IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Cross Domain Sentiment Analysis using AI and ML

Tanmayi Kulkarni, Kashish Bandal, Yash Salvi, Nidhi Zanjal, Prof. P. U. Dere, Department of Electronics and telecommunication engineering, Terna Engineering College, Plot no.12, Sector-22, Opp. Nerul Railway Station, Phase-11, Nerul (W), Navi Mumbai 400706, University of Mumbai, Maharashtra, India.

Abstract: —This paper is about cross-domain sentiment analysis for product reviews. People write reviews for products like electronics, books, and clothes on websites like Amazon. Sentiment analysis helps to find whether these reviews are positive, negative, or neutral. But sometimes, there are not enough labelled reviews in one domain to train the system. So, cross-domain sentiment analysis uses data from one domain to help another domain. In this paper, a web based system is implemented that collects reviews, processes text using NLP techniques, and classifies the sentiment using machine learning models. This helps to understand customer opinions and improve decision-making for products across different domains. Additionally, the system integrates an AI-powered chatbot that interacts with users in real time, answers queries about products, provides sentiment summaries, and guides customers in exploring product feedback efficiently.

(Keywords— AI Chatbot interaction, Cross-domain sentiment analysis, Machine learning models, Natural language processing)

I. INTRODUCTION

People today share their product experiences on online platforms through reviews. These reviews contain useful opinions that help other customers and companies understand product quality and satisfaction levels. Sentiment analysis is a technique used to identify whether a review expresses a positive, negative, or neutral feeling. It uses Natural Language Processing (NLP) and Machine Learning methods to analyze text data. When reviews come from different product categories like electronics, books, and clothing, analyzing them together across domains is called cross-domain sentiment analysis. This technique helps in understanding customer emotions and opinions from various fields more effectively. Cross-domain sentiment analysis reduces the need to build separate models for each category by transferring knowledge from one domain to another. It also helps in improving the accuracy of predictions when there is limited labelled data in a specific domain. By studying customer feedback from multiple areas, businesses can identify overall trends, improve product quality, and enhance user satisfaction. This approach is widely used in e-commerce platforms to make better product recommendations and decision-making. Customers face difficulty understanding thousands of mixed reviews online. Most review systems are domain-specific (only work for one type of product). There is a need for a cross domain sentiment analyzer that can handle multiple product types. The challenge is to maintain accuracy and consistency across domains with different vocabularies. Manual reading of reviews is time-consuming and confusing for buyers. Our system provides a quick and reliable way to understand overall customer opinions. The project aims to reduce confusion and improve the online shopping decision-making process.

II. LITERATURE SURVEY

Sentiment analysis of product reviews has become increasingly important due to its ability to enhance customer experience, inform business decisions, and improve recommendation systems. Recent studies have demonstrated the use of machine learning and pre-trained language models, such as BERT, to improve sentiment classification by capturing semantic meaning and context more effectively than traditional methods. For instance, the study published in Big Data and Cognitive Computing (2024) illustrates how pre-trained large language models enhance cross-domain sentiment analysis by handling domain-specific vocabulary and applying transfer learning for improved accuracy across multiple product categories. Similarly, research in Computers in Human Behavior (2026) highlights the correlation between textual sentiment and numerical product ratings, showing that sentiment aware models can enhance predictive analytics and improve recommendations in generative AI systems. Practical workflows for analyzing product reviews have been explored extensively. Unwrangle (2025) discusses AI tools and processes for review analysis, emphasizing data collection, preprocessing, classification, and visualization. Clean and structured datasets, as well as real time analysis, are crucial for actionable insights. Preprocessing techniques, including tokenization, stop word removal, stemming, lemmatization, and embedding representation, are essential for high-performing models. The work in IEEE Access (2022) and Expert Systems with Applications (2024) emphasizes that deep learning-based models require additional preprocessing steps, such as handling emojis, punctuation, and domain-specific terms, along with embedding normalization, to ensure accurate sentiment interpretation. Moreover, advanced vectorization and tokenization methods, including sub word tokenization and contextual embeddings, significantly improve sentiment detection, especially in noisy real world e-commerce datasets, as shown in Journal of King Saud University - Computer and Information Sciences (2023). Cross-domain sentiment analysis has been a key area of research, particularly when labelled data is limited in specific product categories. A study presented in ACL Proceedings (2023) proposes a bidirectional generative framework for crossdomain aspect-based sentiment analysis, enabling models trained in one domain to generalize effectively to others while extracting fine-grained aspect-level sentiments. The importance of dataset preprocessing for cross-domain applications is further highlighted in Journal of Big Data (2023) and ICDM Workshops (2023), which discuss cleaning, duplicate removal, handling missing values, and aligning feature spaces to ensure that models can transfer knowledge across domains seamlessly. These approaches ensure that sentiment analysis systems remain robust even when working with heterogeneous datasets across electronics, books, and clothing reviews. Equally important is the design and visualization of web-based interfaces. Research in IEEE HCI Proceedings (2023) emphasizes responsive and interactive web design principles for AI powered systems, including dashboards and chatbots. AI chatbots provide real-time interaction, answering queries about products, summarizing sentiment results, and guiding users through the analysis process. Dashboards offer clear visualizations of domain-specific sentiment statistics, overall trends, and product-level insights, enabling both customers and businesses to make informed decisions efficiently. Together, these studies provide a comprehensive foundation for developing a web-based, AI- powered cross-domain sentiment analysis system. Preprocessing techniques, machine learning and deep learning models, cross-domain adaptation strategies, and effective web- based user interfaces all contribute to a robust and user-friendly system. By integrating these best practices, product review analysis can deliver actionable insights, improve recommendation accuracy, and enhance the overall customer experience. The combination of large-scale data processing, advanced modelling, cross-domain generalization, and interactive AI interfaces represents the current state-of-the-art in sentiment analysis research for e-commerce applications.

