

Chatbot Telegram Menggunakan Natural Language Processing

Mhd. Furqan¹, Sriani¹, Muhammad Naufal Shidqi^{1,*}

¹Department of Computer Science, Faculty of Science and Technology,
Universitas Islam Negeri Sumatera Utara, Indonesia

*Penulis Korespondensi: naufalshidqi.fal@gmail.com

Abstract

Social media can help in everyday life to communicate and chat without distance limits. Poltekpar has a website that provides information about university entrance tests and Frequently Ask Questions (FAQ) which aims to help new students find information regarding registration at the Poltekpar. To improve public services, Poltekpar must provide the best service to prospective students. Services that previously used FAQs had problems not being able to answer quickly, so they were considered less efficient, so it will be easier with Chatbots that can answer questions in real time. The research method used is the Waterfall System Development Life Cycle (SDLC) which has four stages, namely analysis, design, coding and testing and uses the User Acceptance Testing (UAT) application testing technique. In this research, the process of making the system uses the Natural Language Processing (NLP) method which is applied to systems built with Telegram Messenger and the Python programming language. The Chatbot results tested got a percentage value of 94%. These results show that the Chatbot system is very feasible and effective in helping students obtain the information they need.

Keywords: Chatbot, Frequently Ask Question, Natural Language Processing, Telegram

Abstrak

Media sosial dapat membantu dalam kehidupan sehari-hari untuk berkomunikasi dan mengobrol tanpa batas jarak. Poltekpar memiliki website yang menyediakan informasi tentang tes masuk universitas dan Frequently Ask Question (FAQ) yang bertujuan membantu mahasiswa baru menemukan informasi terkait pendaftaran di Poltekpar. Untuk meningkatkan pelayanan publik, Poltekpar harus memberikan pelayanan yang terbaik pada calon mahasiswa. Pelayanan yang sebelumnya menggunakan FAQ memiliki kendala tidak dapat menjawab secara cepat, sehingga dinilai kurang efisien, maka akan lebih mudah dengan adanya Chatbot yang dapat menjawab pertanyaan secara realtime. Metode penelitian yang digunakan merupakan System Development

Life Cycle (SDLC) Waterfall yang memiliki empat tahap yaitu analysis, design, coding dan testing dan menggunakan teknik pengujian aplikasi User Acceptance Testing (UAT). Dalam penelitian ini melakukan proses pembuatan sistem dengan menggunakan metode Natural Language Processing (NLP) yang diterapkan pada sistem yang dibangun dengan Telegram Messenger dan bahasa pemrograman Python. Hasil Chatbot yang diuji mendapat nilai persentase sebesar 94%. Hasil ini menunjukkan sistem Chatbot sangat layak dan efektif dalam membantu mahasiswa memperoleh informasi yang dibutuhkan.

Kata kunci: Chatbot, Frequently Ask Question, Natural Language Processing, Telegram

1 Pendahuluan

Memasuki era teknologi saat ini, banyak sekali kebutuhan pengguna akan media informasi yang cepat dan mudah (Budiman et al., 2021). Bentuk pemanfaatan teknologi adalah adanya alat yang dapat membantu komunikasi dalam kehidupan sehari-hari, salah satunya melalui media sosial. Media sosial merupakan aplikasi berbasis internet, dengan pengguna media sosial dapat melakukan chatting tanpa dibatasi oleh jarak (Sudirman, 2016), salah satu contoh media sosial adalah aplikasi Telegram.

Telegram sebagai aplikasi pesan instan diklaim bisa mengatasi kelemahan yang terdapat pada Whatsapp. Telegram adalah aplikasi berbasis cloud dan alat enkripsi. Telegram menyediakan enkripsi end-to-end, penghancuran pesan secara otomatis, dan infrastruktur multi-pusat data (Hamburger, 2014).

Aplikasi Telegram banyak disukai karena multi platform dan ringan, Telegram juga menyediakan Application Programming Interface (API) dengan dokumentasi yang lengkap untuk pengembangan sehingga pembuatan chatbot di Telegram dapat dilakukan dengan baik (Ong et al., 2015). Ada fitur utama yang disediakan Telegram

untuk chatting yaitu Chat Messenger. Beberapa agensi menggunakan Chat Messenger sebagai Layanan Pelanggan.

