

Optimalisasi Pelayanan Akademik Kemahasiswaan Menggunakan Chatbot Berbasis Nlp Dengan Algoritma Naive Bayes

Calvin Putra Samaeri T^a, Wahyu Joni Kurniawan^b

^a Institut Bisnis dan Teknologi Pelita Indonesia, Jl.A.Yani, calvinputrasamaeri@gmail.com

^b Institut Bisnis dan Teknologi Pelita Indonesia, Jl.A.Yani, wahyu.jonikurniawan@lecturer.pelitaindonesia.ac.id

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 18 April 2026

Revisi Akhir: 30 Maret 2026

Diterbitkan Online: 30 April 2026

KATA KUNCI

Chatbot

Pelayanan Akademik

Natural Language Processing

Naive Bayes

Text Classification

KORESPONDENSI

E-mail: calvinputrasamaeri@gmail.com

A B S T R A K

Perkembangan teknologi informasi mendorong perguruan tinggi untuk meningkatkan kualitas layanan akademik kemahasiswaan agar lebih cepat, responsif, dan mudah diakses oleh pengguna. Proses penyampaian informasi akademik yang masih dilakukan secara manual sering menimbulkan keterlambatan informasi serta ketidakkonsistenan jawaban, sehingga diperlukan solusi berbasis teknologi yang mampu memberikan layanan informasi secara otomatis. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem chatbot pelayanan akademik kemahasiswaan berbasis *Natural Language Processing* (NLP) dengan algoritma *Naive Bayes*. Metode pengembangan sistem menggunakan pendekatan Agile yang meliputi tahap analisis kebutuhan, perancangan sistem, pengembangan, pengujian, serta evaluasi. Proses pengolahan data dilakukan melalui tahapan pra-pemrosesan teks yang mencakup *tokenizing*, *stopword removal*, dan *stemming*, kemudian dilakukan proses klasifikasi pertanyaan menggunakan algoritma *Naive Bayes* untuk menentukan intent dan menghasilkan jawaban yang sesuai. Hasil penelitian menunjukkan bahwa sistem chatbot berhasil dikembangkan sebagai aplikasi berbasis web yang mampu memberikan layanan informasi akademik secara otomatis dan responsif. Berdasarkan hasil pengujian model, algoritma *Naive Bayes* menghasilkan nilai akurasi sebesar 80,18%, yang menunjukkan bahwa sistem mampu mengklasifikasikan pertanyaan akademik dengan tingkat ketepatan yang baik. Dengan demikian, sistem chatbot yang dikembangkan dapat menjadi solusi dalam optimalisasi pelayanan akademik kemahasiswaan serta mendukung transformasi digital layanan informasi di lingkungan perguruan tinggi.

1. PENDAHULUAN

Perkembangan transformasi digital yang semakin pesat mendorong perguruan tinggi untuk meningkatkan kualitas layanan akademik melalui pemanfaatan teknologi informasi. Layanan informasi akademik yang cepat, akurat, dan mudah diakses menjadi kebutuhan penting bagi mahasiswa, terutama dalam memperoleh informasi terkait pembayaran kuliah, beasiswa, pengisian Kartu Rencana Studi (KRS), serta penerbitan Kartu Hasil Studi (KHS) [1]. Namun, penyampaian informasi akademik yang masih dilakukan secara manual atau terbatas pada jam layanan tertentu sering menimbulkan keterlambatan informasi, kurangnya keakuratan jawaban, serta menurunkan efektivitas pelayanan akademik.

Untuk mengatasi permasalahan tersebut, diperlukan solusi berbasis teknologi yang mampu memberikan layanan informasi secara otomatis, cepat, dan dapat diakses kapan saja. Salah satu solusi yang dapat diterapkan adalah pengembangan chatbot layanan informasi akademik berbasis web. Chatbot memungkinkan pengguna memperoleh informasi secara

otomatis, interaktif, dan real-time tanpa bergantung pada petugas layanan, sehingga mampu meningkatkan efisiensi serta kualitas pelayanan akademik di perguruan tinggi.

Dalam pengembangannya, chatbot memerlukan teknologi yang mampu memahami bahasa alami yang digunakan pengguna. *Natural Language Processing* (NLP) merupakan teknologi yang berperan penting karena memungkinkan sistem komputer memahami, menganalisis, dan memproses bahasa alami dalam bentuk teks [2]. Melalui tahapan pemrosesan teks seperti normalisasi, tokenisasi, dan pembobotan kata, NLP membantu sistem dalam memahami maksud pertanyaan yang diajukan sehingga chatbot dapat memberikan respons yang lebih relevan dan sesuai dengan kebutuhan pengguna [3].

Selanjutnya, untuk menentukan kategori atau maksud pertanyaan pengguna, digunakan algoritma *Naive Bayes* sebagai metode klasifikasi berdasarkan data teks yang telah diproses menggunakan NLP. Algoritma ini dipilih karena memiliki proses komputasi yang sederhana, efisien dalam pengolahan data teks, serta mampu menghasilkan tingkat akurasi yang baik dalam proses klasifikasi (Herman et al., 2025).

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan sistem chatbot pelayanan akademik kemahasiswaan berbasis web menggunakan pendekatan Natural Language Processing dan algoritma Naive Bayes, sehingga diharapkan mampu meningkatkan kecepatan, ketepatan, dan aksesibilitas layanan informasi akademik bagi mahasiswa.