III .METHODOLOGY

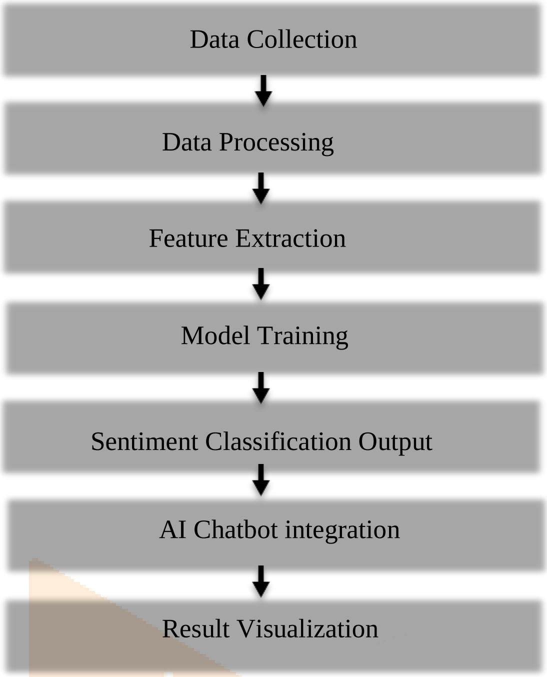

Fig 1: Flowchart

1. User Input & URL Acquisition

The system begins with a user-friendly web interface developed using Streamlit, which allows users to easily interact with the application. In this stage, the user is required to provide the Amazon product review URL as input. The interface validates the URL format to ensure it corresponds to a legitimate Amazon product page. This step is crucial because the quality and accuracy of the analysis depend directly on the correctness of the input URL. By simplifying the input process through a clean and intuitive interface, the system ensures accessibility even for non-technical users. Once the URL is submitted, it is passed to the backend pipeline for further processing and review extraction.

2. Web Scraping of Reviews

After receiving the product URL, the system performs web scraping to collect customer reviews and relevant product details from the Amazon webpage. This process is carried out using ScraperAPI along with BeautifulSoup. ScraperAPI helps bypass common challenges such as bot detection, IP blocking, and dynamic content loading, which are frequently encountered on large e-commerce platforms like Amazon. BeautifulSoup is used to parse the HTML structure of the webpage and extract essential data elements such as review text, ratings, reviewer names, and timestamps. This automated extraction enables the system to efficiently gather large volumes of data in real time, forming the foundation for sentiment analysis and review classification.

3. Data Pre-processing

The raw data collected through web scraping often contains HTML tags, unnecessary whitespace, emojis, special characters, and non-textual elements. Therefore, a dedicated data pre- processing stage is implemented to clean and normalize the extracted reviews. During this phase, the system removes HTML tags, converts text to a standardized format, eliminates duplicate reviews, and filters out irrelevant symbols. Stop words and excessive punctuation may also be removed to improve analysis accuracy. This step ensures that the review data is structured, consistent, and suitable for Natural Language Processing (NLP) tasks. Proper pre-processing significantly enhances the reliability of sentiment classification and analytical insights.

