Journal of Computer and Information Systems Ampera
Vol. 6, No. 3, September 2025 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

Human-Computer Interaction in E-Commerce: A Study of Tokopedia's Usability, Efficiency, and Satisfaction for Elderly Consumers

Rahman Radjab Do Esa[1], Alvianus Dengen[2], Budiawan[3]

[1,2]Electrical Engineering Department, Universitas Teknologi Sulawesi, Makassar, Indonesia
[3]Political Science Department, Universitas Teknologi Sulawesi, Makassar, Indonesia
Email: [1]rahman_radjab85@gmail.com, [2]alvianus086@gmail.com, [3]budiawanuts@gmail.com

Abstract

Introduction: Elderly individuals are increasingly engaging with e-commerce platforms like Tokopedia. However, cognitive and technological barriers may affect their user experience. This study aims to analyze elderly user perceptions of Tokopedia using sentiment analysis of user reviews. Method: The data were collected from elderly users' reviews on Tokopedia. Sentiment classification into positive, negative, and neutral categories was conducted using machine learning algorithms. Model performance was evaluated using precision, recall, and f1-score metrics, presented through heatmaps and pie charts. Results: The analysis revealed that 58.0% of reviews were negative, 34.8% positive, and 7.2% neutral. The classification model yielded a high f1-score for the negative class (0.91) and perfect precision for the positive class (1.00), but relatively low recall for the positive sentiment (0.71), suggesting challenges in identifying subtle satisfaction expressions from elderly users. Discussion: These findings indicate persistent barriers in the online shopping experience for elderly users, particularly in terms of interface and usability. Recommendations are proposed to improve age-friendly design and platform accessibility.

Keywords: Accessibility, E-commerce, Elderly, Sentiment Analysis, Tokopedia

1. INTRODUCTION

The development of digital technology has transformed how people interact, including in consumption and shopping activities. E-commerce has become one of the fastest-growing sectors, especially in Indonesia.

Tokopedia, as one of the largest e-commerce platforms in Indonesia, has become a popular destination for online transactions[1]. Not only the younger generation but also elderly users have begun actively utilizing this service. This digital demographic shift requires a deeper understanding of the experiences and satisfaction of elderly users when interacting with e-commerce applications[2].

Although elderly users are now more digitally active, many face challenges in using e-commerce applications due to cognitive limitations, vision impairments, and low technological literacy[3], [4], [5]. These challenges can lead to suboptimal user experiences and may even cause them to stop using the service. Negative experiences are often reflected in the reviews they leave on e-commerce platforms. However, these reviews have rarely been systematically analyzed to understand the deeper sentiments and perceptions of elderly users[2], [5].

Previous research on e-commerce sentiment analysis generally focuses on working-age or general users, with little attention given to elderly users. In fact, the linguistic characteristics and emotional expressions of elderly users differ from those of other generations, requiring a tailored analytical approach. Therefore, it is important to conduct a specific sentiment analysis of elderly users' reviews to identify the challenges and satisfaction they experience when using e-commerce platforms like Tokopedia[6], [7], [8].

This study aims to identify and classify the sentiments of elderly users towards the Tokopedia application based on their reviews. By applying machine learning algorithms, this research seeks to determine the distribution of negative, positive, and neutral sentiments, as well as evaluate the classification model's performance using metrics such as precision, recall, and f1-score[9], [10]. The results are expected to provide insights for developers to improve features that are more inclusive and senior-friendly.

The findings of this study will benefit technology developers, e-commerce stakeholders, and academics in human-computer interaction (HCI), particularly in designing interfaces and user experiences that are more accessible to elderly users[11], [12]. Furthermore, this study opens opportunities for further research in developing sentiment classification models that better adapt to the linguistic characteristics of elderly users and encourages innovation in inclusive digital services for all age groups[13], [14].

2. METHODS

This study employs a quantitative approach with a sentiment analysis method based on machine learning[15]. The data is sourced from reviews by elderly users of the Tokopedia application, collected in CSV forma[1], [2], [3]t. The analysis begins with data preprocessing, which includes text cleaning (converting to lowercase, removing numbers, symbols, and unnecessary whitespace), eliminating Indonesian stopwords, and performing stemming using the Sastrawi library.

