ISSN 2776-8007

Sentiment Analysis of Doctor's Responses to Patient Inquiries in a Medical Chatbot: A Logistic Regression Approach

Mesra betty Yel[1,*], Rodhiyah[2]

[1,2]*Department of Informatics, STIKOM Cipta Karya Informatika, Jakarta, Indonesia*

(Received: January 15, 2024; Revised: April 20, 2024; Accepted: May 5, 2024; Available online: June 3, 2024)

Abstract

This study addresses the challenge of improving doctor-patient communication in medical chatbot systems by integrating sentiment analysis to classify doctor responses as positive or negative. The primary objective was to develop a model that enhances the emotional intelligence and appropriateness of chatbot interactions using Logistic Regression. The model achieved 98.63% accuracy, 99.68% precision, 95.90% recall, and 97.75% F1-score, demonstrating its high reliability in classifying sentiments with minimal misclassifications. While the model performs well, further improvements could focus on reducing false negatives to increase recall. The implications of this research are significant for digital healthcare, as the model enables chatbots to provide more empathetic, context-aware responses, improving patient engagement and overall communication. The novelty of this study lies in applying sentiment analysis within medical chatbot systems, contributing to the growing field of emotional intelligence in digital healthcare. The findings highlight the potential of sentiment analysis to enhance patient interactions, making medical chatbots more effective and human-like. This study provides a solid foundation for further advancements in healthcare chatbots, demonstrating the potential of machine learning to improve the quality of doctor-patient communication in a digital context.

Keywords: Sentiment Analysis, Medical Chatbot, Logistic Regression, Doctor-Patient

1. Introduction

The development of medical chatbots has progressed rapidly as advancements in artificial intelligence and natural language processing have driven novel solutions for patient care [1], [2]. Early iterations of these systems were designed to offer basic automated responses, while recent hybrid models have demonstrated significant benefits, including reducing hospital readmissions for chronic conditions by up to 25% and increasing patient interactions by approximately 30%, which improves care adherence [1]. Extensive reviews have identified multiple benefits, such as enhanced accessibility and personalized care, while also highlighting challenges relating to data privacy and ethical use [3], [4]. The evolution of these conversational agents reflects an ongoing effort to integrate digital health solutions into routine practice, thus enabling improved chronic disease management and timely patient support [1], [2]. Their growing importance underscores a digital transformation in healthcare delivery that continues to redefine patient interaction and service efficiency [3], [4].

Sentiment analysis involves examining users' text inputs to gauge emotional tone, enabling chatbots to respond with enhanced empathy and contextual relevance [5]. By interpreting subtle emotional cues, AI systems can adjust their conversational strategies to better align with users' feelings, ultimately fostering increased trust and responsiveness in interactions [6]. This application is particularly valuable in healthcare settings, where empathetic dialogue can significantly enhance patient support and intervention efficacy [7]. The integration of sentiment analysis into chatbots transforms standard automated responses into dynamic, human-like interactions that are sensitive to users' emotional states. Consequently, this fusion of natural language processing (NLP) techniques and affective computing not only improves the capacity for personalized care but also contributes to a more engaging and supportive user experience, marking a pivotal evolution in the design and implementation of medical chatbots [5], [6]

*Corresponding author: Mesra betty Yel (mesrabettyyel@stikomcki.ac.id)
DOI: https://doi.org/10.47738/ijaim.v4i2.80

Improving medical chatbot response quality remains a significant challenge, particularly concerning the integration of emotional tone and clinical relevance to address patient needs. Recent studies demonstrated that enhancing instruction prompt tuning and retrieval-augmented methods can generate more accurate free-text responses, thus improving the alignment between automated and human-delivered care [8]. Furthermore, evaluations have revealed that the incorporation of emotional elements enhances empathetic responses, thereby increasing patient satisfaction and trust [8]. Although advancements in conversational dynamics and sentiment analysis have fostered emotionally intelligent interactions, the need to fine-tune these systems to consistently match patients' contextual and emotional cues is evident [9]. Reviews of emotionally intelligent chatbots underscore the importance of ensuring that interactions are not only clinically precise but also emotionally supportive, ultimately contributing to improved patient outcomes [10], [11]. Collectively, these insights highlight the necessity for continued research in refining emotional and contextual accuracy in medical chatbot communications.

The variability of sentiment in doctor-patient chatbot interactions poses notable challenges in ensuring consistently empathetic and contextually appropriate responses. Diverse patient expressions can lead to fluctuations in the chatbot's output, potentially undermining trust and the overall quality of care [12]. Robust sentiment analysis models are therefore essential to detect and address these emotional variances in real time, allowing chatbots to adjust their responses based on the inferred emotional state of the patient[13]. This variability necessitates the development of adaptive algorithms capable of accurately interpreting users' sentiment, reducing the risk of miscommunication, and enhancing therapeutic rapport. A systematic incorporation of sentiment analysis into medical chatbots represents a pivotal step towards improving the dynamic interaction between patients and digital health systems, ultimately supporting better patient engagement and clinical outcomes [13].

The primary objective is to develop a Logistic Regression model tailored for performing sentiment analysis on doctors' responses within a medical chatbot context. Logistic Regression has been demonstrated to achieve competitive performance compared to other machine learning algorithms for sentiment classification tasks [14]. Its established effectiveness in handling multifactorial datasets, including sentiment data, makes it a suitable tool for discerning nuanced emotional cues in doctors' responses [15]. Moreover, its simplicity and computational efficiency facilitate scalability and interpretability in real-time chatbot applications, thus enhancing the overall empathy and responsiveness of the system. By leveraging these advantages, the proposed model intends to quantify sentiment in doctor-patient interactions, thereby contributing to more personalized and contextually appropriate medical care.

The objective is to develop and evaluate a Logistic Regression model that classifies doctor's responses in a medical chatbot as positive, negative, or neutral. This approach leverages established techniques in feature extraction and sentiment classification, ensuring that the model comprehensively addresses the variability of sentiment in doctor–patient interactions [16]. By employing standard evaluation metrics such as accuracy, precision, recall, and F1-score, the model's performance will be robustly assessed using confusion matrices and cross-validation methods [17], [18]. This methodological framework not only aims to quantify sentiment polarity but also to enhance the empathy and responsiveness of chatbot communications. Ultimately, this objective supports the broader goal of improving the quality of automated responses in clinical contexts, thereby fostering a more engaging and reliable patient experience.