4. Sentiment Analysis

Once the reviews are cleaned, the system applies sentiment analysis techniques to determine the emotional tone of each review. Using NLP-based sentiment models, the system calculates a sentiment polarity score, which reflects whether a review expresses a positive, negative, or neutral opinion. Based on predefined thresholds, reviews are categorized into Positive, Negative, or Neutral classes. This classification helps in understanding customer satisfaction, identifying common complaints, and recognizing product strengths. Sentiment analysis plays a vital role in converting unstructured textual data into meaningful insights that can assist consumers and businesses in decision-making.

5. Visualization & Analytics Dashboard

To present insights in an easily understandable manner, the system generates an interactive analytics dashboard using Plotly. Various visualizations such as sentiment distribution bar charts, aspect- based radar charts, and fake vs genuine review pie charts are displayed. These visual tools allow users to quickly grasp overall sentiment trends, identify dominant product aspects (such as quality, price, or durability), and evaluate review authenticity. The interactive nature of the dashboard enables users to explore the data dynamically by hovering, zooming, and filtering results, making the analysis more intuitive and engaging.

6. Interactive User Query Module & Web Deployment

The system also includes an AI-powered interactive assistant that allows users to ask natural-language questions related to the product, such as key advantages, disadvantages, or overall performance analysis. This module enhances user experience by providing personalized and context-aware insights derived from the analysed reviews. Finally, the complete application is deployed as a cloud-based web system using Stream-lit. This deployment ensures real-time access, scalability, and ease of use without requiring local installation. Users can access the platform from any device with an internet connection, making the system practical, efficient, and widely accessible.

System Features:

1	Cross-Domain Sentiment Analysis	The system classifies product reviews from multiple domains such as electronics, books, and clothing as positive, negative, or neutral
2	Knowledge Transfer Across Domains	Cross-domain techniques allow the system to leverage knowledge from one domain to improve predictions in another, especially when labelled data is limited.
3	AI Chatbot Integration	An AI-powered chatbot provides an interactive interface, enabling users to ask questions about specific products, domains, or overall sentiment
4	Dashboard and Visualization	The web-based application includes a dashboard that displays sentiment statistics, trends, and visualizations to help users and businesses easily interpret customer opinions.
5	User Friendly Interface	The system is user-friendly, allowing smooth navigation and efficient exploration of product feedback.

6	Extensibility	It is designed for improvements, with potential for integration of more product categories, multilingual support, and advanced deep learning models like BERT.

IV. IMPLEMENTATION AND RESULTS

The review system is an advanced web-based intelligent system developed to showcase the practical implementation of generative artificial intelligence for automated review analysis and interactive text-based assistance. The application is designed and implemented using the Streamlit framework, which facilitates the rapid development of an intuitive, responsive, and visually accessible web interface, enabling seamless interaction between the user and the system. Python is employed as the primary programming language to implement backend logic, manage user inputs, handle application flow, and coordinate communication with external services. The core intelligence of the system is powered by Google's Gemini Generative AI model, which provides sophisticated natural language processing capabilities, including semantic understanding, contextual reasoning, coherent text generation, summarization, and analytical interpretation of user-provided content.

The proposed system is implemented as an AI-powered web application using **Streamlit** for the frontend interface and **Python** for backend processing. The application performs **automated product review analysis** by combining web scraping, Natural Language Processing (NLP), sentiment scoring, and generative AI.

● First, the system collects real-time product data. When a user enters an Amazon product URL, the application uses **ScraperAPI** along with **BeautifulSoup** to extract the **product title** and **customer reviews** from the webpage. These reviews form the raw textual dataset for analysis.

● Next, the system performs **sentiment analysis** using a model from the NLTK library. Each review is assigned a compound sentiment score ranging from –1 (very negative) to +1 (very positive). Based on this score, reviews are classified into **Positive**, **Negative**, or **Neutral** categories. The average sentiment score is then used to generate an overall product recommendation such as *Must Buy*, *Good Buy*, or *Think Again*.

