

---

## Penerapan *Natural Language Processing* Pada Aplikasi *Chatbot* Info Layanan Kantor Menggunakan *Naïve Baiyes Algorithm*

Kesuma Dwi Ningtyas<sup>1</sup>, Rakhmat Kurniawan<sup>2</sup>, Armansyah<sup>3</sup>

<sup>1,2,3</sup>Ilmu Komputer, Universitas Islam Negeri Sumatera Utara Medan

Email: <sup>1</sup>[dkesuma5@gmail.com](mailto:dkesuma5@gmail.com) <sup>2</sup>[rakhmat.kr@uinsu.ac.id](mailto:rakhmat.kr@uinsu.ac.id), <sup>3</sup>[armansyah@uinsu.ac.id](mailto:armansyah@uinsu.ac.id)

Email Penulis Korespondensi: [dkesuma5@gmail.com](mailto:dkesuma5@gmail.com)

---

### Article History:

Received Dec 12<sup>th</sup>, 2022

Revised Jan 12<sup>th</sup>, 2023

Accepted Jan 21<sup>th</sup>, 2023

### Abstrak

Informasi layanan kantor lurah adalah bagian penting dalam kehidupan sehari-hari, dimana informasi layanan kantor lurah tersebut diperoleh salah satunya dengan konsultasi langsung dengan *customer service*. Berdasarkan hasil wawancara yang dilakukan terhadap beberapa masyarakat, masyarakat memperoleh informasi layanan kantor lurah dengan cara berkunjung ke kantor lurah dan bertanya langsung kepada *customer service*. Peyampaian informasi layanan kantor lurah tersebut dirasa kurang karena keterbatasan oleh waktu kerja pegawai. Pada penelitian ini solusi yang diusulkan untuk masalah tersebut salah satunya dengan cara membangun sebuah aplikasi *chatbot* info layanan kantor dengan penerapan *Natural Language Processing* dengan menggunakan *Naïve Bayes Algorithm*.

**Kata Kunci :** *Chatbot, Natural Language Processing, Naïve Bayes Algorithm, Artificial Intelligence, Informasi Kantor*

---

### Abstract

*Information on village head office services is an important part of everyday life, where information on village head office services is obtained by consulting directly with customer service. Based on interviews conducted with several communities, the community obtained information on the services of the lurah office by visiting the lurah office and asking directly to customer service. Submission of information on the services of the lurah's office is felt to be lacking due to limitations by the employee's working time. In this study, one of the proposed solutions to this problem is by building an office service info chatbot application by implementing Natural Language Processing using the Naïve Bayes Algorithm.*

**Keyword :** *Chatbot, Natural Language Processing, Naïve Bayes Algorithm, Artificial Intelligence, Office Information*

---

## 1. PENDAHULUAN

Penyampaian informasi melalui layanan *customer service* saat ini mulai berkembang pesat. Seperti halnya pada perkembangan layanan *customer service* di media social dunia. Layanan *customer service* dapat dilakukan dengan menjawab langsung pertanyaan oleh *user* atau dapat menggunakan mesin untuk menjawab pertanyaan yang diajukan secara otomatis. Layanan public merupakan suatu media yang disediakan oleh pemerintah untuk dapat memberikan informasi terkini kepada masyarakat.

*Chatbot* merupakan suatu program *computer* yang melakukan percakapan dengan menggunakan metode auditoria atau tekstual [1] *Chatbot* adalah teknologi yang tujuan utamanya adalah untuk berinteraksi dengan pengguna manusia dengan memproses masukan bahasa alami dan menghasilkan keluaran *relative* melalui mesin yang digerakkan oleh aturan atau mesin kecerdasan buatan [2]. Proses *chatbot* dimulai dengan masukan dari pengguna menggunakan bahasa alami dan system akan menjawab dengan respon yang masuk akal atau bisa dikatakan cerdas untuk bahasa yang sebenarnya[3]. *Chatbot* berpotensi disebut sebagai bentuk interaksi manusia-mesin yang paling menjanjikan dan canggih[4].

Kecerdasan buatan atau disebut *Artificial Intelligence*, disingkat AI adalah suatu system atau computer yang memiliki kecerdasan buatan didalamnya, dimana terdapat algoritma dan keputusan-keputusan untuk mengambil sebuah keputusan,

kecerdasan buatan tidak lepas dari campur tangan manusia dan manusia lah yang menanamkan sebuah kecerdasan tersebut ke dalam suatu *system* teknologi dan diatur serta dikembangkan dalam konteks ilmiah dan matematis [5].