Poltekpar memiliki website yang di dalamnya terdapat layanan FAQ yang dirancang untuk membantu mahasiswa baru menemukan informasi terkait pendaftaran di perguruan tinggi. Dalam rangka meningkatkan pelayanan publik, sebagai lembaga pendidikan perlu memberikan pelayanan yang terbaik kepada calon peserta didik. Jika layanan dulu memiliki FAQ memiliki kendala tidak dapat menjawab secara cepat, maka akan lebih mudah dengan chatbot yang dapat menjawab pertanyaan secara real time.

Chatbot adalah program komputer yang dapat mensimulasikan percakapan dengan menggunakan bahasa alami (P dan Afrianto, 2015), pada dasarnya chatbot mengadopsi pengetahuan manusia ke komputer, sehingga komputer dapat melakukan percakapan dengan pengguna (Rosyadi et al., 2020).

Natural Language Processing (NLP) atau sering disingkat NLP adalah salah satu cabang kecerdasan buatan dimana komputer dirancang untuk berkomunikasi dengan manusia menggunakan bahasa alami, seperti bahasa Indonesia (Ali, 2017). Model komputasi seperti ini berguna untuk memudahkan komunikasi antara

manusia dan komputer dalam hal pencarian informasi (Ridha et al., 2020).

Penelitian ini bertujuan untuk membangun Chatbot guna memaksimalkan otomatisasi FAQ di Poltekpar Medan dengan menggunakan metode NLP dan diterapkan pada chat messenger Telegram. Sehingga chatbot yang telah dibangun dapat berkontribusi dalam meningkatkan pelayanan yang mampu menjawab pertanyaan secara otomatis.

2 Kerangka Teori

Natural Language Processing (NLP) adalah bidang kecerdasan buatan yang mempelajari interaksi manusia-komputer menggunakan bahasa alami. Model komputer seperti ini berguna untuk memudahkan komunikasi antara manusia dan komputer yang mencari informasi sehingga keduanya dapat berkomunikasi dalam bahasa alami (Ali, 2017). Beberapa proses preprocessing yang terdapat pada Natural Language Processing, yaitu:

1. Case folding, adalah salah satu proses pengolahan kata. Proses ini mengubah huruf besar menjadi huruf kecil. Selain itu, angka dan simbol khusus seperti tanda seru (!), koma (,), garis miring (/), lebih dari (>), kurang dari (<); dll tidak terlalu penting dalam klasifikasi juga dilaksanakan (Hermawan dan Bellanar Ismiati, 2020).
2. Tokenisasi, menggambarkan proses membagi teks menjadi kata-kata, menggunakan spasi sebagai pemisah, dengan tujuan membuat setiap kata berdiri sendiri, tidak terkait dengan kata lain (Nata dan Yudiasra, 2017).

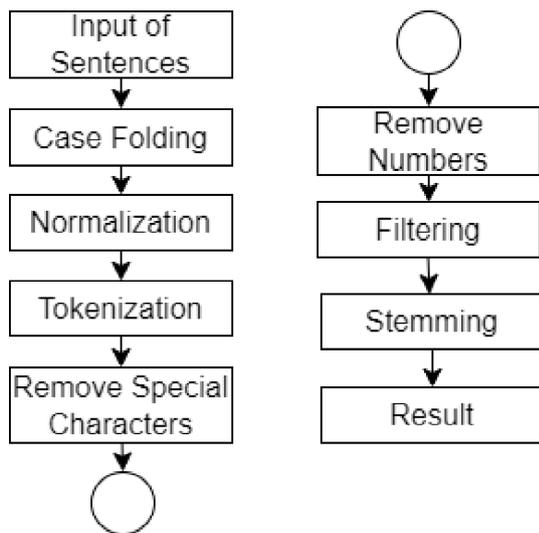
3. Part-of-Speech Tagging, adalah penugasan tag POS ke kalimat atau kata. Pengkodean POS dapat dilakukan dengan memberikan pengenalan pada setiap kata seperti kata kerja, kata benda, kata sifat, dan aturan tata bahasa lainnya.
4. Syntactic Parsing, adalah proses menganalisis sintaksis kalimat berdasarkan teori gramatikal tertentu. Proses screening secara umum dapat dibagi menjadi dua bagian, yaitu kelompok ketergantungan dan kelompok daya tarik.
5. Stemming merupakan tahap menciptakan kata dasar dari sebuah kata dengan menghilangkan sufiks dari kata tersebut seperti "in-", "-nya" dll (Sivakumar dan Gunasundari, 2017).
6. Stopword Removal adalah proses filtering, sebuah kalimat biasanya mengandung beberapa kata yang tidak lagi memiliki makna substantif, seperti "ini", "ini", dll. Oleh karena itu, untuk meningkatkan efisiensi klasifikasi, kata-kata yang sering muncul tetapi tidak berdampak tinggi dihilangkan pada langkah ini (Indraloka dan Santosa, 2017).
7. Stopword merupakan kumpulan kata umum (common words) yang tidak penting namun kerap muncul. Contoh stopwords pada bahasa Indonesia adalah "ke", "di", "yang", dll.

