The 2nd International Innovation Technology Proceedings
Vol 2 No 1 (2024): October
E-ISSN: 3030-9948
https://iite-proceeding.poltekindonusa.ac.id

Implementation of NLP and IR-Based Model on Jateng Gayeng Bot Tourism System (JAGABOTA)

Adityo Putro Wicaksono[1*], Satrio Agung Prakoso[2], Mohamad Fajarianditya Nugroho[3], Stephanus Widjaja[4]
[1,2,3]Universitas AKI, [4]STMIK AKI Pati
*Corresponding author: adityoputrowicaksono@gmail.com

Received:	Revised:	Accepted:	Published:
July 12, 2024	July 19, 2024	July 31, 2024	October 5, 2024

Abstract

Information Retrieval (IR) and Natural Processing Language (NLP) are important components in the development of several applications, one of which is a chat bot. The application of both models has been used for a wide range of cases especially for text data processing and text data automation. On the implementation of a bot application on one of the tourism agencies in Central Java that is being developed, using the NLP architecture and IR-Based Model for the development of a robot application named Jateng Gayeng Bot Tourism (JAGABOTA). This architectural model incorporates the concept of Bag of Words (BoW) for the IR Based Model in the process of collecting text data that will be processed into the question and answer data, then performs the data preprocessing process and performs application of NLP modeling with the Neural Network method so that the results of the question process and answers from the data already created can match and match the combination already given as well as facilitates the chat bot process in finding the combinations between question and response correctly and accurately. From the application of NLP architecture and IR Based model this makes tourism Central Java chat bot can do its job to provide information well and accurately to the user.

Key words: *Architechture, Bot Chat, JAGABOTA, Information Retrieval, Natural Language Process*

1. Introduction

Advanced information technology has driven the use of chatbot applications, which utilize artificial intelligence to perform tasks that would normally require human intelligence. Chatbots allow users to interact naturally and deliver information that meets user requests. Chatbots have become an essential tool in many fields, including customer service and education, thanks to technological advancements. However, despite their widespread use, creating and implementing an efficient chatbot architecture still presents a number of challenges. Ensuring optimal chatbot functionality requires a well-designed chatbot architecture. Natural language processing (NLP), conversation management, and backend system integration are some of the components that make up this architecture. Poor user experience, inability to understand user problems, and difficulty providing relevant answers are just a few of the problems that chatbots may face if they are not designed properly.

The 2nd International Innovation Technology Proceedings
Vol 2 No 1 (2024): October
E-ISSN: 3030-9948
https://iite-proceeding.poltekindonusa.ac.id

The presence of Information Retrieval (IR) can make chatbots able to provide more accurate and relevant answers. The application of IR itself can complement the chatbot's ability to understand questions from users, efficient information searches with the amount of

text data or documents used, and processing search results to find the right solution. The application of IR is divided into several features. One of them is the Bag of Words (BoW)
concept for the process of representing a collection of words for modeling at the chatbot stage.

The development of the IR and NLP concepts is represented in the implementation of the Central Java city tourism chatbot application. The application of this chatbot uses the Information Retrieval architecture combined with the Natural Processing Language architecture. The data in this chatbot application has been designed using json data containing question and answer data about tourist areas in Central Java. Then the IR architecture uses the application of Bag of Words to reduce the dimension of data during the modeling process and the application of the NLP architecture with Neural Network as a model for forming the chatbot that will be used. With the application of this architecture, it is expected that the performance of the chatbot for providing Central Java tourism information will be better in terms of finding answers and the response time will be better for this Central Java tourism chatbot application (JAGABOTA).

2. Architecture Design

The architectural design for the implementation of the Central Java tourism chatbot application is divided into 3 architectural forms shown in Figure 1, namely Knowledge Domain, IR-Based Model, and NLP-Based Model.