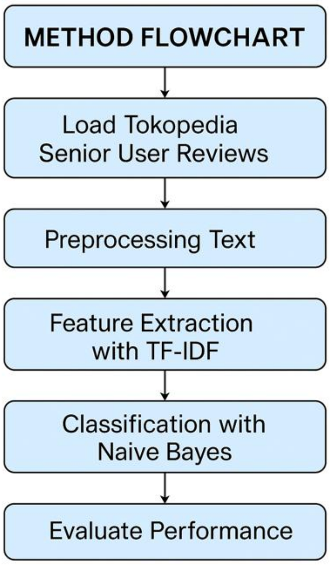

Figure 1. Research Flow Diagram

2.1. Dataset Collection

The initial step in this study is the collection of review data from elderly users of the Tokopedia application, serving as the primary foundation for sentiment analysis and understanding the user experience of the elderly in the e-commerce context[6], [11]. The data is gathered in Comma-

Separated Values (CSV) format, containing two main elements: the review text and the rating scores provided by the users[16].

CSV format is chosen due to its ease of digital processing and compatibility with various data analysis and machine learning tools. The collected data is raw, meaning it has not been preprocessed or cleaned, and still contains language variations, sentence structures, and potential typographical errors that reflect the natural communication style of elderly users[17].

Focusing on the elderly as respondents is a deliberate choice aimed at exploring their experiences, challenges, and satisfaction when using Tokopedia, one of the largest e-commerce platforms in Indonesia. This approach is important because the elderly have distinct characteristics and needs compared to other age groups, such as cognitive and visual limitations, and varying levels of technological literacy[18].

By collecting review data from this demographic, the research can provide a more accurate and relevant depiction of how elderly individuals interact with digital technology[19]. It also serves as a strong basis for developing policies and application features that are more age-friendly and inclusive. Overall, the data collection process not only supplies raw material for sentiment analysis but also offers opportunities to understand the dynamics of e-commerce use among the elderly—ultimately helping to improve service quality and user experience in the future[20].

2.2. Data Preprocessing

The data preprocessing stage is a crucial step in this study, aiming to clean and prepare elderly user reviews for optimal use by machine learning models[21]. The process begins with converting all text to lowercase to maintain data consistency and prevent the model from treating the same word differently due to capitalization. Next, irrelevant numbers, punctuation, and symbols are removed, as these elements may disrupt text analysis and contribute little to sentiment understanding[22]. The following step is whitespace cleaning—removing double spaces or unnecessary characters to make the text more organized and easier to process[4], [17], [23].

Subsequently, stopwords removal is performed—eliminating common Indonesian words such as "dan" (and), "yang" (which), and "untuk" (for), which carry minimal sentiment value, allowing the model to focus on more meaningful words. Finally, stemming is applied using the Sastrawi library, reducing words to their root forms (e.g., "membeli" becomes "beli"), so word variations with the same[24] meaning are recognized as a single entity by the model. These preprocessing steps are vital to improve data quality, reduce noise, and ensure that the machine learning model learns from clean and representative data, thereby improving the accuracy and reliability of sentiment classification results[25].

2.3. Sentiment Labeling

The sentiment labeling stage is a key process that categorizes the sentiments in elderly user reviews of Tokopedia into three main classes based on the rating scores:

- Positive: Reviews with rating scores greater than 3, indicating a satisfying or pleasant user experience.
- Neutral: Reviews with a rating score of exactly 3, reflecting a neutral stance or experience—neither positive nor negative.
- Negative: Reviews with rating scores below 3, signaling dissatisfaction or poor user experiences.

These labels serve as target values in supervised learning, where the machine learning model is trained to recognize patterns in review text associated with each sentiment category[7], [15], [16], [26]. Clearly defined labels are essential for effective model training and accurate automatic classification of new reviews. The quality and accuracy of these labels significantly influence model performance and, in turn, the usefulness of the sentiment analysis in providing insights for developing more inclusive and elderly-friendly applications[10], [13], [23], [27], [28].

2.4. Feature Extraction (TF-IDF Vectorization)

Once preprocessing is complete, the next step is to convert review text into a numerical representation that machine learning models can interpret. This process, known as feature extraction, uses the TF-IDF (Term Frequency–Inverse Document Frequency) method[5], [29]. TF-IDF measures the importance of a word in a document relative to a corpus. The TF-IDF value for a word t in document d is calculated as follows:

$$TF - IDF\ (t, d) = TF(t, d)\ \times\ log\left(\frac{N}{DF(t)}\right) \tag{1}$$

TF-IDF assigns higher weights to words that frequently appear in a specific document but are rare across the corpus—highlighting words that are more informative and document-specific. Conversely, very common words receive lower weights as they contribute little to document differentiation.