Enhanced chatbot responses, characterized by empathetic and contextually appropriate tones, promise to significantly improve user experiences in digital healthcare interfaces. By integrating sentiment analysis with natural language processing (NLP), chatbots can adapt their responses based on patient emotions, thereby increasing trust and engagement. For instance, recognizing subtle cues in patient input allows the system to adjust its tone dynamically, reducing miscommunication and fostering a sense of personalized care. Additionally, incorporating insights from customer feedback further refines chatbot interactions, ensuring that responses are not only clinically correct but also supportive and empathetic [19]. This emphasis on emotional intelligence ultimately bridges the gap between automated services and human care, thus enhancing overall patient satisfaction and system effectiveness [19].

Enhanced healthcare chatbots have the potential to transform patient care by providing timely, personalized, and empathetic interactions that streamline clinical workflows. Integrating rule-based machine learning and AI algorithms facilitates accurate triage, mental health support, and chronic disease management, thereby increasing patient engagement and overall care quality [20], [21]. Furthermore, technology-driven outreach methods can effectively close care gaps in preventive care, improving both response rates and health outcomes [22]. Trust is a crucial factor in the adoption of these digital solutions; studies indicate that patients' confidence in chatbot functionalities and data security directly influences their acceptance and sustained use [23]. Additionally, identifying predictors for the efficient utilization of AI chatbots has shown promise in enhancing self-screening and health counseling, particularly during critical public health challenges [24]. Collectively, these practical implications underscore the value of adopting advanced AI-driven solutions to improve individual patient experiences and optimize healthcare delivery at scale.

2. Literature Review

Previous studies on medical chatbot applications have demonstrated the potential of integrating sentiment analysis to enhance the delivery of empathetic and personalized healthcare. For instance, [7] emphasize the ethical dimensions of incorporating sentiment and emotion analysis in depression intervention, highlighting improved patient outcomes when chatbots adapt to user emotions. Similarly, [25] provide an overview of how sentiment analysis integrated with NLP and ML techniques customizes responses, thereby enhancing patient engagement and support. [5] illustrate that empathic chatbots leveraging sentiment detection can tailor interventions according to patients' emotional cues, leading to a more human-like interaction. However, [26] describe a concept for a therapy chatbot but do not provide empirical evidence of its effectiveness, indicating a need for further research in this area.

The literature consistently underscores that sentiment analysis plays a critical role in enhancing communication between patients and digital assistants by enabling these systems to interpret and respond to users' emotional cues. By employing natural language processing (NLP) techniques, sentiment analysis algorithms can detect nuances in textual data, categorizing language into positive, neutral, or negative sentiments. [25] illustrate that this categorization enhances digital assistants' ability to tailor their responses, thereby fostering more personalized interactions. In healthcare, where emotional context is paramount, such capabilities can bridge the communication gap between patients and automated systems, leading to improved patient satisfaction and adherence to medical advice.

A further dimension of this role is revealed in the challenges of multilingual and culturally diverse settings. Biswas [27] reviews the complexities of multilingual sentiment analysis, noting that advanced AI models are essential for addressing subtleties in linguistic expression and cultural context. These enhanced models not only refine the accuracy of sentiment detection but also improve the voice of digital assistants across varying patient demographics, thereby promoting equitable care. Furthermore, digitally mediated communication benefits from sentiment analysis by rendering digital assistants capable of conveying empathy—a quality that has been empirically associated with higher levels of trust and a more positive patient experience [28],[29].

Additionally, the integration of sentiment analysis into the operational framework of digital assistants enables the monitoring and adaptation of dialog strategies in real time. For instance, [30] demonstrate that incorporating emotional tone adjustments in chatbot responses can significantly reshape service delivery, ensuring that the digital assistant remains responsive to individual patient needs. This dynamic adjustment is crucial for minimizing misunderstandings and reinforcing the therapeutic rapport, which is vital in contexts ranging from chronic disease management to mental health support. Concurrently, [31] emphasize that leveraging state-of-the-art NLP techniques not only streamlines the interpretative process but also enriches the overall quality of patient communication, ensuring that digital assistants supplement and bolster the delivery of healthcare services.

By synthesizing these insights, it becomes evident that sentiment analysis is not merely a technical adjunct but a central component in crafting more empathetic, context-aware digital assistants. These advancements collectively contribute to a paradigm shift in healthcare communication—transforming digital assistants from purely informational tools into interactive, emotionally intelligent partners in patient care.

Common methods for sentiment classification typically involve both traditional machine learning techniques and advanced feature extraction approaches. Logistic Regression, a widely used method, has shown strong performance in text classification tasks due to its computational efficiency and accuracy [32], [14]. Comparative studies indicate that while Support Vector Machines and Naive Bayes classifiers are also effective, Logistic Regression often demonstrates competitive performance in terms of precision and recall when applied to sentiment datasets [32], [33]. Additionally, hybrid models that incorporate techniques such as TF-IDF and word embeddings (e.g., Word2Vec) enhance

 ISSN 2776-8007

classification accuracy by providing richer feature representations [34]. Recent work has confirmed that Logistic Regression yields reliable results, especially in contexts where interpretability and rapid classification are essential, such as in digital assistants and online customer feedback systems [17].

3. Methodology

Figure 1 presented outlining the methodology used for the sentiment analysis process. It provides an overview of the key steps involved, from data collection to model evaluation. This visual representation helps in understanding the systematic approach taken to process and classify sentiments.