2. TINJAUAN PUSTAKA

2.1. Optimalisasi Dan Layanan Akademik

Optimalisasi dalam sistem informasi merupakan proses untuk memperoleh hasil layanan yang lebih efektif dan efisien melalui pemanfaatan sumber daya secara maksimal. Dalam konteks layanan akademik, optimalisasi bertujuan mentransformasi proses pelayanan konvensional yang terbatas oleh waktu dan prosedur manual menjadi layanan yang cepat, akurat, dan dapat diakses kapan saja [5]. Kualitas optimalisasi layanan dapat diukur melalui dimensi keandalan dan daya tanggap, yaitu kemampuan sistem dalam memberikan respons secara cepat serta menyediakan informasi yang konsisten dan akurat.

Pada praktiknya, layanan akademik konvensional masih menghadapi berbagai kendala seperti keterlambatan penyampaian informasi, antrean layanan yang panjang, serta proses administrasi yang masih dilakukan secara manual sehingga berpotensi menurunkan kepuasan pengguna [6]. Oleh karena itu, pemanfaatan teknologi digital menjadi solusi strategis dalam meningkatkan efisiensi dan kualitas layanan akademik melalui otomatisasi dan integrasi sistem informasi. Implementasi sistem layanan berbasis teknologi memungkinkan proses administrasi akademik, seperti pengisian KRS, akses KHS, dan penyampaian informasi akademik, dapat dilakukan secara lebih cepat, akurat, dan fleksibel.

2.2. Pelayanan Akademik Dan Kemahasiswaan

Pelayanan akademik merupakan elemen penting dalam penyelenggaraan administrasi perguruan tinggi karena mendukung kelancaran berbagai aktivitas akademik mahasiswa, seperti registrasi, pengisian Kartu Rencana Studi (KRS), pengelolaan Kartu Hasil Studi (KHS), pembayaran kuliah, serta layanan administrasi akademik lainnya. Kualitas pelayanan akademik memiliki pengaruh langsung terhadap kepuasan dan keberhasilan studi mahasiswa. Namun, dalam praktiknya, sebagian layanan akademik masih bergantung pada interaksi tatap muka atau saluran komunikasi terbatas, sehingga sering menimbulkan antrean panjang, keterlambatan respons, serta keterbatasan waktu layanan [7].

Selain itu, pelayanan akademik juga menghadapi tantangan dalam hal kecepatan respons, keakuratan informasi, dan aksesibilitas layanan. Keterlambatan penyampaian informasi atau ketidakakuratan data akademik dapat menimbulkan kebingungan mahasiswa serta menghambat kelancaran proses akademik [8]. Oleh karena itu, diperlukan inovasi layanan berbasis teknologi digital yang mampu menyediakan informasi akademik secara cepat, akurat, dan dapat diakses kapan saja, sehingga mampu memenuhi kebutuhan dan ekspektasi mahasiswa yang semakin meningkat di era digital.

2.3. Teknologi Chatbot

Teknologi chatbot merupakan salah satu penerapan kecerdasan buatan (Artificial Intelligence/AI) yang dirancang untuk meniru interaksi percakapan manusia melalui media teks maupun suara. Chatbot berfungsi sebagai agen virtual yang mampu memberikan respons otomatis terhadap pertanyaan pengguna secara real-time dan beroperasi selama 24 jam, sehingga efektif dalam menangani pertanyaan yang bersifat rutin dan berulang [9].

Berdasarkan mekanisme kerjanya, chatbot dibedakan menjadi chatbot berbasis aturan (rule-based) dan chatbot berbasis kecerdasan buatan (AI-based). Chatbot rule-based menghasilkan respons berdasarkan pola yang telah ditentukan, sedangkan chatbot AI-based memanfaatkan teknologi Natural Language Processing (NLP) dan Machine Learning untuk memahami bahasa alami pengguna secara lebih fleksibel dan adaptif. Melalui proses pemrosesan teks seperti normalisasi, tokenisasi, dan klasifikasi teks, sistem dapat mengenali maksud pertanyaan pengguna sehingga mampu memberikan respons yang lebih relevan dan akurat.

Penerapan chatbot dalam layanan akademik memberikan berbagai manfaat, antara lain meningkatkan kecepatan penyampaian informasi, mempermudah akses informasi secara real-time, mengurangi beban kerja staf administrasi, serta menjaga konsistensi informasi yang diberikan. Dengan kemampuan melayani banyak pengguna secara bersamaan, chatbot menjadi solusi yang efektif dalam mendukung optimalisasi pelayanan akademik di perguruan tinggi.

2.4. Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang berfokus pada pengembangan sistem komputer agar mampu belajar dari data dan meningkatkan kinerjanya tanpa pemrograman secara eksplisit. Melalui Machine Learning, sistem dapat mengenali pola, melakukan prediksi, serta mengklasifikasikan data berdasarkan informasi yang telah dipelajari sebelumnya [10]. Dalam pengolahan data teks, Machine Learning banyak dimanfaatkan untuk tugas klasifikasi, seperti pengelompokan dokumen, analisis sentimen, serta identifikasi maksud (intent) dari pertanyaan pengguna.

Secara umum, Machine Learning terdiri dari beberapa pendekatan, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Penelitian ini menggunakan pendekatan supervised learning karena memanfaatkan data latih berlabel berupa pasangan pertanyaan mahasiswa dan kategori informasi akademik yang sesuai. Pendekatan ini memungkinkan sistem chatbot mengklasifikasikan pertanyaan pengguna secara lebih terarah dan akurat.

Salah satu algoritma supervised learning yang banyak digunakan dalam klasifikasi teks adalah Naive Bayes, karena memiliki proses komputasi yang sederhana, efisien, serta mampu memberikan performa yang baik pada pengolahan data berbasis teks seperti percakapan chatbot.