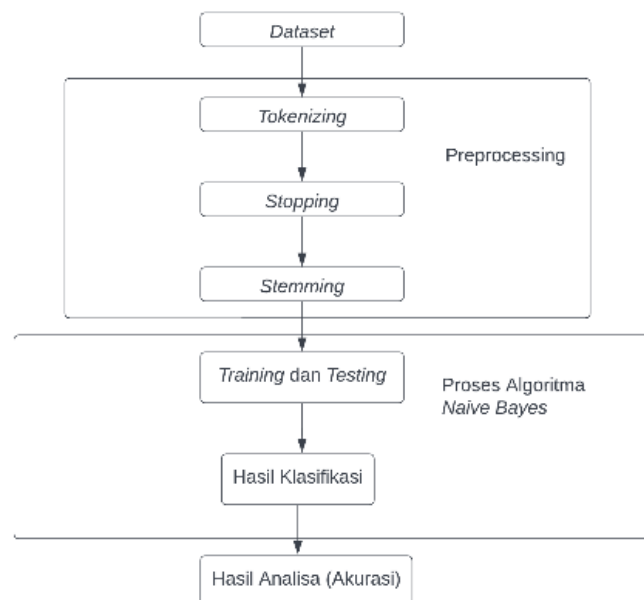
Teorema ini digabungkan dengan “*naïve*” dimana didefinisikan kondisi antar atribut saling bebas [6] *Naïve Bayes Classifier* adalah metode yang populer dan cocok untuk masalah yang diselesaikan dengan algoritma *naïve bayes* [7].

Bahasa natural adalah bahasa yang secara umum digunakan oleh manusia dalam berkomunikasi satu sama lain [8]. Bahasa yang diterima oleh computer butuh untuk diproses dan dipahami terlebih dahulu supaya maksud dari *user* bisa dipahami dengan baik oleh *computer* [9]. Tujuan dari NLP adalah memberikan jawaban atau respon yang sesuai berdasarkan pemahaman makna bahasa manusia yang dilakukan oleh mesin [10]. *Natural Language Processing* menggunakan tahapan *tokenizing*, *filtering* dan analisis serta menerapkan algoritma *nuth morris prrat* [11]. Sehingga dengan NLP pembangunan aplikasi *chatbot* dapat melakukan konversi kata dan memberikan respon sesuai dengan inputan *user* secara alami saat berkomunikasi dengan *chatbot* [12]. Dari *chatbot* ini nanti diharapkan dapat membantu kantor lurah kota Medan untuk berinteraksi dengan masyarakat dalam memberikan informasi kepengurusan surat.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Pada gambar 1 merupakan tahapan penelitian yang dilakukan, sebagai berikut:



Gambar 1. Tahapan Penelitian

Penjelasan langkah-langkah penelitian dari gambar 1 adalah sebagai berikut:

a. Dataset

Pada tahap ini penelitian dilakukan dengan teknik observasi yaitu proses wawancara dan melakukan pengamatan secara langsung ke tempat penelitian.

b. *Preprocessing*

Setelah proses pengumpulan data terselesaikan lalu melakukan *scanning* terhadap dokumen . kemudian memilah dokumen menjadi kalimat. Pemilahan data dilakukan menggunakan memecah *string* teks dari pertanyaan yg panjang menjadi kalimat-kalimat menggunakan fungsi *split()*, menggunakan indikasi titik ".", tanda tanya "?" dan pertanda seru "!" menjadi delimiter untuk memotong *string* pertanyaan. selesainya proses memecah *string* terselesaikan dilanjutkan dengan pemilahan kalimat menjadi perkata (*tokenizing*). lalu menggunakan kata yg tidak krusial (*stoplist*) yang di akhir mengembalikan kedalam bentuk dasar (*stemming*). Proses Algoritma *Naive Bayes* Proses algoritma *Naive Bayes* ini dilakukan setelah melakukan tahapan *preprocessing* pada data, hasil dari inputan data tersebut kemudian dimasukkan kedalam perhitungan *Naive Bayes*. Pada tahapan ini, hasil dari inputan yang sudah melalui tahap *preprocessing*, data akan di *training*, dan hasil yang lebih tinggi akan menjadi data *testing*.

c. Hasil Klasifikasi *Naive Bayes*

Hasil klasifikasi di dapat pada waktu melakukan proses pelatihan serta testing terlebih dahulu kemudian di proses menggunakan prosedur pemecahan *Naive Bayes*, maka sesudah melakukan proses tadi data akan terklasifikasi.

d. Hasil Analisis (Akurasi)

Dapat diketahui klasifikasi atau pengelompokan data berdasarkan masing-masing kategori data pertanyaan.

