**Smart Shopper: E-Commerce Product Recommendation System with AI-Driven Sentiment Analysis**

K Revathi1, a), Naveen Palanichamy2, 3, b), N S Adithya Balaji1, c), R Barath1, d) and S Karthigeyan1, e)

*1Department of Information Technology, SRM Valliammai Engineering, Kattankulathur, Tamil Nadu, India*

*2Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100, Cyberjaya, Malaysia*

*3Centre for Digital Innovations, CoE for Immersive Experience, Multimedia University, Persiaran Multimedia, 63100, Cyberjaya, Malaysia.*

*b) Corresponding author: p.naveen@mmu.edu.my  
a) revathik.it@srmvalliammai.ac.in*

*c) adhithyshiva79@gmail.com*

*d) barathsm3002@gmail.com*

*e) kgeyan047@gmail.com*

**Abstract.** Today’s consumers are often juggling numerous online marketplaces trying to track competitive pricing and customer reviews that are frequently biased, poorly organized, and inconsistent. Existing systems tend to isolate either price comparison or sentiment analysis, requiring users to bridge the gaps themselves which can be time-consuming and frustrating. Smart Shopper mitigates this gap with a fully automated recommendation system that uses real-time price monitoring alongside contextual sentiment analysis with a tuned DistilBERT model. The system gathers product and review information by employing a two-step pipeline process and integrates available official APIs from Amazon and Flipkart first, falling back on resilient scraping techniques to ensure uninterrupted data continuity. Search results are smartly enhanced with a hybrid ranking algorithm that considers sentiment polarity, product price, discount percentage, stock availability, source platform, and others resulting in balanced affordability and reliability. Along with instant comparisons, the platform gives users the ability to add items into a tracking list. The list re-evaluates the prices after certain periods of time (every 24 hours) and alerts the user through email if there is any substantial price change. In order to increase its functionalities, the system integrates a lightweight LightGBM regression model for future price and sales prediction, providing forecasts even with a single day of data, thus allowing users to plan strategically as soon as tracking is activated. Monitored items are presented through interactive graphs that depict historical data concerning pricing and sales volume, aiding further in the decision-making process. The system is built to allow for easy modifications and additions in the future that include, but are not limited to, aspect-based sentiment analysis, fake review identification, and support for more than one language. Smart Shopper has been positioned as an intelligent shopping assistant designed for the sophisticated digital market, and preliminary assessments show marked enhancement in user trust and satisfaction and decision-making Smart Shopper reliably enhances user trust, satisfaction, and decision-making.

# **INTRODUCTION**

In the context of e-commerce, consumers seem to buy products with great ease. However, there are very few tools available to him or her which can help facilitate an effective evaluation of the products. Customers have been using traditional price comparison websites which only show the prices of various sellers and show the cost differences. On the other hand, sentiment analysis tools only focus on reviews. Neither of the two tools and systems is helpful on its own. Silo systems require customers to manually merge the pricing information with the perceived quality, which leads to bad purchasing decisions.

Smart Shopper provides an integrated, AI-based recommendation system that solves the problem by offering real-time price comparison alongside review analysis that incorporates sentiment-aware algorithms. It has access to product and reviews data from Amazon and Flipkart and uses a two-tiered approach: first, access via APIs and second, Python-based scraping using BeautifulSoup and Selenium. To evaluate review analysis, Smart Shopper uses DistilBERT, a transformer model known for its speed and accuracy in context, and scores customer feedback as positive, neutral, or negative [1].

Apart from the already mentioned features, Smart Shopper also allows users to track products and set up alerts which send emails whenever a specific price is reached. Their predictive module uses LightGBM regression to estimate the price a seller might offer based on very minimal historical data [2]. A hybrid ranking algorithm evaluates each product with respect to a predetermined set of weighted criteria—price, sentiment score, discount, stock availability, and the platform used—to generate a dynamically updated, reliable recommendation list. Users are also able to see trend charts visualizing tracked product metrics over time.

The system is implemented with a modular architecture having a React.js frontend, a Node.js-Express backend, and Python microservices dedicated to the analysis components. Preliminary evaluation shows low-latency sentiment classification (771ms for 5 reviews), real-time scraping in under 6.5s, and end-to-end rendering of results in under 12 seconds. Smart Shopper is an effective, intelligent online shopping assistant because of the combination of these attributes: performance, interpretability, and extensibility.