2.5. Natural Language Processing (NLP)

Natural Language Processing (NLP) merupakan cabang kecerdasan buatan yang berfokus pada kemampuan komputer untuk memahami, memproses, dan menghasilkan bahasa manusia

dalam bentuk teks maupun suara. Dalam sistem chatbot, NLP berperan penting dalam menganalisis pertanyaan pengguna, memahami maksud percakapan, serta menghasilkan respons yang relevan dan sesuai konteks, sehingga interaksi antara manusia dan sistem menjadi lebih alami dan efektif [11].

Penerapan NLP memungkinkan sistem memproses bahasa manusia yang beragam melalui beberapa tahapan pemrosesan teks, seperti tokenisasi, stemming, dan entity extraction. Tokenisasi berfungsi memecah teks menjadi unit kata, stemming digunakan untuk mengubah kata menjadi bentuk dasarnya, sedangkan entity extraction bertujuan mengidentifikasi entitas penting dalam teks sesuai konteks percakapan. Melalui tahapan tersebut, chatbot mampu memahami maksud pertanyaan pengguna secara lebih akurat dan memberikan informasi yang relevan, sehingga mendukung peningkatan kualitas layanan informasi akademik berbasis digital [3].

2.6. Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu metode klasifikasi yang termasuk dalam kategori supervised learning pada Machine Learning dan didasarkan pada Teorema Bayes untuk menghitung probabilitas suatu kelas berdasarkan data yang diamati. Algoritma ini disebut *naive* karena mengasumsikan bahwa setiap fitur bersifat independen satu sama lain, namun tetap memiliki kinerja yang baik dan efisien dalam berbagai permasalahan klasifikasi, terutama pada pengolahan data teks [12].

Dalam sistem chatbot, Naive Bayes digunakan untuk menentukan kategori atau maksud (intent) dari pertanyaan pengguna berdasarkan kata-kata yang terdapat pada teks input. Proses klasifikasi dilakukan dengan menghitung probabilitas posterior setiap kelas menggunakan probabilitas awal kelas dan probabilitas kemunculan kata terhadap kelas tersebut, kemudian memilih kelas dengan nilai probabilitas tertinggi sebagai hasil klasifikasi.

Penerapan algoritma Naive Bayes diawali dengan tahap pelatihan menggunakan data latih berupa kumpulan pertanyaan yang telah diberi label kategori, sehingga sistem dapat membangun model probabilitas klasifikasi. Selanjutnya, pada tahap pengujian, pertanyaan pengguna yang telah diproses melalui tahapan Natural Language Processing digunakan sebagai input untuk menentukan kelas pertanyaan secara otomatis. Pemilihan algoritma Naive Bayes dalam penelitian ini didasarkan pada kemudahan implementasi, efisiensi komputasi, serta kemampuannya dalam menangani data teks dengan jumlah fitur yang besar, sehingga cocok diterapkan pada sistem chatbot informasi akademik berbasis web [13].

2.7. Metode Agile

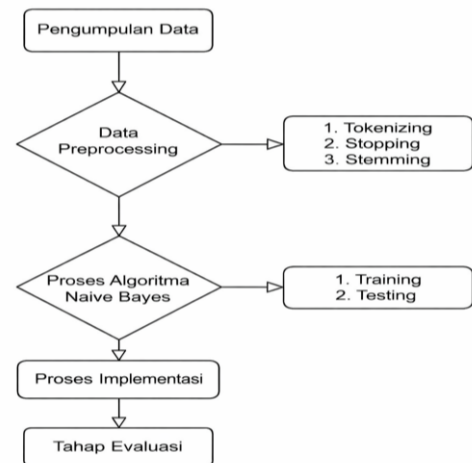
Metode Agile merupakan pendekatan pengembangan perangkat lunak yang bersifat iteratif dan fleksibel, di mana sistem dikembangkan secara bertahap melalui siklus pengembangan singkat (sprint). Setiap siklus menghasilkan fitur yang dapat langsung diuji dan dievaluasi, sehingga memungkinkan proses pengembangan yang lebih adaptif terhadap perubahan kebutuhan pengguna [14].

Pendekatan Agile menekankan kolaborasi tim, evaluasi berkelanjutan, serta perbaikan sistem secara bertahap berdasarkan umpan balik pengguna. Dengan karakteristik tersebut, metode Agile mampu mempercepat proses pengembangan, meningkatkan kualitas sistem yang dihasilkan,

serta memastikan bahwa fitur yang dikembangkan sesuai dengan kebutuhan pengguna.

3. METODOLOGI

Metode penelitian yang digunakan dalam penelitian ini adalah metode Agile, yaitu pendekatan pengembangan perangkat lunak yang bersifat iteratif, adaptif, dan fleksibel. Metode ini dipilih karena sesuai dengan pengembangan sistem chatbot informasi akademik yang memerlukan kemampuan beradaptasi terhadap perubahan kebutuhan serta melibatkan pengguna dalam proses evaluasi sistem.



Gambar 1. Kerangka Penelitian

3.1. Pengumpulan Data

Dilakukan identifikasi terhadap jenis pertanyaan yang sering diajukan mahasiswa terkait layanan akademik kemahasiswaan. Pertanyaan-pertanyaan tersebut kemudian dikumpulkan dan disusun menjadi dataset awal yang merepresentasikan kebutuhan informasi pengguna, yang mencakup berbagai pertanyaan layanan akademik beserta jawaban sesuai dengan ketentuan dan informasi resmi institusi. Penyusunan dataset dilakukan dengan memperhatikan penggunaan bahasa yang umum digunakan oleh mahasiswa, sehingga data yang dihasilkan dapat mendukung proses pelatihan sistem chatbot agar mampu memahami pertanyaan pengguna secara lebih alami dan menghasilkan respons yang sesuai.