Menurut Pendekatan Natural Language Processing percakapan dalam program chatbot seperti percakapan antar manusia. Obrolan dilakukan melalui koneksi Internet, sehingga pengguna dapat mengobrol dengan chatbot kapan saja, di mana saja. Analisis entitas dianggap sebagai

masalah pemrosesan Natural Language Processing (NLP) (Hartono et al., 2016).

3 Metode

Pada penelitian ini, NLP diimplementasikan pada chat messenger Telegram. Proses yang dilakukan terlihat pada Gambar 1.



Gambar 1. Kerangka Kerja

Dari perencanaan pada Gambar 1 dapat dijelaskan tahap-tahapnya sebagai berikut:

1. Input Kalimat, Input pada sistem ini tentunya berupa kata kunci dengan kalimat sederhana yang akan disampaikan oleh pengguna ke sistem proses selanjutnya yang akan dijalankan dengan menggunakan metode natural language processing dalam pengecekan kalimat yang akan dimasukkan .
2. Case Folding, Pada tahap ini, Anda akan mengubah keseluruhan kalimat dari huruf besar menjadi huruf kecil (Hermawan dan Bellanar Ismiati, 2020).
3. Normalisasi, Pada tahap ini mereka akan mengubah kata tidak

baku menjadi kata baku atau kata biasa.

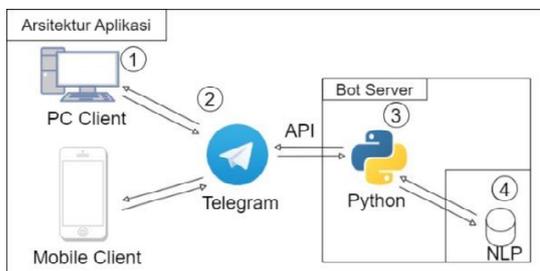
4. Tokenization, yaitu proses yang digunakan untuk membagi input data atau teks menjadi token kata tunggal (Ab. Nasir et al., 2020).
5. Hapus Karakter Khusus. Tahap ini menghilangkan simbol dan tanda baca dalam kalimat. Pada penelitian ini, perintah-perintah yang disediakan oleh Python dapat digunakan untuk membersihkan tanda baca.
6. Hapus Angka, Pada tahap ini akan menghilangkan angka dalam kalimat (Nata dan Yudiasra, 2017).
7. (Stopword Removal), tahap ini berfungsi untuk memilah kata-kata yang penting pada token yang dihasilkan oleh proses sebelumnya. Daftar stopwords dicatat dalam tabel menggunakan daftar stopwords yang digunakan dalam penelitian ini oleh Tala (2003), yaitu stopwords bahasa Indonesia, yang meliputi kata-kata seperti: ini, itu, yang, ke, di, di, ke, dan seterusnya (Indraloka dan Santosa, 2017).
8. Stemming, Stemming merupakan proses pembentukan kata dasar. Istilah-istilah yang diperoleh dari langkah eliminasi stopwords diturunkan. Proses sebenarnya adalah menghasilkan kata kunci karena kata kunci ini memberikan dokumen/informasi yang lebih relevan (Sivakumar dan Gunasundari, 2017).
9. Hasil yang akan dikeluarkan pada sistem ini adalah jawaban atas pertanyaan pengguna berupa jawaban benar atau salah (tidak ditemukan jawaban).