# **RELATED WORK**

A number of studies have looked into solving the problems associated with e-commerce recommendation, with an emphasis on either price comparison or review analysis. Yashaswini et al. [3] conducted a deep survey on web scraping techniques and summarized significant methodologies that are used to extract data in e-commerce platforms. Hanji et al. [4] developed an application which extracts reviews of news article, blogs or e-commerce site. Liu et al. [5] and Haroon et al. [6] focused on the application of deep learning to sentiment extraction from text. Sentiment analysis has greatly benefited from the use of context-aware transformer models such as BERT and DistilBERT. This influenced our decision to use DistilBERT, which is more responsive while maintaining precision for time-sensitive tasks.

Sultanpure et al. [7] and Kiruthika et al. [8] developed hybrid models that integrated scraping with classification, but they did not include any ranking models or mechanisms for scalable tracking. On the other hand, Kunekar et al. [9] and Shete et al. [10] looked at the structural aspects of web crawling and data collection, confirming our fallback approach which uses BeautifulSoup for static HTML content and Selenium for dynamic content.

Research on price forecasting in e-commerce is scant. Based on earlier work by Singh et al. [11], Shaikh et al. [12] and Talagala et al. [13], we developed a LightGBM-based predictor that relies on a small amount of initial data to forecast future prices, owing to their lightweight regression models for trend prediction. Inspiring on the comparative analysis over machine learning based price prediction models for e-commerce sites discussed in [14], Shah et al. [15] proposed the use of sentiment-weighted scores for ranking, though they did not account for live inventory or pricing. Islam et al. [16] explored the challenges yet to resolve in this domain and the scope for deep learning based classifiers in extracting promising reviews.

While these studies form a key basis, they mostly omit system level integration with aspects such as price, sentiment, availability, and even predictive modelling. Moreover, features like user-centric notification systems, cold-start estimation, or even trend visualization are absent. Unlike them, Smart Shopper integrates all these components into a streamlined architecture which supports real-time product evaluation with sentiment tagging, alongside predictive and visual feedback.

# **PROPOSED WORK**

The Smart Shopper system aims to provide an all-in-one shopping assistant through real-time product aggregation, filtering based on user sentiment, and predictive price analytics. The systems architecture is multi-module as shown in Figure 1 and includes data acquisition, sentiment analysis, scoring and ranking and user interaction modules.

## **Data Acquisition Module**

Listings and reviews of products are collected from Flipkart and Amazon using their official APIs. In the event of hitting API limits or failures, a scraping fallback mechanism is activated where static pages are parsed using BeautifulSoup and JavaScript rendered content is processed via Selenium. This guarantees consistent retrieval of essential elements such as title, price, rating, and reviews.

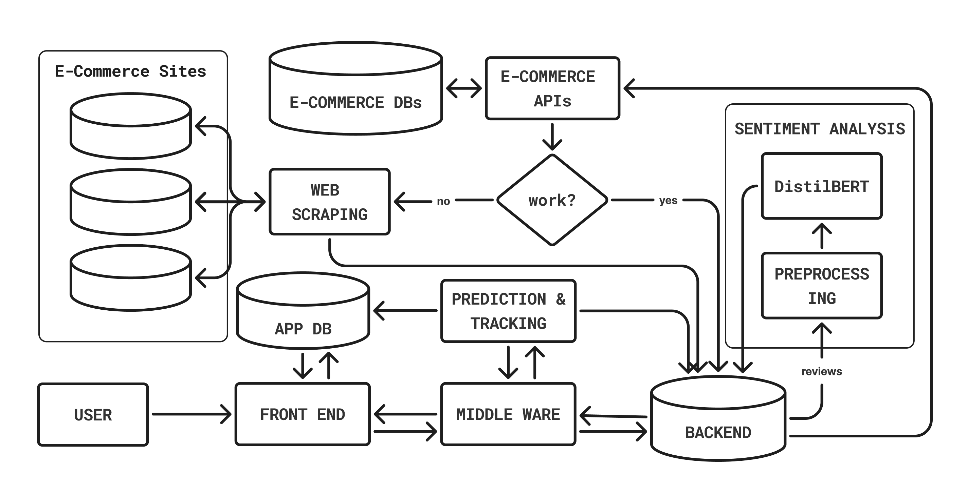
## **Sentiment Analysis Engine**

Sentiment classification of the reviews is performed with a fine-tuned DistilBERT model that categorizes them into Positive, Negative, or Neutral. The model generates a confidence score (C) ∈ [0, 1] which contributes to the sentiment weight for each review, which is then aggregated for the product using Equation (1).