Tabel 1 : Contoh Data

No	Pertanyaan	Jawaban
1	Bagaimana cara mengambil surat keterangan aktif kuliah?	Surat keterangan aktif kuliah dapat diajukan dengan cara mengambalnya secara langsung di bagian BAAK di kampus 1, yang beralamat di Jalan Jenderal Ahmad Yani No. 78-88, Pekanbaru.
2	Bagaimana cara lihat KTM?	Kartu Tanda Mahasiswa (KTM) dapat dilihat dengan mengunduh aplikasi PI Mobile di Play Store, kemudian <i>login</i> menggunakan NIM dan <i>password</i>

	SIPOPI. KTM tersebut juga dapat dicetak sesuai kebutuhan
--	--

Tahapan berikutnya dilakukan dengan mengklasifikasikan setiap pertanyaan ke dalam *intent* atau label yang telah ditentukan. Selain itu, ditambahkan pula data percakapan umum guna meningkatkan kemampuan chatbot dalam memberikan respons yang lebih alami dan interaktif. Contoh hasil pengelompokan data tersebut disajikan pada Tabel dibawah.

Tabel 2 : Contoh Data Dengan Label

No	Pertanyaan	Jawaban	Intent
1	Bagaimana cara mengambil surat keterangan aktif kuliah?	Surat keterangan aktif kuliah dapat diajukan dengan cara mengambilnya secara langsung di bagian BAAK di kampus 1, yang beralamat di Jalan Jenderal Ahmad Yani No. 78–88, Pekanbaru.	surat_aktif
2	Bagaimana cara lihat KTM?	Kartu Tanda Mahasiswa (KTM) dapat dilihat dengan mengunduh aplikasi PI Mobile di Play Store, kemudian <i>login</i> menggunakan NIM dan <i>password</i> SIPOPI. KTM tersebut juga dapat dicetak sesuai kebutuhan	ktm_info

Data pada diatas selanjutnya disusun dalam satu berkas *Excel* agar lebih terstruktur dan mudah diolah menggunakan *Google Colaboratory*. Format ini memudahkan proses pengolahan data dan pelatihan model sehingga chatbot dapat merespons pertanyaan dengan lebih akurat. Struktur dataset ditunjukkan pada Tabel dibawah ini.

Tabel 3 : Struktur Dataset

No	Atribut	Keterangan
1	Intent	Label atau Kelas dari kelompok pertanyaan
2	Pertanyaan	Daftar Pertanyaan yang ditanyakan
3	Jawaban	Daftar jawaban dari pertanyaan

3.2. Data Preprocessing

Preprocessing adalah tahapan awal dalam pengolahan data yang bertujuan membersihkan dan menyiapkan data teks agar lebih terstruktur serta siap digunakan dalam proses klasifikasi menggunakan algoritma Naive Bayes. Dataset pertanyaan dan jawaban yang telah disusun sebelumnya diolah melalui tahapan

preprocessing yang meliputi tokenizing, stopwords removal, dan stemming. Tahapan ini dilakukan untuk meningkatkan kualitas data sehingga proses klasifikasi dapat menghasilkan hasil yang lebih akurat.

a. Tokenizing

Tokenizing merupakan proses memecah teks pertanyaan menjadi unit-unit kata (token) agar lebih mudah diproses oleh sistem. Proses ini bertujuan untuk memisahkan setiap kata dalam kalimat sehingga dapat dianalisis secara individual. Contoh hasil proses *tokenizing* ditunjukkan pada Tabel dibawah.

Tabel 4 : Contoh Proses Tokenizing

Teks	Hasil
"bagaimana cara mendapatkan surat keterangan aktif kuliah"	["bagaimana", "cara", "mendapatkan", "surat", "keterangan", "aktif", "kuliah"]

b. Stopword Removal

Stopword removal adalah proses menghapus kata-kata yang sering muncul namun tidak memiliki makna penting dalam proses klasifikasi, seperti “yang”, “dan”, “di”, serta “ke”. Tahapan ini bertujuan untuk mengurangi noise pada data teks sehingga hanya kata-kata yang relevan yang diproses lebih lanjut. Contoh hasil *stopword removal* disajikan pada dibawah.

Tabel 5 : Contoh Proses Stopword Removal

Teks	Hasil
"bagaimana cara mendapatkan surat keterangan aktif kuliah"	["mendapatkan", "surat", "keterangan", "aktif", "kuliah"]

c. Stemming

Stemming merupakan proses mengubah kata ke dalam bentuk dasarnya untuk menghindari variasi kata yang dapat memengaruhi hasil klasifikasi. Dengan melakukan *stemming*, kata-kata yang memiliki makna sama akan diperlakukan sebagai satu bentuk kata dasar. Perubahan kata yang terjadi merupakan hasil dari proses *stemming* otomatis menggunakan algoritma *stemming* bahasa Indonesia. Contoh hasil proses *stemming* dapat dilihat pada Tabel berikut.

Tabel 6 : Contoh Proses Stemming

Teks	Hasil
"bagaimana cara mendapatkan surat keterangan aktif kuliah"	["dapat", "surat", "terang", "aktif", "kuliah"]

3.3. Proses Algoritma Naive Bayes

Proses klasifikasi menggunakan algoritma Naive Bayes dilakukan setelah tahap preprocessing teks selesai. Pada tahap ini, model memanfaatkan data latih untuk mempelajari pola kemunculan kata pada setiap kategori pertanyaan sehingga sistem dapat mengenali hubungan antara teks pertanyaan dan kelas yang sesuai.