**2.2.1 Teknik Pengumpulan Data**

1. Observasi  
Di Kantor Lurah Medan Tenggara di Jl. Menteng VII, Medan Tenggara, Kec. Medan Denai, Kota Medan, Sumatera Utara 20228
2. Studi Pustaka (Informasi dari beberapa Jurnal).

**3. HASIL DAN PEMBAHASAN**

**3.1 Tahap Simulasi**

Di tahap simulasi ini dijelaskan bagaimana cara mendapatkan data *input* berdasarkan pengumpulan data yang sudah dilakukan dan bagaimana cara mengolah data tersebut untuk mendapatkan kategori pertanyaan, data pertanyaan dan kata kunci. Ada beberapa langkah untuk mendapatkan kategori pertanyaan dalam penelitian ini adalah:

- a. Mengumpulkan pertanyaan-pertanyaan yang sering ditanyakan masyarakat kepada *customer service* di Kantor Lurah, seperti yang terdapat pada tabel 1 berikut.

Tabel 1. Kategori Pertanyaan dan Pertanyaan

Kategori Pertanyaan	Pertanyaan (P)
Pengurusan KTP	Buk apa saja syarat pengurusan KTP? Apa saja yang harus dibawa untuk pembuatan KTP?
Pengurusan Akta Kelahiran	Apa saja syarat untuk buat akta kelahiran? Bu, anak saya lahir diluar kota, tapi saya asli di kelurahan ini, apa anak saya bisa mengurus akta kelahirannya di kelurahan ini?

- b. Melakukan analisis terhadap informasi apa yang ingin diambil dari pertanyaan masyarakat untuk setiap kategori, seperti yang terdapat pada tabel 2 berikut.

Tabel 2. Kategori Pertanyaan Berdasarkan Kata Kunci

Kategori Pertanyaan	Kata Kunci
Pengurusan KTP	pengurusan KTP
Pengurusan Akta Kelahiran	akta kelahiran

- c. Membuat *rule* jawaban yang digunakan *chatbot* untuk menjawab pertanyaan, sesuai pada gambar 3.

Tabel 3. Pengenalan Etitas

Kategori pertanyaan	Kata Kunci	Rule
Pegurusan KTP	Pengurusan KTP	K1R1
Pegurusan KTP	Pengurusan KTP	K1R2
Pengurusan Akta Kelahiran	Akta Kelahiran	K2R1

- d. Pengenalan pertanyaan
  1. *Text Preprocessing* yaitu langkah penting sebelum melakukan proses klasifikasi kategori pertanyaan
  2. Pembobotan kata pada penelitian ini menggunakan pembobotan TF-IDF

Perhitungan manual TF  

$$TF(d, t) = f(d, t) \tag{1}$$

Ket:  
 F (d,t) : frekuensi kemunculan *term* t pada data d

Perhitungan manual IDF  

$$IDF(t) = \log \left( \frac{Nd}{df(t)} \right) \tag{2}$$

Ket:  
 Nd : Jumlah seluruh data  
 Df (t) : Jumlah data yang mengandung *term* t

Perhitungan manual bobot (w)  

$$W(t, d) = TF(d, t) \times IDF$$

$$= TF(d, t) \times \left( \log \left( \frac{Nd}{df(t)} \right) \right) \tag{3}$$

Keterangan:

W (d,t) : Bobot *term* t pada data d

TF (d,t) : Total kemunculan *term* data d

Nd : Jumlah seluruh data

Df(t) : jumlah data yang memiliki *term* t

3. Tahapan *naive bayes* dilakukan setelah *prerocessing* untuk menentukan kategori pertanyaan. Tahapan ini memiliki dua bagian, yaitu *training* dan *testing*.

