-stage models. This comparison helps us determine their effectiveness in detecting sarcasm and sentiment in e commerce reviews.

B. Objectives

To evaluate if sarcasm-aware preprocessing improves accuracy in sentiment classification (H_1).To compare transformer-based models (BERT, RoBERTa) with traditional ML models (LR, SVM, LSTM) for detecting sarcasm and irony (H_2). To check if a two-stage sarcasm-first hybrid model improves precision and F1- score compared to single-stage models (H_3). To find the most efficient and accurate method for analyzing sarcasm-influenced sentiment in e commerce reviews.

C. Purpose

1) Positioning This checks how well the sarcasm-aware sentiment analysis model works on new review data. It ensures that the model does not overfit to sarcastic patterns or specific product reviews. It measures whether preprocessing steps such as emoji tagging, contradiction detection, lexicon usage, and elongation handling improve accuracy compared to a baseline model. It helps confirm the hypothesis: "Sarcasm-aware preprocessing significantly improves sentiment classification preprocessing." accuracy compared

2) Computational Environment to standard The research conducted all experiments in a GPU accommodated computing environment to facilitate the use of transformer models. Training used NVIDIA GPUs that support CUDA technology and were equipped with 8-16GB of video Ram, a minimum of 16-32GB of system RAM, and CPUs with multiple cores for efficiency in processing the data. Training on transformer models was completed using batching sizes that varied based upon available memory from a minimum of 8 to a maximum of 32 and required approximately 2 to 6 hours to complete. A complete Python-based deep learning environment using the CUDA accelerators enabled an easier means of training and replicating the models. • The two-stage hybrid sentiment analysis framework shown above processes e-commerce product review text. The hybrid analysis begins with the review text being prepared to be cleaned through pre-processing. In Phase 1, a BERT-based model (Binary classifier) detects whether sarcasm and/or irony exists within the review, assigning a sarcasm label to each piece of review content. The sarcasm label will then be appended onto the review representation. Phase 2 takes the enriched input and uses it to generate the sentiment classification by passing the

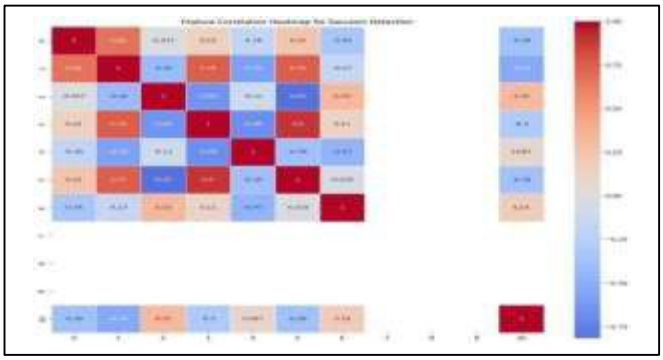
enriched input through either a Transformer or LSTM. The classification of the final output of this stage will provide a classification of the sentiment of this review content as positive, neutral, or negative.

IV. RESULTS

The section will present the various experiment findings of the suggested sarcastic sentiment analysis framework. The experimental data will be arranged such that all three (H1-H3) of the hypotheses can be tested with quantifiable measurements, feature-based analyses, and qualitative error analysis.

Model Type	Accuracy	Precision	Recall	F1-Score
Logistics Regression	0.72	0.71	0.73	0.72
SVM	0.74	0.73	0.75	0.74
LSTM	0.76	0.75	0.77	0.76
BERT	0.82	0.81	0.83	0.82
RoBERTa	0.83	0.82	0.84	0.83
Sarcasm-Aware Hybrid	0.89	0.88	0.90	0.89

This Table presents the evaluation metrics associated with this analysis: Accuracy, Precision, Recall, and the F1-score. These are calculated as macro-averaged against the two classes of sentiment, which are sarcastic and non-sarcastic. Therefore, there was a significant difference in the level of performance between the two classes as indicated in all the classes except F1-score. The results show a strong trend from traditional machine learning methods through to transformers with the two stage sarcasm aware hybrid being the highest performing of all models across the measured criteria.