Setelah proses pelatihan (training), model digunakan pada tahap pengujian (testing) untuk mengklasifikasikan pertanyaan pengguna secara otomatis. Hasil klasifikasi tersebut menjadi dasar bagi sistem chatbot dalam menentukan respons yang relevan, seperti penyampaian informasi akademik atau layanan yang sesuai dengan kebutuhan pengguna.

3.4. Proses Implementasi

Proses implementasi dilakukan dengan mengintegrasikan model klasifikasi yang telah dibangun ke dalam sistem chatbot berbasis web. Pada tahap ini, model Natural Language Processing dan algoritma Naive Bayes yang telah melalui proses pelatihan digunakan untuk memproses setiap pertanyaan yang diajukan oleh pengguna. Pertanyaan pengguna terlebih dahulu melalui tahap preprocessing, kemudian diklasifikasikan untuk menentukan kategori atau maksud pertanyaan.

Selanjutnya, hasil klasifikasi tersebut digunakan oleh sistem untuk menentukan dan menampilkan respons yang sesuai berdasarkan basis pengetahuan yang telah disiapkan sebelumnya. Proses ini berlangsung secara otomatis dan real-time sehingga pengguna dapat memperoleh informasi akademik dengan cepat, akurat, dan mudah diakses melalui antarmuka chatbot.

3.5. Tahap Evaluasi

Evaluasi performa model dilakukan untuk mengetahui tingkat akurasi klasifikasi yang dihasilkan. Evaluasi ini bertujuan mengukur sejauh mana model mampu memprediksi data uji secara tepat dengan membandingkan hasil prediksi model terhadap label sebenarnya menggunakan metrik *accuracy*.

Nilai akurasi yang diperoleh menunjukkan persentase prediksi yang sesuai dengan data sebenarnya, sehingga dapat menggambarkan tingkat kinerja model dalam melakukan klasifikasi teks.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini disimpan dalam format Microsoft Excel dan berisi sekitar 3.500 pola pertanyaan (patterns). Setiap pertanyaan dalam dataset telah dikategorikan ke dalam tag atau label tertentu sesuai dengan jenis layanan informasi akademik yang diwakilinya.

Data tersebut kemudian digunakan sebagai data latih dalam proses pengembangan model klasifikasi, sehingga sistem chatbot dapat mengenali pola pertanyaan pengguna dan memberikan respons yang sesuai dengan kategori informasi yang telah ditentukan.

Tabel 7 : Dataset Excel

INTENT	QUESTION	ANSWER
surat_aktif	Bagaimana prosedur pengajuan surat aktif kuliah?	Surat keterangan aktif kuliah dapat diajukan dengan cara mengambalnya secara langsung di bagian BAAK di kampus 1, yang beralamat di Jalan Jenderal Ahmad Yani No. 78–88, Pekanbaru.
cuti_syarat	Bagaimana proses resmi pengajuan cuti kuliah?	Untuk mengajukan cuti kuliah, mahasiswa diwajibkan telah melunasi pembayaran uang semester yang lalu. Selanjutnya, mahasiswa melakukan konsultasi dengan dosen Pembimbing Akademik (PA) serta pihak Program Studi. Setelah mendapatkan persetujuan, mahasiswa menemui dosen yang bersangkutan untuk mengajukan cuti dan memperoleh formulir berhenti studi sementara.
no_fo	Nomor resmi Front Office berapa?	Nomor kontak Front Office (FO) adalah 0823-8178-3939.

Dari dataset yang tersedia, data pertanyaan (patterns) beserta label atau kategori (tag) digunakan dalam proses pemodelan untuk melatih model klasifikasi. Sementara itu, data jawaban (responses) dimanfaatkan sebagai basis respons sistem yang akan ditampilkan kepada pengguna sesuai dengan hasil klasifikasi tag yang dihasilkan oleh model.

4.2. Data Preprocessing

Setelah dataset pertanyaan dan jawaban berhasil dikumpulkan dalam format Microsoft Excel, langkah selanjutnya adalah melakukan proses preprocessing data teks sebelum digunakan dalam proses klasifikasi. Tahapan preprocessing bertujuan membersihkan dan menyiapkan data agar lebih terstruktur sehingga dapat meningkatkan kualitas proses pembelajaran model.

```
def text_preprocessing(text):
    # 1. Case Folding (mengubah teks menjadi huruf kecil)
    text = text.lower()

    # 2. Cleaning (menghapus angka dan tanda baca)
    text = re.sub(r"[^a-zA-Z\s]", '', text)

    # 3. Stopword Removal (menghapus kata umum bahasa Indonesia)
    text = stopword_remover.remove(text)

    # 4. Stemming (mengubah kata ke bentuk dasar)
    text = stemmer.stem(text)

    # 5. Tokenizing (memecah kalimat menjadi kata)
    tokens = word_tokenize(text)

    return tokens

print(df[['QUESTION', 'clean_question']].head(10))
```

	QUESTION	clean_question
0	Apa arti Kartu Rencana Studi	[apa, arti, kartu, rencana, studi]
1	Apa itu KRS	[apa, krs]
2	Apa itu KRS mahasiswa	[apa, krs, mahasiswa]
3	Apa yang dimaksud dengan KRS	[apa, maksud, krs]
4	Apa yang disebut KRS	[apa, sebut, krs]
5	Arti KRS	[arti, krs]
6	Arti KRS dalam sistem perkuliahan	[arti, krs, sistem, kuliah]
7	Definisi KRS	[definisi, krs]

Gambar 1. Implementasi Pra-Pemrosesan Teks

Tahapan preprocessing yang dilakukan dalam penelitian ini meliputi tokenizing, stopwords removal, dan stemming. Tokenizing dilakukan dengan memecah teks pertanyaan menjadi unit kata, kemudian stopwords removal digunakan untuk

menghapus kata-kata yang tidak memiliki makna penting dalam proses klasifikasi. Selanjutnya, stemming dilakukan untuk mengubah kata berimbuhan menjadi bentuk dasarnya sehingga sistem dapat mengenali makna kata secara lebih konsisten.