Selanjutnya akan dilakukan

perencanaan sistem, di mana metode dan data yang telah diperoleh akan dirancang, sehingga selama proses pembangunan sistem akan lebih mudah untuk diimplementasikan ke dalam aplikasi. Kemudian akan dilakukan pengembangan sistem berdasarkan desain sistem yang telah dibuat. Desain yang telah selesai kemudian diimplementasikan ke dalam sebuah aplikasi.

Identifikasi sistem adalah contoh model skematik yang memperlihatkan tata letak keadaan suatu sistem dengan memperlihatkan komponen perangkat lunak yang beroperasi pada perangkat keras. Ponsel dan komputer (komputer) yang menggunakan salah satu fungsi Telegram Messenger yaitu bot digunakan sebagai perangkat keras. Bot ini kemudian akan menjawab atau menanggapi pertanyaan lebih lanjut dari pengguna atau pelanggan.

Arsitektur sistem ditunjukkan pada Gambar 2. Pengguna mengirim pesan teks ke akun bot melalui server Telegram dan meneruskannya ke server bot. Server bot memproses pesan tersebut sehingga dapat memberikan respon yang benar kepada pengguna dalam bentuk pesan teks. Tanggapan dikirim ke pengguna melalui server Telegram.



Gambar 2. Arsitektur Sistem

Arsitektur sistem dalam Gambar 2 dapat dijelaskan sebagai berikut:

1. Pengguna atau Klien mengirim pesan frase (kueri) dari desktop atau seluler.
2. Pesan diterima oleh server Telegram, yang kemudian diteruskan melalui server bot API Telegram.
3. Server bot berbasis Python menerima pass dari API Telegram dan kemudian beralih ke sisi Python untuk melakukan NLP pada input data.
4. Data yang sudah diolah menggunakan NLP kemudian dilakukan pengecekan database untuk melihat apakah data yang diterima sesuai dengan data yang ada di database (Oktafiani, 2020).

Penelitian ini akan menggunakan teknik pengumpulan data berikut:

1. Penelitian kepustakaan adalah teknik pengumpulan data yang dilakukan dengan cara mencari informasi dari buku, jurnal ilmiah dan sumber lain, yang secara alami berkaitan dengan penelitian yang dilakukan oleh peneliti yaitu mengenai pembuatan chatbot menggunakan metode NLP.
2. Observasi adalah teknik yang akan digunakan untuk mengumpulkan data dengan cara mengamati secara langsung objek penelitian, akan dikumpulkan informasi dan pengetahuan berupa pertanyaan tentang pendaftaran mahasiswa baru yang akan menjadi fokus penelitian. Peneliti mengambil data berupa informasi dan berupa pertanyaan di Poltekpar yang terletak di Jalan Rumah Sakit Haji No 12 Medan.

Penggunaan sistem ini diawali dengan pencarian jawaban atas pertanyaan input dari pengguna. Proses

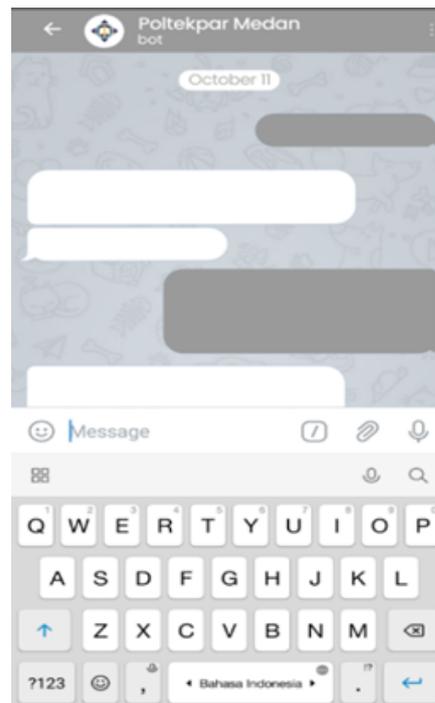
pertama yang dilakukan Terdapat 3 langkah dalam proses klasifikasi teks untuk memproses input berupa kalimat teks. Text preprocessing adalah tahapan yang paling penting dari klasifikasi teks. Daftar umum untuk basis chatbot. Ada 2 langkah dalam pengolahan teks, yaitu tokenization dan stemming. Langkah tokenizing merupakan langkah pemisahan string input berdasarkan masing-masing kata penyusunnya (Ratniasih et al., 2017) dengan memakai tanda baca dan spasi untuk pemisah. Perintah tokenisasi yang dipakai dalam penelitian ini yaitu kata token() yang telah tersedia pada Natural Language Toolkit (NLTK). Fungsi NLTK sebagai berikut:

1. Mendeteksi tanda baca sebagai karakter terpisah.
2. Pisahkan koma dan tanda petik jika diikuti dengan spasi.
3. Pisahkan titik-titik yang muncul di akhir baris. Selain itu juga menggunakan API yang dibuat dengan menggunakan bahasa pemrograman Python, Penelitian ini bertujuan untuk membantu pengguna dalam menemukan informasi terkait Poltekpar secara real time.

4 Hasil dan Pembahasan

Pada penelitian ini data berupa dataset diambil dan dikumpulkan dari pihak Poltekpar dan website yaitu poltekparmedan.ac.id, data tersebut berupa FAQ yang berisi pertanyaan dan informasi seputar Poltekpar yang sering ditanyakan. Untuk menjalankan sistem, pengguna menginputkan kalimat ke dalam aplikasi chat Telegram. Antarmuka chatbot yang telah dibangun ditunjukkan pada Gambar 3.

Hasil dari sistem Chatbot tersebut adalah proses pencarian jawaban dan percakapan yang dilakukan. Beberapa skenario dibahas. Skenario digunakan untuk mencocokkan input dan output yang diharapkan. Skenario tersebut ditunjukkan dalam Tabel 1.



Gambar 3. Antarmuka Chatbot

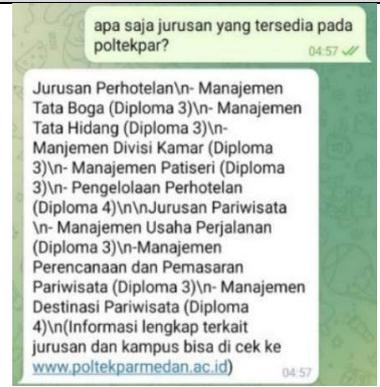
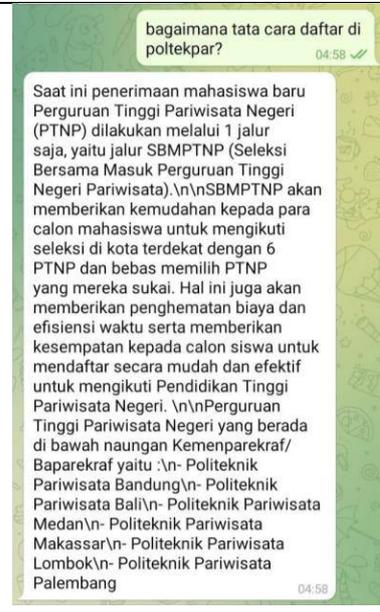
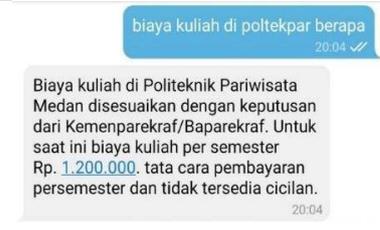
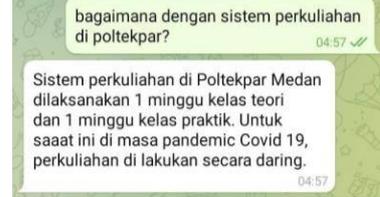
Berdasarkan hasil chat yang dibangun, diperoleh sistem yang mampu memberikan jawaban berdasarkan input dari pengguna. Saat pengguna menggunakan Chatbot untuk pertama kalinya. Pertanyaan yang dapat dijawab oleh chatbot terbatas pada pendaftaran perguruan tinggi, informasi terkait Poltekpar, dan sistem perkuliahan.