- *Training*

Menghitung nilai *prior probabilities* P© dengan menggunakan *maximum likelihood* dan Langkah selanjutnya menghitung nilai *conditional propabilities* dengan menggunakan *laplace smoothing*.

$$P(C) = \frac{N_c}{N} \tag{4}$$

Ket:

P© : Probabilitas kemunculan kata pada data

N<sub>c</sub> : jumlah data dalam kelas c

N : jumlah data yang dipilih

Tahapan *laplace smoothing* bertujuan untuk menghilangkan nilai 0 (nil) pada akhir proses pelatihan

$$P(tk|c) = \frac{W_{ct}+1}{(\sum W' \epsilon V W_{ct}') + B'} \tag{5}$$

Ket:

P(tk|c) : *conditional probabilities term* yang memiliki kategori

W<sub>ct</sub> : nilai pembobotan TF-IDF atau bobot dari kata t pada kelas C

∑W' εV W<sub>ct</sub>' : jumlah bobot dari seluruh kata yang ada di kelas C

B' : jumlah bobot kata unik di semua kelas

- *Testing*

Setelah hasil data *text preprocessing* didapat, selanjutnya melakukan Langkah *matvhing* dan menghitung probabilitas tertera pada tabel 4.

Tabel 4. Data *Testing*

Jenis Pertanyaan	Isi Pertanyaan
?	Jika KTP hilang, apa yang dibutuhkan untuk mengurusnya?

Hasil *text preprocessing* untuk data *testing* dapat dilihat pada tabel 5.

Tabel 5. Hasil *Text Preprocessing* Data *Testing*

Testing
ktp
hilang
butuh
urus

Setelah hasil data *text preprocessing* didapat, selanjutnya melakukan Langkah *matching* dan menghitung probabilitas dapat dilihat pada tabel 6.

Tabel 6. Hasil *Matching Data Training* Dan *Data Testing*

Testing
ktp
butuh
urus

Selanjutnya terlihat pada tabel 7 merupakan data perhitungan manual probabilitas *testing*.

Tabel 7. Data Perhitungan Manual Probabilitas *Testing*

Term	TF	LS					
		Pengurusan KTP	Pengurusan Akta Kelahiran	Pengurusan Surat Keterangan Tidak Mampu	Pengurusan Surat Keterangan Kematian	Pengurusan Surat Keterangan Usaha	Pengurusan Surat Pengantar Kartu Keluarga
ktp	1	0,134417483	0,071394289	0,069892179	0,071394289	0,071394289	0,069892179
butuh	1	0,075593954	0,071394289	0,069892179	0,126949845	0,071394289	0,069892179
urus	1	0,098350001	0,071394289	0,090931821	0,071394289	0,092886112	0,069892179

Hasil data perhitungan manual pemangkatan LS dan TF dapat dilihat pada tabel 8 berikut.

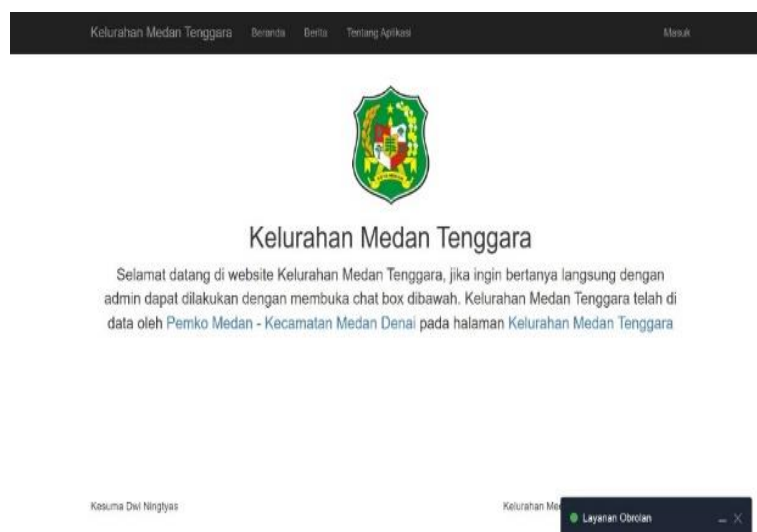
Tabel 8. Data Perhitungan Manual Pemangkatan LS dan TF

TF	TF	LS					
		Pengurusan KTP	Pengurusan Akta Kelahiran	Pengurusan Surat Keterangan Tidak Mampu	Pengurusan Surat Keterangan Kematian	Pengurusan Surat Keterangan Usaha	Pengurusan Surat Pengantar Kartu Keluarga
ktp	1	0,134417483	0,071394289	0,069892179	0,071394289	0,071394289	0,069892179
butuh	1	0,075593954	0,071394289	0,069892179	0,126949845	0,071394289	0,069892179
urus	1	0,098350001	0,071394289	0,090931821	0,071394289	0,092886112	0,069892179
Hasil perkalian		0,000999349	0,000363907	0,000444194	0,000647082	0,000473454	0,000341417
Perkalian <i>prior probability</i>		0,000166891	6,07725E-05	7,41805E-05	0,000108063	7,90668E-05	5,70167E-05

