

Implementation of Chatbot System on Tourism Objects in Banyumas Regency with AIML and Chatterbot

Adzan Bari Naufal¹, Sudianto Sudianto^{1*}, Moh Aminullah Al Fachri¹

¹Department of Informatics, Institut Teknologi Telkom Purwokerto
D.I. Panjaitan Street No. 128, Purwokerto, Banyumas, Middle East, Indonesia
E-mail: sudianto@ittelkom-pwt.ac.id

Received: 16 Juni 2023; Accepted: 05 Agustus 2023; Published: 28 Agustus 2023

ABSTRACT

Abstract - Information technology can be applied to all fields, including tourism. Tourism object information media can be implemented into the chatbot system to make the information search process more flexible. Currently, searching for tourist spot information is often done manually; this makes tourist information services limited in time, while the need for tourism information must always be available. This research aims to build a chatbot system using Artificial Intelligence Markup Language (AIML) and ChatterBot methods. Both methods are accessed from libraries in Python using input in the form of natural language that has been processed into certain patterns. The pattern determination process is carried out by classifying a collection of questions on the chatbot using the Support Vector Machines (SVM) method. Then the classification is divided into five attributes, namely address, ticket price, facilities, description, and access. The SVM model built obtained an accuracy rate of 88%. Based on the testing results on both models that have been tested, the approach with AIML results in an accuracy rate in answering questions correctly of 90%, while ChatterBot is 40%.

Keywords: AIML, Chatbot, Chatterbot, Classification, SVM, Tourism.

Copyright © 2023 Jurnal Teknik Elektro dan Komputasi (ELKOM)

1. INTRODUCTION

As time goes by, human life is inseparable from technology. Technology develops not only in the form of hardware that can be seen with the naked eye, but programs that run the hardware are also developing, one of which is appropriate technology based on artificial intelligence. Artificial intelligence can help human activities perform their duties through training and testing [1], [2] as well as image-based solutions [3]–[5] and text [6]–[9].

One form of artificial intelligence is chatbots which are part of NLP (Natural Language Processing). NLP works by processing human language or natural language that is processed so that it can be understood by computers so that communication can be established between the two [10]. NLP can be applied to many things, one of which is the field of tourism. According to Wahyono (Dinporabudpar) Banyumas Regency, there are 109 tourism objects managed by the Regional Government, SOEs, Private Sector, and Tourism Villages. However, of the many tourist attractions, not all tourists know about attractions in the Banyumas area.

In previous research [11], the application of Acehnese chatbots using the AIML concept. The purpose of this study is to find out the accuracy of the chatbot built. The questionnaire data were circulated to users and then analyzed with Likert scales. Based on the test results, the accuracy of the chatbot was 84.4%. Chatbot system about tourist information system in East Java with the concept of AIML, researchers use data on the list of tourist objects such as origin, location, ticket prices, and existing facilities. The author conducted this research intending to make it easier for tourists to find information about tourism objects in East Java. Based on the five-question test results, the chatbot system could answer four questions correctly, while 1 question could not be answered [12]. In addition, the school information center chatbot system with AIML-based concept of Android. This research uses the concept of AIML as chatbot logic. This research is complemented by processing in the form of audio. The data used by the researcher is all information about SMK Otomotif Al Husna [13]. From previous research, it has been able to implement chatbots with the AIML approach. However, it has not been compared with other approaches. In addition to AIML, other approaches can use the ChatterBot library. Therefore, a comparison between the two approaches needs to be done to compare the level of accuracy in answering questions so that a better chatbot system is obtained.

Based on existing problems, a system is needed that can provide information in the form of tourist attractions in Banyumas using a chatbot system. This study aims to implement a chatbot system with AIML and ChatterBot approaches. It will compare the accuracy of the two proposed approaches. The benefits of this research are expected to make it easier for tourists to search for information on tourist attractions in Banyumas. With this system, tourists can get tourist attraction information by doing questions and answers to the system like a tourist guide.

2. RESEARCH METHOD

This research uses two python libraries to develop a chatbot system, namely the AIML (Artificial Intelligence Markup Language) library and the ChatterBot library. More details on the completion of this system chatbot, conduct this study, the flow of the method is depicted on the flow chart in Figure 1.