● To provide deeper insight beyond simple polarity, the system computes a **Product DNA score**. Reviews are grouped according to keyword-based dimensions such as **Quality, Value, Usability, Durability, and Service**. Sentiment is calculated for each dimension and scaled to a 0–10 range. These scores are visualized using a **radar chart**, allowing users to understand strengths and weaknesses of the product at a glance.

For intelligent interpretation, the system integrates **Google Gemini AI** through the google-generative ai API. Gemini is used for:

● Extracting structured product metadata (Company, Model, Category)

● Generating human-like insights such as top pros, cons, and custom analytical responses based on user queries

The results are displayed through an interactive dashboard with:

- Metric cards (company, model, category, recommendation)
- Sentiment distribution pie chart
- Product DNA radar visualization
- AI-powered analyst chat interface
- Review-level sentiment table

Therefore, the implementation successfully integrates web scraping, sentiment analysis, generative AI, and interactive visualization into a unified system for intelligent product review evaluation. By automating the extraction of reviews, applying NLP techniques for sentiment scoring, and leveraging AI for contextual insights, the system transforms unstructured customer feedback into meaningful, decision-support information. The modular design ensures smooth data flow from acquisition to visualization, while the interactive dashboard enables users to easily interpret both overall sentiment and feature-specific performance. This implementation demonstrates how AI-driven tools can enhance product analysis by combining computational accuracy with human-like interpretation.

Figure 1 *shows the user interface of the Sentiment Analysis system, where users can paste a live Amazon product URL. The system fetches real-time review data from Amazon, analyses sentiments across product categories such as electronics, books, and clothing, and displays the overall customer opinion.*

Electronics (IPhone)

Figure 2 *shows the product page of the Sentiment Analysis system displaying the analysed results for a selected Amazon product. It presents key product details such as brand, model, category, and recommendation status, along with a visual representation of sentiment scores. The radar chart illustrates ratings for quality, value, usability, durability, and service, helping users quickly understand the overall product performance.*

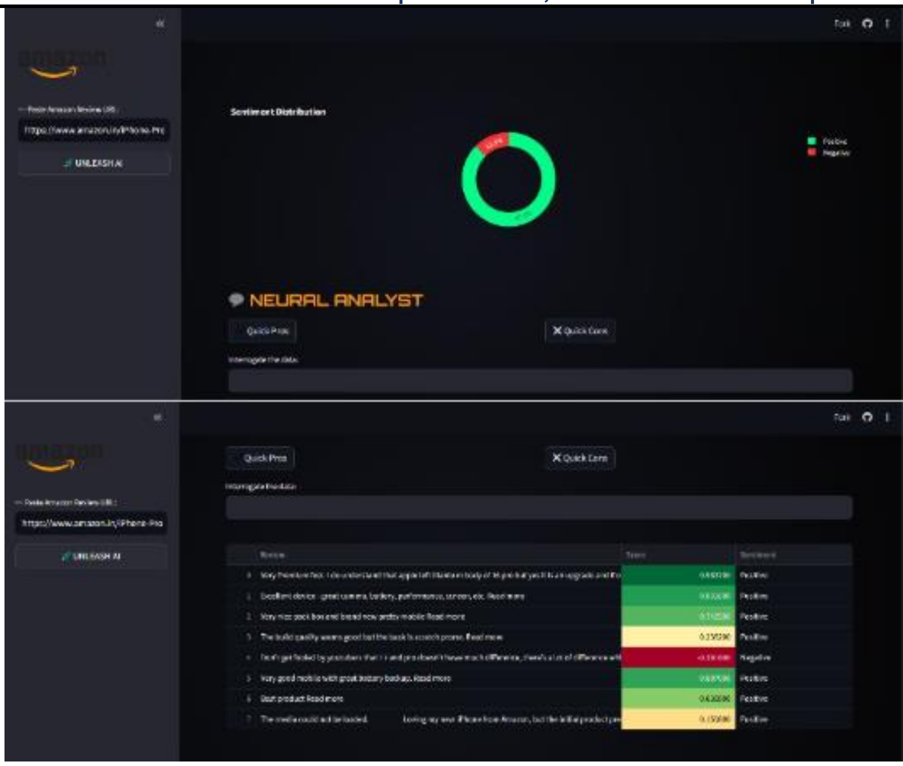

***Figure 3** illustrates the sentiment distribution of customer reviews, highlights top user reviews with their sentiment scores, and includes an interactive chatbot that provides quick insights and analysis based on the extracted review data.*

