

Pembangunan Aplikasi dan Klasifikasi Pertanyaan Chatbot Informasi Akademik Menggunakan Metode Cosine Similarity dan Naïve Bayes

**Rosida Nur Aziza^{1*)}; Efy Yosrita¹; Rahma Farah Ningrum¹; Tiara Sukma Ardanti¹;
Syeh Rafhil Arafazin¹**

1. Teknik Informatika, Institut Teknologi PLN, Menara PLN, Jl. Lingkar Luar Barat,
Duri Kosambi, Cengkareng, Jakarta Barat, DKI Jakarta 11750, Indonesia

**)Email: rosida@itpln.ac.id*

Received: 26 Januari 2023 | Accepted: 16 Oktober 2023 | Published: 29 Januari 2024

ABSTRACT

The provision of precise and fast academic information for students who need it is one of the problems faced by the Informatics Engineering Program of IT PLN. Therefore, an Academic Information chat bot was created that operates on the Telegram chat application and is based on Cosine Similarity. This writing will discuss two parts, namely the application development part and the application part of the classification method. In the testing part, several questions asked by students through the chatbot application have been classified into 3 groups to find out what information is most requested by students. Before being classified using the Naïve Bayes method, incoming questions must go through text processing stages first, such as case folding, tokenizing, filtering, stemming, and weighting using TF/IDF techniques. Based on the test results, the average cosine similarity value is 82.4% and the accuracy of the classification based on 137 test data is 89%.

Keywords: chatbot, academic information, cosine similarity

ABSTRAK

Penyediaan informasi akademik yang tepat dan cepat bagi mahasiswa yang membutuhkan menjadi salah satu masalah yang dihadapi oleh Prodi Teknik Informatika IT PLN. Oleh karena itu dibuatlah chat bot Informasi Akademik yang beroperasi pada aplikasi chat Telegram dan berbasis Cosine Similarity. Penulisan ini akan membahas dua bagian, yaitu bagian pengembangan aplikasi dan bagian penerapan metode klasifikasinya. Pada bagian pengujian, beberapa pertanyaan yang diajukan mahasiswa melalui aplikasi chatbot yang telah kemudian diklasifikasikan menjadi 3 kelompok untuk mengetahui informasi apa yang paling banyak diajukan oleh mahasiswa. Sebelum diklasifikasikan menggunakan metode Naïve Bayes, pertanyaan-pertanyaan yang masuk harus melalui tahapan pemrosesan teks terlebih dahulu, seperti tahapan case folding, tokenizing, filtering, stemming, dan pembobotan menggunakan teknik TF/IDF. Diharapkan hasil dari penelitian ini bisa memberikan rekomendasi bagi Prodi Teknik Informatika ITPLN mengenai jenis informasi yang disebarkan kepada mahasiswanya. Berdasarkan hasil pengujian, diperoleh nilai rata-rata cosine similarity 82.4% dan nilai akurasi hasil klasifikasi menggunakan confusion matrix untuk 37 data uji adalah 89%.

Kata kunci: chatbot, informasi akademik, cosine similarity

1. PENDAHULUAN

Chatbot adalah aplikasi yang secara efisien menyediakan layanan percakapan kepada pengguna dengan menggunakan metode *Artificial Intelligence* (AI) [1]. Aplikasi ini dirancang untuk membantu pengguna dalam merespon pertanyaan yang masuk secara otomatis dan *real-time*, serta dapat dimanfaatkan dalam berbagai bidang, seperti pemasaran[2], pendidikan[3], hiburan[4], layanan kesehatan[5], dan sektor layanan publik [6]. Di bidang pendidikan, *chatbot* dapat digunakan untuk mendukung peserta didik dalam proses belajar mengajar, seperti yang terlihat dalam penelitian [7] yang mengembangkan *chatbot* untuk membantu siswa dalam kegiatan laboratorium. Selain itu, teknologi ini juga dapat dimanfaatkan untuk mengakses informasi akademik di perguruan tinggi[8]. Penggunaan teknik pencocokan adalah salah satu langkah dalam pengembangan *chatbot*. Salah satu metode yang dapat digunakan untuk mencocokkan pertanyaan dengan basis pengetahuan *chatbot* dengan cepat dan akurat adalah metode *cosine similarity* [9], [10]. Aplikasi *chatbot* Informasi Akademik pada penelitian ini diimplementasikan menggunakan aplikasi chat Telegram. Aplikasi ini dipilih karena termasuk aplikasi *chat* dengan jumlah pengguna yang terbanyak dan telah dilengkapi dengan fasilitas untuk pembangunan *chatbot* [11].