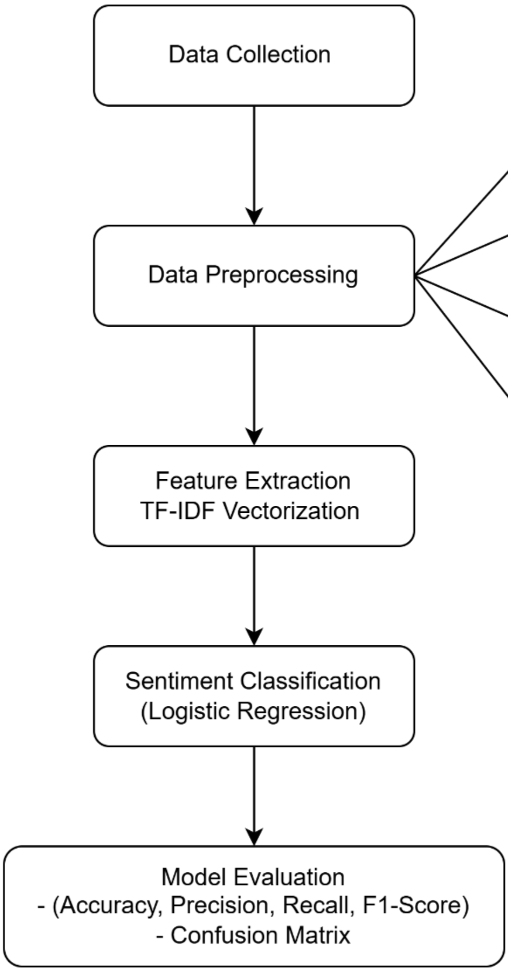

Figure 1. Research Methodology

3.1.Data Collection

The dataset consists of interactions between patients and doctors within a medical chatbot system. It includes three main columns: Description, Patient, and Doctor. The Description column captures the patient's query or concern, typically outlining symptoms, medical inquiries, or requests for advice. This column reflects the nature of the patient's question, which can range from simple health-related issues to more complex medical conditions. The Patient column contains the full text of the patient's message, providing additional context and details such as age, gender, or specific health information. It is an unstructured message from the patient to the doctor, which can include informal language and descriptions of symptoms or experiences. The Doctor column holds the doctor's response, offering medical explanations, advice, or guidance related to the patient's query. The doctor's answer is typically structured and aims to address the patient's concern in a clear and informative way.

Together, these three columns form a comprehensive dataset of patient queries and doctor responses. The Description and Patient columns work together to establish the context of the inquiry, with the Doctor column providing the

 ISSN 2776-8007

necessary information to address the patient's question. This structure allows for detailed analysis, such as sentiment classification, where the doctor's responses can be categorized into positive, negative, or neutral sentiments. Additionally, the Description column can be useful for understanding the specific medical concerns being raised, while the Patient column provides important context that might influence how the doctor responds.

3.2. Data Preprocessing

Data preprocessing in sentiment analysis is critical for improving classification accuracy and reducing computational complexity. The process begins with text cleaning, which encompasses case folding (converting text to lowercase), punctuation removal, and elimination of numbers to standardize the input data [35]. Tokenization then segments text into individual words or tokens, facilitating further analysis [36]. Following this, stopword removal eliminates commonly used words that do not contribute significantly to the sentiment, thus reducing noise in the dataset [36]. Stemming further refines the text by reducing words to their root forms, ensuring that semantically similar tokens are treated uniformly[35][37]. These steps, implemented efficiently using libraries such as NLTK, decrease training time and improve the performance of sentiment classifiers by ensuring that only meaningful features are retained in the dataset[37].

3.3. Sentiment Classification

The sentiment classification approach involves converting preprocessed text into numerical features using vectorization techniques such as TF-IDF, which capture significant lexical information. The extracted feature vectors are then input into a multinomial Logistic Regression model that computes the probability of each response belonging to one of three sentiment categories (positive, negative, or neutral) [18]. The logistic regression model uses the following equation to calculate the probability of a class for a given input x:

$$P(y = 1|x) = \frac{1}{1+e^{-(\omega^{\tau}x+b)}} \tag{1}$$

Where:

$P(y = 1|x$ is the probability of the positive class (sentiment).

ω is the vector of model weights.

x is the vector of features from the input data.

b is the bias term.

e is Euler's number (base of the natural logarithm).

The output of the logistic regression model is a probability, and based on the decision threshold (usually 0.5), the sentiment is classified into one of the categories: positive, negative, or neutral. Parameters of Logistic Regression:

Weights (w) : These are the learned parameters that correspond to the importance of each feature.

Bias (b) : The offset used in the decision boundary of the model.

Learning Rate : Controls how much the model updates its weights during training. Typically used in optimization techniques such as gradient descent.

Regularization : This parameter helps prevent overfitting by penalizing large coefficients.

L2 regularization (Ridge) : Adds a penalty proportional to the sum of the squared weights.

L1 regularization (Lasso) : Adds a penalty proportional to the absolute sum of the weights.

Max Iterations : Defines the maximum number of iterations the optimization algorithm should run.

The model leverages weight coefficients to discern subtle differences in emotional tone, thereby ensuring that even neutral sentiments are effectively identified. To optimize its performance, cross-validation and hyperparameter tuning are incorporated, while performance is assessed using standard metrics like precision, recall, F1-score, and accuracy

ISSN 2776-8007

[38]. This methodology, by combining lexicon-based features (e.g., VADER) with machine learning techniques, has been proven to improve the robustness and generalizability of sentiment classification in diverse text datasets [18].

3.4. Model Evaluation

Model evaluation involves assessing the performance of the Logistic Regression classifier using standard metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix to quantify its effectiveness in sentiment classification. These metrics provide insights into both overall classifier performance and class-specific prediction quality. For instance, [17] compares various classifiers including Logistic Regression and provides performance metrics that help in understanding the strengths and weaknesses of different models. Similarly, [15] discusses how detailed metric analysis, including confusion matrices, aids in recognizing misclassification trends, which is vital for iterative model enhancements. Together, these metrics facilitate a robust evaluation framework that drives refinements in model training and feature selection, ensuring enhanced accuracy and reliability in classifying sentiments as positive, negative, or neutral.

4. Results and Discussion

4.1. Model Evaluation Result

The evaluation results for the sentiment analysis model are summarized in Table 1. The model achieves high performance across all key metrics, including accuracy, precision, recall, and F1-score.