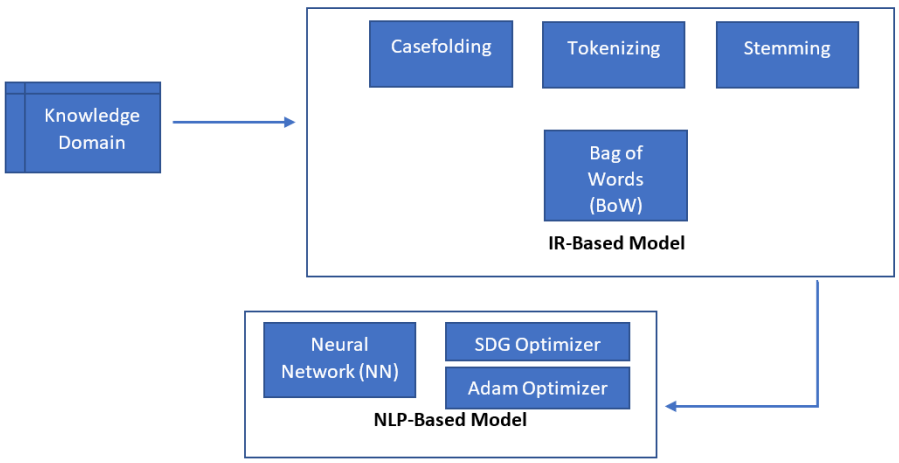

Figure 1. Chatbot Architecture

1. Knowledge Domain

At the Knowledge Domain stage, the determination of the data used and the design of the chatbot to be created are determined first. The data used is data originating from respondent questions called patterns, and the results of the chatbot's answers are obtained from the official tourism website of Central Java Province called responses. This dataset is then created manually

with the json type and the data used are tag, pattern, and response. From the data used, this dataset includes 2072 tag, 114506 pattern, and 2072 response. Then text data modeling will be carried out on the IR-Based Model.

2. IR-Based Model

Information Retrieval has become one of the models for processing text data, either to find information or as a feature for optimizing text data by reducing data dimensions before modeling. In the IR-Based Model used by the JAGABOTA chatbot, Bag of Words is used

for the process of reducing text data that is not needed for chatbot modeling. The stages carried out in the IR-Based Model process with Bag of Words include casefolding,
tokenizing, stemming, and Bag of Words. Casefolding is a stage for changing the text data structure to lowercase. While tokenizing is used to separate words in a document into a smaller form called a token. Stemming functions to change the basic form of a word into a basic word, thereby reducing the form of a word. This stage is the architecture of the IR-Based Model for the text data preprocessing process before the chatbot modeling process is carried out with the NLP-Based Model on the JAGABOTA chatbot.

3. NLP-Based Model

The application of Natural Language Process for the modeling process has been widely used for text data processing. But in this case, it will be applied to the JAGABOTA chatbot modeling. Neural Network is chosen for the chatbot modeling process that will be used based on the availability of data and the desired results. In this study, the Neural Network functions to determine the chatbot prediction process to provide answers or responses based on the tags and patterns that have been created. For the performance of the Neural Network model to provide a better response, the addition of an optimizer for the Neural Network model is added to it. The addition of this optimizer helps in the process of selecting the best prediction based on the parameters used. Some of the parameters used in the NN model in the NLP-Based Model are num_epoch, batch_size, learning_rate, hidden_size, and the Relu activation model in this chatbot modeling architecture.

3. **Implementation**

1. Dataset

In the implementation of this chatbot system, there is a dataset that has been created for the chatbot application based on the explanation of the chatbot architecture above. This dataset is public and obtained from several sources. Among them are using questionnaire answers from respondents and tourist attraction information from the official tourism website of Central Java Province. This data is formed and made into a json data structure with a dictionary data structure for intent, tag, pattern, and response. Intent is a question or request made by the user. While the tag is used as a category or label given to each intent. Pattern is a combination of words or phrases used by users when interacting with the chatbot to better understand the user's intent, while response is the answer or response given by the chatbot based on intent. This dataset was collected collectively and contains 507 tourist attractions in Central Java in 34 districts and cities. From this collective data collection, it was designed by creating several questionnaires consisting of 7 questions displayed in Table 1.

The 2nd International Innovation Technology Proceedings
Vol 2 No 1 (2024): October
E-ISSN: 3030-9948
https://iite-proceeding.poltekindonusa.ac.id

Table 1. Question List

No	Question	Answer
1	Apakah mudah untuk memperoleh informasi seputar objek wisata Jawa Tengah?	• Sangat Susah • Susah • Mudah • Sangat Mudah
2	Apakah sarana teknologi didalam pemberian informasi mengenai objek wisata di Jawa Tengah sudah akurat?	• Sangat Tidak Akurat • Tidak Akurat • Akurat • Sangat Akurat
3	Apa informasi yang dibutuhkan jika anda pergi ke sebuah objek wisata yang baru dikunjungi pertama kali?	
4	Jika anda menggunakan *chatbot*, apa yang akan ditanyakan ketika ingin mengetahui **informasi** objek wisata yang dikunjungi? Tuliskan minimal 3 pertanyaan.	
5	Jika anda menggunakan *chatbot*, apa yang akan ditanyakan ketika ingin mengetahui **harga tiket** objek wisata yang dikunjungi? Tuliskan minimal 3 pertanyaan.	
6	Jika anda menggunakan *chatbot*, apa yang akan ditanyakan ketika ingin mengetahui **alamat** objek wisata yang dikunjungi? Tuliskan minimal 3 pertanyaan.	