Pengelompokan teks menggunakan metode-metode klasifikasi pada *machine learning* merupakan salah satu tugas dari NLP (*Natural Language Processing*). Klasifikasi teks dapat ditemui pada tugas pengelompokan dokumen dan pengelompokan pertanyaan [12]. Penelitian dari Mohammed dan Omar menjelaskan mengenai penggunaan klasifikasi pertanyaan berdasarkan taksonomi Bloom untuk mengakses kemampuan siswa menggunakan *K-Nearest Neighbour*, *Support Vector Machine*, dan *Logistic Regression* [13]. Riset lain mengusulkan suatu *framework* untuk klasifikasi pertanyaan menggunakan pendekatan berbasis tata bahasa (GQCC, *Grammar-based Question Categorization and Classification*) yang mengeksplorasi struktur kalimat dari pertanyaan yang dianalisis [14]. Hasil penelitian menunjukkan bahwa menggunakan kategori sintaksis yang terkait dengan berbagai jenis *Common Nouns*, *Numerical Numbers*, dan *Proper Nouns*, memungkinkan algoritma *machine learning* untuk membedakan antara berbagai jenis pertanyaan dengan lebih baik. Salah satu metode *machine learning* yang banyak digunakan untuk klasifikasi pertanyaan pada sistem tanya jawab adalah Naïve Bayes [15], [16]. Naïve Bayes *classifier* yang banyak digunakan untuk klasifikasi teks dalam *machine learning* didasarkan pada probabilitas kondisional fitur milik kelas, yang fitur-fiturnya dipilih dengan metode tertentu. Keunggulan dari metode ini adalah membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Selain itu, Naive Bayes seringkali bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks daripada yang diharapkan [17].

Penelitian ini akan membuat aplikasi *chatbot* untuk menjawab kebutuhan mahasiswa Teknik Informatika ITPLN supaya mendapat tanggapan yang cepat apabila ada pertanyaan terkait informasi akademik. Pada tahapan uji coba, pertanyaan-pertanyaan yang masuk akan diklasifikasi untuk menentukan apakah menanyakan informasi dosen (kelas 1), permintaan pembuatan surat atau dokumen tertentu (kelas 2), atau mengenai jadwal akademik (kelas 3).

2. METODE PENELITIAN

2.1. Tahapan Penelitian

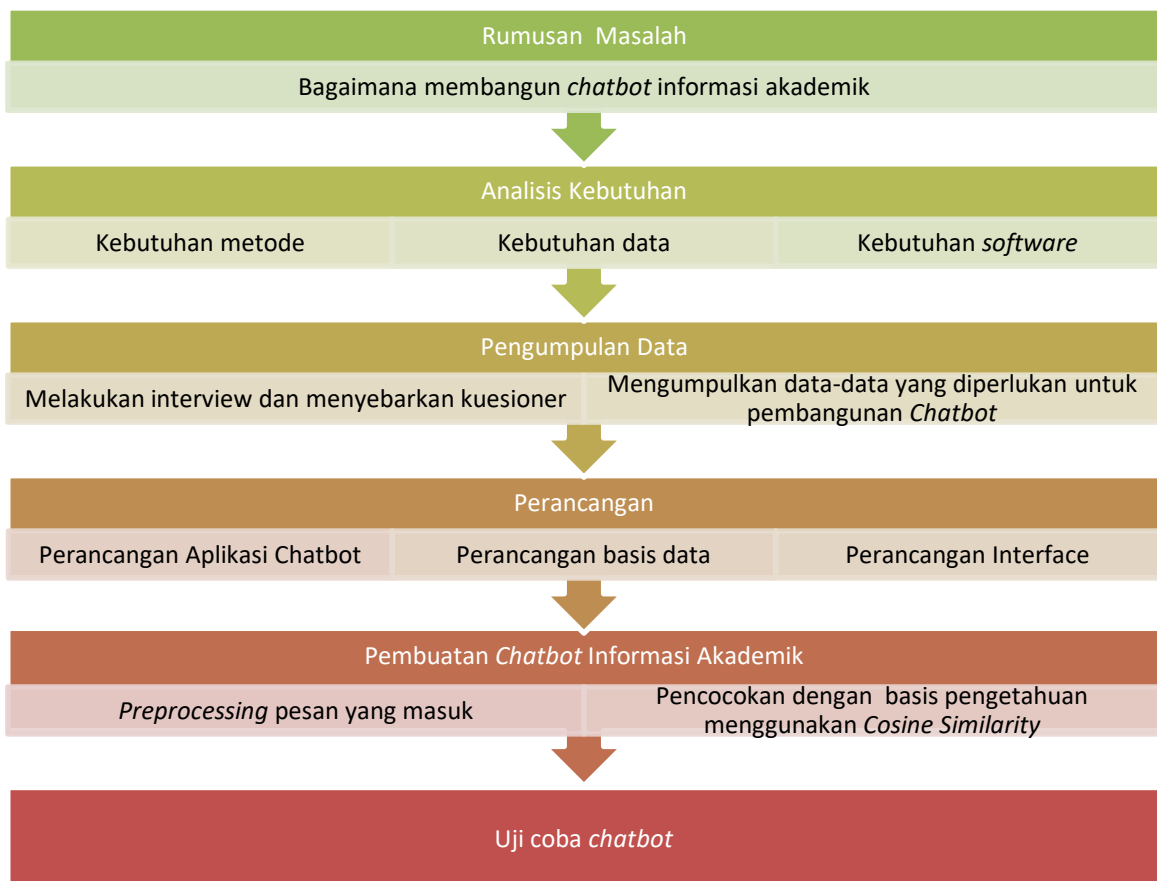
Gambar 1 menunjukkan garis besar tahapan pelaksanaan penelitian ini, yang dimulai dengan pembuatan *chatbot* informasi akademik dan diikuti dengan pelaksanaan klasifikasi berbagai pertanyaan yang diajukan mahasiswa Teknik Informatika ITPLN ke *chatbot*. Tahapan penelitian

diawali dengan pembuatan aplikasi *chatbot* informasi akademik, dengan langkah-langkah pembangunan aplikasi ditunjukkan pada gb.2.