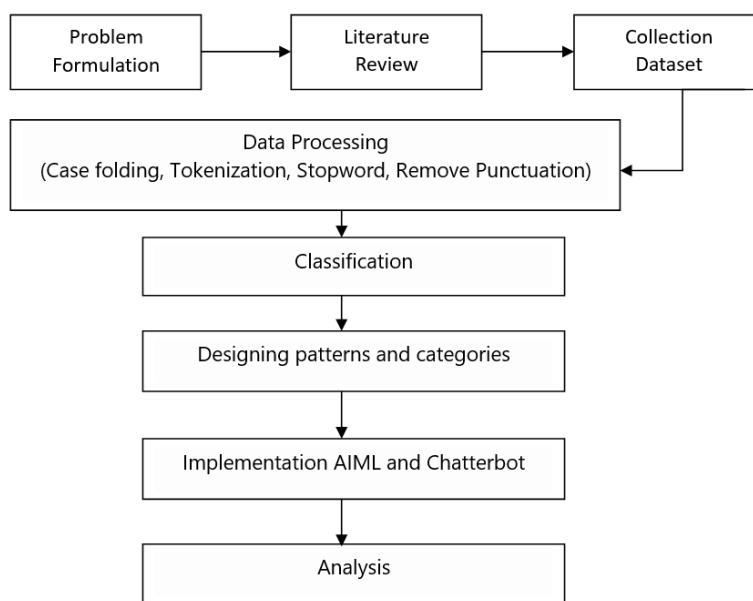


Figure 1. The method flows

2.1. Data Collection

Two types of data are used in this study, namely question data and data for patterns and categories. *Question data* contains the possibility that users will submit to the chatbot system. This study used 84 questions sourced from crawling data on internet media. Data patterns and categories are data used for databases in each library. This pattern and category data was obtained from 20 tourism objects in Banyumas with five categories: address, ticket price, facilities, access, and description; this was chosen because it represents the large number of words needed in searching for tourist attractions. A comparison between the five categories can be seen in Figure 2. This data will make the chatbot system able to answer questions from users.

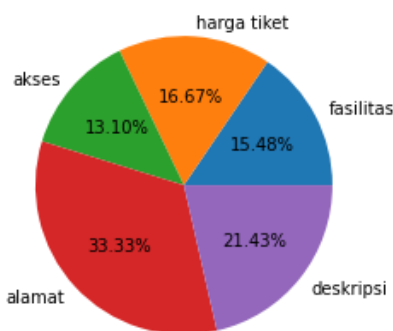


Figure 2. Comparison of five question categories

2.2. Data Processing

The data processing process is divided into two stages: question data processing and pattern and category data processing. First, question data is classified by the SVM method; the purpose of classifying the data is to find patterns of user questions when chatbots are used.

a. Preprocessing

Before data classification, the preprocessing stage is carried out using the Python Natural Language Toolkit (NLTK) library, which consists of the following:

1) Case folding

Case folding is converting all letters in a string into lowercase.

2) Tokenization

The tokenization stage will divide a string into certain parts. For example, there is a string with the content "Hey, try dong show the address of the village", then after tokenization will be ['heii', ',', 'try', 'dong', 'show', 'location', 'the', 'village', '!'] [14].

3) Stopword

The stopword stage will eliminate tokens considered unimportant. It cannot be processed by the system, one of which is hyphens. For example, the previous tokenization result after a stopword would be " heii coba tampilkan lokasi the village " [8], [9].

4) Remove Punctuation

This process will remove punctuation from a string. From the previous stage, after removing punctuation, it will be " heii coba tampilkan lokasi the village ".

b. Classification

The next data processing is to classify data into 5 clusters. The classification stage is used SVM method. The data that has been processed previously is divided into 2 parts with a composition of 80% as training data and 20% as testing data. Data testing is used to find specific patterns known as models. The model will be used as logic in determining chatbot patterns based on user input. Based on accuracy tests using the confusion matrix, the model created an accuracy rate of 88%; the formulas used in SVM include Sigmoid Kernel. Sigmoid Kernel is a development of artificial neural networks, Sigmoid Kernel equations as in equations [15].

$$K(\vec{X}_i, \vec{X}_j) = \tanh(\vec{a}x_i + X_j + \beta) \quad (1)$$

Data patterns and categories are arranged according to the needs of each library. For example, the AIML library requires an XML file, while the ChatterBot library requires a file with a .txt extension. The library has different file formats, but both files have the same content and treatment.

c. AIML and Chatterbot Implementations

AIML [16], [17] and ChatterBot libraries [18] require input to be taken into consideration for chatbot responses. The input is obtained from data users to match the pattern based on the training data at the SVM stage to find the core pattern. From this core pattern, each library will give its best response based on the algorithm of each library.

d. Analysis

The analysis phase was carried out by testing each library with 20 questions, 8 of which were not contained in the training data. This stage is carried out to determine how the comparison of the two methods is used.