Same results for the following -

Electronics (Micromax)

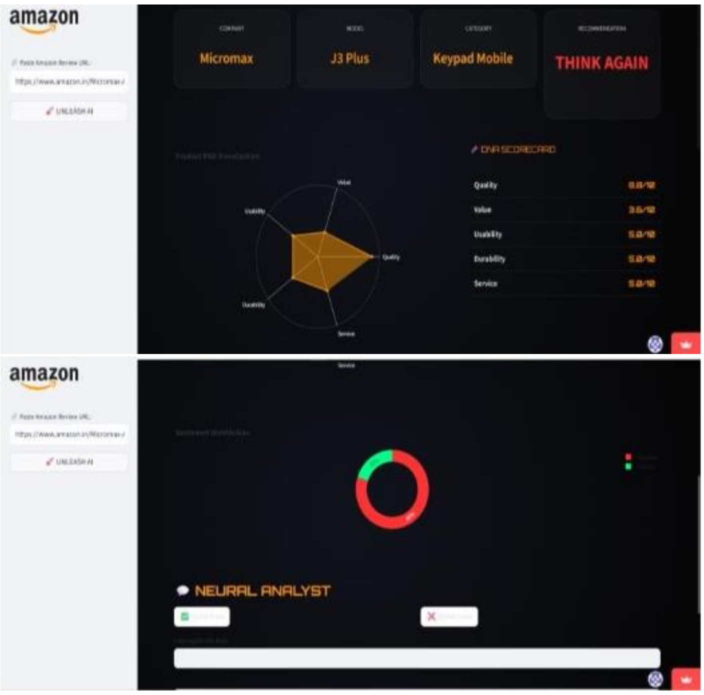

Figure 4 and 5 display details for Micromax phone

Books (Harry Potter and the Philosopher's Stone)

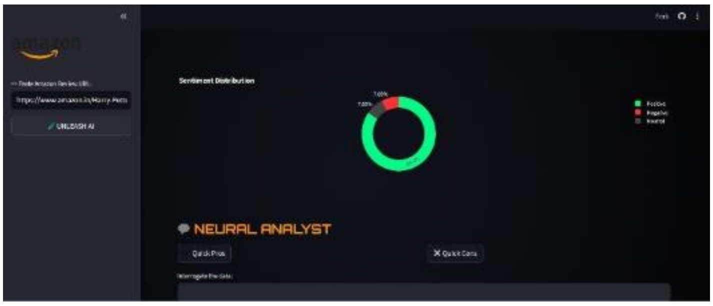

Figure 6, 7 and 8 display details for the book

Books (The Worst Book Ever Written)

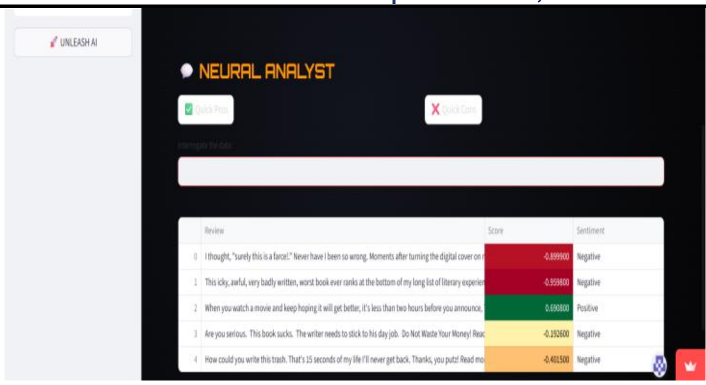

Figure 9, 10 and 11 display details for the book

Clothes (Branded)