Hasil dari proses preprocessing kemudian disimpan sebagai data teks yang telah diproses dan digunakan sebagai input pada tahap pemodelan menggunakan algoritma Naive Bayes. Proses ini membantu model dalam mengenali pola teks secara lebih efektif sehingga dapat meningkatkan akurasi klasifikasi pertanyaan pada sistem chatbot.

4.3. Proses Algoritma Naive Bayes

Proses Algoritma *Naive Bayes* dilakukan setelah data pertanyaan melalui tahapan pra-pemrosesan teks. Pada tahap ini, data teks yang telah dibersihkan diubah ke dalam bentuk numerik agar dapat diproses oleh algoritma klasifikasi. Proses ini dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk merepresentasikan setiap pertanyaan dalam bentuk vektor numerik. Selanjutnya, algoritma *Multinomial Naive Bayes* digunakan untuk melakukan proses pelatihan model berdasarkan data latih yang telah disiapkan.

```
from sklearn.feature_extraction.text import TfidfVectorizer

# 1. Inisialisasi Vectorizer
# Kita simpan di variabel 'tfidf_vectorizer' agar nanti bisa dipanggil di Server Flask
tfidf_vectorizer = TfidfVectorizer()

# 2. UBAH Data Latih (Text -> Angka)
fit_transform: Pelajari kosa kata DARI DATA LATIH, lalu ubah jadi angka
X_train_tfidf = tfidf_vectorizer.fit_transform(X_train)

# 3. UBAH Data Uji (Text -> Angka)
transform: Ubah jadi angka pakai rumus yang sudah dipelajari tadi (JANGAN pakai fit lagi)
X_test_tfidf = tfidf_vectorizer.transform(X_test)

print("Sukses mengubah Teks menjadi Angka (TF-IDF)!")
print(f"Ukuran Matrix Training: {X_train_tfidf.shape}")

from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, classification_report

# 1. Inisialisasi Model Naive Bayes (MultinomialNB cocok untuk teks)
model = MultinomialNB()

# 2. Latih Model
model.fit(X_train_tfidf, y_train)

# 3. Prediksi Data Test
y_pred = model.predict(X_test_tfidf)
```

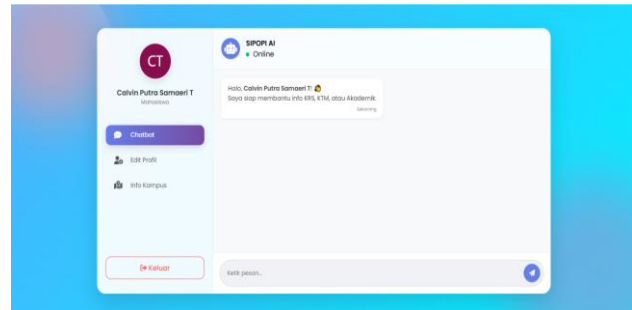
Gambar 2. Implementasi Algoritma Naive Bayes

Potongan kode di atas menunjukkan proses implementasi algoritma *Naive Bayes* yang dimulai dengan pembentukan vektor *TF-IDF* dari data latih dan data uji. Setelah proses vektorisasi dilakukan, model *Multinomial Naive Bayes* dilatih menggunakan data latih untuk mempelajari pola keterkaitan antara pertanyaan dan *intent*. Model yang telah dilatih kemudian digunakan untuk melakukan prediksi terhadap data uji, sehingga sistem mampu mengklasifikasikan pertanyaan akademik berdasarkan *intent* yang sesuai. Hasil dari proses klasifikasi ini selanjutnya digunakan pada tahap pengujian untuk mengevaluasi kinerja dan akurasi model.

4.4. Proses Implementasi

Implementasi sistem menampilkan antarmuka utama yang terdiri dari menu chatbot sebagai media interaksi pengguna dalam memperoleh informasi akademik, serta dashboard admin yang berfungsi untuk mengelola data, memantau aktivitas sistem, dan memperbarui basis pengetahuan chatbot.

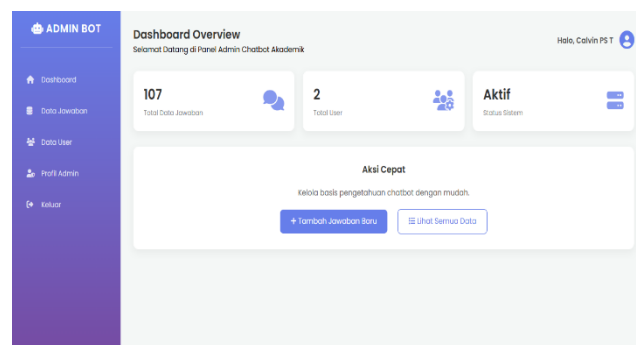
4.4.1. Implementasi Menu Chatbot



Gambar 3. Implementasi Menu Chatbot

Menu Chatbot merupakan halaman utama untuk interaksi antara pengguna dan sistem chatbot. Pengguna mengirim pesan yang diproses untuk menentukan *intent* dan menghasilkan jawaban dari basis pengetahuan, dengan tampilan percakapan secara real-time. Menu ini juga menampilkan identitas pengguna, status chatbot, serta navigasi ke fitur lain, sehingga pengguna dapat memperoleh informasi akademik secara interaktif dan responsif.