7	Jika anda menggunakan *chatbot*, apa yang akan ditanyakan ketika ingin mengetahui **jam operasional** objek wisata yang dikunjungi? Tuliskan minimal 3 pertanyaan.	

2. Evaluation

Evaluation of this chatbot application is carried out to see the results of the accuracy of the chatbot in answering and responding to user questions accurately and well. In the evaluation applied to this chatbot architecture by looking at the comparison of the performance of the Adams and SGD optimizers with several parameters used in this application. The results seen are based on the accuracy and loss of the training data used with the number of hidden_sizes 8 and 1024, epochs of 100 and 200 times, and learning_rates of 0.001 and 0.01. By looking at the results of this evaluation, it is expected to provide optimal and accurate results for the chatbot that has been created at this time and become the evaluation material for the development of this application.

4. Result and Explaination

From the results of the creation of the JAGABOTA chatbot application based on the architecture that has been created, here are the results of the application of the architecture for chatbots with Neural Networks for tourism with the use of the Adam and SDG optimizers that have been created. The optimizer comparison table is shown in tables 2 and 3.

Table 2. Comparison of Neural Network Performance with Adam Optimizer

No	Epoch	Batch Size	Learning Rate	Hidden Size	Accuracy Training Data	Loss Train Data
1	100	8	0.01	8	46.4	3.75
2	100	8	0.001	8	99.14	0
3	200	8	0.001	8	97.03	0
4	200	8	0.001	1024	99.92	0
5	100	8	0.001	1024	98.14	3.51

Table 3. Comparison of Neural Network Performance with SDG Optimizer

No	Epoch	Batch Size	Learning Rate	Hidden Size	Accuracy Training Data	Loss Train Data
1	100	8	0.01	8	86.88	0.17

The 2nd International Innovation Technology Proceedings
Vol 2 No 1 (2024): October
E-ISSN: 3030-9948
https://iite-proceeding.poltekindonusa.ac.id

2	100	8	0.001	8	97.98	0
3	200	8	0.001	8	97.31	0
4	200	8	0.001	1024	98.16	0
5	100	5	0.001	1024	98.12	0

Referring to tables 2 and 3, it is clear that the best results from the application of chatbots with IR-Based Model and NLP-Based Model for Neural Network with the application of Optimizer are by using Optimizer Adam. From the results of the experiment with epoch of 200, batch_size with a value of 8, learning_rate with a value of 0.001, and hidden_size of 1024, it has the highest accuracy results for training data, which is 99.92% and there is no loss in the training data that has been tried. While in the SGD Optimizer for improving prediction results, the highest value is in the fourth test with an accuracy value of 98.16%. So overall, Optimizer Adam is used as an optimizer for the Neural Network optimization process in the chatbot prediction process.

After testing the Chatbot model, this model is deployed in a website-based application. On the website page, information about tourist attractions in Central Java and chatbot services are displayed. The implementation of this web-based chatbot application can be seen in Figures 2.