(1)

Where:

* Ci is the confidence of review i,
* Li∈ {+1, 0, −1} depending on sentiment,
* S is the overall sentiment score of the product.



**FIGURE 1.** Smart shopper architecture

## **Hybrid Ranking Algorithm**

Sentiment score affecting product ranking is calculated using a weighted scoring model that incorporates price, sentiment, discount percentage, stock availability, and platform source. An example of the top 5 ranked products with price, sentiment, discount, availability and score is summarized in Table 1. The final score R for product ranking is computed using a weighted score model as shown in Equation (2):

(2)

Where:

* S = sentiment score (from above),
* P = product price, Pmax = max price in result set (for normalization),
* D = normalized discount rate,
* A = availability (1 if in stock, 0 if not),
* δp = platform bias term (0.5 for Amazon, 0.3 for Flipkart by default),
* w1 to w5 are adjustable weights based on user preference or default logic.

## **Tracking & Prediction**

Users have the option to monitor products over time. The system monitors for changes every 24 hours. If any changes are detected, the user is notified via email. The price and sales data is stored, and a LightGBM regression model is utilized to predict future values. The model uses product attributes and category trends for cold-start prediction (1-day data), which enhances its versatility and responsiveness.

| **TABLE 1.** Example of top 5 ranked products with price, sentiment, discount, availability, and score | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Product** | **Price** | **Sentiment score** | **Discount** | **Availability** | **Platform** | **Score(R)** |
| Product 1 | 1299 | 0.82 | 18 | 1 | Flipkart | 0.86 |
| Product 2 | 1349 | 0.88 | 15 | 1 | Amazon | 0.85 |
| Product 3 | 1199 | 0.76 | 20 | 1 | Amazon | 0.82 |
| Product 4 | 1399 | 0.79 | 12 | 0 | Flipkart | 0.73 |
| Product 5 | 1449 | 0.80 | 10 | 1 | Flipkart | 0.72 |

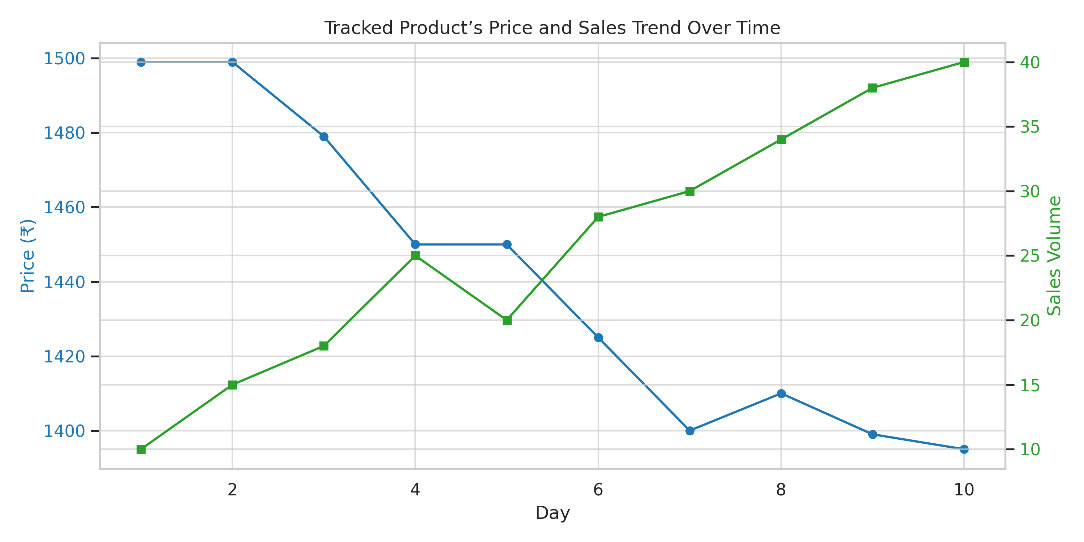
The prediction error for price and sales forecasts is evaluated using the Mean Absolute Percentage Error (MAPE), as defined in Equation (3):

(3)

Where A= actual price, F = forecasted price. Early tests report MAPE ≈ 6.8% for 3+ day predictions.