Gambar 1. Tahapan Penelitian *Chatbot* Informasi Akademik

Aplikasi *chatbot* Informasi Akademik yang dibangun diimplementasikan menggunakan aplikasi chat Telegram. Pada awal penelitian, pengumpulan data dilakukan dengan mengadakan wawancara dengan pihak Prodi dan menyebarkan kuesioner kepada mahasiswa sebagai instrumen kebutuhan awal penelitian.



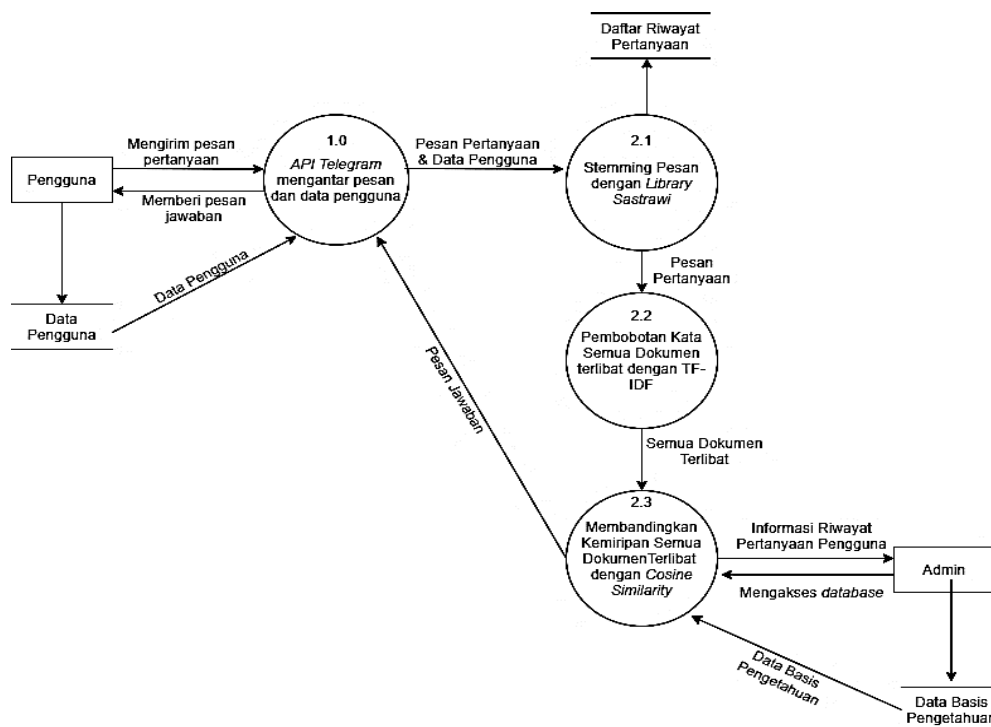
Gambar 2. Tahapan Penelitian *Chatbot* Informasi Akademik

Data yang digunakan dalam pembangunan *chatbot* merupakan data yang didapatkan dari catatan internal Prodi S1 Teknik Informatika IT PLN. Data ini meliputi data dosen, termasuk nama dosen, NIDN (Nomor Induk Dosen Nasional), NIP (Nomor Induk Pegawai), kalender akademik, *link* surat menyurat, pelaksanaan remedial, jadwal kegiatan, pertanyaan-pertanyaan yang pernah diajukan ke prodi, dan lain sebagainya. Jumlah pertanyaan yang digunakan sebagai basis pengetahuan ada 63

pertanyaan. Sedangkan pada penyebaran kuesioner, terdapat beberapa pertanyaan yang diberikan dan responden memberikan tanggapan dengan jawaban: Sangat Tidak Setuju (STS), Tidak Setuju (TS), Kurang Setuju (KS), Setuju (S), atau Sangat Setuju (SS).

2.2. Perancangan Aplikasi Chatbot Informasi Akademik

Perancangan *chatbot* informasi akademik ini dijelaskan dengan menggunakan diagram DFD (*Data Flow Diagram*) level 0, level 1, dan 2. DFD level 0 menggambarkan diagram konteks, yang menggambarkan sistem secara umum. Gambar 3 merupakan diagram rinci ke dalam DFD level 2. Diagram tersebut memecah proses yang ada pada proses 2.0 menjadi beberapa bagian. Pemecahan proses dimulai dari memasukkan pertanyaan pengguna melalui *API Telegram* ke sistem *chatbot*, kemudian terjadi proses *stemming* pesan dengan *Library Sastrawi Stemmer* (proses 2.1). Setelah itu, pertanyaan tersebut diproses dengan semua data yang ada pada sistem basis pengetahuan untuk selanjutnya diberi pembobotan melalui teknik TF-IDF (proses 2.2). Kemudian setelah pembobotan semua dokumen atau data yang terlibat akan dibandingkan menggunakan *Cosine Similarity* (proses 2.3). Setelah semua proses berjalan, maka sistem akan memberikan pesan balasan sesuai dengan pertanyaan yang diajukan oleh pengguna.