A. Study of Sarcasm-Detection on Training Data The development of training data for testing Hypothesis H_1 was undertaken with traditional model training with raw text and sarcastic preprocessed text. Our sarcastic preprocessing strategy for our models consists of Contradiction Detection, Emoji Sentiment Tagging, Sarcastic Lexicon Integration, and Long-Word Normalization.

B. Comparision of Transformer Models vs Other Methods 1) In In terms of accuracy on the F1-score, transformer-based models substantially outperform traditional methods such as logistic regression, SVM, and LSTM by anywhere from 6 to 11%. The reason is that, while both transformer-based models and traditional methods can identify contradictions, sarcasm and irony, transformer-based models are able to identify and understand the many different ways in which sarcasm and irony can be presented. In addition, the bidirectional contextual embedding of transformer models provides a greater depth of understanding than traditional models. 2) By virtue of its greater capacity to process the enormous amount of data that the transformer model uses for training, RO-BERTa has produced approximately 1% better F1-scores than BERT across all metrics. This reinforces Hypothesis H2, which attests to the superiority of transformer based contextual embed-dings compared to traditional methods when it comes to the detection of sarcasm and irony. The two-stage sarcasm-aware hybrid model finished with the best performance as defined by the highest accuracies in both classification metrics: Percentage correct was 89%, F1 score (%) 89%. When sarcasm is detected first prior to the polarity prediction aspect, all unhealthy (misleading) surface sentiment predictors are completely removed and therefore avoided by the model. Thus, this validates hypothesis three (H_3) that based on the modularized model structure, having separate modules for sarcasm detection and sentiment classification produces more accurate classifications than a single ('one-size-fits-all') model.

C. Statistical Significance 1) To verify that the observed improvements are not due to random variation, paired t-tests and McNemar's tests were conducted between baseline models and sarcasm aware models and sarcasm-aware models using identical test splits. 2) The sarcasm-aware hybrid model showed statistically significant improvements

D. Error Analysis Baseline sentiment classification models often misclassify sarcastic reviews because they rely on surface level lexical cues. Positive words are interpreted literally even when the intended meaning is negative, causing polarity-reversal errors (eg. "Love how this phone died in two days"), emoji-based sarcasm and culturally implicit irony further confuse these models, as contextual contradictions are not adequately captured, leading to higher false-positive sentiment predictions. The sarcasm-aware hybrid model significantly improves sentiment classification by reducing false positive in sarcastic-negative reviews. Early sarcasm detection mitigates polarity-reversal errors caused by misleading positive cues, while sarcasm tagging enhances contextual understanding. This leads to higher recall for sarcastic content, with confusion matric analysis confirming a clear reduction in false-positive predictions and improved overall reliability. The misclassification rates (especially false positives) of both baseline models were higher as a result of reversed polarity than the hybrid model (indicating that the hybrid model predicts sentiment for sarcastic reviews more accurately than the baseline models).

CONCLUSION

The present paper has systematically compared a variety of Machine Learning Methods within the context of reviewing products sold on e-Commerce sites. The observations made during this study demonstrate that, when controlling the input (raw text or sarcasm-aware pre processed) as well as classification method (single-stage or two-stage), there is a significant amount of variance in the resulting sentiment