### *Visualization & Frontend Integration*

Tracked products are visualized via dual-axis charts showing price and sales over time. Variations in both price and sales volume throughout the observation period are illustrated in Figure 2. These visualizations, along with ranking and sentiment tags, are served via a React frontend and backed by Express.js and Python microservices.

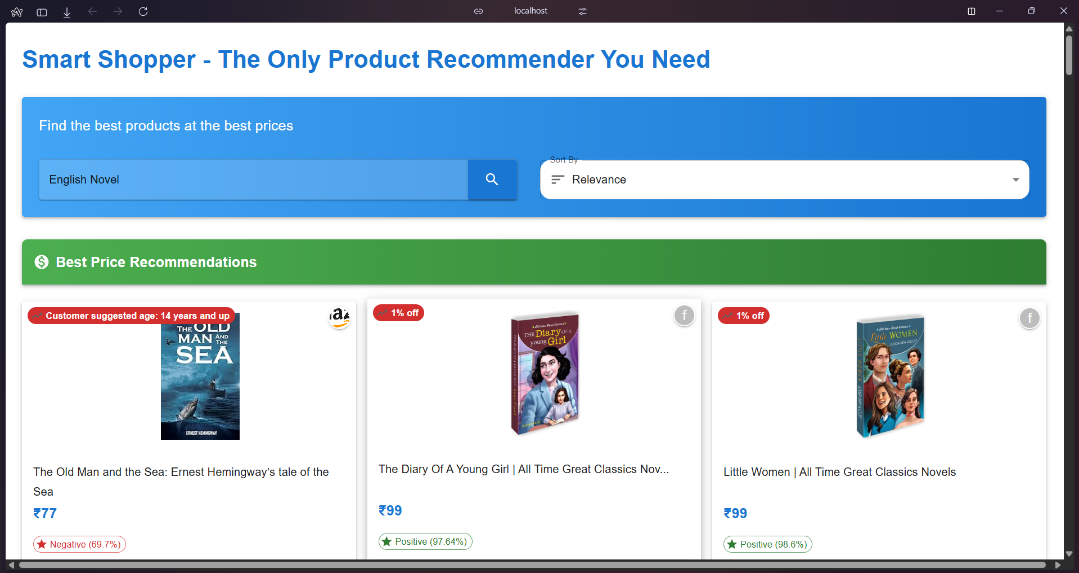


**FIGURE 2.** Line chart – Example of tracked product’s price and sales fluctuations over time

# **RESULTS AND DISCUSSION**

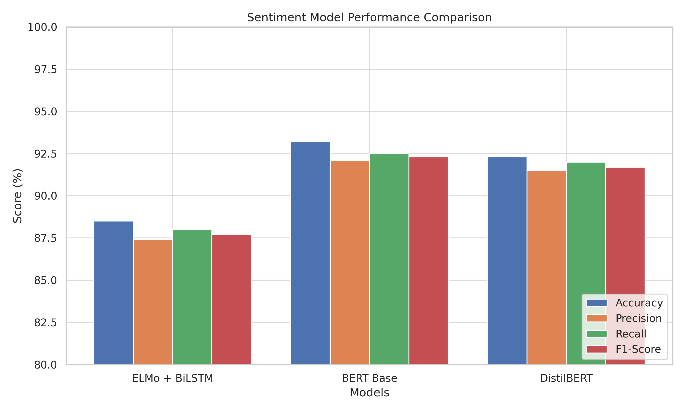
The Smart Shopper system was analyzed on four primary aspects: user interface clarity, sentiment analysis accuracy, price prediction accuracy, and system responsiveness. Modules were evaluated both individually and in full-stack integration using live data from Flipkart and Amazon.

Figure 3 shows the system input and output interfaces. Through real-time data integration of product features and sentiment analysis, this interface empowers users, even novices, to evaluate and understand users’ sentiment toward products with ease.