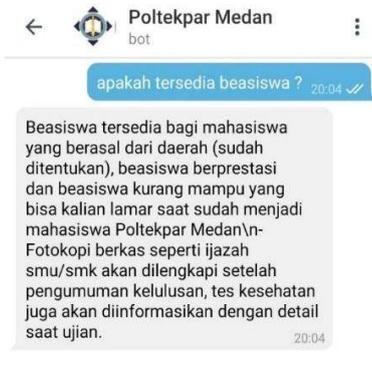
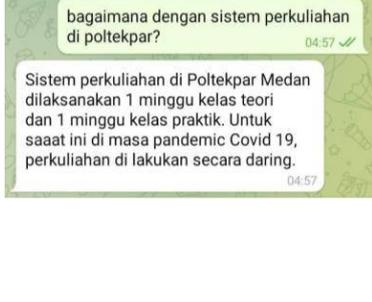
Pengembangan Chatbot menggunakan metode natural language processing membuahkan hasil yang baik. Sistem dapat yang telah dibuat dan jawaban yang telah ditentukan.

Sistem chatbot yang dibangun pada penelitian ini gagal merespon ketika terjadi salah ketik atau salah ketik pada input yang dimasukkan oleh pengguna.

Tabel 1. Pengujian Chatbot

No	Masukan	Tanggapan	Keterangan	Status
1	/Star		Pengujian pada pertama kali dengan Chatbot, untuk dapat melihat tampilan seperti gambar diatas adalah yaitu mengakses https://t.me/poltekparmedan.bot .	Berhasil
2	Permisi saya ingin tau tentang Poltekpar?		Langkah awal dalam menggali informasi seputar kampus adalah menanyakan mengenai profil kampus. Dengan memberikan masukan berupa kalimat “permisi saya ingin tau tentang Poltekpar?”, maka Chatbot akan memberikan informasi mengenai profil Poltekpar.	Berhasil
3	Di mana alamat Poltekpar?		Pertanyaan mengenai alamat kampus adalah pertanyaan umum, maka peneliti memilih kalimat masukan “di mana alamat poltekpar?” kemudian Chatbot akan memberikan jawaban berupa informasi alamat lengkap Poltekpar.	Berhasil
4	Apa saja jurusan yang tersedia		Pemilihan kalimat masukan “apa saja jurusan yang tersedia pada Poltekpar?” karena umumnya calon mahasiswa akan melakukan	Berhasil

No	Masukan	Tanggapan	Keterangan	Status
	pada Poltekpar?		riset mengenai jurusan yang tersedia pada kampus tujuan, maka Chatbot akan menampilkan hasil pencarian dari pertanyaan yang telah diajukan.	
5	Bagaimana tata cara daftar di Poltekpar?		Pemilihan kalimat masukan "saya ingin bertanya mengenai alur daftar?" karena merupakan pertanyaan umum yang akan ditanyakan calon mahasiswa yang ingin mendaftar di Poltekpar. Dengan balasan pesan dari Chatbot mengenai tahap pendaftaran di Poltekpar.	Berhasil
6	Biaya kuliah di Poltekpar		Biaya kuliah masih menjadi salah satu faktor utama sebagian mahasiswa dalam menyambung studinya. Maka calon pertanyaan biaya kuliah akan sering sering ditanyakan.	Berhasil
7	Bagaimana dengan sistem perkuliahan di poltekpar?		Sistem perkuliahan menjadi salah satu faktor utama yang akan dipertimbangkan calon mahasiswa dalam memilih kampus, maka masukan pertanyaan "Bagaimana dengan sistem perkuliahan di poltekpar?" dipilih untuk	Berhasil

No	Masukan	Tanggapan	Keterangan	Status
			Chatbot dapat memberikan informasi sistem perkuliahan di Poltekpar.	
8	Apakah beasiswa tersedia?		Umumnya calon mahasiswa mencari tahu apakah kampus menyediakan beasiswa yang relevan, maka dipilih masukan “Apakah tersedia beasiswa?” dengan balasan pesan yang berhasil direspons oleh bot yaitu memberikan sebuah informasi beasiswa yang tersedia pada Poltekpar.	Berhasil
9	Cek lulus		Untuk mempermudah melihat status pendaftaran calon mahasiswa yaitu dengan memilih kalimat masukan “Cek lulus”, dengan balasan pesan yang berhasil direspons oleh bot yaitu memberikan sebuah <i>link</i> untuk melihat daftar kelulusan.	Berhasil
10	Kapan jadwal smm?		Pertanyaan mengenai jadwal smm umum ditanyakan oleh mahasiswa, maka kalimat masukan “Kapan jadwal smm?” dipilih untuk Chatbot dapat memberikan informasi berupa <i>link</i> untuk mengetahui jadwal secara keseluruhan.	Berhasil

