

Implementasi NLP Dalam Pembuatan Chatbot Customer Service Publisher Jurnal Studi Kasus LARISMA

Mutiara Akbar Nasution¹, Anisa Fitri², Khesya Sabila Rizwine³, Vetryc Styphen Silaban⁴, Fihi Khoirani⁵

^{1,2,3,4,5}Digital Business Study Program, Universitas Negeri Medan, Indonesia

ABSTRAK

This research aims to develop a Customer Service Chatbot for Larisma Journal Publisher that can effectively answer questions and reduce response time to questions that are often similar. This research will design a Chatbot system that can facilitate users in interacting and finding information related to journal publishing, by applying Natural Language Processing. This chatbot system uses training data collected from various questions commonly asked by authors to produce accurate information, with data analysis methods involving preprocessing, training and model building. Testing the accuracy of this chatbot was carried out by the designer through a terminal in Google Colab by viewing and assessing the suitability between the questions and the answers generated. The test results state that this chatbot is able to provide effective responses by evaluating answers based on keywords in the chatbot, so as to provide the right answers.

Keyword: Natural Language Programming; Chatbot; Artificial Intelligence

Corresponding Author:

Mutiara Akbar Nasution,
Universitas Negeri Medan,
Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan,
Kabupaten Deli Serdang, Sumatera Utara 20221, Indonesia
Email: mubarton.7213250020@mhs.unimed.ac.id



1. PENDAHULUAN

Chatbot merupakan salah satu produk kecerdasan buatan yang sangat banyak digunakan di berbagai aspek, karena chatbot mampu menjawab berbagai pertanyaan yang akan ditanyakan pengguna (R dkk., 2018). Dalam dunia bisnis di masa kini, banyaknya user dapat membuat kita kesulitan untuk menjawab pertanyaan yang masuk terhadap bisnis kita, maka dari itu chatbot di ciptakan untuk menjawab pertanyaan-pertanyaan umum yang sering ditanyakan para pengguna agar dapat mengurangi keterlambatan respons.

Natural Language Processing yang disingkat NLP pada sebuah aplikasi chatbot merupakan cara bagaimana komputer mengerti dan memanipulasi teks bahasa alami untuk mendapatkan informasi tertentu (Prasetyo dkk., 2021). Menurut Jurafsky dan Martin (2008) Natural Language Processing adalah bidang ilmu komputer yang mempelajari bagaimana komputer dapat memahami dan menghasilkan bahasa alami. Dengan menggunakan bahasa alami, manusia bisa berinteraksi lebih mudah bersama komputer. NLP pada chatbot digunakan dalam memproses pesan yang masuk dan menganalisis dokumen sumber untuk memberikan jawaban terbaik dan yang paling sesuai.

Larisma atau Lembaga Riset Mutiara Akbar merupakan lembaga yang bergerak dalam publisher jurnal atau sering disebut sebagai Rumah Jurnal Larisma, memiliki 7 jurnal aktif yang bergerak di berbagai bidang. Sejauh ini Admin Larisma mampu mengatasi pertanyaan yang masuk walau terkadang mengalami delay dalam menjawab karena adanya kesibukan lainnya. Diharapkan dengan adanya chatbot maka mampu menangani pertanyaan penulis, mengarahkan mereka pada pedoman submission yang tepat, dan menjawab pertanyaan umum seputar proses penerbitan jurnal. Dengan pengaplikasian respon terhadap pertanyaan-pertanyaan umum dari para penulis dapat terkendali dan tidak mengecewakan.

Pengaplikasian Chatbot Larisma menggunakan NLP ini akan menggunakan Bahasa pemrograman Python di Google Collab. Python adalah bahasa pemrograman yang menggunakan kode-kode yang mudah digunakan atau dimodifikasi oleh pemrogram perangkat lunak (Mohamad, 2022). Python adalah bahasa pemrograman populer yang banyak digunakan untuk analisis dan pemodelan data, dan menjadi pilihan utama

bagi para ilmuwan dan insinyur (Lin dkk., 2022). Python adalah bahasa pemrograman yang digunakan untuk menulis kode yang logis dan jelas untuk berbagai proses, seperti Data Analyst, Data Science, visualisasi, pengambilan data tidak terstruktur, pemrosesan gambar, dan pembuatan model Machine Learning (Gholizadeh, 2022). Pemilihan Bahasa python merupakan pilihan yang sangat baik karena python juga memiliki banyak library yang dapat digunakan.