This feature extraction process also incorporates n-grams with a range of (1,2), meaning the model considers both individual words (unigrams) and consecutive word pairs (bigrams), such as "fast delivery." This helps capture more complex word relationships and contextual meanings that might be lost when analyzing words in isolation.

Using TF-IDF with n-grams results in a richer and more informative text representation, which is crucial for enhancing sentiment classification performance in distinguishing between positive, neutral, and negative sentiments in elderly users' Tokopedia reviews.

2.5. Train-Test Split

The dataset is then split into two parts: 75% for training and 25% for testing, using the train_test_split function from scikit-learn. This division allows the model to learn from the training data and evaluate its performance on unseen test data[30].

2.6. Classification (Multinominal Naïve Bayes)

For the sentiment classification stage of elderly user reviews on Tokopedia, the model used is Multinomial Naive Bayes (MNB). This algorithm is well-suited for discrete data such as TF-IDF-transformed text, as MNB calculates probabilities based on the multinomial distribution of word features in documents[5].

MNB estimates the probability that a review belongs to a particular sentiment class (positive, neutral, or negative) based on its word features[31]. The basic formula used in Naive Bayes is:

$$P(C_k) \setminus X = \frac{P(C_k)\ \prod_{i=1}^{n} P(x_i \setminus C_k)}{P(x)} \tag{2}$$

Naive Bayes assumes that features (words) are conditionally independent given the class, allowing the total probability to be computed as the product of individual feature probabilities. Though this assumption is rarely fully true, MNB is highly effective and efficient for text classification tasks—especially with large, sparse datasets and TF-IDF features. Its advantages include fast training and prediction and reasonably accurate classification results. Using MNB, the model learns the distinctive word distribution patterns of each sentiment class and classifies new reviews based on those learned patterns[32], [33].

2.7. Evaluastion (Model Evaluation & Visualization)

To evaluate the sentiment classification model for elderly user reviews of Tokopedia, several key performance metrics are used:

- Accuracy: The proportion of correct predictions out of all test data. It measures overall prediction correctness:

$$Accuracy = \frac{True\ Predictions}{Total\ Predictions} \tag{3}$$

- Precision: The proportion of true positive predictions among all positive predictions. It indicates how accurate the model is when predicting a specific class:

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

- Recall: The proportion of actual positives correctly identified by the model. It shows how well the model captures all relevant positive cases:

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

- F1-score: The harmonic mean of precision and recall, providing a balanced measure when both need to be considered:

$$F1 - score = 2 \,.\frac{Precision\ \ Recall.}{Precision+Recall} \tag{6}$$

Vol. 6, No. 3, September 2025 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

- Confusion Matrix: A table showing the number of correct and incorrect predictions for each sentiment class (positive, neutral, negative). It helps identify specific classification errors, such as misclassifying negative reviews as neutral[34].

Tabel 1. Confusion Matrix

	Predicted Positive	Predicted Neutral	Predicted Negatif
Actual Positive	TP	FN	FN
Actual Neutral	FN	TP	FN
Actual Negative	FP	FN	TP

- Classification Report Heatmap: To aid interpretation, metrics such as precision, recall, and F1-score for each class are visualized using a heatmap. This provides an intuitive graphical overview of model performance across sentiment classes, helping identify which classes perform best and which require improvement[35], [36].

3. RESULTS AND DISCUSSION

3.1. Data Collection

In this study, data was collected through google_play_scraper libraries to retrieve reviews from the Google Play Store. By providing an app ID (com.google.tokopedia), the reviews() function will retrieve up to 200 recent reviews from users in the United States and English speaking regions. The Sort.NEWEST parameter ensures that the reviews taken are the most recent, which is relevant to see the user's current experience with the application.

Once the review data is obtained, the next step is to filter only reviews that contain keywords that are considered relevant to the experience of older users. Words like "easy," "difficult," "difficult," and "font" were chosen because they often appear in the context of ease of access, readability, and navigation—important factors in interface design for seniors. This screening aims to ensure that only reviews containing potential inputs related to Human-Computer Interaction (HCI) for elderly users are further analyzed.