Jadi jika terjadi salah ketik, chatbot akan menjawab pertanyaan pengguna yang berulang-ulang. Hal ini dikarenakan mesin chatbot menganalisis kata masukan yang sesuai dengan kata kunci yang ditentukan dan tidak menganalisis memberikan jawaban yang sesuai dengan input yang masukkan. Input dicocokkan terhadap kata kunci kesamaan antara kata input dan kata kunci. Jika kata yang tidak cocok maka terdeteksi salah, melihat kesamaan kata membuat Chatbot mendeteksi kata tersebut benar sesuai dengan skor kesamaan yang tinggi.

Karena adanya kecerdasan buatan dalam bentuk Chatbots, diharapkan dapat melakukan hal tersebut dimanfaatkan oleh pihak Poltekpar dalam upaya memberikan kemudahan dan kecepatan bagi mahasiswa untuk mendapatkan informasi yang mereka butuhkan.

Selanjutnya dilakukan pengujian dengan metode UAT untuk mengetahui keefektifan Chatbot yang telah dibangun. Pengujian dilakukan dengan memberikan 5 pertanyaan kepada 30 responden (mahasiswa) dimana terdapat 5 pilihan yaitu Sangat Tidak

Setuju, Tidak Setuju, Cukup, Setuju, Sangat Setuju. Hasil pengujian ditunjukkan pada Tabel 2.

Tabel 2. Hasil UAT

No	Pertanyaan	Hasil
1	Apakah Chatbot mudah dioperasikan?	93%
2	Apakah fitur Chatbot mudah dikenali?	94,6%
3	Apakah isi Chatbot mudah dipahami?	96,6%
4	Apakah jawaban yang disajikan sudah sesuai dengan kebutuhan?	94%
5	Apakah jawaban yang disajikan mudah dipahami?	92,6%

Berdasarkan hasil UAT pada Tabel 2, diperoleh nilai rata-rata sebesar $(93+94,6+96,6+94+92,6)\%/5=94,16\%$. Dengan demikian tingkat penerimaan pengguna terhadap Chatbot ini adalah

94%. Hasil pengujian dengan tingkat keberhasilan sebesar 94% menunjukkan bahwa Chatbot sangat layak dan efektif untuk digunakan.

5 Kesimpulan

Berdasarkan pengujian yang telah dilakukan, diperoleh kesimpulan bahwa dengan menerapkan proses NLP pada pengolahan kata, data yang akan diolah dapat disaring dengan cepat dan tepat sehingga menghasilkan satuan kata yang baik sehingga program dapat mengidentifikasi atau memahami pertanyaan yang diajukan oleh mahasiswa, antara lain mengenai pendaftaran perkuliahan.

Chatbot bekerja dengan baik sesuai dengan alur yang telah ditetapkan, yaitu dimulai dengan membuka pesan, setelah itu pengguna bisa bertanya sesuai dengan keinginan pengguna, kemudian Chatbot menjawab berdasarkan kata kunci yang telah ada sebelumnya. Pada pengujian UAT didapatkan hasil bahwa pengguna cukup puas dengan Chatbot, dan Chatbot sangat layak dan efektif dengan persentase sebesar 94%.

Referensi

- Ab. Nasir, A. F., Seok Nee, E., Sern Choong, C., Shahrizan Abdul Ghani, A., Abdul Majeed, A. P. P., Adam, A. dan Furqan, M. (2020), 'Text-based emotion prediction system using machine learning approach', *IOP Conference Series: Materials Science and Engineering* **769**(1), 012022. doi: 10.1088/1757-899X/769/1/012022.
URL: <https://iopscience.iop.org/article/10.1088/1757-899X/769/1/012022>
- Ali, A.-R. (2017), 'Pengenalan Natural Language Toolkit (NLTK)'.
URL: <https://code.tutsplus.com/id/tutorials/introducing-the-natural-language-toolkit-nltk-cms-28620>
- Budiman, I., Saori, S., Anwar, R. N., Fitriani dan Pangestu, M. Y. (2021), 'ANALISIS PENGENDALIAN MUTU DI BIDANG INDUSTRI MAKANAN (Studi Kasus: UMKM

Mochi Kaswari Lampion Kota Sukabumi)', *Jurnal Inovasi Penelitian* **1**(10), 2185–2190.