4.4.2. Implementasi Menu Admin



Gambar 4. Implementasi Menu Admin

Implementasi Dashboard Admin merupakan realisasi dari rancangan antarmuka yang telah dibuat sebelumnya. Dashboard ditampilkan sebagai halaman utama setelah admin berhasil *login* dan menyajikan informasi ringkas terkait data jawaban, data pengguna, serta status sistem chatbot. Selain itu, tersedia menu navigasi untuk mengakses fitur pengelolaan data, profil admin, dan keluar dari sistem. Implementasi dashboard ini bertujuan untuk memudahkan admin dalam memantau dan mengelola sistem chatbot secara efektif.

4.5. Tahap Evaluasi

Tahap evaluasi dilakukan untuk mengukur kinerja model klasifikasi yang dibangun menggunakan algoritma Naive Bayes dalam mengklasifikasikan pertanyaan akademik ke dalam kategori yang sesuai. Evaluasi bertujuan untuk mengetahui tingkat ketepatan model dalam memberikan hasil prediksi berdasarkan data uji yang telah disiapkan sebelumnya.

Pengukuran performa model dilakukan menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score, yang dihitung berdasarkan perbandingan antara hasil prediksi model dan label data sebenarnya. Hasil evaluasi kemudian disajikan dalam bentuk perhitungan nilai metrik serta visualisasi grafik guna memberikan gambaran yang lebih jelas mengenai performa model yang dikembangkan.

```
import seaborn as sns
import pandas as pd
from sklearn.metrics import precision_recall_fscore_support

# Hitung rata-rata metrics (weighted average)
precision, recall, f1, _ = precision_recall_fscore_support(y_test, y_pred, average='weighted')
accuracy = accuracy_score(y_test, y_pred)

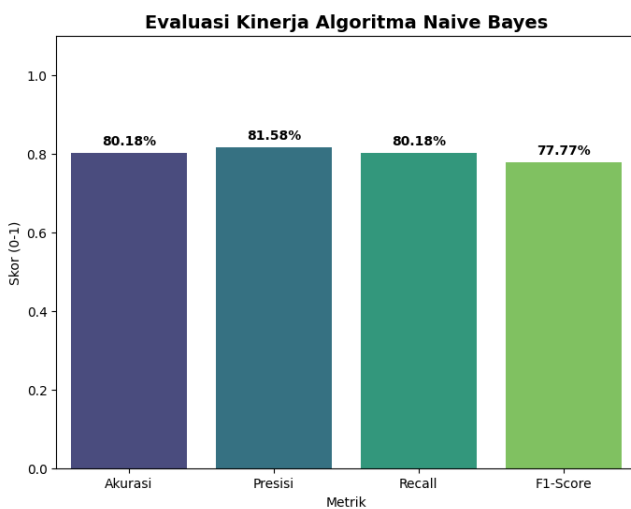
# Buat DataFrame untuk plotting
metrics_data = pd.DataFrame({
    'Metrik': ['Akurasi', 'Presisi', 'Recall', 'F1-Score'],
    'Nilai': [accuracy, precision, recall, f1]
})

plt.figure(figsize=(8, 6))
ax = sns.barplot(x='Metrik', y='Nilai', data=metrics_data, palette='viridis')

# Menambahkan angka di atas batang
for p in ax.patches:
    ax.annotate(f'{p.get_height():.2f}%',
                (p.get_x() + p.get_width() / 2., p.get_height()),
                ha='center', va='center',
                xytext=(0, 9), textcoords='offset points', fontweight='bold')

plt.ylim(0, 1.1) # Batas atas y
plt.title('Evaluasi Kinerja Algoritma Naive Bayes', fontsize=14, fontweight='bold')
plt.xlabel('Skor (0-1)')
plt.show()
```

Gambar 5. Perhitungan Akurasi Pada Model



Gambar 6. Visualisasi Bar Chart Hasil Perhitungan

Berdasarkan grafik evaluasi kinerja algoritma Naive Bayes di atas, diperoleh nilai akurasi sebesar 80,18%, presisi sebesar 81,58%, recall sebesar 80,18%, dan F1-score sebesar 77,77%. Hasil tersebut menunjukkan bahwa model mampu mengklasifikasikan pertanyaan akademik dengan tingkat ketepatan yang cukup baik dan konsisten. Nilai presisi yang lebih tinggi mengindikasikan bahwa sistem memiliki kemampuan yang baik dalam memberikan jawaban yang relevan sesuai dengan *intent* yang dikenali, sedangkan nilai *recall* menunjukkan bahwa

sebagian besar pertanyaan pengguna dapat teridentifikasi dengan benar. Secara keseluruhan, hasil pengujian ini membuktikan bahwa algoritma Naive Bayes layak digunakan sebagai metode klasifikasi pada sistem chatbot pelayanan akademik kemahasiswaan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil perancangan, implementasi, dan pengujian sistem chatbot pelayanan akademik kemahasiswaan berbasis *Natural Language Processing* dengan algoritma Naive Bayes, maka dapat disimpulkan sebagai berikut:

1. Sistem chatbot pelayanan akademik kemahasiswaan berhasil dikembangkan sebagai aplikasi berbasis web yang bisa memberikan layanan informasi akademik secara otomatis dan responsif. Sistem ini akan mempermudah mahasiswa dalam memperoleh informasi akademik tanpa harus melalui proses manual, sehingga mendukung optimalisasi layanan akademik.
2. Penerapan *Natural Language Processing* melalui tahapan pra-pemrosesan teks, yaitu *tokenizing*, *stopword removal*, dan *stemming*, memungkinkan sistem memahami pertanyaan mahasiswa dalam bahasa alami serta mendukung proses klasifikasi pertanyaan secara lebih efektif.
3. Algoritma Naive Bayes yang digunakan menunjukkan kinerja yang baik dalam mengklasifikasikan pertanyaan akademik, dengan nilai akurasi sebesar 80,18%, presisi sebesar 81,58%, recall sebesar 80,18%, dan F1-score sebesar 77,77%, sehingga sistem dinilai mampu memberikan respons yang relevan dan andal sesuai dengan kebutuhan pengguna.