Dataset yang digunakan berbentuk JSON, JSON merupakan format data standar terbuka yang digunakan untuk pertukaran data terstruktur di internet, dengan status standar diperoleh pada tahun 2013 (Barlas & Stefaneas, 2022). JSON juga merupakan format pertukaran dan penyimpanan yang digunakan dalam aplikasi bisnis dan analitis (Belloni dkk., 2022).

2. METODE PENELITIAN

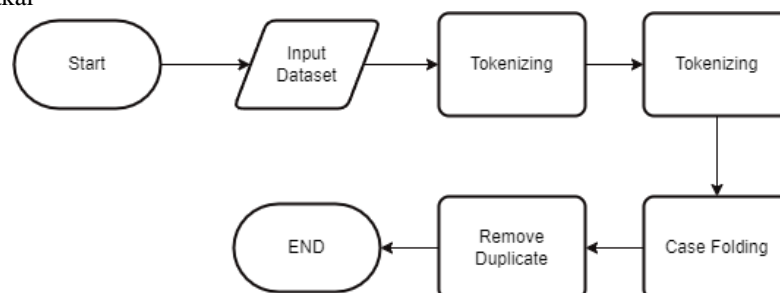
Data dikumpulkan dari berbagai pertanyaan yang ditanyakan para penulis yang ingin melakukan publikasi ke jurnal Larisma, pertanyaan-pertanyaan umum ini dirangkum dan dikumpulkan agar menjadi data training chatbot. Data tersebut dijadikan sebuah dataset berformat JSON yang terbagi dari tag, patterns, serta respons.



Gambar 1 Bagian dari Dataset

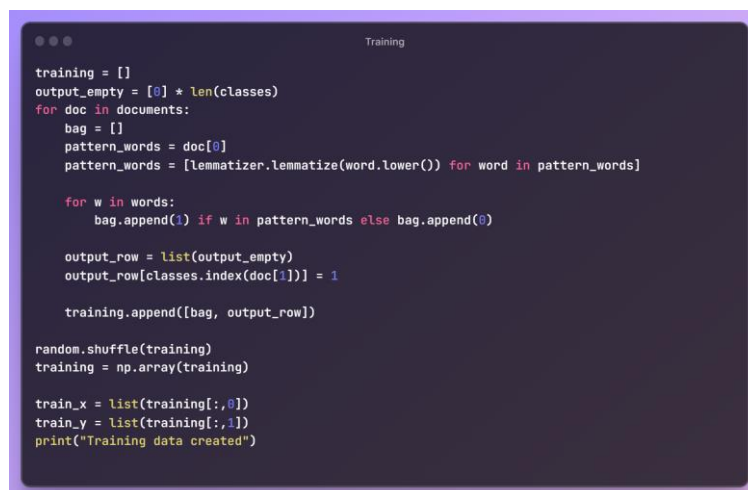
Metode Analisis data terhadap dataset pertanyaan ini dimulai dari:

- 1) PRE-PROCESSING yaitu mengekstrak informasi berharga dan nontrivial dari data teks tidak terstruktur, yang penting dalam bidang penelitian seperti Pengambilan Informasi, Pemrosesan Bahasa Alami, dan penambahan teks (Kashid dkk., 2022). Sebelum data di proses harus dilakukan pre-processing agar data yang ada dapat digunakan sesuai dengan sistem yang telah dibangun. Pre-processing ini dilakukan dimulai dari Tokenizing, Lemmatize, LowerCase, dan Remove Duplicate:
 - a. Tokenizing yakni penghapusan data yang tidak penting dan mengoptimalkan padding token untuk mengurangi waktu pelatihan sambil menjaga akurasi (Dang dkk., 2022).
 - b. Lemmatisasi dalam NLP adalah metode yang digunakan untuk mengekstrak makna kata dari sebuah teks, meningkatkan penalaran dan terjemahan mesin (Chai, 2023).
 - c. Case Folding, yaitu proses mengubah seluruh huruf pada pertanyaan menjadi huruf kecil (Hikmah dkk., 2023).
 - d. Remove Duplicate, di proses ini menghapus berbagai kata yang mengalami duplikat dan tidak terpakai