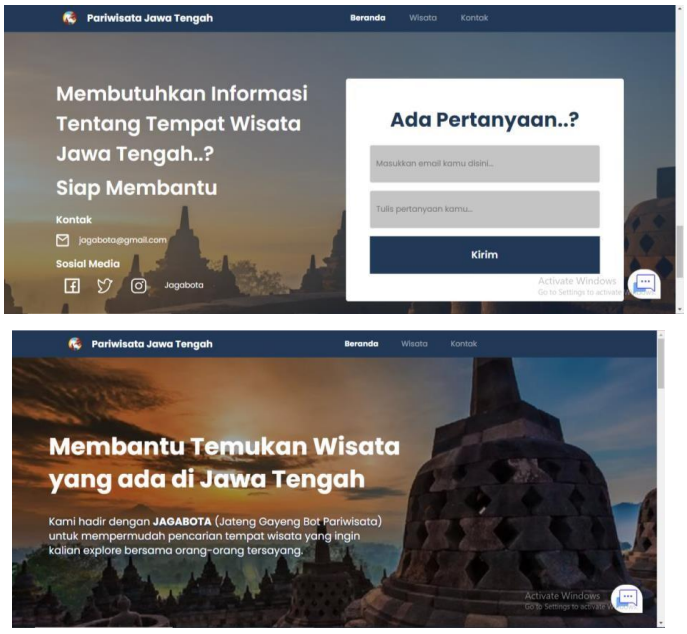

Figure 2. User Interface Chatbot Application

5. Conclusion

This paper presents the architectural design and implementation review of the JAGABOTA chatbot application. The purpose of this review is to provide an overview of design practices and implementation strategies in chatbot systems. Based on the implementation of the chatbot application, it is in accordance with the architecture and parameters that are adjusted to meet the requirements of each system. With this, the JAGABOTA chatbot application can be used as an application to assist in providing information to tourism users in Central Java.

Reference

Aggarwal, C.C. (2022) *Neural Networks and Deep Learning*, *Springer International Publishing AG*. Available at: https://doi.org/10.1007/978-3-031-03758-0_5.

Al-Faruq, M.N.M., Nur'aini, S. and Aufan, M.H. (2022) 'Perancangan UI/UXSemarang Virtual Tourism Dengan Figma', *Walisongo Journal of Information Technology*, 4(1), pp. 43–52. Available at: https://doi.org/10.21580/wjit.2022.4.1.12079.

Amrizal, V. and Aini, Q. (2013) *Kecerdasan Buatan*, *Kecerdasan Buatan*. Jakarta:Halaman Moeka Publishing. Available at: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/44538/2/naskah kecerdasan buatan.pdf.

Anggoro, D.A. and Aziz, N.C. (2021) 'Implementation of K-Nearest Neighbors Algorithm for Predicting Heart Disease Using Python Flask', *Iraqi Journalof Science*, 62(9), pp. 3196–3219. Available at: https://doi.org/10.24996/ijs.2021.62.9.33.

Azimah, F. and Wardani, K.R.N. (2022) 'Sistem Pendeteksi Gejala Awal Covid-19 dengan Penggunaan Metode Al Project Cycle', *Journal Locus Penelitian danPengabdian*, 4(3). Available at: https://doi.org/10.58344/locus.v1i6.135.

Badan Pusat Statistik Provinsi Jawa Tengah (2021). Available at: https://jateng.bps.go.id/indicator/153/613/1/luas-wilayah-menurut- kabupaten-kota.html.

Bell, P. and Beer, B. (2015) *Introducing GitHub*. First, *O'Reilly Media Inc.* First. Edited by M. Blanchette. United States of America. Available at: https://doi.org/10.1007/978-1-4302-6602-0_7.

Chan, J. (2014) *Learn Python in one day and learn it well: Python for beginners with hands-on project: the only book you need to start coding in Python immediately*, *Learn coding fast*.

Chaudhuri, A.B. (2020) *Flowchart and Algorithm Basics: The Art of Programming*. David Pallai.

Disporapar (2023) *Dinas Kepemudaan, Olahraga dan Pariwisata Provinsi JawaTengah*. Available at: https://disporapar.jatengprov.go.id/.

Embarak, O. (2018) *Data Visualization BT - Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems*. Abu Dhabi. Available at: https://doi.org/10.1007/978-1-4842-4109-7.

Ertel, W. (2017) *Introduction of Artificial Intelligence*. Second, *Springer International Publishing AG*. Second. Edited by I. Mackie. Switzerland:Springer-Verlag London Limited. Available at: https://doi.org/10.1002/9781119762010.ch6.

Firmansyah, D. and Dede (2022) 'Teknik Pengambilan Sampel Umum dalamMetodologi Penelitian: Literature Review', *Jurnal Ilmiah Pendidikan Holistik (JIPH)*, 1(2), pp. 85–114. Available at: https://doi.org/10.55927/jiph.v1i2.937.

Goyal, P., Pandey, S. and Jain, K. (2018) *Deep Learning for Natural Language Processing*, *Lecture Notes in Networks and Systems*. Bangalore, Karnataka: Springer Science+Business Media New York. Available at: https://doi.org/10.1007/978-1-4842-3685-7.