Berdasarkan hasil penelitian yang telah dilaksanakan, terdapat beberapa saran yang dapat dijadikan acuan untuk pengembangan sistem chatbot pelayanan akademik kemahasiswaan agar dapat berjalan lebih optimal di masa mendatang, yaitu:

1. Pengembangan chatbot dapat ditingkatkan dengan menambah jumlah dataset, variasi intent, dan cakupan layanan akademik, sehingga sistem mampu mengenali lebih banyak jenis pertanyaan dan memberikan respons yang semakin akurat.
2. Performa klasifikasi chatbot dapat ditingkatkan dengan melakukan eksplorasi dan perbandingan algoritma lain selain Naive Bayes, serta melakukan optimasi pada tahapan pra-pemrosesan teks untuk meningkatkan akurasi dan keandalan sistem.

3. Untuk pengembangan selanjutnya, sistem chatbot dapat diperluas dengan integrasi ke berbagai platform lain, seperti aplikasi mobile atau media pesan instan, sehingga akses terhadap layanan akademik menjadi lebih fleksibel dan mudah dijangkau oleh pengguna.

Terap., vol. 7, no. 2, pp. 300–313, 2021, doi: 10.35143/jkt.v7i2.5116.

DAFTAR PUSTAKA

- [1] R. Andi Kambau, “Proses Transformasi Digital pada Perguruan Tinggi di Indonesia,” *J. Rekayasa Sist. Inf. dan Teknol.*, vol. 1, no. 3, pp. 126–136, 2024, doi: 10.59407/jrsit.v1i3.481.
- [2] M. F. Faqih, W. Lestari, and M. Muhtarom, “Pengembangan Arsitektur Chatbot untuk Layanan Informasi Akademik Berbasis Natural Language Processing (NLP),” pp. 284–290, 2025.
- [3] A. Dwiyo Widianoro Mustafid Ridwan Sanjaya, *Pengantar Nlp Dan Topik Model Lda Sampul Dalam*. 2024.
- [4] F. H. Unisba, K. Teks, and L. Akademik, “Aplikasi chatbot berbasis nlp untuk layanan akademik studi kasus fakultas hukum unisba,” vol. 20, pp. 89–97, 2025.
- [5] S. I. Akbar, “Kualitas Layanan Jasa Pendidikan: Analisis Quality Function Deployment,” *J. Gama Soc.*, vol. 2, no. 2, p. 121, 2019, doi: 10.22146/jgs.48866.
- [6] C. A. Oktavia, “4. Full Paper - Implementasi Chatbot,” *Implementasi Chatbot*, 2019.
- [7] A. Setiawan, Z. F. Ikatrinasari, and H. A. Prabowo, “Analisis Kualitas Pelayanan Pendidikan Dengan Menggunakan Metode Eduqual dan Importance Performance Analysis Assesment (PISA) pada tahun 2019 , posisi 71 diantara 77 negara lain (Organization for Economic Cooperation menempatkan Indonesia pada peringkat,” vol. 10, no. 2, pp. 89–98, 2023.
- [8] K. Publik, D. Manajemen, F. Ekonomi, and U. Andalas, “Tersedia Online di <http://jurnal.umb.ac.id/index.php/JMPKP> ISSN Print : 2685-7499 ISSN Online : 2745-8660,” vol. 3, pp. 1–13, 2021.
- [9] P. Dewonoto Laut Santoso, I. Riski, N. Kholik, M. Raffi Akbar, and A. Saifudin, “Penerapan Artificial Intelligence dalam Aplikasi Chatbot sebagai Media Informasi dan Pembelajaran mengenai Kebudayaan Bangsa,” *J. Inform. Univ. Pamulang*, vol. 6, no. 3, pp. 579–589, 2021, [Online]. Available: <http://openjournal.unpam.ac.id/index.php/informatika579>
- [10] M. Mustaqim, A. Gunawan, Y. B. Pratama, and I. Zaliman, “PENGEMBANGAN CHATBOT LAYANAN PUBLIK MENGGUNAKAN,” vol. 1, no. 1, pp. 1–4, 2023.
- [11] M. Qalimaturrahmah and D. B. Santoso, “Aplikasi Layanan dan Informasi Akademik Berbasis Chatbot Telegram Menggunakan Natural Language Processing,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 2, pp. 434–443, 2024, doi: 10.35870/jtik.v8i2.1887.
- [12] N. A. Ionendri, F. Candra, and A. Rizal, “JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) News Classification using Natural Language Processing with TF-IDF and Multinomial Naïve Bayes,” vol. 6, no. 1, pp. 37–45, 2025.
- [13] W. C. Wahyudin, “Penerapan Algoritma Naive Bayes Pada Sistem Chatbot Persewaan Kos,” vol. 22, no. 1, pp. 53–62, 2025, doi: 10.30595/sainteks.v22i1.25962.
- [14] M. sari Zulvi, “Systematic Literature Review Penerapan Metodologi Agile Dalam Berbagai Bidang,” *J. Komput.*