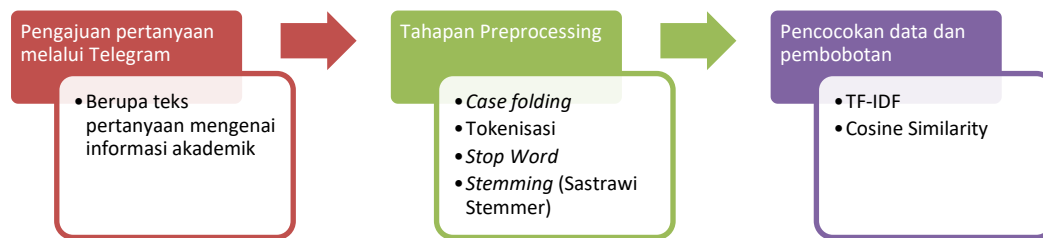
Gambar 3. Diagram Rinci 2.0 level 2

2.3. Tahapan Pemrosesan Teks pada Chatbot Telegram Informasi Akademik

Tahapan pemrosesan teks pada sistem *chatbot* ini dimulai dari penerimaan pesan melalui *instant messaging* (Telegram), kemudian diteruskan ke tahapan *preprocessing* untuk mengubah pesan menjadi data yang sesuai dengan format pemrosesan yang diperlukan. Tahapan selanjutnya adalah proses pencocokan pola dengan kata dasar yang ada di dalam basis pengetahuan (*knowledge base*) menggunakan metode *Cosine Similarity*. Tahapan ini diilustrasikan pada gambar 4. Pada tahapan *Preprocessing*, pesan yang masuk akan diubah menjadi kata dasar melalui proses *case folding*, *tokenisasi*, *stop word*, dan *stemming* dengan bantuan *Library Sastrawi*. Sastrawi

merupakan pustaka PHP yang digunakan untuk mereduksi kata-kata infleksi Bahasa Indonesia ke bentuk kata dasar [11]. Adapun penjelasan dari tiap proses tersebut adalah sebagai berikut:

- 1) *Case Folding*, untuk mengubah semua huruf dalam teks menjadi huruf kecil dan menghapus karakter yang tidak diperlukan.
- 2) *Tokenisasi*, yaitu proses memecah aliran teks menjadi kata, frasa, simbol, atau elemen bermakna lainnya yang disebut token atau *term*.
- 3) *Stop Word*, yaitu tahapan menghapus kata berulang pada dokumen yang tidak memiliki makna karena berfungsi sebagai penghubung kata-kata pada kalimat.
- 4) *Stemming*, yaitu proses menggabungkan beberapa *term* menjadi suku kata.



Gambar 4. Bagan Alir *Text Processing*

Adapun tabel 1 menunjukkan contoh perubahan pada kalimat pertanyaan yang diberikan ke *chatbot* setelah melalui tahapan-tahapan pemrosesan teks. Setelah tahapan dilanjutkan dengan perhitungan TF/IDF (*Term Frequency- Inverse Document Frequency*). Proses *Stop Word* pada tabel 1 menghapus kata berulang dan kata yang tidak memiliki makna serta angka untuk kemudian dideklarasikan pada program. Kata yang akan dihilangkan saat tahapan *Stop Word*, antara lain: 'yang', 'untuk', 'pada', 'ke', 'para', 'dsb', 'dst', 'dll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'mari', 'nant', 'ok', '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '!', '@', '#', '\$', '%'.

Tabel 1. Contoh Hasil *Pre-Processing* Pesan

Proses	Pertanyaan 1	Pertanyaan 2
Tanpa <i>Pre-Processing</i> pesan	Kapan finalisasi dan validasi krs semester genap 2022/2023 oleh Dosen Pembimbing?	Membuat surat aktif kuliah bagaimana caranya?
<i>Case Folding</i>	kapan finalisasi dan validasi krs semester genap 2022/2023 oleh dosen pembimbing?	membuat surat aktif kuliah bagaimana caranya?
<i>Tokenisasi</i>	kapan , finalisasi , dan , validasi , krs , oleh dosen , pembimbing	membuat , surat , aktif , kuliah , bagaimana caranya
<i>Stop Word</i>	finalisasi , validasi , krs , semester , genap, dosen , pembimbing	surat , aktif , kuliah
<i>Stemming</i>	final validasi krs semester genap dosen bimbing	surat aktif kuliah

Library Sastrawi, metode *TF-IDF*, dan metode *Cosine Similarity* diterapkan dalam menjawab pertanyaan seputar informasi akademik yang ditanyakan oleh pengguna *Chatbot Telegram* Informasi

Akademik Program Studi S1 Teknik Informatika ITPLN. Dimisalkan pertanyaan yang diajukan adalah “Saya Mau Membuat Surat Riset?” (Q). Maka penerapan metode direalisasikan melalui beberapa tahap sebagai berikut:

1. *Processing* Pesan

Pertama menentukan nilai N yaitu jumlah semua dokumen yang terlibat, termasuk dengan dengan pertanyaan (Q), “Saya Mau Membuat Surat Riset?”. Kemudian dilakukan beberapa tahapan *preprocessing* pesan, yaitu: *Case Folding*, *Tokenisasi*, *Stop Word*, dan *Stemming*.