Hasugian, P.S. (2018) 'Perancangan Website Sebagai Media Promosi Dan Informasi', *Journal Of Informatic Pelita Nusantara*, 3(1), pp. 82–86.

Hellmann, D. (2017) *The Python 3 Standard Library by Example*. Pearson Education, Inc. Available at: https://www.ptonline.com/articles/how-to-get- better-mfi-results.

Hutama, R.C., Fauziah and Komalasari, R.T. (2021) 'Aplikasi Chatbot Berbasis Teks Menggunakan Algoritma Naive Bayes Classifier FAQ GrabAds', *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 6(1), pp. 90–97. Available at: https://doi.org/10.30998/string.v6i1.9919.

Kim, P. (2017) *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*. First, *Springer Science+Business MediaNew York*. First. Seoul. Available at: https://doi.org/10.1007/978-1-4842-2845-6.

Linge, S. and Langtangen, H.P. (2020) *Programming for Computations - Python*. Second, *Springer Open*. Second. Edited by T.J. Barth et al. Porsgrunn,Norway: Springer Nature Switzerland AG. Available at: https://doi.org/10.1007/978-3-030-16877-3.

Lubis, M.S.Y. (2021) 'Implementasi Artificial Intelligence Pada System Manufaktur Terpadu', *Seminar Nasional Teknik (SEMNASTEK) UISU*, 4(1),pp. 1–7. Available at: https://jurnal.uisu.ac.id/index.php/semnastek/article/view/4134.

Marlena, N. and Sasongko, D. (2010) 'Pembuatan Website Profil Pada Sekolah Menengah Pertama (SMP) Negeri 2 Kartasura', *Journal Speed-Sentra Penelitian Engineering dan Edukasi*, 2(3), p. 7.

Maulid, R. (2022) *Easy Way Array Python untuk Analisis Korelasi*, *https://www.dqlab.id/*. Available at: https://www.dqlab.id/easy-way-array- python-untuk-analisis-korelasi (Accessed: 12 January 2023).

Meng Lee, W. (2019) *Python Maching Learning*. Edited by D. Mahugh. John Wiley& Sons, Inc.

Mishra, P. (2019) *PyTorch Recipes*, *Apress Media*. Bangalore, Karnataka: Apress Media. Available at: https://doi.org/10.1007/978-1-4842-4258-2.

Moolayil, J. (2019) Learn Keras for Deep Neural Networks: A Fast-Track Approach to Modern Deep Learning with Python, Apress. Edited by W. Spahr et al. Vancouver. Available at: https://doi.org/10.1007/978-1-4842-4240-7.

Mueller, J.P. (2018) Beginning Programming with Python for dummeis ® 2nd Edition. second. Canada: John Wiley & Sons, Inc.

MUKLIS, A. (2021) 'Pengaruh Harga dan Kualitas Produk Terhadap Keputusan Pembelian Sepeda Motor Yahama Tipe MX King (Studi pada Pengguna Sepeda Motor Yamaha Tipe MX King Di Sangatta)', Jurnal Administrasi Bisnis Fisipol Unmul, 9(4), p. 314. Available at: https://doi.org/10.54144/jadbis.v9i4.6005.

Mukrodin and Mega Sasmita, N. (2021) 'Artificial Inteligence Dalam Apilkasi Chatbot Sebagai Helpdesk Obyek Wisata Dengan Permodelan Natural Language Processing (Studi Kasus: Kabupaten Cilacap)', Smart Comp, 10(1), pp. 7–14. Available at:

https://doi.org/10.30591/smartcomp.v10i1.2135.

Mustakim, F., Fauziah and Hayati, N. (2021) 'Algoritma Artificial Neural Network pada Text-based Chatbot Frequently Asked Question (FAQ) Web Kuliah Universitas Nasional', Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 5(4), pp. 438–446. Available at: https://doi.org/10.35870/jtik.v5i4.261.

Nafi'ah, H. (2021) AI Project Cycle, Medium.com. Available at: https://medium.com/@hannnfh/ai-project-cycle-ccd67c3dd21d.

Nelli, F. (2018) Python Data Analytics, Python Data Analytics. Available at: https://doi.org/10.1007/978-1-4842-3913-1.

Ningtyas, D.F. and Setiyawati, N. (2021) 'Implementasi Flask Framework pada Pembangunan Aplikasi Purchasing Approval Request', Jurnal Janitra Informatika dan Sistem Informasi, 1(1), pp. 19–34. Available at: https://doi.org/10.25008/janitra.v1i1.120.