Table 1: Model Evaluation Results

Metric	Value
Accuracy	98.63%
Precision	99.68%
Recall	95.90%
F1-Score	97.75%

The accuracy of 98.63% (0.9863) indicates that the model correctly classified the sentiment of 98.63% of the responses. This high accuracy demonstrates the model's overall reliability and effectiveness in distinguishing between positive and negative responses. Given this high accuracy, it suggests that the model performs well across the board with few errors in classification.

The precision of 99.68% (0.9968) is particularly impressive, highlighting that when the model predicts a positive sentiment, it is almost always correct. Precision is an important metric for evaluating how many of the positive predictions made by the model are accurate, and with this high precision, the model minimizes the likelihood of false positives, ensuring that negative responses are not mistakenly classified as positive.

The recall value of 95.90% (0.9589) shows that the model correctly identified 95.90% of the actual positive responses. Recall measures how effectively the model detects positive sentiments, and while this value is slightly lower than precision, it still indicates strong performance. The slight difference between precision and recall suggests that the model occasionally misses some positive responses, but it captures the majority of them, reflecting good overall performance in identifying positive sentiment.

The F1-score of 97.75% (0.9775) indicates a strong balance between precision and recall. The F1-Score is especially useful when both false positives and false negatives need to be minimized. This balanced score shows that the model is not biased toward one type of misclassification and is effective in managing both precision and recall. A high F1-Score confirms that the model performs well in both identifying positive sentiment correctly and reducing errors in sentiment classification.

The model demonstrates exceptional performance in classifying sentiments, achieving high accuracy, precision, recall, and F1-score. These results highlight the effectiveness of the sentiment analysis model, making it a reliable tool for applications, especially in medical chatbots, where accurate classification of doctor-patient communication is crucial for enhancing user experience and patient engagement.

In Figure 2, a bar chart is presented, showing the distribution of sentiment responses within the dataset. The chart illustrates the number of responses classified into Negative and Positive sentiment categories. As depicted in the figure, the Negative sentiment responses significantly outnumber the Positive sentiment responses, with the negative responses reaching over 175,000 while positive responses are around 75,000. This indicates an imbalance in the dataset, with a higher frequency of negative sentiments compared to positive ones.

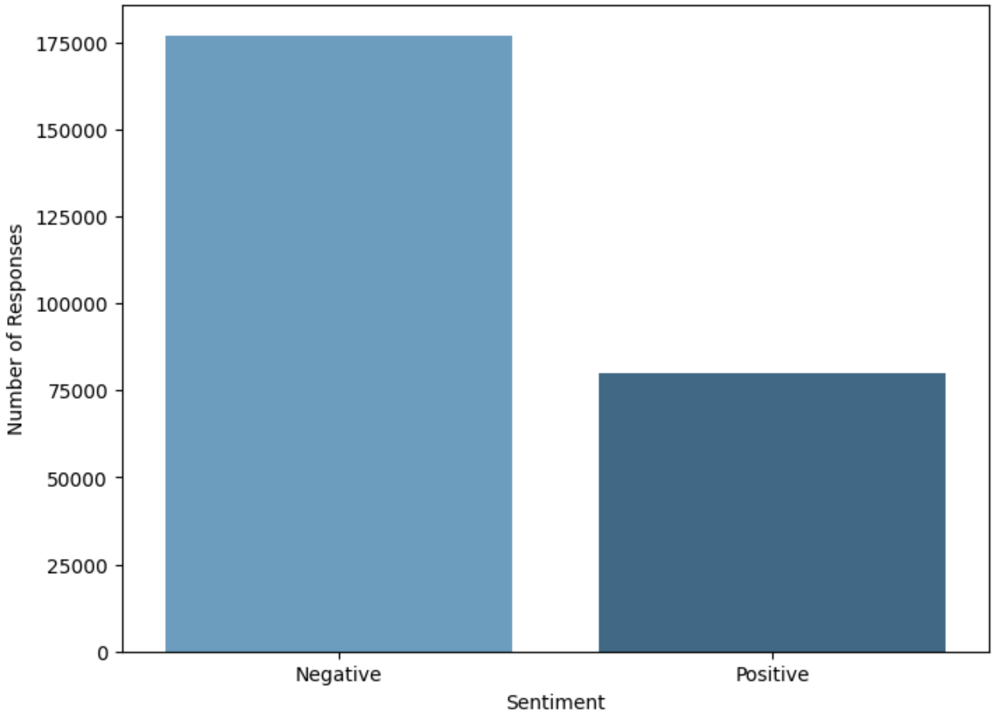

Figure 2. Sentiment Distribution (Positive vs Negative)

Figure 3 displays the 20 most common words in doctor's responses within a medical chatbot system. The chart provides insights into the most frequently used terms, highlighting the prominence of words related to medical advice, treatment, and patient engagement.

The word "take" ranks highest in the frequency list, likely due to its usage in advice related to medications or actions a patient should take. Other common terms like "help," "may," "get," and "like" suggest a conversational and supportive tone often employed by doctors in their responses. Words such as "pain," "consult," and "treatment" point to the medical nature of the responses, emphasizing the focus on addressing health-related concerns.

Terms like "answer", "would", and "hope" show the interactive nature of doctor-patient communication, where the doctor may offer potential solutions, suggest next steps, or express empathy. Additionally, more casual words like "hi", "also", and "hello" indicate the presence of conversational elements in the chatbot interactions.