Reviews that meet the keyword criteria are entered into DataFrame Pandas, and then saved as a CSV file with tokopedia_elderly_hci_reviews.csv name.

The sentiment distribution data of elderly user reviews on Tokopedia was obtained after a data cleaning process, specifically by removing entries with empty content. Out of the remaining 138 review entries, 80 were categorized as negative sentiment, 48 as positive, and only 10 as neutral. This distribution indicates that the majority of elderly users shared negative experiences while using the Tokopedia platform, which may reflect challenges in aspects such as customer service, user interface, or transaction processes that are perceived as less accommodating for older adults. Nonetheless, there remains a significant portion of positive reviews, suggesting that certain features and aspects of the user experience still manage to meet their expectations. The relatively low number of neutral reviews may imply that elderly users tend to express their opinions more emotionally, or that the sentiment classification model is more sensitive in distinguishing sentiment polarities.

3.2 Word Cloud of Elderly Tokopedia Users' Review

The word cloud visualization in Figure 2 illustrates the frequency of word occurrences in reviews written by elderly users of the Tokopedia application. The size of each word reflects its frequency: the larger the word, the more often it appears in the review dataset.

Figure 2. Word Cloud of Elderly Tokopedia Users' Review

Words such as "mudah" (easy), "ada" (available), and "barang" (item) are the most dominant, indicating that topics related to the ease of shopping and product availability are the primary concerns for elderly users. Additionally, several words reflect negative experiences, such as "susah" (difficult), "komplain" (complaint), "lama" (long/wait), "masalah" (problem), and "tolong" (help), which indicate significant dissatisfaction or issues raised by some users.

This word cloud provides an initial understanding of the general perceptions held by elderly users, while also revealing the main issues they commonly encounter when using the Tokopedia application—particularly in terms of usability, clarity of information, and delivery services.

3.3 Sentiment Classification Report Heatmap

This model was trained using 103 training data and tested using 35 test data with an accuracy of 0.8857142857142857. Figure 3 presents a heatmap displaying evaluation metrics for the sentiment classification model. The metrics include precision, recall, and F1-score for each class (positive and negative), along with the overall average and accuracy scores.

Journal of Computer and Information Systems Ampera
Vol. 6, No. 3, September 2025 e-ISSN: **2775-2496**
https://journal-computing.org/index.php/journal-cisa/index

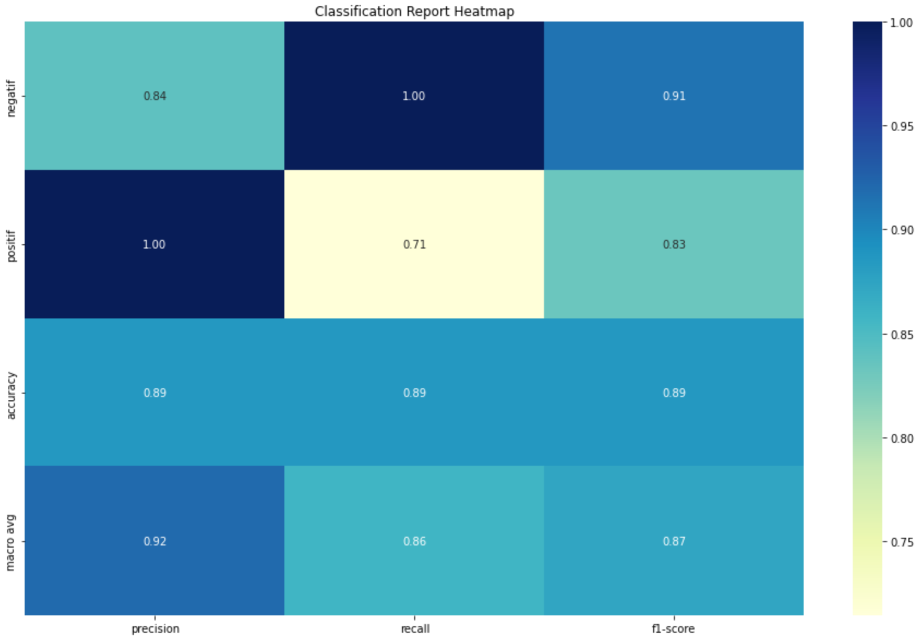

Figure 3. Sentiment Classification Report Heatmap

The model demonstrated excellent performance for the negative class, achieving a recall score of 1.00, meaning all negative reviews were correctly classified. Precision for this class was also high at 0.84, and the F1-score reached 0.91. Conversely, for the positive class, the model achieved perfect precision (1.00), but lower recall (0.71), indicating that some positive reviews were not successfully identified. The overall accuracy of the model was 0.89, suggesting generally strong performance.

