



Telegram FAQ Chatbot Design for Budi Mulia Lawang Junior High School

Adecya Jalu Mahadwija¹, Fadhli Almu'iini Ahda²

^{1,2}, Informatics Engineering, Asia Institute of Technology and Business Malang, Rembeksari 1A, Indonesia

Article Information

Received: 21-11-2024

Revised: 28-11-2024

Published: 05-12-2024

Keywords

Information; system; technology; software; chatbot

*Correspondence Email:

jaludwija37@gmail.com

Abstract

This study effectively created and assessed an academic services chatbot that employs Natural Language Processing (NLP) and Artificial Neural Networks (ANN). It demonstrates a thorough methodology that includes data collection, preprocessing, model training, and performance evaluation. The dataset was meticulously derived from frequently asked questions at the Institute of Business and Technology Asia Malang, ensuring its relevance and applicability to the target audience. Thorough preprocessing methods, such as tokenization, stemming, and the removal of stop words, were utilized to elevate the quality of the input data for the ANN model, leading to a notable enhancement in its performance. The training outcomes demonstrated a significant relationship between the number of training epochs and the accuracy of the chatbot's responses, suggesting that more extensive training resulted in improved performance. Additionally, the measurement tool employed for user feedback demonstrated confirmed validity and reliability, evidenced by a Cronbach's Alpha coefficient of 0.965, highlighting the strength of the data gathered. A study conducted with 37 students revealed a significant level of satisfaction regarding the chatbot's performance, which attained an impeccable accuracy score of 100%. The results underscore the promise of chatbots powered by natural language processing to improve academic information services, efficiently responding to student questions and markedly alleviating the burden on academic personnel. This study provides a significant framework for educational institutions looking to adopt AI-driven solutions to enhance their academic support services and enrich the student experience.

1. Introduction

The swift progression of information technology has profoundly influenced multiple sectors, particularly education. The capacity to obtain information efficiently is crucial for improving the educational experience for students, educators, and parents. Junior High Schools (SMP) play a crucial role in developing the younger generation, tasked with delivering timely and pertinent information services to address the evolving demands of the educational community.

Budi Mulia Lawang Junior High School, recognized for its dedication to educational quality, encounters difficulties in efficient information transmission. The school's administration frequently faces a substantial number of requests from students and parents regarding class schedules, school policies, and major events, resulting in communication bottlenecks. This frequently leads to delayed replies that may obstruct the learning process and diminish parental satisfaction.

In the current rapid environment, concerns regarding information retrieval might hinder the educational process, causing confusion among stakeholders. The use of a chatbot could effectively resolve these difficulties. A chatbot is a software application that emulates dialogue with users, utilizing Natural Language Processing (NLP) technology to comprehend and provide responses in natural language. The application of NLP in chatbots facilitates more human-like interactions, permitting prompt replies to common inquiries (Mittal et al., 2021).

Recent research has demonstrated the effectiveness of employing deep learning techniques to develop chatbots specifically designed for handling inquiries related to new student admissions. For instance, Heryati et al. (2023) utilized a deep learning model that achieved high accuracy in understanding and responding to user queries, significantly improving the efficiency of information dissemination. This approach not only streamlined the communication process but also enhanced user satisfaction by providing timely and accurate responses to frequently asked questions (Heryati et al., 2023). Additionally, McGrath et al. (2024) provide a comprehensive review of empirical studies on the use of generative AI chatbots in higher education, highlighting their potential to enhance student engagement and streamline administrative process (McGrath et al., 2024). Similarly, Fütterer et al. (2023) explore global reactions to ChatGPT in educational contexts, emphasizing both the opportunities it presents and the challenges it poses for educators (Fütterer et al., 2023). Khosravi et al. (2023) assess the effectiveness of ChatGPT in answering complex questions related to genetics, demonstrating its potential as a valuable educational tool (Khosravi et al., 2024). Furthermore, Yan (2023) investigates the impact of ChatGPT on second language learners during writing practices, revealing its role in supporting language acquisition (Yan, 2023). These studies collectively underscore the transformative potential of AI chatbots in education, aligning with the objectives of this research to improve communication and information dissemination at Budi Mulia Lawang Junior High School.