2. Pembobotan setiap *term* dari total N dokumen, dengan menerapkan metode *Term Frequency – Inverse Document* atau TF-IDF.

3. Menghitung kemiripan dokumen antara pertanyaan yang diajukan (Q) dengan data pada basis pengetahuan (D) menggunakan rumus metode *cosine Similarity*.

Dengan pertanyaan yang masuk adalah “Saya Mau Membuat Surat Riset?”, kemudian dinormalisasikan menjadi “saya mau buat surat riset “. Pertanyaan yang mirip pada dokumen basis pengetahuan adalah “Membuat surat riset “. Maka jawaban yang diberikan (oleh *chatbot*) adalah “Isi tautan berikut untuk membuat surat aktif kuliah <https://bit.ly/SuratAktifKuliahInformatika> pembuatan surat diproses dalam 3 hari setelah pengisian data, kemudian surat dapat diunduh pada tautan berikut <http://bit.ly/SuratMahasiswaInfor> “.

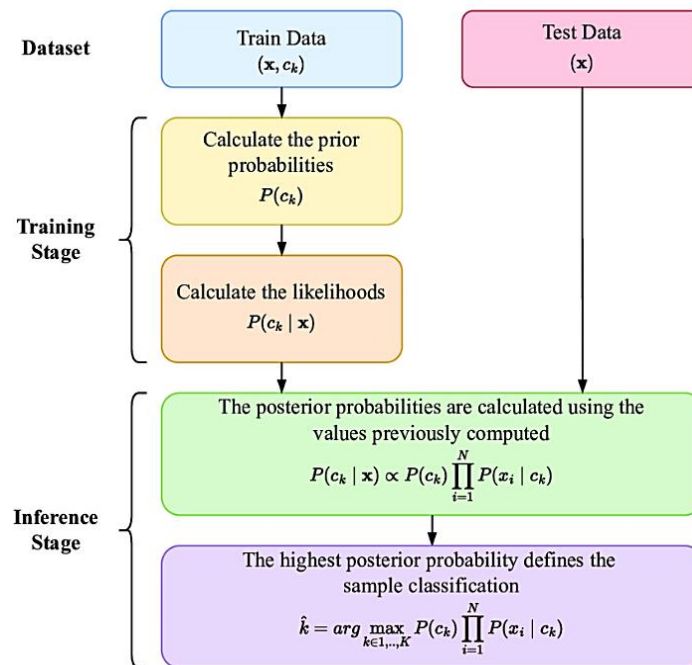
2.4. Penerapan Klasifikasi

Metode Naïve Bayes digunakan untuk mengklasifikasikan pertanyaan-pertanyaan yang masuk ke *chatbot* pada tahapan ujicoba. Pertanyaan yang masuk akan dibagi menjadi 3 kelas, yaitu kelas pertanyaan mengenai informasi dosen (kelas 1), permintaan pembuatan surat atau dokumen tertentu (kelas 2), dan kelas pertanyaan mengenai jadwal akademik (kelas 3). Tabel 2 menunjukkan beberapa contoh pertanyaan dan label kelas dari pertanyaan tersebut.

Tabel 2. Contoh Pertanyaan dan Kelas

No.	Pertanyaan	Kelas
1	Berapa NIDN bu Rosida?	1
2	Pembayara SKS mulai tanggal berapa?	3
3	Kapan wisuda Angkatan 40?	3
4	Surat diterima magang bisa didapatkan di mana?	2
5	Cara mengisi form remedial?	2
6	Bagaimana cara membuat surat riset?	2
7	Kapan pembayaran bpp genap?	3
8	KHS semester ganjil bisa dilihat kapan?	3
9	Pengumpulan revisi skripsi gelombang?	3
10	Berapa lama surat diproses?	2

Tahapan penerapan metode Naïve Bayes dapat dilihat pada gb. 5. Data pertanyaan informasi akademik yang digunakan untuk klasifikasi adalah 685 pertanyaan, dengan pembagian 80% untuk data latih (*train data*) dan 20% untuk data uji (*test data*). Implementasi klasifikasi menggunakan bahasa pemrograman Python menggunakan Naïve Bayes Multinomial.