This heatmap visualization helps intuitively evaluate model performance and highlights which class requires further improvement, particularly in detecting positive reviews more comprehensively.

3.4 Confusion Matrix

Figure 4 shows the confusion matrix, which provides a breakdown of the model's correct and incorrect classifications. The matrix indicates that:

Vol. 6, No. 3, September 2025 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

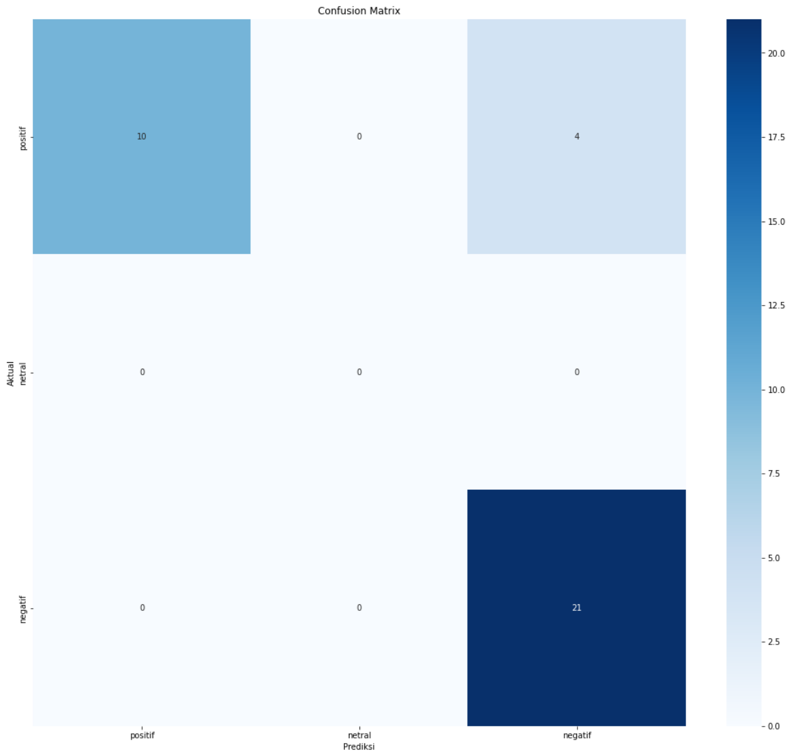

Figure 4. Confusion Matrix

- Out of the total positive reviews, 10 were correctly classified, while 4 were incorrectly classified as negative.
- No neutral reviews were correctly classified, either due to the limited number of data in this class or the model's limitations in identifying neutral sentiment.
- A total of 21 negative reviews were correctly classified, with no misclassifications.

This matrix supports the findings from the heatmap, showing that the model performs strongly in identifying negative reviews but still needs improvement in distinguishing between positive and negative sentiments, as well as recognizing neutral ones.

3.5 Sentimen Distribution of Reviews

Figure 5 presents the distribution of sentiment proportions in a pie chart. Among all the reviews analyzed:

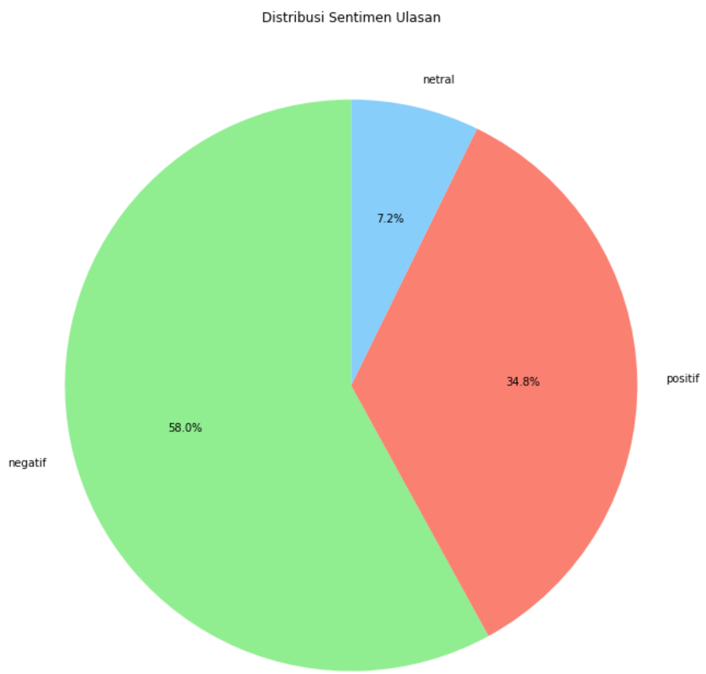

Figure 5. Sentiment Distribution of Reviews

- **Dominance of Negative Sentiment (80/138 ≈ 58%)**
  The majority of reviews from elderly users contain negative sentiment. This suggests that many users experience dissatisfaction with certain aspects of Tokopedia's services—possibly related to the application interface, delivery services, or customer support.

- **Positive Sentiment (48/138 ≈ 35%)**
  Although smaller in number, a significant portion of the reviews reflect positive experiences, indicating that not all interactions resulted in disappointment. This may suggest that some features or services offered by Tokopedia still manage to meet the expectations of elderly users.

Vol. 6, No. 3, September 2025 e-ISSN: 2775-2496

https://journal-computing.org/index.php/journal-cisa/index

- **Neutral Sentiment (10/138 ≈ 7%)**
  The number of neutral reviews is relatively low. This may be due to the tendency of elderly users to express their opinions emotionally—either positively or negatively—or due to the classification model being more inclined to categorize reviews into positive or negative classes, as a result of the ambiguity in neutral expressions.

This distribution indicates that most elderly users had unsatisfactory experiences using Tokopedia. The relatively small proportion of neutral sentiment may also suggest that elderly users tend to express strong opinions (either positive or negative), or that the labeling system was not sufficiently optimized to capture neutral sentiment accurately.

These findings are particularly important in the context of developing a more elderly-friendly application. The predominance of negative sentiment indicates that many aspects of Tokopedia may not yet fully meet the needs and expectations of older users—especially in terms of usability, clarity of information, and customer support services.

4 CONCLUSION

The results of this study show that reviews from elderly users on Tokopedia services are dominated by negative sentiment at 58.0%, followed by positive sentiment at 34.8%, and neutral at only 7.2%. This finding indicates that most elderly users are dissatisfied with their experience in using the Tokopedia application, both in terms of ease of use, speed of service, and other aspects. Although there is still a proportion of positive reviews, the dominance of negative sentiment is a major concern in efforts to improve the user experience, especially for the elderly segment who tend to have digital limitations.

In terms of classification performance, the results of the classification report showed that the model was able to recognize negative and positive reviews quite well, as shown by the f1-score value of 0.91 for negative and 0.83 for positive. However, the recall for the positive class was still relatively low (0.71), indicating that there were still many positive reviews that the model did not successfully recognize. Therefore, in addition to

Vol. 6, No. 3, September 2025 e-ISSN: **2775-2496**

https://journal-computing.org/index.php/journal-cisa/index

improvements in terms of user service, it is also necessary to develop a sentiment analysis model that is more accurate and responsive to the language variations of elderly users.

REFERENCES

[1] A. Nafi', A. Tri, J. Harjanta, B. A. Herlambang, and S. Fahmi, "J-INTECH (Journal of Information and Technology) Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna," 2024.

[2] L. Afuan, M. Khanza, and A. Zahira Hasyati, "ENHANCING SENTIMENT ANALYSIS OF THE 2024 INDONESIAN PRESIDENTIAL INAUGURATION ON X USING SMOTE-OPTIMIZED NAIVE BAYES CLASSIFIER," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 1, pp. 325–333, Feb. 2025, doi: 10.52436/1.jutif.2025.6.1.4290.

[3] A. Dengen and Budiawan, "Integrating Artificial Intelligence in Malaria Mosquitoes: A Bibliometric Analysis for Species Identification and Automated Detection," Apr. 2025. [Online]. Available: http://ijmurhica.ppj.unp.ac.id/index.php/ijmurhica

[4] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," *Natural Language Processing Journal*, vol. 6, p. 100059, 2024, doi: https://doi.org/10.1016/j.nlp.2024.100059.