The use of an NLP-based chatbot is expected to enhance communication between Budi Mulia Lawang Junior High School's administration and its stakeholders. By automating solutions to frequently asked questions, the chatbot can enhance the quality of information services while alleviating the workload of the teaching staff. Moreover, employing Telegram as the communication medium leverages its extensive usage among students and parents, guaranteeing convenient access to information at any time and from any location. This research seeks to develop and deploy an NLP-driven FAQ chatbot for Budi Mulia Lawang Junior High School over Telegram, thereby optimizing communication processes and improving the overall educational experience.

1.1 Literature Review

1.1.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a branch of artificial intelligence that focuses on the interaction between computers and humans using natural language. In the domain of chatbots, NLP functions as a conceptual framework that allows systems to understand, interpret, and respond to user queries efficiently. Theories in natural language processing, encompassing syntactic and semantic analysis, support the development of effective chatbots. Recent improvements in natural language processing have resulted in the creation of conversational agents that offer mental health care, including chatbots aimed at aiding patients with depression and encouraging good behaviors (Ahda et al., 2024; Aramaki et al., 2022). These applications demonstrate the capacity of NLP to augment user engagement and boost patient outcomes in healthcare environments.

1.1.2 Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is a machine learning framework modeled after the human brain, significantly improving chatbot capabilities. In clinical applications, artificial neural networks (ANNs) are employed to analyze unstructured clinical notes, enhancing chatbots' comprehension and responsiveness to

user queries. Recent studies indicate that deep learning models, such as artificial neural networks, may proficiently interpret clinical narratives to extract pertinent medical information, hence enhancing the accuracy of responses in healthcare chatbots (Sheikhalishahi et al., 2019). This adaptability enables chatbots to learn from interactions, so increasing their responsiveness to customer requests and improving the overall user experience.

1.1.3 Chatbot

Chatbots are automated systems engineered to replicate dialogue with human users, especially online. They employ diverse technologies, such as natural language processing (NLP) and machine learning, to comprehend customer inquiries and deliver pertinent solutions. In education, AI chatbots have become essential tools that improve teaching efficacy and efficiency by serving as virtual assistants, delivering tailored tutoring, and providing prompt homework and study support (Labadze et al., 2023). This integration enhances student engagement and learning results while reducing the administrative workload on educators, enabling them to concentrate more on instructional planning and student interaction. The advancement of technology significantly enhances the potential influence of AI chatbots on education, with the promise of transforming learning experiences and educational methodologies (Labadze et al., 2023).

1.1.4 Tokenization and Stemming

Tokenization and stemming are fundamental techniques in natural language processing (NLP) that enable the study of textual material. Tokenization is dividing text into smaller parts, or tokens, which may consist of words, phrases, or symbols. This stage is essential for comprehending the text's structure and facilitates the manipulation of discrete components. In Indonesian language processing, tokenization can be executed using tools such as the Natural Language Toolkit (NLTK), which facilitates case folding, tokenization, and the elimination of stop words (Rianto et al., 2021). Stemming is the process of reducing words to their root forms, hence standardizing variants of a word to enhance information retrieval accuracy (Ashari et al., 2024; Sulistyono et al., 2023). For example, the terms "berlari" (running) and "menulis" (writing) are derived from the roots "lari" (run) and "tulis" (write), respectively. Although conventional stemming techniques such as Sastrawi are proficient for formal Indonesian, they encounter difficulties with informal language, which frequently incorporates slang and acronyms. Recent innovations, as the Incorbiz stemming approach, seek to tackle these issues by integrating a more extensive vocabulary that encompasses both formal and informal terms, thus improving the precision of text categorization models (Rianto et al., 2021).

1.1.5 Bag of Word (BoW)

The Bag of Words (BoW) model is a fundamental technique in text processing and natural language processing (NLP) that represents text data by disregarding grammar and word order while maintaining word frequency. Widely used in applications like sentiment analysis and document classification, the BoW model converts text into a numerical format, enabling effective processing by machine learning algorithms. It involves preprocessing steps such as tokenization, stop word removal, and the creation of a master dictionary, which simplify documents but may degrade meaning. Variations like n-grams can help preserve contextual relationships, while feature values can be optimized using term count, term frequency (TF), and term frequency-inverse document frequency (TF-IDF). Understanding the strengths and limitations of the BoW model is essential for professionals, including radiologists, to enhance communication with data scientists and improve the implementation of NLP techniques in specific tasks (Juluru et al., 2021).