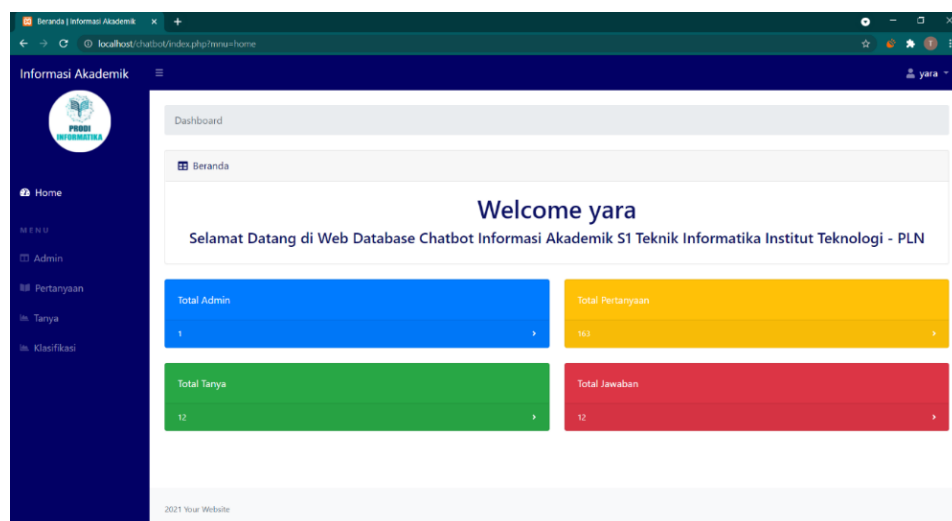


Gambar 5. Algoritma Metode Naïve Bayes

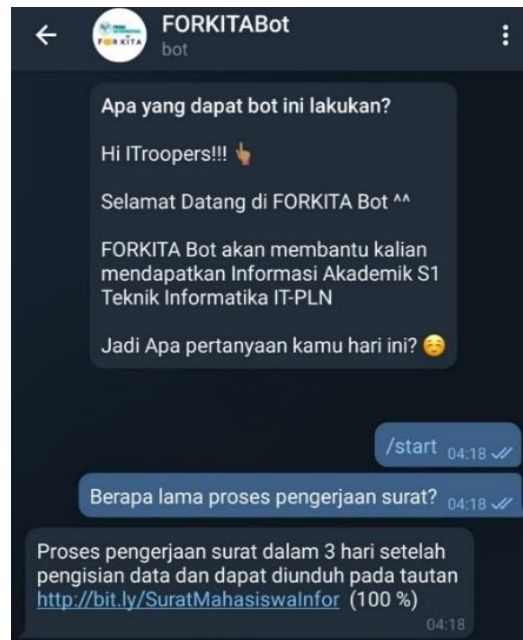
3. HASIL DAN PEMBAHASAN

3.1. Tampilan Aplikasi Chatbot

Aplikasi web yang dibangun ini ditujukan untuk admin untuk pengelolaan *chatbot* Telegram Informasi Akademik di Prodi S1 Teknik Informatika IT PLN. Gambar 6 merupakan tampilan halaman beranda untuk *Web Database Chatbot* yang menampilkan total admin yang terdaftar, total pertanyaan dalam basis pengetahuan, total tanya dari pengguna, dan total jawaban yang dijawab oleh sistem. Gambar 7 adalah tampilan saat aplikasi *chatbot* informasi akademik, yang diberi nama *FORKITA*, dipanggil melalui aplikasi *chat* Telegram. Pada gambar tersebut terlihat tampilan dari aplikasi *chatbot* ketika diberi pertanyaan, yang menanyakan mengenai waktu pengerjaan surat. Berdasarkan perhitungan kesamaan menggunakan *Cosine Similarity*, maka diperoleh jawaban yang sesuai dengan dokumen basis pengetahuan dengan nilai akurasi kemiripan sebesar 100%.



Gambar 6. Tampilan Halaman Beranda (Home)



Gambar 7. Tampilan Halaman Input Data Pertanyaan

Pengujian akurasi awal untuk aplikasi *Chatbot Telegram* Informasi Akademi S1 Teknik Informatika IT-PLN dilakukan dengan metode *Confusion Matrix*, yaitu dengan menghitung nilai *true positive*, *true negative*, *false positive*, dan *false negative* dari pertanyaan-pertanyaan yang diberikan sebagai data uji. Untuk pengujian tersebut, digunakan 10 pertanyaan yang umum diajukan oleh mahasiswa, yaitu sebagai berikut:

1. Q1 = Berapa lama proses pengerjaan surat?
2. Q2 = Saya Mau Membuat Surat Riset?
3. Q3 = Info pelaksanaan sidang magang
4. Q4 = Kapan pengumpulan proposal skripsi
5. Q5 = Info mengenai penentuan dosen pembimbing skripsi
6. Q6 = Dimana mendapatkan surat diterima magang?
7. Q7 = Minta *form* remedial?
8. Q8 = Kapan Pelaksanaan remedial
9. Q9 = Membuat surat keterangan aktif kuliah
10. Q10 = Berapa NIP dosen bu dewi arianti?

TahapanSelanjutnya adalah menghitung rata-rata akurasi kemiripan pertanyaan Q1 sampai dengan Q10 dengan dokumen pertanyaan pada basis pengetahuan menggunakan metode *Cosine Similarity* ditunjukkan pada tabel di bawah ini.