3. RESULTS AND DISCUSSION

The automatic answering service, in this case, is that Chatbot provides optimal service and more flexible service times. In this study, Chatbot implementation was carried out through two testing approaches, namely with AIML and testing with Chatterbot. Testing is done by asking 20 questions, as shown in Table 1.

Tabel 1. Test results against both algorithmic approaches

No	Questions	Pattern	Data Training	Chatterbot	AIML
1	Wisata curug Gomblang wisata seperti apa?	deskripsi curug gomblang	No	0	1
2	Di bukit Agaran itu bayarnya berapa?	tiket bukit agaran	No	0	1
3	Tolong jelaskan the village itu wisata apa?	deskripsi the village	No	0	1
4	Apa saja fasilitas di curug nangga?	fasilitas curug nangga	Yes	1	1
5	Bagaimana akses menuju ke the village?	akses the village	Yes	0	1

No	Questions	Pattern	Data Training	Chatterbot	AIML
6	Masuk ke bukit agaran bayar berapa?	tiket bukit agaran	Yes	1	1
7	Apakah fasilitas di bukit tranggulasih itu lengkap?	fasilitas bukit tranggulasih	Yes	0	1
8	Akses ke curug telu bagaimana ya?	akses curug telu	Yes	0	1
9	Tolong dong tampilin dimana curug bayan?	alamat curug bayan	No	1	1
10	Tempat limpakuwus ada di mana?	alamat limpakuwus	Yes	1	0
11	Apa bisa di tunjukkan alamat curug jenggala?	alamat curug jenggala	No	1	1
12	Ada apa saja di wisata pancuran pitu?	fasilitas pancuran pitu	Yes	1	1
13	Hai wibot, tampilkan lokasi hutan pinus limpakuwus!	alamat limpakuwus	Yes	1	1
14	Fasilitas The Village bagus gak ya?	fasilitas the village	Yes	0	1
15	Masuk ke telaga sunyi mahal gak?	tiket telaga sunyi	Yes	0	1
16	Ceritakan deskripsi Baturaden <i>adventure forest</i> dong!	deskripsi baturaden adventure forest	No	0	1
17	Apa aja si fasilitas di bukit Agaran?	fasilitas bukit agaran	No	0	1
18	Masuk ke bukit tranggulasih bayarnya berapa ya?	tiket bukit tranggulasih	No	0	1
19	Di museum jendral soedirman ada apa saja?	fasilitas museum jendral soedirman	Yes	0	0
20	Alamat lengkap curug Gomblang dimana?	alamat curug gomblang	Yes	1	1
Results of testing				40%	90%

In the testing process of the 20 questions, the SVM algorithm was used to classify patterns belonging to predefined categories. After that, the susceptibility of answers was tested with AIML and Chatterbot. Questions from users become vocabulary, which will create Chatbot's responses. The SVM algorithm finds question patterns so the Chatbot system can understand user questions. Out of 20 questions, the ChatterBot library could answer a total of 6 questions correctly, and 2 were questions outside of the training data. While AIML consists of 20 questions, it can answer some 18 questions correctly, and eight questions outside of the training data can be answered correctly.

Test results differ because AIML writes more structured code with a provided tagging process. The AIML library has pattern tags and template tags; The pattern tag is used as a match against user input and will respond with the content of the template tag. At the same time, the ChatterBot library does not differ in terms of patterns and templates.

4. CONCLUSION

The chatbot system has the advantage of being able to respond quickly without paying attention to the many users who use it. Chatbots also make the atmosphere of searching for information feel more natural. Based on the test results, the AIML approach has the highest accuracy rate of 90% compared to ChatterBot, which only reaches 40% accuracy.