ISSN 2776-8007

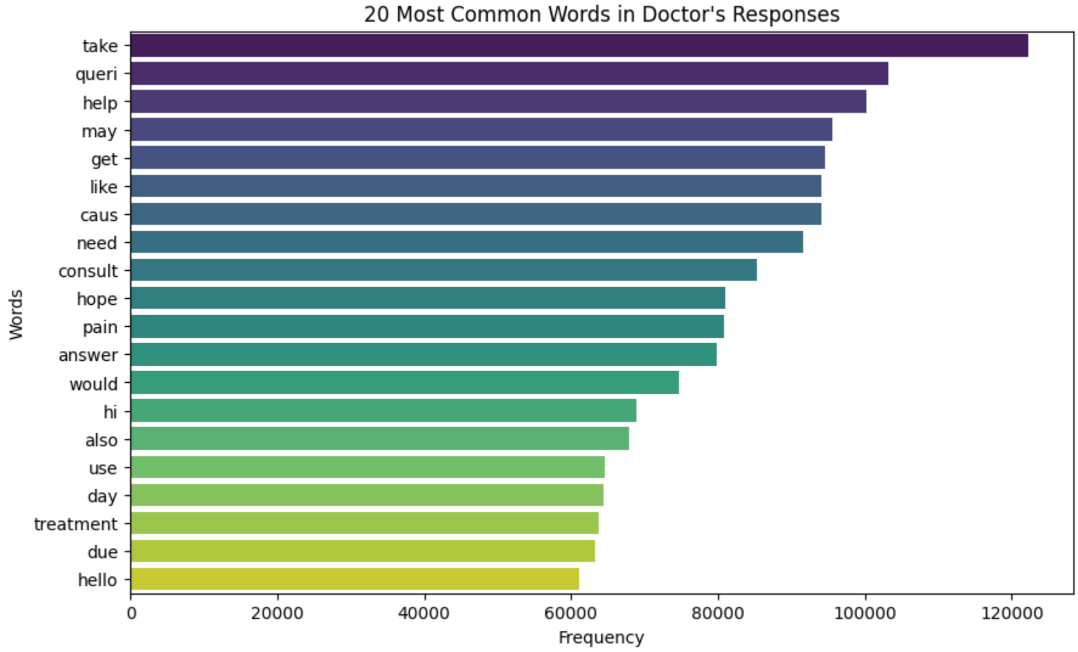

Figure 3. Common Words

This analysis can provide valuable insights into the communication style of doctors and how chatbot responses are structured to provide both medical and emotional support to patients.

Figure 4, a word cloud is presented, showing the most frequently occurring words in doctor's responses within the medical chatbot system. The size of each word in the cloud represents its frequency, with larger words being used more often.

Key terms such as "good," "health," and "care" dominate the cloud, reflecting the positive and supportive tone that doctors often convey in their responses. Words like "wish," "thank," and "help" suggest the empathetic and caring nature of the responses, emphasizing the emotional support provided alongside medical advice.

Other notable words include "answer," "understand," and "concern," indicating that the chatbot responses focus not only on providing medical information but also on ensuring that the patient's concerns and questions are addressed. The presence of words like "query," "need," and "worry" further highlights the interaction between the patient and doctor, where the doctor aims to alleviate any concerns or confusion.

This word cloud offers a visual representation of the language used by doctors, revealing the balance between offering medical advice and emotional reassurance in the interactions.

Figure 4. Word Cloud Positive Sentiment

Figure 5, a word cloud is displayed, showing the most frequently used words in the doctor's responses within the medical chatbot system. Larger words, such as "hope," "answer," "assist," and "know", indicate a higher frequency of occurrence, demonstrating that these words are central to the responses provided by the doctor.

The presence of words like "query," "regard," and "let" suggests that the doctor's responses aim to address the patient's questions or concerns in a clear and empathetic manner. Terms like "take care", "help," and "inform" show the supportive tone of the interaction, emphasizing the doctor's role in guiding the patient and providing reassurance.

Additionally, words such as "doctor," "consult," and "physician" highlight the medical nature of the responses, underscoring the professional and consultative aspect of the chatbot. The variety of words in this word cloud reflects both the medical advice and the empathetic, patient-centered communication style used by doctors.

Figure 5. Word Cloud Negative Sentiment

Figure 6 shows the confusion matrix for the sentiment analysis model, which provides a clear visualization of how well the model performs in classifying positive and negative responses. The confusion matrix is an essential tool for understanding the model's classification behavior, as it illustrates the actual and predicted values of each sentiment class (positive and negative).

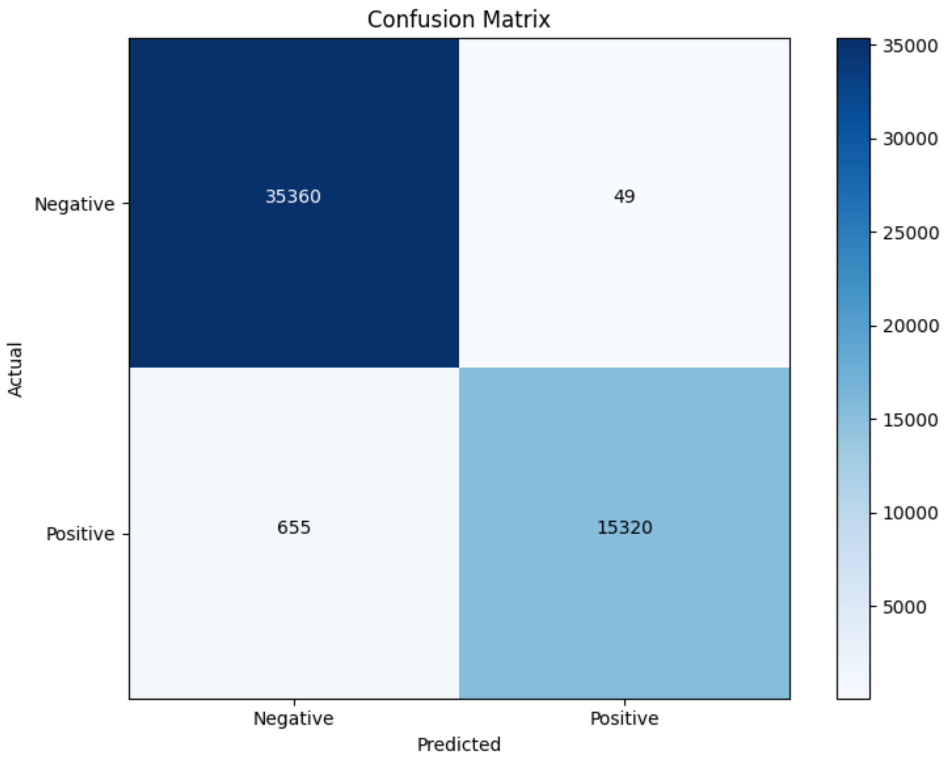

Figure 6. Confusion Matrix

The confusion matrix provides a comprehensive visual representation of the model’s performance in classifying positive and negative sentiments. It highlights the number of correct and incorrect predictions made by the model, serving as an essential tool for evaluating classification performance. The matrix is divided into four quadrants: True Positives, True Negatives, False Positives, and False Negatives.