URL: <https://stp-mataram.e-journal.id/JIP/article/view/419>

Hamburger, E. (2014), 'Why Telegram has become the hottest messaging app in the world'.

URL: <https://www.theverge.com/2014/2/25/5445864/telegram-messenger-hottest-app-in-the-world>

Hartono, Abdullah, D., Hartama, D., Roslina, Furqan, M., Zarlis, M. dan Situmorang, Z. (2016), SENTIMENT ANALYSIS USING CONTEXT BASED FUZZY LINGUISTIC HEDGES, in 'Proceeding International Conference of Computer, Environment, Social Science, Engineering and Technology (ICEST) 2016', Medan, Indonesia, pp. 160–162.

Hermawan, L. dan Bellanier Ismiati, M. (2020), 'Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval', *Jurnal Transformatika* **17**(2), 188. doi: 10.26623/transformatika.v17i2.1705.

URL: <https://journals.usm.ac.id/index.php/transformatika/article/view/1705>

Indraloka, D. S. dan Santosa, B. (2017), 'Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia', *Jurnal Sains dan Seni ITS* **6**(2). doi: 10.12962/j23373520.v6i2.24419.

URL: http://ejournal.its.ac.id/index.php/sains_seni/article/view/24419

Nata, G. N. M. dan Yudiastra, P. P. (2017), Preprocessing Text Mining Pada Email Box Berbahasa Indonesia, in 'E-Proceedings KNS&I STIKOM Bali', Bali, Indonesia, pp. 479–483.

URL: <http://knsi.stikom-bali.ac.id/index.php/e proceedings/article/view/88>

Ong, S. P., Cholia, S., Jain, A., Brafman, M., Gunter, D., Ceder, G. dan Persson, K. A. (2015), 'The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles', *Computational Materials Science* **97**, 209–215. doi: 10.1016/j.commatsci.2014.10.037.

URL: <https://linkinghub.elsevier.com/retrieve/pii/S0927025614007113>

P, E. N. S. C. dan Afrianto, I. (2015), 'RANCANG BANGUN APLIKASI CHATBOT INFORMASI OBJEK WISATA KOTA BANDUNG DENGAN PENDEKATAN NATURAL LANGUAGE PROCESSING', *Komputa : Jurnal Ilmiah Komputer dan Informatika* **4**(1), 49–54. doi: 10.34010/komputa.v4i1.2410.

URL: <https://ojs.unikom.ac.id/index.php/komputa/article/view/2410>

Ridha, M. P., Nurhayatunnufus, L. dan Maulid, H. (2020), 'Lappybot : Aplikasi Chatbot Untuk Informasi Pemilihan Laptop Menggunakan Metode Natural Language Processing (nlp)', *eProceedings of Applied Science* **6**(2), 2586–2594.

URL: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/appliedscience/article/view/>

Rosyadi, H. E., Amrullah, F., Marcus, R. D. dan Affandi, R. R. (2020), 'Rancang Bangun Chatbot Informasi Lowongan Pekerjaan Berbasis Whatsapp dengan Metode NLP (Natural Language Processing)', *Briliant: Jurnal Riset dan Konseptual* **5**(3), 619–626. doi: 10.28926/briliant.v5i3.487.

URL: <https://jurnal.unublitar.ac.id/index.php/briliant/article/view/487>

Sivakumar, A. dan Gunasundari, R. (2017), 'A Survey on Data Preprocessing Techniques for Bioinformatics and Web Usage Mining', *International Journal of Pure and Applied Mathematics* **117**(20), 785–794.

URL: <https://acadpubl.eu/jsi/2017-117-20-22/articles/20/68.pdf>

Sudirman (2016), 'ANALISIS KOMUNIKASI DATA DENGAN XML DAN JSON PADA WEBSERVICE', *Computer Engineering, Science and System Journal* **1**(2), 1–6. doi: 10.24114/cess.v1i2.4066.

URL: <http://jurnal.unimed.ac.id/2012/index.php/cess/article/view/4066>