REFERENCES

- [1] R. M. S. Adi and S. Sudianto, "Prediksi Harga Komoditas Pangan Menggunakan Algoritma Long Short-Term Memory (LSTM)," vol. 4, no. 2, pp. 1137–1145, 2022, doi: 10.47065/bits.v4i2.2229.
- [2] S. Sudianto, "Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi," vol. 4, no. 2, pp. 1027–1034, 2022, doi: 10.47065/bits.v4i2.2261.
- [3] Sudianto, Y. Herdiyeni, and L. B. Prasetyo, "Machine learning for sugarcane mapping based on segmentation in cloud platform," presented at the The 3rd International Conference on Engineering, Technology and Innovative Researches, Purwokerto, Indonesia, 2023, p. 020001. doi: 10.1063/5.0132180.
- [4] S. Sudianto and R. D. Wahyuningrum, "Identifikasi Sebaran Nitrogen pada Tanaman Padi Berbasis Pengetahuan Fenologi dan Remote Sensing," *Jurnal Nasional Pendidikan Teknik Informatika (Janapati)*, vol. 11, no. 3, pp. 166–175, 2022.
- [5] Sudianto, Y. Herdiyeni, A. Haristu, and M. Hardhienata, "Chilli quality classification using deep learning," in *2020 International Conference on Computer Science and Its Application in Agriculture, ICOSICA 2020*, 2020. doi: 10.1109/ICOSICA49951.2020.9243176.
- [6] S. Sudianto, J. A. Marseli, N. Nugroho, R. W. A. Rumpoko, and Z. Akhmad, "Comparison of Support Vector Machines and K-Nearest Neighbor Algorithm Analysis of Spam Comments on YouTube Covid Omicron," *JTI*, vol. 15, no. 2, pp. 110–118, doi: https://doi.org/10.15408/jti.v15i2.24996.

- [7] W. Afandi, S. N. Saputro, A. M. Kusumaningrum, H. Ardiansyah, M. H. Kafabi, and S. Sudioanto, "Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM," *Jurnal Pengembangan IT*, vol. 7, no. 2, pp. 85–89, 2022.
- [8] S. Sudioanto, A. D. Sripamuji, I. Ramadhanti, R. R. Amalia, J. Saputra, and B. Prihatnowo, "Penerapan Algoritma Support Vector Machine dan Multi-Layer Perceptron pada Klasifikasi Topik Berita," vol. 11, no. 2, pp. 84–91, 2022.
- [9] S. Sudioanto, P. Wahyuningtias, H. W. Utami, U. A. Raihan, and H. N. Hanifah, "Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis (Case Study : Internet Selebgram Rachel Vennya Escape From Quarantine) Perbandingan Metode Random Forest Dan Support Vector Machine Pada Analisis Sentimen Twitt," *Jutif*, vol. 3, no. 1, pp. 141–145, 2022.
- [10] S. Chandra Ayunda Apta, N. Trivetisia, N. A. Winanti, D. P. Martiyaningsih, T. W. Utami, and S. Sudioanto, "Analisis Komparasi Algoritma Machine Learning untuk Sentiment Analysis (Studi Kasus: Komentar YouTube 'Kekerasan Seksual')," *Jurnal Pengembangan IT*, vol. 7, no. 2, pp. 80–84, 2022.
- [11] Muhajirsyah, Mursyidah, and Jamilah, "Pembuatan Chatbot Bahasa Aceh Menggunakan Artificial Intelligence Markup Language," *Jurnal Infomedia : Teknik Informatika, Multimedia, dan Jaringan*, vol. 4, no. 1, pp. 42–49, 2019.
- [12] D. Suryani and E. L. Amalia, "Aplikasi Chatbot Objek Wisata Jawa Timur Berbasis AIML." 2017.
- [13] R. Z. Ramadhani, H. Rusdianto, and V. Yahya, "Penerapan Chatbot Menggunakan Aiml Berbasis Android Pada Smk," *Jurnal Teknik Informatika*, vol. 1, no. 1, pp. 27–33, 2019.
- [14] W. Afandi, S. N. Saputro, A. M. Kusumaningrum, H. Ardiansyah, M. H. Kafabi, and S. Sudioanto, "Klasifikasi Judul Berita Clickbait menggunakan RNN-," *Jurnal Pengembangan IT*, vol. 7, no. 2, pp. 85–89, 2022.
- [15] I. Muhammad and Z. Yan, "Supervised Machine Learning Approaches: a Survey," *ICTACT Journal on Soft Computing*, vol. 05, no. 03, pp. 946–952, 2015, doi: 10.21917/ijsc.2015.0133.
- [16] B. Setiaji, E. Utami, and H. Al Fatta, "Membangun Chatbot Berbasis AIML dengan Arsitektur Pengetahuan Modular," *Semnasteknomedia Online*, vol. 1, no. 1. pp. 18–15, 2013.
- [17] N. Science, G. Bishwabidyalay, and S. Al Mamun, "Review of Integrated Applications with AIML Based Chatbot," *Ist International Conference on Computer & Information Engineering*, pp. 87–90, 2015.
- [18] O. V. Deryugina, "Chatterbots," *Scientific and Technical Information Processing*, vol. 37, no. 2, pp. 143–147, 2010, doi: 10.3103/S0147688210020097.