The top-left cell of the confusion matrix represents True Negatives (TN), where the model correctly identified negative responses as negative. This demonstrates that the model is effective in accurately distinguishing negative sentiment, which is crucial to ensure that negative responses are not mistakenly classified as positive. The bottom-right cell represents True Positives (TP), where the model correctly identified positive responses as positive. This highlights the model’s strong ability to classify positive sentiment accurately, which is especially important in healthcare applications where empathy and positive responses are key.

The top-right cell represents False Positives (FP), where the model incorrectly classified negative responses as positive. Although the number of false positives is relatively low, minimizing these misclassifications is important, particularly in healthcare contexts, as misclassifying negative responses as positive could lead to misunderstandings or incorrect conclusions about a patient’s condition. Similarly, the bottom-left cell represents False Negatives (FN), where the model failed to identify positive responses, classifying them as negative. While the number of false negatives is also relatively small, it suggests that the model missed some positive sentiments, which could be improved by refining the model’s sensitivity to positive responses. Addressing these false negatives would further enhance the model’s accuracy in detecting positive sentiment.

The confusion matrix shows that the model is highly effective in classifying both positive and negative sentiments, with a low number of misclassifications. The relatively small numbers in the false positive and false negative cells indicate strong performance, but there is still room for improvement, particularly in reducing false negatives. These insights provide guidance for future improvements to the model, ensuring even greater accuracy in classifying sentiments in doctor-patient interactions.

4.2. Discussion

The sentiment analysis model's performance, as summarized in Table 1, reflects impressive results across all evaluation metrics, including accuracy, precision, recall, and F1-ccore. These metrics provide a comprehensive understanding of the model's effectiveness in classifying positive and negative sentiments in doctor responses to patient queries. The accuracy of 98.63% indicates that the model correctly predicted the sentiment of 98.63% of the responses, which shows its overall reliability. Given the high accuracy, it suggests that the model performs well in distinguishing between positive and negative responses in the medical chatbot context.

The precision value of 99.68% is particularly noteworthy, as it signifies that when the model predicted a positive sentiment, it was almost always correct. This is critical in applications where false positives incorrectly classifying a negative sentiment as positive must be minimized, such as in patient care where miscommunication could lead to misunderstandings. Similarly, the recall value of 95.90% shows that the model correctly identified 95.90% of the actual positive responses. While slightly lower than precision, this value still reflects strong performance in capturing most of the positive sentiments. The slightly lower recall suggests that the model occasionally misses some positive responses but, overall, performs well in identifying them.

The F1-score of 97.75% indicates a strong balance between precision and recall. The F1-score is particularly important when both false positives and false negatives are significant concerns, as it combines the performance of both precision and recall into a single measure. A high F1-Score confirms that the model strikes a good balance between minimizing both types of misclassifications and is not biased toward one over the other. In the context of medical chatbot applications, this balanced performance is essential for ensuring the system remains both accurate and reliable when classifying sentiment in doctor-patient conversations.

The confusion matrix further complements these results by visually representing the model's classification behavior. It shows that the model performs well in distinguishing between positive and negative sentiments, with a high number of true positives (15,320) and true negatives (35,360). The relatively low numbers in the false positive (49) and false negative (655) cells indicate that the model makes very few misclassifications. This suggests that the model has been effectively trained to identify sentiment accurately, with only a small percentage of responses being incorrectly classified.

However, the presence of false positives and false negatives does highlight areas for potential improvement. For instance, although the number of false positives is minimal (49), reducing such misclassifications further would ensure even more accurate predictions. Similarly, the false negatives (655) indicate that a small portion of actual positive sentiments were missed, which could be addressed by refining the model's sensitivity to positive sentiments. Nevertheless, the overall performance, as shown in both the metrics and the confusion matrix, demonstrates that the model is highly effective in classifying sentiment in doctor responses.

The model's high precision, recall, and F1-Score confirm its suitability for use in real-world medical chatbot applications, where understanding the sentiment of doctor-patient interactions is critical. By correctly classifying positive and negative sentiments, the model can help enhance the patient experience by ensuring that responses from the chatbot are appropriately framed and empathetic, improving communication and patient satisfaction.

Previous studies have shown that integrating sentiment analysis in medical chatbots can significantly enhance the quality of patient interactions. For example, Denecke & Gabarrón [7] highlighted the importance of sentiment and emotion analysis in depression interventions, where personalized, empathetic responses from chatbots led to improved patient outcomes. Similarly, S. S et al. S. S et al., (2024) discussed how sentiment analysis, combined with NLP and machine learning techniques, could tailor responses and improve patient engagement. The findings from Lahoz-Beltrá & López Lahoz-Beltrá & López, (2021) also support this by demonstrating that empathic chatbots can adjust their responses based on emotional cues, creating more human-like interactions with patients.

In the context of multilingual and culturally diverse settings, Biswas [27] emphasized the challenges of sentiment analysis across different languages and cultural contexts. Advanced AI models are necessary to detect linguistic subtleties, ensuring that digital assistants provide more equitable care. This is particularly important in healthcare, where accurate sentiment detection can bridge the communication gap between patients and automated systems, enhancing trust and patient satisfaction. The integration of sentiment analysis, as shown in this study, further supports the notion that sentiment analysis not only enriches the chatbot's ability to engage users emotionally but also ensures that responses are culturally sensitive and contextually appropriate.

The integration of sentiment analysis into medical chatbots aligns with findings from studies such as those by Chakravarthy & S [31], who explored how advanced NLP techniques improve the quality of patient communication in healthcare settings. Moreover, studies by Sestino et al. [30] demonstrated that incorporating emotional tone adjustments can dynamically reshape chatbot service delivery, ensuring that responses are responsive to individual patient needs, especially in sensitive areas such as mental health support and chronic disease management.