Gambar 2 Pre-Processing

- 2) Training, Setelah Pre-Processing dilakukan, maka hasil dari pre-processing tersebut harus dilatih/training agar komputer dapat mengerti maksud dari setiap kata antara pertanyaan dan jawaban yang berada di dalam dataset pertanyaan yang kita miliki, pengkodeannya dapat dilihat dibawah.



```

training = []
output_empty = [0] * len(classes)
for doc in documents:
    bag = []
    pattern_words = doc[0]
    pattern_words = [lemmatizer.lemmatize(word.lower()) for word in pattern_words]

    for w in words:
        bag.append(1) if w in pattern_words else bag.append(0)

    output_row = list(output_empty)
    output_row[classes.index(doc[1])] = 1

    training.append([bag, output_row])

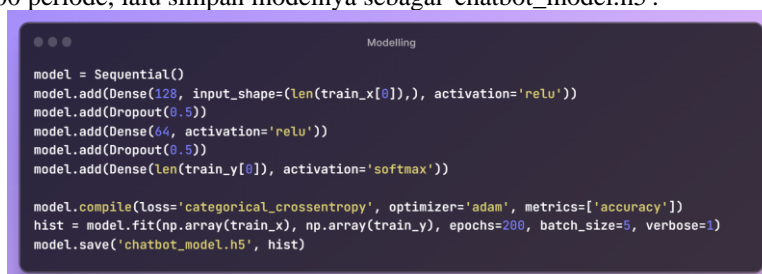
random.shuffle(training)
training = np.array(training)

train_x = list(training[:,0])
train_y = list(training[:,1])
print("Training data created")

```

Gambar 3 Kode Python Training Data

- 3) Pembuatan Model, Data training kita telah siap, sekarang kita akan membangun deep neural network yang memiliki 3 lapisan. Kami menggunakan API Sequentiel Keras untuk ini. Setelah melatih model selama 200 periode, lalu simpan modelnya sebagai 'chatbot_model.h5'.



```

model = Sequential()
model.add(Dense(128, input_shape=(len(train_x[0]),), activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(len(train_y[0]), activation='softmax'))

model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
hist = model.fit(np.array(train_x), np.array(train_y), epochs=200, batch_size=5, verbose=1)
model.save('chatbot_model.h5', hist)

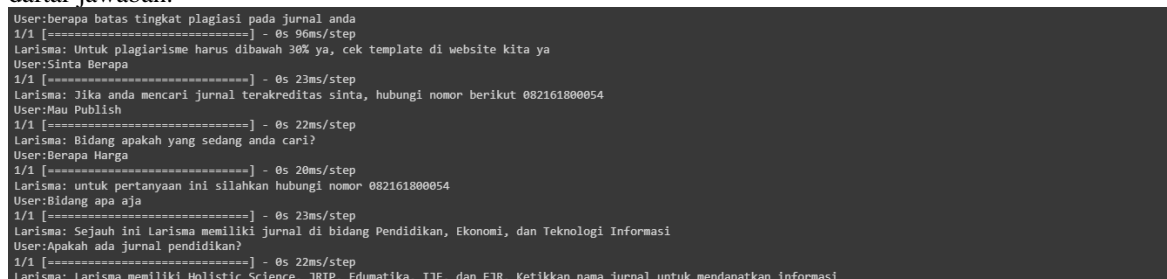
```

Gambar 4 Pemodelan Bot

3. HASIL DAN PEMBAHASAN

Setelah selesai pengolahan data, maka didapatkan model dan akan dijalankan di terminal python, model tersebut di load menggunakan library Keras. Kami akan memuat model training dan kemudian menggunakan antarmuka pengguna grafis yang akan memprediksi respons dari bot. Model hanya akan memberi tahu kita kelas yang dimilikinya, jadi kita akan mengimplementasikan beberapa fungsi yang akan mengidentifikasi kelas tersebut dan kemudian mengambil respons acak dari daftar respons. Sekali lagi kami mengimpor paket yang diperlukan dan memuat file acak 'words.pkl' dan 'classes.pkl' yang telah kami buat saat melatih model kami.