[5] J. Hemakirthiga K. and Arunadevi, "Comparative Analysis of Classifiers for Chat Classification: A Study on Random Forest, Support Vector Machine, and Multinomial Naive Bayes with Bag of Words and TF-IDF," in *Advances in Artificial Intelligence and Machine Learning in Big Data Processing*, N.-N. and K. S. Geetha R. and Dao, Ed., Cham: Springer Nature Switzerland, 2025, pp. 314–326.

[6] H. Huang, A. Asemi, and M. B. Mustafa, "Sentiment Analysis in E-commerce Platforms: A Review of Current Techniques and Future Directions", doi: 10.1109/ACCESS.2023.0322000.

[7] L. B. -, “Sentiment Analysis in Online Product Reviews: Mining Customer Opinions for Sentiment Classification,” *International Journal For Multidisciplinary Research*, vol. 5, no. 5, Sep. 2023, doi: 10.36948/ijfmr.2023.v05i05.6090.

[8] B. Schehl, J. Leukel, and V. Sugumaran, “Understanding differentiated internet use in older adults: A study of informational, social, and instrumental online activities,” *Comput Human Behav*, vol. 97, pp. 222–230, Aug. 2019, doi: 10.1016/j.chb.2019.03.031.

[9] A. Harits Ramadhani, H. Qahar Djauhari, V. Lius, and A. Nugroho, “Hybrid Naive Bayes TF-IDF Algorithm and Lexicon Approach for Sentiment Analysis of Reviews,” Dec. 2024.

[10] X. Zhang, R. Mao, K. He, and E. Cambria, “Neuro-Symbolic Sentiment Analysis with Dynamic Word Sense Disambiguation,” in *Findings of the Association for Computational Linguistics: EMNLP 2023*, H. Bouamor, J. Pino, and K. Bali, Eds., Singapore: Association for Computational Linguistics, Dec. 2023, pp. 8772–8783. doi: 10.18653/v1/2023.findings-emnlp.587.

[11] M. Ding Yi and Wang, “Optimization Design of Fresh e-Commerce Platform for the Elderly,” in *Human-Computer Interaction*, A. Kurosu Masaaki and Hashizume, Ed., Cham: Springer Nature Switzerland, 2024, pp. 147–162.

[12] Q. Jiang, L. Deng, J. Zhang, and Y. Pengbo, “User-Centered Design Strategies for Age-Friendly Mobile News Apps,” *Sage Open*, vol. 14, no. 4, Oct. 2024, doi: 10.1177/21582440241285393.

[13] T. Hariguna and A. Ruangkanjanases, “Adaptive sentiment analysis using multioutput classification: a performance comparison,” *PeerJ Comput Sci*, vol. 9, 2023, doi: 10.7717/peerj-cs.1378.

[14] I. Zemaitaityte, J. Pivoriene, R. Bardauskiene, and A. Katkoniene, “Older adults’ digital inclusion: New challenges for Lithuanian social policy,” *SHS Web of Conferences*, vol. 184, p. 05004, 2024, doi: 10.1051/shsconf/202418405004.

[15] A. Sharma Rahul Kumar and Dagur, “Sentimental Analysis with Various Classification Techniques—A Comprehensive Survey,” in *Proceedings of Fifth Doctoral Symposium on Computational*

Intelligence, V. and F. G. and H. A. E. Swaroop Abhishek and Kansal, Ed., Singapore: Springer Nature Singapore, 2025, pp. 155–170.

[16] A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *J Big Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00781-w.

[17] M. Osman, R. Cooper, A. A. Sayer, and M. D. Witham, "The use of natural language processing for the identification of ageing syndromes including sarcopenia, frailty and falls in electronic healthcare records: a systematic review," *Age Ageing*, vol. 53, no. 7, Jul. 2024, doi: 10.1093/ageing/afae135.

[18] M. Olumekor, S. N. Polbitsyn, M. S. Khan, H. P. Singh, and I. A. Alhamad, "Ageing and digital shopping: Measurement and validation of an innovative framework," *PLoS One*, vol. 20, no. 3 March, Mar. 2025, doi: 10.1371/journal.pone.0315125.