### 3.2 Pengujian

#### 1. Halaman utama *website*

Dari tampilan awal gambar 2 dibawah ini merupakan tampilan dari halaman utama sebuah website yang merupakan halaman yang akan pertama kali dilihat saat membuka website Kelurahan Medan Tenggara.



Gambar 2. Halaman Utama Aplikasi

#### 2. Halaman otentikasi *admin*

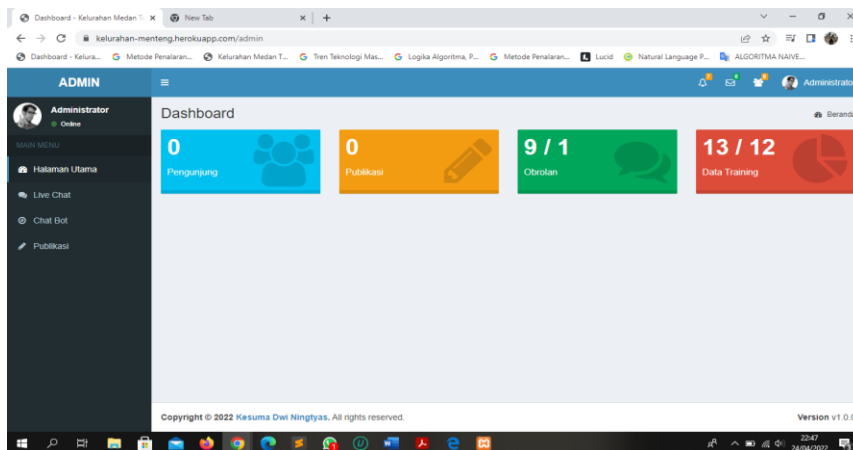
Pada gambar 3 dibawah menampilkan halaman otentikasi admin.



Gambar 3. Halaman Otentikasi Admin

3. Halaman beranda admin

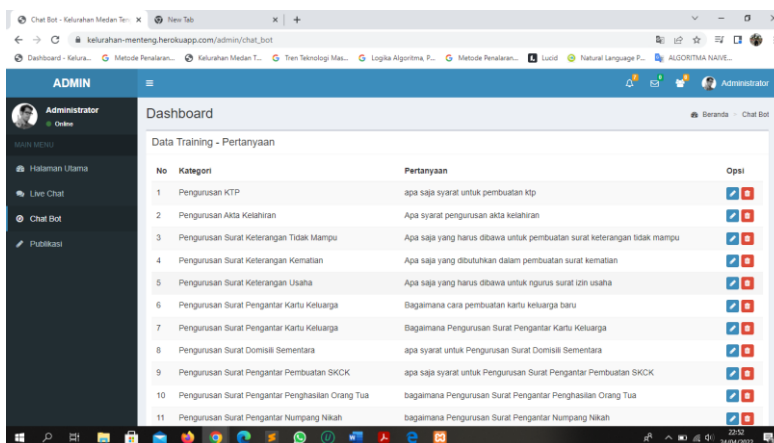
Setelah berhasil melakukan otentikasi maka tampilan gambar 4, pengguna mendapatkan izin untuk mengakses halaman admin.



Gambar 4. Halaman Dashboard Admin

4. Tampilan pengujian data training/ pelatihan bot

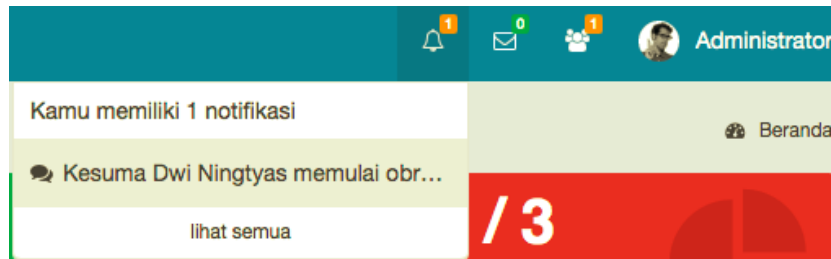
Pada gambar 5 dibawah menampilkan pop-up berisi form untuk menambahkan data training berupa input pertanyaan dan kategori pertanyaan.