Untuk memprediksi kelas, kita perlu memberikan masukan dengan cara yang sama seperti yang kita lakukan saat pelatihan. Jadi kita akan membuat beberapa fungsi yang akan melakukan preprocessing teks dan kemudian memprediksi kelasnya. Setelah memprediksi kelasnya, kita akan mendapatkan respon acak dari daftar jawaban.



```

User: berapa batas tingkat plagiasi pada jurnal anda
1/1 [=====] - 0s 96ms/step
Larisma: Untuk plagiarisme harus dibawah 30% ya, cek template di website kita ya
User: Sinta Berapa
1/1 [=====] - 0s 23ms/step
Larisma: Jika anda mencari jurnal terakreditasi sinta, hubungi nomor berikut 082161800054
User: Mau Publish
1/1 [=====] - 0s 22ms/step
Larisma: Bidang apakah yang sedang anda cari?
User: Berapa Harga
1/1 [=====] - 0s 20ms/step
Larisma: untuk pertanyaan ini silahkan hubungi nomor 082161800054
User: Bidang apa aja
1/1 [=====] - 0s 23ms/step
Larisma: Sejauh ini Larisma memiliki jurnal di bidang Pendidikan, Ekonomi, dan Teknologi Informasi
User: Apakah ada jurnal pendidikan?
1/1 [=====] - 0s 22ms/step
Larisma: Larisma memiliki Holistic Science, JRIIP, Edumatika, IJE, dan EJR. Ketikkan nama jurnal untuk mendapatkan informasi

```

Gambar 5 Beberapa Hasil Respon Chatbot

Pengujian akurasi chatbot ini dilakukan oleh perancang melalui terminal di Google Colab dengan melihat dan menilai kesesuaian antara pertanyaan dan jawaban yang dihasilkan. Perhitungan pada pengujian ini dilakukan dengan menggunakan rumus sebagai berikut:

$$\text{Akurasi} = (\text{Jawaban Benar} / \text{Pertanyaan}) \times 100\% \quad (1)$$

Hasil Uji Bot Larisma ini dapat dilihat di tabel 1 yang terdiri dari pertanyaan jawaban beserta keterangan apakah jawaban tersebut sesuai atau tidak dengan yang diharapkan

Table 1 Hasil Uji Bot

No	Pertanyaan	Jawaban	Keterangan
1	berapa batas tingkat plagiasi pada jurnal anda	Untuk plagiarisme harus dibawah 30% ya, cek template di website kita ya	VALID
2	Sinta Berapa	Jika anda mencari jurnal terakreditasi sinta, hubungi nomor berikut 082161800054	VALID
3	Mau Publish	Bidang apakah yang sedang anda cari?	VALID
4	Berapa Harga	untuk pertanyaan ini silahkan hubungi nomor 082161800054	VALID
5	Bidang apa aja	Sejauh ini Larisma memiliki jurnal di bidang Pendidikan, Ekonomi, dan Teknologi Informasi	VALID
6	Apakah ada jurnal pendidikan?	Larisma memiliki Holistic Science, JRIP, Edumatika, IJE, dan EJR. Ketikkan nama jurnal untuk mendapatkan informasi	VALID
7	Artikel saya Bidang Ekonomi	Larisma memiliki EJEB, Economic journal and business yang publish di januari, april, juli, oktober. Ejeb belum terakreditasi sinta, apakah anda berminat?	VALID
8	plagiarisme berapa persen	Untuk plagiarisme harus dibawah 30% ya, cek template di website kita ya	VALID
9	Apakah ada bidang komputer?	Larisma memiliki Saintek, jurnal sains, teknologi & komputer yang publish di Februari, Juni, dan Oktober. Saintek belum terakreditasi sinta, apakah anda berminat?	VALID
10	Saya berminat	jika anda berminat langsung upload artikel anda atau hubungi nomor 082161800054 untuk lebih lanjut, terima kasih	VALID

Berdasarkan data tabel diatas yang berisi 10 pertanyaan yang paling sering ditanyakan kepada admin larisma, chatbot ini mampu menjawab seluruh pertanyaan yang diajukan dan menghasilkan jawaban benar dengan akurasi 100%. Dapat disimpulkan bahwa chatbot ini dapat menjawab pertanyaan yang ditanyakan oleh para penulis yang ingin melakukan publish ke rumah jurnal larisma.