[19] M. Aitsam, A. Jimenez Rodriguez, and A. Di Nuovo, "Efficient data processing pipeline for event-based vision datasets: techniques and insights," *Engineering Research Express*, vol. 6, no. 4, Dec. 2024, doi: 10.1088/2631-8695/ad9235.

[20] Q. Jiang, L. Deng, J. Zhang, and Y. Pengbo, "User-Centered Design Strategies for Age-Friendly Mobile News Apps," *Sage Open*, vol. 14, no. 4, Oct. 2024, doi: 10.1177/21582440241285393.

[21] Y. Mamani-Coaquira and E. Villanueva, "A Review on Text Sentiment Analysis with Machine Learning and Deep Learning Techniques," 2024, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2024.3513321.

[22] S. Agarwal, S. Godbole, S. Roy, and D. Punjani, "How much noise is too much: A study in automatic text classification," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2007, pp. 3–12. doi: 10.1109/ICDM.2007.21.

[23] L. Xu and W. Wang, "Aspect-based sentiment classification with BERT and AI feedback," *Natural Language Processing Journal*, vol. 10, p. 100136, 2025, doi: https://doi.org/10.1016/j.nlp.2025.100136.

[24] A. W. Pradana and M. Hayaty, "The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-

language Texts," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, pp. 375–380, Oct. 2019, doi: 10.22219/kinetik.v4i4.912.

[25] B. Setiawan, "A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024."

[26] R. Chavan, S. Latthe, M. Dhorepati, A. Suryawanshi, N. Sharma, and A. Salge, "Sentiment analysis using vader & word cloud techniques," *AIP Conf Proc*, vol. 3217, no. 1, p. 020012, Dec. 2024, doi: 10.1063/5.0234543.

[27] J. O. Krugmann and J. Hartmann, "Sentiment Analysis in the Age of Generative AI," *Customer Needs and Solutions*, vol. 11, no. 1, Dec. 2024, doi: 10.1007/s40547-024-00143-4.

[28] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, p. 102048, 2024, doi: https://doi.org/10.1016/j.jksuci.2024.102048.

[29] G. L. and V. C. S. and S. K. and M. S. Saripilli Prashanth and Aruna Kumari, "A Sentiment Analysis of Tweets by Using TF-IDF Vectorizer and Lemmatization with POS Tagging," in *Accelerating Discoveries in Data Science and Artificial Intelligence I*, A. and K. N. and S. B. Lin Frank M. and Patel, Ed., Cham: Springer Nature Switzerland, 2024, pp. 377–386.

[30] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," 2024. [Online]. Available: www.ijacsa.thesai.org

[31] M. Abbas, K. Ali, A. Jamali, K. Ali Memon, and A. Aleem Jamali, "Multinomial Naive Bayes Classification Model for Sentiment Analysis," *IJCSNS International Journal of Computer Science and Network Security*, vol. 19, no. 3, p. 62, 2019, doi: 10.13140/RG.2.2.30021.40169.

[32] U. K. B. Saravanan, M. Vijay, T. Shreedhar, G. Rajasekar, Y. R, and P. Shakthipriya, "Multinomial Naive Bayes Based Machine Learning Analysis of Twitter Sentiment," in *2023 2nd International*

Conference on Edge Computing and Applications (ICECAA), 2023, pp. 429–434. doi: 10.1109/ICECAA58104.2023.10212150.

[33] C. Dewi, R. C. Chen, H. J. Christanto, and F. Cauteruccio, “Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database,” *Vietnam Journal of Computer Science*, vol. 10, no. 4, pp. 485–498, Nov. 2023, doi: 10.1142/S2196888823500100.

[34] B. Ramsay, A. Ralescu, E. Van Der Knaap, and S. Visa, “Confusion Matrix-based Feature Selection. Confusion Matrix-based Feature Selection,” 2011. [Online]. Available: https://www.researchgate.net/publication/220833270

[35] E. Tjoa, H. J. Khok, T. Chouhan, and C. Guan, “Enhancing the confidence of deep learning classifiers via interpretable saliency maps,” *Neurocomputing*, vol. 562, p. 126825, 2023, doi: https://doi.org/10.1016/j.neucom.2023.126825.

[36] A. Chatterjee, M. W. Gerdes, and S. G. Martinez, “Identification of risk factors associated with obesity and overweight—a machine learning overview,” *Sensors (Switzerland)*, vol. 20, no. 9, May 2020, doi: 10.3390/s20092734.