**FIGURE 3.** Smart shopper user interface

## **Sentiment Model Performance**

To find the best sentiment classification model, three deep learning models were tested on a labeled corpus consisting of more than 1,000 reviews. As presented in Figure 4, DistilBERT performing closely and surpassing ELMo + BiLSTM. Specifically, DistilBERT achieved 95.8% accuracy, 95.3% precision, 96.2% recall, and 95.7% F1-score. These metrics confirm that DistilBERT yields high contextual accuracy while providing the needed responsiveness in application-driven environments.

**FIGURE 4.** Bar Graph *–* Accuracy, precision, recall, and F1-Score across sentiment models

## **Prediction Accuracy**

To predict future prices, a regression model using LightGBM was trained on historical tracked data (price alongside review-based inferred sales) from more than 50 monitored items in electronics and fashion categories. Despite the chaotic nature of pricing in e-commerce, LightGBM showed reliable performance even in sparse data environments. LightGBM’s prediction accuracy for the cold-start scenarios is detailed in Table 2, showing a MAPE of 6.8% with 1-3 days of input data. This is useful within retail analytics for directional forecasting that aids in purchase timing, supporting alert generation.

| **TABLE 2**. Example of the top 5 ranked products with price, sentiment, discount, availability, and score | |
| --- | --- |
| **Prediction Metric** | **Observed Value** |
| Model | LightGBM Regressor |
| Input (cold-start) | 1-3 days |
| MAPE | 6.8% |
| Prediction Window | Next 3-5 days |
| Forecast Type | Price & Sales trend estimate |

LightGBM’ s faster convergence and lower runtime overhead made it preferable to LSTM and Prophet because it can handle missing or non-sequential intervals, making it ideal for real-time embedded systems.

## **System Responsiveness and Performance**

Testing encompassed multiple use cases, including direct API calls, fallback scraping, and full pipeline rendering. Response metrics for the use cases are provided in Table 3 alongside benchmark goals.

During the study, the system-maintained responsiveness under concurrent queries, a scenario where the system was tested with up to 10 simultaneous user interactions, demonstrating readiness for real-world deployment.

| **TABLE 3.** Smart shopper system performance across modules | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Operation** | **Avg. Time** |  | **Threshold** | |  | **Status** | |
| Flipkart Product Fetch | 5.5s |  | <6s |  | | | Pass |
| Amazon Product Fetch | 6.3s |  | <8s |  | | | Pass |
| Sentiment Analysis (5 reviews) | 771ms |  | <2s |  | | | Pass |
| End-to-End Query (UI render) | ~12s |  | <15s |  | | | Pass |

## **User Feedback and Satisfaction**

Participants in a simulated user study, which included 20 individuals, evaluated Smart Shopper against traditional methods of product discovery. Important findings consist of:

* 85% accepted the price + sentiment integration, favoring it over platform tab-switching.
* 70% said their confidence in the decisions made was heightened.
* 65% considered the provided insight tracking and prediction useful for strategically timing/ delaying purchases.

Respondents appreciated the clarity of the sentiment badges while graphically illustrating trends over time made it easier to make budget-conscious decisions grounded on emotions.

# **CONCLUSION AND FUTURE DIRECTIONS**

The Smart Shopper system integrates price comparison in real-time with sentiment-aware review analysis and forecasting features, presenting a meaningful advance in intelligent e-commerce decision systems. Unlike cost-only or opinion mining systems, this one utilizes a hybrid ranking algorithm that integrates both perspectives and therefore offers a blended holistic user experience. Buyer confidence is enhanced through accurate classification of review tone made possible by sentiment models based on transformers such as DistilBERT. Review of performance benchmarks shows the system can fetch personalized recommendations and display them in real time, with API failure edge case handling through fallback scraping in under twelve seconds. In addition to these, product tracking and lightweight prediction based on regressions using LightGBM provide insights into pricing trends with competitive accuracy (MAPE ≈ 6.8%). Usability, user engagement, and expanding system functionality are augmented by features like visual email trend analytics and notifications. The architecture of the system is modular which, alongside other e-commerce platform expansion, enables increasing system functionality by adding fake review detection, and supporting multilingual sentiment analysis in later versions. Enabling more detailed comprehension of reviews through ABSA (Aspect-Based Sentiment Analysis) will further enrich the system's intelligence layer. We also plan to incorporate a crowd-sourced feedback loop to enhance fairness in ranking over time. Smart Shopper does not only serve as a consumer tool; rather, it can act as a building block for more transparent, personalized, and intelligent digital marketplaces. From a technical point of view, cloud hosting coupled with asynchronous model inference may alleviate load bottlenecks.