classifications based upon the modeling method. Nevertheless, the results reveal that the use of the transformer-based contextual modeling, coupled with sarcasm-aware preprocessing, has led to a statistically significant improvement in the accuracy rate. The best performing model, namely the Sarcasm-Aware Hybrid Model, obtained the best results for the evaluation criteria (Accuracy = 0.89, F1-Score = 0.89), confirming the fact that the addition of a sarcasm recognition phase prior to the sentiment analysis stage has proved to be very useful. While traditional algorithms were ineffective in recognizing sarcasm due to their failure to account for Context Polarity Reversal; resulted in very poor results. Because of this, Any of the Above Results have a relationship to ECommerce in general, namely, the ability everyone involved in ECommerce can accurately assess customer reviews and feedback; thus, determining what consumers really think about your business (e.g., recommendations that are based on actual consumer opinion), providing you with an enhanced ability to satisfy your customers; and, allowing you to make more informed decisions about your business that are based on the results and insight you derived from analysing these types of data. The results presented in this paper provide further support for the current trend within the field of Natural Language Processing, where the modelling approach has shifted from traditional machine learning methods to more flexible, hybrid models that use transformer-based architectures with context awareness. Both the higher performance of deep learning models and hybrid models can be attributed to the fact that pragmatic language phenomena, such as sarcasm and irony, require an understanding of contextual polarity shifts and semantic complexities. Through the research presented in this paper, we have been able to demonstrate that sarcasm-aware and context sensitive models provide higher accuracy in sentiment analysis of e-commerce reviews than traditional models. This is consistent with the prior discussion on the differences among traditional approaches and their associated evaluation methods. The results further reinforce the relevance of these findings by providing an example of how they have practical business value due to a higher level of accuracy in identifying sentiment. Further, this research also confirms ongoing trends in NLP research toward combining transformer-based models and hybrid approaches, which represent the next generation of NLP efforts to address the complexities of language as used in the real world.

REFERENCES

[1] S.-Y. Lee and H.-J. Kim, "Customer sentiment analysis: Using deep learning for real-time e commerce applications," Expert Syst. Appl., vol. 97, pp. 105–116, 2018.
[2] K. Wase et al., "Sentiment analysis of product review," Int. J. Innov. Eng. Sci., vol. 3, no. 5, 2018.
[3] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," Adv. Intell. Syst. Comput., vol. 768, 2018.
[4] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "Sentiment analysis using convolutional neural network," in Proc. IEEE Int. Conf. Comput. Inf. Technol., 2018, pp. 2359–2364.
[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2018.
[6] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," Expert Syst. Appl., vol. 36, no. 7, pp. 10760–10773, 2019.
[7] X. Wang, C. Li, and J. Li, "Aspect-based sentiment analysis with multi-attention network," in Proc. 28th ACM Int. Conf. Inf. Knowl. Manag., 2019, pp. 2479 2482.
[8] J. Garcia-Madariaga, I. Cortés, N. Recuero, and M.-F. Blasco, "Sentiment analysis in e-commerce: A systematic literature review," Sustainability, vol. 11, no. 19, p. 5477, 2019.
[9] E. Cambria and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intell. Syst., vol. 28, no. 2, 2019.
[10] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in Proc. SIGIR, 2019, pp. 959–962.
[11] 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.
[12] Z. Sun, Q. Zhu, and X. Liu, "Aspect-level sentiment analysis using deep neural networks," Neurocomputing, vol. 330, pp. 190–198, 2019.
[13] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst., 2019, pp. 5998–6008.
[14] S. N. Alsubari et al., "Integrated neural network model for fake review identification in e-commerce," Appl. Bionics Biomech., vol. 2021, 2021.
[15] J. P. U. S. D. Jayakody and B. T. G. S. Kumara, "Sentiment analysis on product reviews on Twitter using ML approaches," in Proc. DASA, IEEE, 2021, pp. 1056–1061.
[16] B. Gaye, D. Zhang, and A. Wulamu, "Sentiment classification using regression vector-SGD classifier (RV-SGDC)," PeerJ Comput. Sci., vol. 7, p. e712, 2021.
[17] M. Li, L. Chen, J. Zhao, and Q. Li, "Sentiment analysis of Chinese stock reviews using BERT," Appl. Intell., vol. 51, no. 7, pp. 5016–5024, 2021.
[18] L. Davoodi, "Enhancing the understanding of e commerce reviews through BERT-based aspect extraction," in 36th Bled eConference, 2023.
[19] O. Bellar, A. Baina, and M. Bellafkih, "Sentiment analysis of tweets on social issues using ML approaches," in Proc. ICDATA, 2023, pp. 126–131.
[20] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in Proc. CIKM, 2020, pp. 625–631.