Gambar 5. Halaman Training Chatbot

5. Tampilan pengujian system notifikasi

Pada gambar 6 dibawah menampilkan notifikasi jika ada tamu yang memulai obrolan kepada *admin*, dan jika *admin* tidak sedang membuka halaman *live chat* maka akan memunculkan notifikasi pesan baru.



Gambar 6. Notifikasi Aplikasi

6. Tampilan pengujian system obrolan

Pada gambar 7 dibawah menampilkan kotak obrolan atau *chat box* saat uji coba melakukan obrolan teks.



Gambar 7. Pengujian system chatbot

## 4. KESIMPULAN

Berdasarkan pembahasan dan hasil yang dijelaskan maka didapat kesimpulan *chatbot* hanya dapat menjawab pertanyaan berdasarkan kategori dan kata kunci yang telah disimpan didalam *knowledge base chatbot*. *Natural Language Processing* dan *Naïve Bayes algorithm* dapat membantu *chatbot* untuk memahami maksud pertanyaan berdasarkan bahasa alami dengan nilai akurasi klasifikasi kategori pertanyaan sebesar 0,000166891.

## DAFTAR PUSTAKA

- [1] E. W. and P. Pebriantara, “Rancangan Bangun Aplikasi Pembelajaran dengan Memanfaatkan Chatbot API Dialogflow dan Moodle Berbasis Android Pada SMA IT ALIA Tangerang,” vol. 3, hal. 328–335, 2018.
- [2] F. Z. M. Indrayani, L. M., Amalia, R. M., & Hakim, “Emotive Expressions on Social Chatbot,” *J. Sositologi*, vol. 18, hal. 509, 2020.
- [3] S. J. S. Tyas, M. Febianah, F. Solikhah, A. L. Kamil, dan W. A. Arifin, “Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan,” *J. Teknol. Inf. Dan Komun.*, vol. 8, no. 1, hal. 86–99, 2021.
- [4] F. Battineni, G., Chintalapudi, N., & Amenta, “AI Chatbot Design during an Epidemic like the Novel Coronavirus,” *Healthcare*, vol. 8, hal. 154, 2020.
- [5] N. Ratama, *Sistem Penunjang Keputusan Dan Sistem Pakar Dengan Pemahaman Studi Kasus*. Tangerang Selatan: Penerbit Uwais Inspirasi Indonesia, CV, 2019.
- [6] D. Ginantra, Ni luh Wiwik Sri Rahayu., Arifah, Fatimah Nur., Wijaya, Anggi Hadi., Septarini, Ri Sabti., *Data Mining dan Penerapan Algoritma*. Medan: Yayasan Kita Menulis, 2021.
- [7] A. Khan, R., & Das, “Build Better Chatbots,” 2018.



- [8] A. Y. Chandra, D. Kurniawan, dan R. Musa, "Perancangan Chatbot Menggunakan Dialogflow Natural Language Processing (Studi Kasus: Sistem Pemesanan pada Coffee Shop)," *J. Media Inform. Budidarma*, vol. 4, no. 1, hal. 208, 2020, doi: 10.30865/mib.v4i1.1505.
- [9] B. et al Sugara, "Perbandingan Akurasi Algoritma C4.5 dan Naïve Bayes untuk Deteksi Dini Gangguan Autisme pada Anak," *J. IKRA-ITH Inform.*, vol. 3, no. 1, hal. 119–128, 2019.
- [10] M. Aleedy, M., Shaiba, H., & Bezbradica, "Generating and Analyzing Chatbot Responses using Natural Language Processing," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, 2019.
- [11] A. Amrizal, V., Munandar, A., & Arini, "Identifikasi Matan Hadits Menggunakan Natural Language Processing Dan Algoritma Knuth Morris Pratt Berbasis Web," *J. CoreIT J. Has. Penelit. Ilmu Komput. Dan Teknol. Inf.*, vol. 5, hal. 56, 2019.
- [12] Y. Yunefri dan Y. Ersan Fadrial, "Chatbot Pada Smart Cooperative Oriented Problem Menggunakan Natural Language Processing Dan Naive Bayes Classifier Chatbot on Smart Cooperative Oriented Problems Using Natural Language Processing and Naive Bayes Classifier," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2, hal. 2021, 2021.