Oktarina, Y. and Mulyani, S. (2022) 'Pengembangan Media Pembelajaran Berbasis Android pada Mata Kuliah Keperawatan Medikal Bedah III', Jurnal Ilmiah Universitas Batanghari Jambi, 22(1), p. 651. Available at: https://doi.org/10.33087/jiubj.v22i1.1976.

Purwitasari, N.A. and Soleh, M. (2022) 'Implementasi Algoritma Artificial Neural Network Dalam Pembuatan Chatbot Menggunakan Pendekatan Natural Language

Parocessing', Jurnal IPTEK, 6(1), pp. 14–21. Available at: https://doi.org/10.31543/jii.v6i1.192.

PyTorch (2023) DATASETS & DATALOADERS. Available at: https://pytorch.org/tutorials/beginner/basics/data_tutorial.html (Accessed: 24 July 2023).

Rahayu, D., Mukrodin, M. and Hariyono, R. (2020) 'Penerapan Artificial Intelligence Dalam Aplikasi Chatbot Sebagai Helpdesk Objek Wisata Dengan Permodelan Simple Reflex-Agent (Studi Kasus : Desa Karangbenda)', *Smart Comp*, 9(1), pp. 7–21. Available at: https://doi.org/10.30591/smartcomp.v9i1.1813.

Raj, S. (2019) Building Chatbots with Python, Building Chatbots with Python. Bangalore, Karnataka: Springer Science+Business Media New York. Available at: https://doi.org/10.1007/978-1-4842-4096-0.

Ridwan, Lubis, H. and Kustanto, P. (2020) 'Implementasi Algoritma Neural Network dalam Memprediksi Tingkat Kelulusan Mahasiswa', Jurnal Media Informatika Budidarma, 4(2), pp. 286–293. Available at: https://doi.org/10.30865/mib.v4i2.2035.

Rochmawati, N. et al. (2021) 'Analisa Learning Rate dan Batch Size pada Klasifikasi Covid Menggunakan Deep Learning dengan Optimizer Adam', Journal of *Information Engineering and Educational Technology*, 5(2), pp. 44–48. Available at: https://doi.org/10.26740/jieet.v5n2.p44-48.

Ruder, S. (2016) 'An overview of gradient descent optimization algorithms', pp. 1–

14. Available at: http://arxiv.org/abs/1609.04747.

Sari, A.K. et al. (2021) 'Pengembangan Aplikasi Chat Bot untuk Rekomendasi Wisata', Jurnal Informatika Universitas Pamulang, 6(3), pp. 550–554. Available at: https://doi.org/10.32493/informatika.v6i3.11630.

Septiani, Y., Aribbe, E. and Diansyah, R. (2020) 'Analisis Kualitas Layanan Sistem Informasi Akademik Universitas Abdurrab Terhadap Kepuasan Pengguna Mengguna Metode Sevqual (Studi Kasus : Mahasiswa Universitas Abdurrab Pekanbaru)', Jurnal Teknologi Dan Open Source, 3(1), pp. 131–143. Available at: https://doi.org/10.36378/jtos.v3i1.560.

Del Sole, A. (2021) Visual Studio Code Distilled, Apress Media. Edited by W. Spahr et al. Cremona. Available at: https://doi.org/10.1007/978-1-4842- 4224-7.

Triawan, A. and Siboro, A.R.Y. (2021) 'Penerapan Application Programming Interface (API) Pada Push Notification Untuk Informasi Monitoring Stok Barang Minim', Teknois : Jurnal Ilmiah Teknologi Informasi dan Sains, 11(2), pp. 107–114. Available at: https://doi.org/10.36350/jbs.v11i2.120.

Verdaningroem, N.J.M. and Saifudin, A. (2018) 'Penerapan Kamus Dasar pada Algoritma Porter untuk mengurangi kesalahan Stemming Bahasa Indonesia', Jurnal Teknologi, 10(2), pp. 103–112. Available at: https://doi.org/10.24853/jurtek.10.2.103-112.

Yunitasari, E., Triningsih, A. and Pradanie, R. (2019) 'Analysis of Mother Behavior Factor in Following Program of Breastfeeding Support Group in the Region of Asemrowo Health Center Surabaya', NurseLine Journal, 4(2), pp. 94–102. Available at: https://doi.org/10.19184/nlj.v4i2.11515.