These insights collectively confirm that sentiment analysis is a critical component of improving digital assistants' empathetic capabilities, making them not just informational tools but also emotionally intelligent partners in patient care.

5. Conclusion

This study developed and evaluated a sentiment analysis model for classifying doctor responses in a medical chatbot system using Logistic Regression. The model demonstrated excellent performance with an accuracy of 98.63%, precision of 99.68%, recall of 95.90%, and an F1-score of 97.75%. These results show that the model effectively distinguishes between positive and negative sentiments, making it highly reliable for real-world applications.

While the model performs well with minimal misclassifications, further improvements could focus on reducing false negatives to enhance recall. The findings align with existing research, which highlights the benefits of integrating sentiment analysis in medical chatbots to improve patient engagement, trust, and overall healthcare outcomes.

The sentiment analysis model provides a solid foundation for developing more empathetic and contextually aware chatbots. With further refinement, such models have the potential to enhance digital healthcare solutions, improving communication and patient care in the future.

6. Declarations

6.1. Author Contributions

Conceptualization: M.B.Y. and R.; Methodology: R.; Software: M.B.Y.; Validation: M.B.Y. and R.; Formal Analysis: M.B.Y. and R.; Investigation: M.B.Y.; Resources: R.; Data Curation: R.; Writing—Original Draft Preparation: M.B.Y. and R.; Writing—Review and Editing: R. and M.B.Y.; Visualization: M.B.Y. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Kulkarni and D. R. C. Jaiswal, "A Survey on AI Chatbots," *Int. J. for Research in Applied Science and Engineering Technology*, vol. 11, no. 3, 2023. doi: 10.22214/ijraset.2023.55911.

[2] G. Jelić and D. M. Tartalja, "How Can AI-Powered Solutions Improve Communication in Healthcare?," *in Proc. Int. Scientific Conf. ITEMA*, vol. 7, 2023, pp. 81–88. doi: 10.31410/itema.s.p.2023.81.

[3] G. Sun and Y. Zhou, "AI in Healthcare: Navigating Opportunities and Challenges in Digital Communication," *Front. Digit. Heal.*, vol. 5, no. 1291132, pp. 01-05, 2023, doi: 10.3389/fdgth.2023.1291132.

[4] L. Balcombe, "AI Chatbots in Digital Mental Health," *Informatics*, vol. 10, no. 4, 2023. doi: 10.3390/informatics10040082.

[5] R. Lahoz-Beltrá and C. C. López, "LENNA (Learning Emotions Neural Network Assisted): An Empathic Chatbot Designed to Study the Simulation of Emotions in a Bot and Their Analysis in a Conversation," *Computers*, vol. 10, no. 12, pp. 170, 2021, doi: 10.3390/computers10120170.

[6] W. Trzebiński, T. Claessens, J. Buhmann, A. D. Waele, G. Hendrickx, P. V. Damme, W. Daelemans, and K. Poels, "The Effects of Expressing Empathy/Autonomy Support Using a COVID-19 Vaccination Chatbot: Experimental Study in a Sample of Belgian Adults," *JMIR Form. Res.*, vol. 7, no. e41148, pp. 1-16, 2023, doi: 10.2196/41148.

[7] K. Denecke, A. A. Abd-alrazaq, and M. Househ, "Artificial Intelligence for Chatbots in Mental Health: Opportunities and Challenges," in Multiple Perspectives on Artificial Intelligence in Healthcare, Springer, 2021, pp. 207–222. doi: 10.1007/978-3-030-67303-1_10.

[8] H. Chin et al., "The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study," *J. Med. Internet Res.*, vol. 25, 2023. doi: 10.2196/51712.

[9] S. Narynov, Z. Zhumanov, A. Gumar, M. Khassanova, and B. Omarov, "Chatbots and Conversational Agents in Mental Health: A Literature Review," *in Proc. 2021 Int. Conf. on Control, Automation and Systems (ICCAS)*, pp. 353–358, 2021. doi: 10.23919/ICCAS52745.2021.9649855.

[10] G. Bilquise, S. Ibrahim, and K. Shaalan, "Emotionally Intelligent Chatbots: A Systematic Literature Review," *Hum. Behav. Emerg. Technol.*, vol. 2022, no. 1, pp. 1–23, 2022, doi: 10.1155/2022/9601630.

[11] J. J. Bird and A. Lotfi, "Generative Transformer Chatbots for Mental Health Support: A Study on Depression and Anxiety,"

in Proc. 16th Int. Conf. on Pervasive Technologies Related to Assistive Environments, 2023. doi: 10.1145/3594806.3596520.

[12] A. X. Nguyen, X.-V. Trinh, S. Y. Wang, and A. Y. Wu, "Determination of Patient Sentiment and Emotion in Ophthalmology: Infoveillance Tutorial on Web-Based Health Forum Discussions," *J. Med. Internet Res.*, vol. 23, no. 5, pp. e20803, 2021, doi: 10.2196/20803.

[13] F. J. Ramírez-Tinoco, G. Alor-Hernández, J. L. Sánchez-Cervantes, M. del P. Salas-Zárate, and R. Valencia-García, "Use of Sentiment Analysis Techniques in Healthcare Domain," in Studies in Computational Intelligence, Springer, 2019, pp. 199–225. doi: 10.1007/978-3-030-06149-4_8.

[14] S. Ahmadov and A. Boyacı, "Multilingual Text Mining Based Open Source Emotional Intelligence," *Turkish J. Sci. Technol.*, vol. 17, no. 2, pp. 161–166, 2022, doi: 10.55525/tjst.1113832.

[15] E. a. S. Roja, "Performance of Machine Learning Models in Predicting Sentiments of Post-Covid Patients," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 10, pp. 2324–2329, 2023, doi: 10.17762/ijritcc.v11i10.8953.

[16] N. Khamphakdee and P. Seresangtakul, "An Efficient Deep Learning for Thai Sentiment Analysis," *Data*, vol. 8, no. 5, pp. 90, 2023, doi: 10.3390/data8050090.

[17] H. Jadia, "Comparative Analysis of Sentiment Analysis Techniques: SVM, Logistic Regression, and TF-IDF Feature Extraction," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 5, no. 1, pp. 1182-1187, 2023, doi: 10.56726/irjmets45265.