4. KESIMPULAN

Chatbot merupakan salah satu hasil dari perkembangan kecerdasan buatan, hal ini dapat kita lihat dengan berbagai perkembangan teknologi yang ada, dengan memahami bahasa manusia menggunakan Natural Language Programming (NLP), komputer dapat belajar menjawab pertanyaan dengan lebih akurat, dengan dilatih menggunakan dataset yang dimiliki. Chatbot merupakan sebuah sistem komputer yang dibangun untuk menjawab berbagai pertanyaan yang dilontarkan terhadap sistem tersebut dengan harapan chatbot tersebut mampu menjawab pertanyaan sesuai dengan pertanyaan yang diberikan. Chatbot yang dibangun menggunakan python dan berbagai library didalamnya terlihat sangat baik tanpa ada kendala, hal ini membuktikan bahwa python merupakan bahasa yang sangat baik dalam pengolahan data. Pemanfaatan chatbot untuk publisher larisma ini diharapkan mampu mempermudah admin larisma dalam menjawab pertanyaan-pertanyaan yang muncul secara berkala yang sifatnya berulang dan sudah dilatih untuk sistem chatbot yang dibangun

DAFTAR PUSTAKA

- Barlas, K., & Stefaneas, P. (2022). An Algebraic Specification/Schema for JSON. *Journal of Engineering Research and Sciences*, 1(5), 243–250. <https://doi.org/10.55708/js0105025>
- Belloni, S., Ritter, D., Schröder, M., & Rörup, N. (2022). DeepBench. *Proceedings of the 2022 workshop on 9th International Workshop of Testing Database Systems*, 1–9. <https://doi.org/10.1145/3531348.3532176>
- Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3), 509–553. <https://doi.org/10.1017/S1351324922000213>

- Dang, T., Sakai, Y., Tabaru, T., & Kasagi, A. (2022). Regularizing Data for Improving Execution Time of NLP Model. *The International FLAIRS Conference Proceedings*, 35. <https://doi.org/10.32473/flairs.v35i.130672>
- Gholizadeh, S. (2022). Top Popular Python Libraries in Research. *Journal of Robotics and Automation Research*, 3(2). <https://doi.org/10.33140/JRAR.03.02.02>
- Hikmah, A., Azmi, F., & Nugrahaeni, R. A. (2023). Implementasi Natural Language Processing Pada Chatbot Untuk Layanan Akademik. *eProceedings of Engineering*, 371–382.
- Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing* (2 ed.). Upper Saddle River.
- Kashid, S., Kumar, K., Saini, P., Negi, A., & Saini, A. (2022). Approach of a Multilevel Secret Sharing Scheme for Extracted Text Data. *2022 IEEE Students Conference on Engineering and Systems (SCES)*, 1–5. <https://doi.org/10.1109/SCES55490.2022.9887697>
- Lin, J. W.-B., Aizenman, H., Manette Cartas Espinel, E., Gunnerson, K., & Liu, J. (2022). *An Introduction to Python Programming for Scientists and Engineers*. Cambridge University Press. <https://doi.org/10.1017/9781108571531>
- Mohamad, O. A. (2022). New Virtual Environment Based on Python Programming. *Iraqi Journal for Computer Science and Mathematics*, 111–118. <https://doi.org/10.52866/ijcsm.2022.02.01.012>
- Prasetyo, V. R., Benarkah, N., & Chrisintha, V. J. (2021). Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya. *Teknika*, 10(2), 114–121. <https://doi.org/10.34148/teknika.v10i2.370>
- R, A. D., Imamah, F., S, Y. M. A., & Does, A. (2018). Aplikasi Chatbot (Milki Bot) Yang Terintegrasi Dengan Web Cms Untuk Customer Service Pada UKM Minsu. *Jurnal Cendikia*, 16(1), 100–106.