1.1.6 Telegram

Telegram is a cloud-based messaging program that provides a multifaceted platform for communication, encompassing features like group chats, channels, and bots. The incorporation of chatbots in Telegram enables enterprises and educational organizations to automate customer contacts, disseminate information, and improve user engagement. (Hartati & Manullang, 2024) assert that the deployment of a Telegram chatbot utilizing Natural Language Processing (NLP) in creative entrepreneurship education markedly enhances student engagement and learning results, with 97.5% of participants deeming the chatbot beneficial. The chatbot provides learning materials in several formats and use NLP to enhance human-machine interaction,

promoting vital abilities like creativity and critical thinking in pupils (Hartati & Manullang, 2024). This illustrates how Telegram's API may be utilized to design advanced chatbots that improve user experience and foster the growth of students' creative and imaginative skills.

2. Research Methods

2.1 Data Preparation

This research employs data derived from interviews with the administrative staff concerning common inquiries posed by students or parents at Budi Mulia Lawang Junior High School. The subsequent step entails converting this data into a dataset, which will be entered via the admin interface. This dataset will include essential questions to facilitate the system's recognition of diverse inquiry patterns and their respective target responses. The dataset will be structured into three columns:

- **Tags:** This column will classify the questions into distinct themes or subjects. Tags function as keywords that assist the system in discerning the context of a question, facilitating more precise responses. Examples of tags may encompass categories like "admissions," "curriculum," or "extracurricular activities."
- **Pattern:** The pattern column will contain the precise wording of the questions as typically posed by students or parents. This encompasses variances in phrasing and syntax that illustrate how individuals commonly articulate their questions. Through the analysis of these patterns, the system can enhance its comprehension and interpretation of user inquiries, resulting in improved communication.
- **Responses:** This column will encompass the predetermined replies associated with each question pattern. Responses must be lucid, succinct, and informative, delivering the requisite information sought by users. The objective is to guarantee that the chatbot provides precise and pertinent responses based on the recognized patterns and tags.

This dataset will ultimately be transformed into JSON format and solicited by the chatbot via a predetermined API endpoint. This systematic method not only improves the chatbot's capacity to address customer inquiries but also facilitates the updating and management of the dataset when new questions emerge. This system aims to enhance the communication experience for students and parents at Budi Mulia Lawang Junior High School.