[18] M. Arief and N. A. Samsudin, "Hybrid Approach With VADER and Multinomial Logistic Regression for Multiclass Sentiment Analysis in Online Customer Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 12, pp. 312-320, 2023, doi: 10.14569/ijacsa.2023.0141232.

[19] A. S. Murwati and L. Aldianto, "Exploring Voice of Customers to Chatbot for Customer Service With Sentiment Analysis," *Asian J. Technol. Manag.*, vol. 15, no. 2, pp. 141–153, 2022, doi: 10.12695/ajtm.2022.15.2.4.

[20] Prof. M. B. Mundhe, "Health Care Chatbot," *Int. J. Sci. Res. Eng. Manag.*, vol. 8, no. 6, pp. 1–5, 2023. doi: 10.55041/ijsrem26399.

[21] J. Chow and L. Xu, "Chatbot for Healthcare Communication using Artificial Intelligence and Machine Learning (Preprint)," *JMIR Preprints,* 2021. doi: 10.2196/preprints.27850.

[22] R. Surve, T. Purohit, R. Joseph, and P. Shaikh, "HealthCare Chatbot Using Machine Learning and NLP," *in Proc. 2023 Int. Conf. on Advances in Science and Technology (ICAST),* pp. 411–416, 2023. doi: 10.1109/ICAST59062.2023.10455027.

[23] G. Jelić and D. M. Tartalja, "How Can AI-Powered Solutions Improve Communication in Healthcare?," *in Proc. Int. Scientific Conf. ITEMA,* vol. 7, pp. 81–88, 2023. doi: 10.31410/itema.s.p.2023.81.

[24] C. Bhola, F. Afzal, and V. Kumar, "Identification of Predictors for Utilization of Artificial Intelligence Powered COVID -19 Chatbot for Self-Screening and Health Counselling," *Indian J. Sci. Technol.*, vol. 16, no. 32, pp. 2540–2547, 2023, doi: 10.17485/ijst/v16i32.1003.

[25] Nivethan and S. Sankar, "Sentiment Analysis and Deep Learning Based Chatbot for User Feedback," *in Proceedings of Intelligent Communication Technologies and Virtual Mobile Networks,* 2019. doi: 10.1007/978-3-030-28364-3_22.

[26] J. Ragaventhiran, S. Manickavasagam, N. Prasath, M. K. B, and M. Islabudeen, "Mobile Phone Recommender Using Multi Criteria Decision Making Algorithm," *Malaysian J. Comput. Sci.*, vol. 2022, n0. Dec, pp. 43–52, 2022, doi: 10.22452/mjcs.sp2022no2.4.

[27] Y. Xu, H. Cao, W. Du, and W. Wang, "A Survey of Cross-Lingual Sentiment Analysis: Methodologies, Models and Evaluations," *Data Science and Engineering,* vol. 7, pp. 279–299, 2022. doi: 10.1007/s41019-022-00187-3.

[28] M. Rostami and S. Navabinejad, "Artificial Empathy: User Experiences With Emotionally Intelligent Chatbots," *Aitechbesosci*, vol. 1, no. 3, pp. 19–27, 2023, doi: 10.61838/kman.aitech.1.3.4.

[29] M. Sakumoto and A. Joshi, "Digital Empathy 2.0: Connecting With Patients Using the Written Word," *Telehealth and Medicine Today,* vol. 8, 2023. doi: 10.30953/thmt.v8.433.

[30] R. Surve, T. Purohit, R. Joseph, and P. Shaikh, "HealthCare Chatbot Using Machine Learning and NLP," *in Proc. 2023 Int. Conf. on Advances in Science and Technology (ICAST),* pp. 411–416, 2023. doi: 10.1109/ICAST59062.2023.10455027.

[31] B. Bhattacharya and V. S. Pissurlenkar, “Assistive Chatbots for Healthcare: A Succinct Review,” *arXiv preprint*, arXiv:2308.04178, 2023. doi: 10.48550/arXiv.2308.04178.

[32] M. S. Hossain and M. F. Rahman, “Customer Sentiment Analysis and Prediction of Insurance Products’ Reviews Using Machine Learning Approaches,” *Fiib Bus. Rev.*, vol. 12, no. 4, pp. 386–402, 2022, doi: 10.1177/23197145221115793.

[33] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, “Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling,” *J. Inf. Syst. Informatics*, vol. 5, no. 3, pp. 915–927, 2023, doi: 10.51519/journalisi.v5i3.539.

[34] E. M. Alshari, A. Azman, S. Doraisamy, N. Mustapha, and M. Alksher, “Senti2vec: An Effective Feature Extraction Technique for Sentiment Analysis Based on Word2vec,” *Malaysian J. Comput. Sci.*, vol. 33, no. 3, pp. 240–251, 2020, doi: 10.22452/mjcs.vol33no3.5.

[35] F. Rustam, I. Ashraf, R. Shafique, A. Mehmood, S. Ullah, and G. S. Choi, “Review Prognosis System to Predict Employees Job Satisfaction Using Deep Neural Network,” *Comput. Intell.*, vol. 37, no. 2, pp. 924–950, 2021, doi: 10.1111/coin.12440.

[36] C. P. Chai, “Comparison of Text Preprocessing Methods,” *Nat. Lang. Eng.*, vol. 29, no. 3, pp. 509–553, 2022, doi: 10.1017/s1351324922000213.

[37] H. A. Rosyid, A. Y. H. Putra, M. Akbar, and F. A. Dwiyanto, “Can Multinomial Logistic Regression Predicts Research Group Using Text Input?,” *Knowl. Eng. Data Sci.*, vol. 5, no. 2, pp. 150, 2022, doi: 10.17977/um018v5i22022p150-159.

[38] N. K., S. S., H. K. Hemanandhini, S. Haritha, and A. S., “Machine Learning Based Twitter Sentiment Analysis on COVID-19,” *in Proc. 2021 5th Int. Conf. on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1745–1749, 2021. doi: 10.1109/ICECA52323.2021.9676113.