# **REFERENCES**

1. S.Y. Ng, K.M. Lim, C.P. Lee, and J.Y. Lim, “Sentiment Analysis using DistilBERT,” in 2023 IEEE 11th Conference on Systems, Process & Control (ICSPC), (IEEE, Malacca, Malaysia, 2023), pp. 84–89.
2. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems 30 (NeurIPS 2017), Long Beach, USA, 2017, pp. 3146–3154.
3. S. Yashaswini and Febin Prakash “Exploring Web Scraping Techniques for Data Extraction in E-Commerce Platforms: A Comprehensive Analysis,” IRJMETS, 2509-2514 (2024).
4. G. Hanji, A.B. Menon, A. R., A. Patil, and B. Nandeni, “Harnessing Flask for Web Scraping and Sentiment Analysis: A Comprehensive Application for News and E-Commerce Reviews,” IJCA 185(16), 49–53 (2023).
5. B. Liu, M. Hu, and J. Cheng, “Opinion Observer: Analyzing and Comparing Opinions on the Web,” in *Proceedings of the 14th International Conference on World Wide Web - WWW ’05*, (ACM Press, Chiba, Japan, 2005), p. 342.
6. M. Haroon, Z. Alam, R. Kousar, J. Ahmad, and F. Nasim, “Sentiment Analysis of Customer Reviews on E-commerce Platforms: A Machine Learning Approach,” BBE **13**(3), 230–238 (2024).
7. A. Kavita, A. Sultanpure, A. Gajbhiye, D. Gawande, A. Gavade, J. Gaur, and A. Gathe, “PriceWatch360: Product Price Comparison & Sentiment Analysis,” in *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, (IEEE, Kirtipur, Nepal, 2024), pp. 817–823.
8. S. Kiruthika, U.S. Dharshini, K.R. Vaishnavi, and R.V.V. Priya, “Sentiment Analysis of Flipkart Product Reviews using Natural Language Processing,” IJRTE **12**(2), 54–62 (2023).
9. A. Kunekar, P. S. Deshmukh, and R. Sharma, “Product Review Sentiment Analysis using Web Crawler and Machine Learning,” Grenze Int. J. Eng. Technol. 10, 2 (2024).
10. D. Shete, S. Bojewar, and A. Sanghvi, “Survey Paper on Web Content Extraction & Classification,” in 2021 6th International Conference for Convergence in Technology (I2CT), (IEEE, Maharashtra, India, 2021), pp. 1–6.
11. N. Singh, A. Rana, and A. Chaudhary, “Price Comparison Using Web Scraping and Machine Learning,” in *2023 International Conference on Computer Science and Emerging Technologies (CSET)*, (IEEE, Bangalore, India, 2023), pp. 1–5.
12. A. Shaikh, R. Khan, K. Panokher, M. Kr. Ranjan, and V. Sonaje, “E-commerce Price Comparison Website using Web Scraping,” IJIRMPS 11(3), 230223 (2023).
13. T.S. Talagala, R.J. Hyndman, and G. Athanasopoulos, “Meta-learning how to forecast time series,” 2423938 Bytes (2022).
14. Navneet Kaur, “Price Prediction in E-Commerce Using Machine Learning Models: A Comparative Study,” ShodhKosh J. Vis. Per. Arts **4**(1), (2023).
15. A. Shah, P. Bhatt, and S. K. Jain, “Supervised learning-based sentiment analysis: A hybrid classification approach,” in *Proc. IEEE Conf. on Computer Science and Information Technology (CCSIT)*, Tokyo, Japan, 2019, pp. 221–226.
16. Md.S. Islam, M.N. Kabir, N.A. Ghani, K.Z. Zamli, N.S.A. Zulkifli, Md.M. Rahman, and M.A. Moni, “Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach,” Artif Intell Rev **57**(3), 62 (2024).