2.2 Model Training

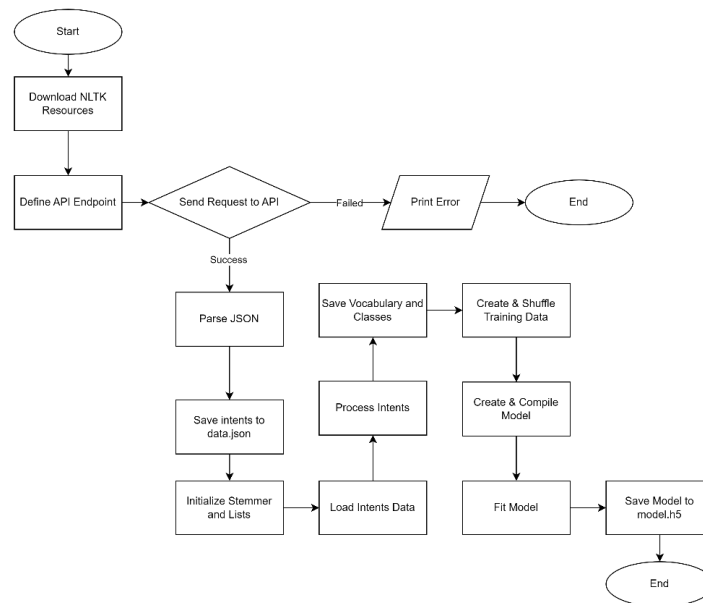


Fig. 1 Flowchart model training

The training of the chatbot model, developed in Python, follows a systematic approach that integrates the Natural Language Toolkit (NLTK) and a Laravel-based API for data acquisition. The process begins with downloading essential NLTK resources and defining the API endpoint, which serves as the source for retrieving relevant data. After sending a request to the API, the data is parsed from JSON format, and the intents are saved to a data.json file. This step is crucial for ensuring that the model can effectively recognize and categorize user inputs. The intents are processed, and a stemmer is initialized to enhance the model's understanding of language variations.

```

{
  "intents": [
    {
      "tag": "admission",
      "patterns": [
        "Bagaimana cara mendaftar di sekolah ini?",
        "Apakah ada pendaftaran siswa baru?",
        "Bagaimana proses pendaftarannya?",
        "Bagaimana cara saya mendaftar di SMPK Budi Mulia?",
        "Apa persyaratan untuk mendaftar di sekolah ini?"
      ],
      "responses": [
        "Untuk mendaftar di SMPK Budi Mulia Lawang, Anda dapat mengunjungi situs resmi kami atau datang langsung ke sekolah untuk informasi lebih lanjut.",
        "Pendaftaran siswa baru dapat dilakukan melalui website sekolah atau dengan datang langsung ke SMPK Budi Mulia di Lawang.",
        "Silakan kunjungi halaman pendaftaran di situs web SMPK Budi Mulia atau datang langsung ke sekolah untuk informasi lebih lanjut."
      ]
    },
    {
      "tag": "tuition_fee",
      "patterns": [
        "Berapa biaya sekolah di sini?",
        "Apa biaya SPP di SMPK Budi Mulia?",
        "Berapa uang sekolahnya?",
        "Berapa biaya pendidikan di SMPK Budi Mulia?",
        "Bagaimana struktur biaya sekolah di sini?"
      ],
      "responses": [
        "Untuk informasi biaya sekolah di SMPK Budi Mulia, Anda dapat menghubungi bagian administrasi sekolah atau mengunjungi situs resmi kami.",
        "Biaya pendidikan di SMPK Budi Mulia bervariasi, silakan hubungi administrasi sekolah untuk detail lebih lanjut.",
        "Untuk informasi lengkap mengenai biaya SPP dan biaya lainnya, Anda bisa menghubungi kantor sekolah atau melihat di situs web kami."
      ]
    }
  ]
}

```

Fig. 3 Data of data.json

Once the preprocessing stage is complete, the process moves into the deep learning model, where the dataset will be trained using an Artificial Neural Network algorithm. During this phase, the accuracy of the chatbot will be measured. After preparing the training data, the model is created and compiled, followed by fitting it with the training dataset. This dataset is shuffled to improve the model's generalization capabilities. Upon successful training, the model is saved to a file named model.h5, allowing for easy deployment in real-time user interactions.

2.3 Chatbot Handle

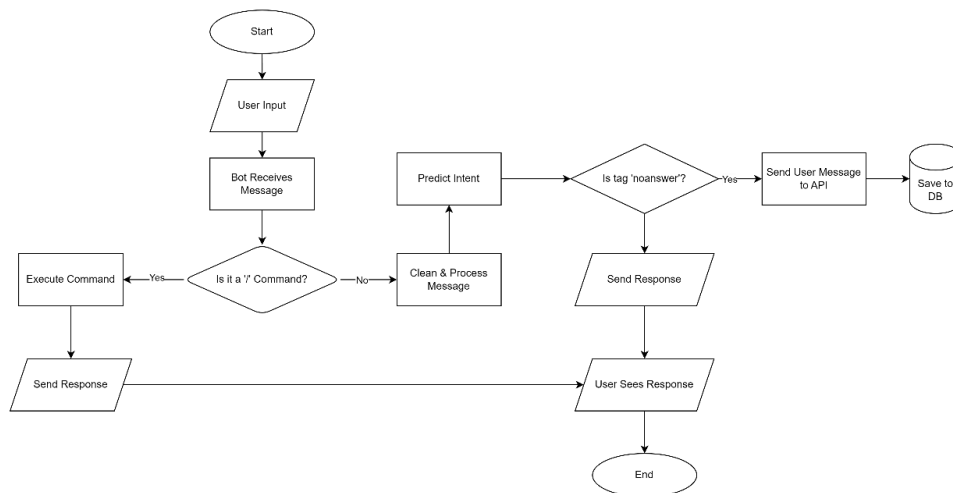


Fig. 4 Flowchart Telegram Chatbot

We utilized a Telegram bot specifically created to enhance communication about Budi Mulia Lawang Junior High School. The principal technique entails users entering inquiries or commands into the Telegram platform, where the bot analyzes these inputs to discern the underlying intents. Through the examination of user interactions, we seek to comprehend the particular demands and inquiries pertaining to the school, including academic programs, extracurricular activities, and administrative details. This method enables us to obtain significant insights into the community's interests and issues.