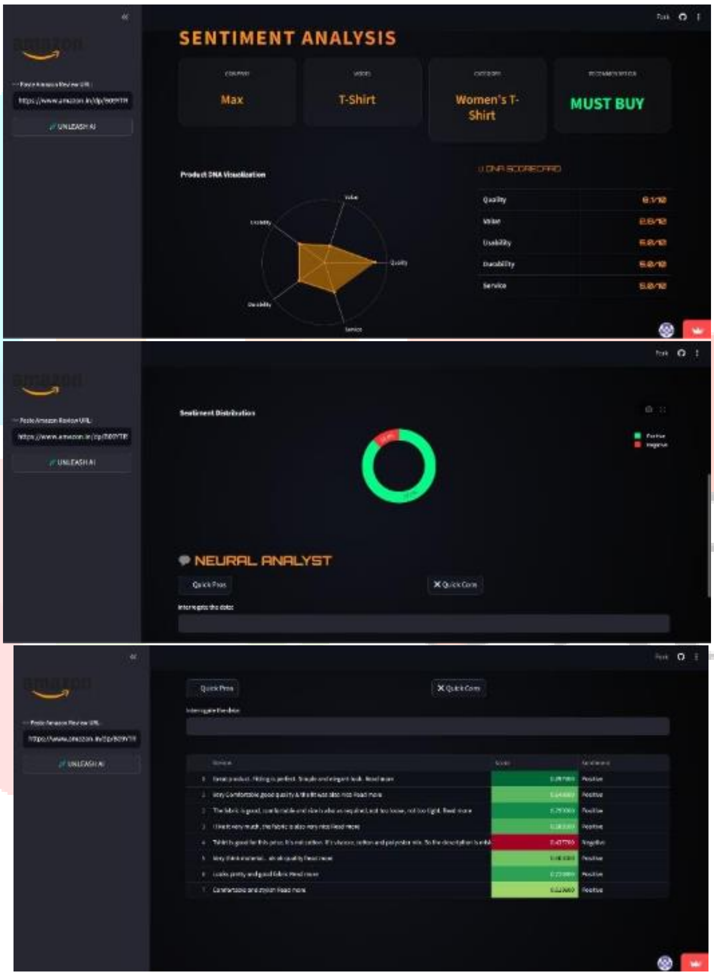

Figure 12, 13 and 14 display details for the cloth

Clothes (Non- Branded)

Figure 15,16 and 17 display details for the cloth

V. CONCLUSION

The project successfully demonstrates how sentiment analysis can simplify online shopping. It analyses and classifies customer reviews across multiple domains, helping both customers in product selection and businesses in understanding feedback. The system integrates data visualization to enable smarter AI-driven decision making and build trust in ecommerce by highlighting genuine customer opinions. The developed AI- powered chatbot provides an interactive interface for exploring product reviews across multiple domains in real time. It combines sentiment analysis and visualization to make insights easily accessible and understandable for users. Overall, the chatbot efficiently transforms review data into actionable insights that benefit both customers and businesses.

VI. REFERENCES

1. Meng, Y. (2025). How Review Sentiment Influences Ratings in Generative AI Systems. Computers in Human Behavior, 96, 339– 349.
2. Miyatra, S., et al. (2025). Review Analysis: AI Tools & Process for Sentiment & Product Reviews. *Unwrangle Blog*, April 28, 2025. Retrieved from https://www.unwrangle.com/blog/review-analysis
3. Ghatora, P. S., et al. (2024). Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained Large Language Models. Big Data and Cognitive Computing, 8(12), 199.
4. L. Chen et al., "Text Preprocessing for Deep Learning-Based Sentiment Analysis: Techniques and Best Practices," Expert Systems with Applications, vol. 220, p. 119715, 2024.
5. Y. Deng et al., "Bidirectional Generative Framework for Crossdomain Aspect-Based Sentiment Analysis," in Proc. 2023 Annu. Meeting Assoc. Comput. Linguistics, 2023, pp. 686–696.
6. [M. R. Islam and M. H. Kabir, "Dataset Preprocessing Techniques for Machine Learning and Sentiment Analysis of E-Commerce Reviews," Journal of Big Data, vol. 10, no. 1, pp. 1–19, 2023.
7. Y. Liu and B. Wang, "Cross-Domain Data Preprocessing for AspectBased Sentiment Analysis," in Proc. 2023 IEEE Int. Conf. Data Mining Workshops (ICDMW), pp. 512–519.
8. S. Kumar and A. K. Singh, "Advanced Tokenization and Text Vectorization Techniques in NLP for Sentiment Analysis," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 2, pp. 1445–1458, 2023.
9. R. K. Sharma and S. Verma, "Responsive and Interactive Web Design for AI-Powered

Customer Service Systems," in Proc. 2023 IEEE Int. Conf. Human-Computer Interaction (HCI), 2023, pp. 2

10. P. Kaur and A. Kaur, "Text Preprocessing Techniques for Sentiment Analysis: A Survey," IEEE Access, vol. 10, pp. 58210– 58234, 2022.