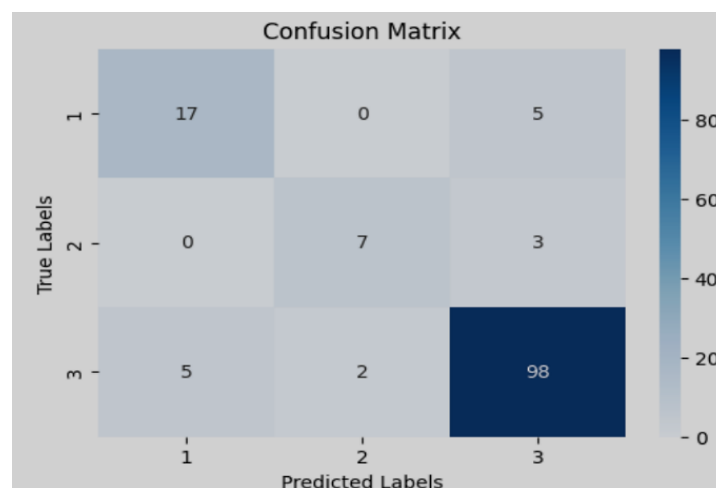
Tabel 3. Rata-Rata Akurasi Kemiripan Dokumen dari Data Uji Pertanyaan

Data Uji Ke-	Akurasi
Q1	100%
Q2	79,32%
Q3	63,09%
Q4	86,62%
Q5	49,96%

Q6	100%
Q7	73,38%
Q8	100%
Q9	81,28%
Q10	90,59%
Rata-Rata	82,424%

Dari 10 kali uji atau 10 pertanyaan data uji didapatkan hasil akurasi rata-rata *Cosine Similarity* sebesar 82,424%. Dari hasil ini dapat disimpulkan bahwa nilai akurasi termasuk nilai yang cukup baik. Ada beberapa penyebab kenapa terdapat akurasi yang bernilai di bawah 80% seperti adanya pertanyaan yang tidak lengkap keterangannya sehingga terdapat dua dokumen yang menjadi prediksi jawaban namun ditampilkan dokumen dengan nilai yang paling mendekati, contohnya dengan tidak menyebutkan keterangan gelombang atau semester. Penyebab lainnya adalah adanya pertanyaan baru yang belum terdapat pada basis pengetahuan dan terdapat beberapa kata yang tidak termasuk dalam proses *stemming*.

Pengujian yang kedua dilakukan dengan memberikan berbagai pertanyaan seputar informasi akademik ke aplikasi *chatbot* selama beberapa 1 bulan. Pertanyaan yang masuk beragam, dari pertanyaan mengenai NIDN dosen sampai permintaan pembuatan surat menyurat, dan berjumlah kurang lebih 450. Dari total 456 pertanyaan yang masuk, diperoleh 396 jawaban benar dan 60 jawaban yang salah. Jadi berdasarkan uji coba secara langsung, diperoleh tingkat kesalahan sebesar 19,74 % atau tingkat akurasi sebesar 80,26%.



Gambar 8. *Confusion Matrix* untuk Klasifikasi Pertanyaan Menggunakan Naïve Bayes

Sedangkan pengujian ketiga, yaitu menguji akurasi dari pelaksanaan klasifikasi dari pertanyaan-pertanyaan yang diajukan. Dengan menggunakan total 685 pertanyaan dan perbandingan sebesar 4:1 untuk data latih dan data uji, diperoleh hasil *confusion matrix* seperti pada gb.8. Adapun nilai metrik akurasi, presisi, *recall*, dan F1 score adalah sebagai berikut:

$$\text{Accuracy} = \text{Total True Positive (TP)} / \text{Total Data} = (17 + 7 + 98) / 137 = 0,89$$

$$\text{Precision Kelas 1} = \text{True Positive} / (\text{True Positive} + \text{False Positive})$$

$$= 17 / (17 + 0 + 5) = 0,77$$

$$\text{Precision Kelas 2} = 7 / (7 + 0 + 3) = 0,7$$

$$\begin{aligned}
 \text{Precision Kelas 3} &= 98/(5+2+98) = 0,933 \\
 \text{Precision (All)} &= ((17/22) + (7/10) + (98/105))/3 = 0,803 \\
 \text{Recall Kelas 1} &= 17/(17+0+5) = 0,77 \\
 \text{Recall Kelas 2} &= 7/(7+0+2) = 0,778 \\
 \text{Recall Kelas 3} &= 98/(98+3+5) = 0,925 \\
 \text{Recall (All)} &= ((17/22) + (7/9) + (98/106))/3 = 0,825 \\
 \text{F1 score} &= 2 * 0,803 * 0,825 / (0,803 + 0,825) = 0,813
 \end{aligned}$$

4. KESIMPULAN DAN SARAN

Dari serangkaian proses perencanaan pembuatan, penerapan metode *Cosine Similarity* dengan *Library Sastrawi Stemmer* hingga pengujian akurasi pada penelitian ini dapat diambil beberapa kesimpulan yaitu:

1. Dalam pembuatan *Chatbot Telegram* Informasi Akademik program studi S1 Teknik Informatika Institut Teknologi PLN penulis melalui beberapa proses mulai dari pembuatan *Data Flow Diagram* (DFD), perancangan serta pembuatan basis data, perancangan tampilan tatap muka, pelatihan data latih pada database, pembuatan *bot*, dan penghubungan *web database* dengan *bot*.
2. Penerapan dari metode *Cosine Similarity* dengan *Library Sastrawi Stemmer* pada *Chatbot Telegram* memiliki hasil akurasi ketepatan yang cukup tinggi yaitu sebesar 90% dengan rata-rata akurasi kemiripan dokumen pada setiap data uji sebesar 82,424%. Hal ini dipengaruhi oleh penulisan pertanyaan oleh pengguna dan isi dari basis pengetahuan.
3. Dari pemberian 456 pertanyaan yang beragam pada aplikasi *chatbot* diperoleh jawaban yang benar atau sesuai dengan pertanyaan sebanyak 396 (80,26%).
4. Dari proses klasifikasi pertanyaan mengenai informasi akademik yang masuk ke *chatbot* menggunakan metode *Naïve Bayes Multinomial* diperoleh nilai akurasi 89%, presisi 80,3%, *recall* sebesar 82,5%, dan *F1 score* sebesar 81,3%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada LPPM Institut Teknologi PLN yang telah memberi dukungan dana hibah untuk membantu pelaksanaan penelitian dan penulisan artikel.

DAFTAR PUSTAKA

- [1] A. Heryandi, "Developing Chatbot for Academic Record Monitoring in Higher Education Institution," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 879, no. 1, 2020, doi: 10.1088/1757-899X/879/1/012049.
- [2] E. Adamopoulou and L. Moussiades, *An Overview of Chatbot Technology*, vol. 584 IFIP, no. June. Springer International Publishing, 2020.
- [3] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100033, 2021, doi: 10.1016/j.caeai.2021.100033.
- [4] S. Garcia-Mendez, F. De Arriba-Perez, F. J. Gonzalez-Castano, J. A. Regueiro-Janeiro, and F. Gil-Castineira, "Entertainment Chatbot for the Digital Inclusion of Elderly People without Abstraction Capabilities," *IEEE Access*, vol. 9, pp. 75878–75891, 2021, doi: 10.1109/ACCESS.2021.3080837.
- [5] N. Bhirud, S. Tataale, S. Randive, and S. Nahar, "A Literature Review On Chatbots In Healthcare Domain," *Int. J. Sci. Technol. Res.*, vol. 8, no. 7, pp. 225–231, 2019.
- [6] T. Makasi, A. Nili, K. C. Desouza, and M. Tate, "A Typology of Chatbots in Public Service

- Delivery,” *IEEE Softw.*, vol. 39, no. 3, pp. 58–66, 2022, doi: 10.1109/MS.2021.3073674.
- [7] T. Nadu, “Education,” no. Icicv, pp. 1317–1322, 2021.
- [8] J. R. Carrizales, Y. J. Ramirez, J. A. Armas, and E. E. Grandón, “Cognitive services to improve user experience in searching for academic information based on chatbot,” pp. 1–4, 2019.
- [9] M. Kowsher, M. A. Alam, M. J. Uddin, M. R. Islam, N. Pias, and A. R. M. Saifullah, “Bengali Informative Chatbot,” *5th Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2019*, pp. 1–7, 2019, doi: 10.1109/IC4ME247184.2019.9036585.
- [10] R. Rismanto, Y. Yunhasnawa, and R. A. Bhakti, “Penerapan Metode Cosine Similarity Dalam Aplikasi Chatbot Layanan Wisata Di Wilayah Malang,” *Semin. Inform. Apl. Polinema*, pp. 1–8, 2019.
- [11] I. K. P. Pinajeng, I. M. Sukarsa, and I. M. S. Putra, “Perbaikan Kata pada Sistem Chatbot dengan Metode Jaro Winkler,” *JITTER J. Ilm. Teknol. dan Komput.*, vol. 1, no. 2, pp. 86–95, 2020.
- [12] T. Seidakhmetov, “Question Type Classification Methods Comparison,”
- [13] M. Mohammed, N. Omar N, “Question Classification Based on Bloom’s Taxonomy Cognitive Domain using Modified TF-IDF and word2vec,” *PLoS ONE*, vol. 15, no.3, e0230442, 2020. <https://doi.org/10.1371/journal.pone.0230442>.
- [14] A. Mohasseb, M. Bader-El-Den, M. Cocea, “Question Categorization and Classification Using Grammar-based Approach,” *Information Processing and Management*, vol.54, no.6, pp.1228-1243, 2018.
- [15] W. Zhang, F. Gao, “An Improvement to Naïve Bayes for Text Classification,” *Procedia Engineering*, vol. 15, pp. 2160-2164, 2011. [Online] An Improvement to Naive Bayes for Text Classification - ScienceDirect
- [16] M. M. Temesgen, D. T. Lemma, “A Scalable Text Classification Using Naïve Bayes with Hadoop Framework,” *Proceeding of Information and Communication Technology for Development for Africa (ICT4DA) 209*, pp. 291-300, Communications in Computer and Information Science, vol 1026. Springer, Cham. https://doi.org/10.1007/978-3-030-26630-1_25
- [17] A. C. Herlingga, I. P. E. Prismana, D. R. Prehanto, and D. A. Dermawan, “Algoritma Stemming Nazief & Adriani dengan Metode Cosine Similarity untuk Chatbot Telegram Terintegrasi dengan E-layanan,” *J. Informatics Comput. Sci.*, vol. 2, no. 01, pp. 19–26, 2020, doi: 10.26740/jinacs.v2n01.p19-26.