The bot utilizes natural language processing methods to effectively predict user intents. Each user input is evaluated to ascertain its alignment with established categories, including inquiries regarding school events or requests for academic materials. Categorizing these intentions will boost the bot's responses and elevate user happiness. The data gathered from user interactions will be crucial in enhancing the bot's functioning and ensuring it aligns with the changing requirements of the Budi Mulia Lawang Junior High School community.

3. Result and Discussion

This study surveyed 96 respondents, including students, parents, and staff of Budi Mulia Lawang Junior High School, to evaluate current communication practices and the potential integration of a chatbot as an official source of school information. The findings provide valuable insights into stakeholder perceptions regarding information accessibility and communication effectiveness.

3.1 Survey Result

- **Social Media Engagement:** The average score for the frequency of social media use as a source of school information was 3.7, indicating a positive perception among respondents. This suggests that social media is an effective tool for keeping the community informed, and the school should consider enhancing its presence on these platforms.
- **Website Utilization:** Respondents reported an average score of 2.8 for visiting the official school website, reflecting infrequent engagement. This finding highlights the need for improved content, user experience, and regular updates to encourage more frequent visits.
- **Information Accessibility:** The average score of 2.5 for the difficulty in finding important information indicates significant challenges faced by respondents. Enhancing the organization and accessibility of information across various platforms is crucial for improving stakeholder experiences.
- **Interaction with Administrative Staff:** With an average score of 3.2, interactions with administrative staff were perceived positively. However, there is room for improvement in responsiveness and clarity, which could enhance overall stakeholder satisfaction.
- **Use of Messaging Apps:** The average score of 3.5 for messaging apps, such as WhatsApp and Telegram, indicates a strong preference for these platforms as effective communication tools. The school should continue to leverage these channels for timely communication.
- **Need for Quick Responses:** Respondents expressed a moderate need for quick answers to school-related questions, reflected in an average score of 3.6. Implementing a chatbot could effectively address this need by providing instant responses to frequently asked questions.
- **Clarity of Information:** The average score of 2.9 for the clarity of information suggests that while communications are somewhat clear, there is significant potential for improvement. Simplifying language and ensuring comprehensibility will enhance stakeholder understanding.
- **Desire for 24/7 Access:** Respondents expressed a desire for 24/7 access to school information, with an average score of 3.4. The integration of a chatbot could fulfill this need, allowing stakeholders to access information at any time.
- **Awareness and Comfort with Technology:** The moderate awareness of artificial intelligence (AI) (3.5) and comfort with new technologies (3.1) suggest that while there is some familiarity, further education and training are necessary to increase confidence in using these tools.

- **Interest in Providing Feedback:** The average score of 3.4 for interest in providing feedback indicates a willingness among stakeholders to engage in the improvement of school services. Establishing structured feedback channels could foster a culture of continuous improvement.

3.2 Discussion

The results highlight both strengths and areas for improvement in communication and information accessibility at Budi Mulia Lawang Junior High School. The positive engagement with social media and messaging apps suggests that these platforms are effective for communication, while the low scores for website visits and information accessibility indicate critical areas needing attention.

The findings underscore the importance of timely communication, as evidenced by the moderate need for quick responses. Implementing a chatbot could significantly enhance the school's ability to provide immediate answers to frequently asked questions, thereby improving stakeholder satisfaction.

Furthermore, the desire for 24/7 access to information and the moderate comfort with new technologies suggest that stakeholders are open to adopting innovative solutions. The school should consider providing training and resources to increase familiarity with AI and chatbot technology, ensuring that all community members feel confident in utilizing these tools.

4. Conclusions

This study successfully design and develop a Natural Language Processing (NLP)-based FAQ chatbot on the Telegram platform, aimed at enhancing communication between Budi Mulia Lawang Junior High School and its stakeholders, including students, parents, and staff. The methodology involved systematic data preparation, where information was gathered from interviews with administrative staff to create a comprehensive dataset of common inquiries. This dataset facilitated the training of the chatbot using an Artificial Neural Network (ANN), achieving a training accuracy of 100% at several epochs.

The results demonstrated that the chatbot significantly improved user engagement and satisfaction, as evidenced by positive feedback regarding its responsiveness and the quality of information provided. It effectively addressed frequently asked questions, thereby reducing the administrative workload and allowing staff to focus on more complex inquiries. Survey results indicated a strong preference for using messaging apps like Telegram for communication, underscoring the importance of accessibility and convenience in information dissemination.

Overall, the implementation of the FAQ chatbot not only modernizes communication practices at Budi Mulia Lawang Junior High School but also enhances the educational experience by providing timely and accurate information. Future improvements could focus on expanding the dataset to include a wider variety of inquiries, further increasing the chatbot's effectiveness and user satisfaction.

5. References

- Ahda, F. A., Wibawa, A. P., Prasetya, D. D., & Sulisty, D. A. (2024). Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation. *JOIV : International Journal on Informatics Visualization*, 8(1), 231–238. <https://doi.org/10.62527/joiv.8.1.1818>
- Aramaki, E., Wakamiya, S., Yada, S., & Nakamura, Y. (2022). Natural Language Processing: From Bedside to Everywhere. *Yearbook of Medical Informatics*, 31(1), 243–253. <https://doi.org/10.1055/s-0042-1742510>
- Ashari, M., Sulisty, D. A., & Ahda, F. A. (2024). STEMMING IN MADURESE LANGUAGE USING NAZIEF AND ADRIANI ALGORITHM. *Jurnal Teknik Informatika (Jutif)*, 5(4), Article 4. <https://doi.org/10.52436/1.jutif.2024.5.4.2012>

- Fütterer, T., Fischer, C., Alekseeva, A., Chen, X., Tate, T., Warschauer, M., & Gerjets, P. (2023). ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, 13(1), 15310. <https://doi.org/10.1038/s41598-023-42227-6>
- Hartati, R., & Manullang, E. B. (2024). Implementation of Telegram Chatbot AI with Natural Language Processing (NLP) in Learning Creative Entrepreneurship to Develop Students' Creative and Innovative Competence. *Talenta Conference Series: Local Wisdom, Social, and Arts (LWSA)*, 7(2), Article 2. <https://doi.org/10.32734/lwsa.v7i2.2055>
- Heryati, D., Zulkifli, & Fajri, R. M. (2023). APLIKASI CHATBOT UNTUK PENERIMAAN MAHASISWA BARU UNIVERSITAS INDO GLOBAL MANDIRI MENGGUNAKAN DEEP LEARNING. *Journal of Intelligent Networks and IoT Global*, 1(1), Article 1. <https://doi.org/10.36982/jinig.v1i1.3073>
- Juluru, K., Shih, H.-H., Keshava Murthy, K. N., & Elnajjar, P. (2021). Bag-of-Words Technique in Natural Language Processing: A Primer for Radiologists. *Radiographics: A Review Publication of the Radiological Society of North America, Inc*, 41(5), 1420–1426. <https://doi.org/10.1148/rg.2021210025>
- Khosravi, T., Al Sudani, Z. M., & Oladnabi, M. (2024). To what extent does ChatGPT understand genetics? *Innovations in Education and Teaching International*, 61(6), 1320–1329. <https://doi.org/10.1080/14703297.2023.2258842>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>
- McGrath, C., Farazouli, A., & Cerratto-Pargman, T. (2024). Generative AI chatbots in higher education: A review of an emerging research area. *Higher Education*. <https://doi.org/10.1007/s10734-024-01288-w>
- Mittal, M., Battineni, G., Singh, D., Nagarwal, T., & Yadav, P. (2021). Web-based chatbot for Frequently Asked Queries (FAQ) in Hospitals. *Journal of Taibah University Medical Sciences*, 16(5), 740–746. <https://doi.org/10.1016/j.jtumed.2021.06.002>
- Rianto, Mutiara, A. B., Wibowo, E. P., & Santosa, P. I. (2021). Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation. *Journal of Big Data*, 8(1), 26. <https://doi.org/10.1186/s40537-021-00413-1>
- Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. *JMIR Medical Informatics*, 7(2), e12239. <https://doi.org/10.2196/12239>
- Sulistyo, D. A., Wibawa, A. P., Prasetya, D. D., & Ahda, F. A. (2023). LSTM-Based Machine Translation for Madurese-Indonesian. *Journal of Applied Data Sciences*, 4(3), Article 3. <https://doi.org/10.47738/jads.v4i3.113>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28(11), 13943–13967. <https://doi.org/10.1007/s10639-023-11742-4>