

APLIKASI CHATBOT BERBASIS NLP UNTUK LAYANAN AKADEMIK STUDI KASUS FAKULTAS HUKUM UNISBA

Herman Syaefuloh¹, Castaka Agus Sugianto²

Teknik Informatika, Politeknik TEDC Bandung¹

email: hermansyaefuloh16@gmail.com¹

Teknik Informatika, Politeknik TEDC Bandung²

email: castaka@poltektedc.ac.id²

Abstrak

Fakultas Hukum Universitas Islam Bandung (Unisba) melayani ratusan mahasiswa aktif dengan kebutuhan informasi akademik yang beragam. Saat ini, layanan masih bergantung pada interaksi langsung atau menggunakan aplikasi WhatsApp yang terbatas pada satu operator sesuai jam kerja, sehingga respons sering terlambat. Kondisi ini berpotensi menghambat mahasiswa memperoleh informasi penting secara tepat waktu. Penelitian ini menawarkan solusi berupa chatbot berbasis Natural Language Processing (NLP) untuk memberikan layanan akademik cepat, responsif, dan otomatis. Chatbot dirancang agar mampu memahami dan merespons pertanyaan mahasiswa secara real-time dengan tingkat akurasi tertentu. Algoritma Naive Bayes Classifier dipilih untuk mengklasifikasikan intent pertanyaan. Dataset terdiri atas 250 data pertanyaan, jawaban, dan intent akademik, yang melalui tahapan praproses teks yaitu tokenization, stopword removal, stemming serta representasi fitur menggunakan TF-IDF. Hasil pengujian menunjukkan akurasi model sebesar 96%, dengan macro average precision 83%, recall 82%, dan f1-score 82%. Sementara itu, weighted average precision, recall, dan f1-score masing-masing mencapai 95%, menandakan kinerja model stabil di berbagai kelas. Dengan demikian, Naive Bayes terbukti efektif mendukung chatbot layanan akademik Fakultas Hukum Unisba, sehingga mampu menjawab pertanyaan mahasiswa dengan cukup akurat.

Kata Kunci: Naive Bayes, NLP, Chatbot, Klasifikasi Teks, Layanan Akademik

Abstract

The Faculty of Law at Universitas Islam Bandung (Unisba) serves hundreds of active students with diverse academic information needs. Currently, the service relies on face-to-face interaction or WhatsApp communication, which is limited to a single operator during working hours, often resulting in delayed responses. This condition may hinder students from obtaining important information in a timely manner. This study proposes a solution in the form of a chatbot based on Natural Language Processing (NLP) to provide fast, responsive, and automated academic services. The chatbot is designed to understand and respond to student inquiries in real time with a certain level of accuracy. The Naive Bayes Classifier algorithm was chosen to classify question intents. The dataset consists of 250 academic-related questions, answers, and intents, processed through text preprocessing stages including tokenization, stopword removal, stemming, and feature representation using TF-IDF. Experimental results show a model accuracy of 96%, with macro average precision of 83%, recall of 82%, and f1-score of 82%. Meanwhile, the weighted average precision, recall, and f1-score each reached 95%, indicating stable performance across classes. Therefore, Naive Bayes proves effective in supporting the academic service chatbot at Unisba's Faculty of Law, enabling accurate responses to student queries.

Keywords: Naive Bayes, NLP, Chatbot, Text Classification, Academic Services

1. PENDAHULUAN

Perguruan Tinggi merupakan institusi penyelenggara pendidikan tinggi yang memiliki peran strategis dalam mencetak sumber daya manusia unggul, berdaya saing, serta mampu menjadi agent of change dalam masyarakat (Kementerian Pendidikan dan Kebudayaan, 2020). Universitas Islam Bandung (Unisba), yang berdiri sejak tahun 1958, adalah salah satu perguruan tinggi swasta (PTS) di Indonesia yang telah memperoleh akreditasi “Unggul” dari Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT, 2023).

Fakultas Hukum Unisba merupakan salah satu unit akademik yang menyelenggarakan pendidikan pada jenjang Sarjana (S1), Magister (S2), hingga Doktor (S3). Fakultas ini memiliki visi “Terwujudnya Fakultas Hukum yang mandiri, maju, terkemuka, dan profesional berasaskan nilai-nilai Islam dan mampu menghadapi tantangan di Asia Tahun 2034” (UNISBA, 2023). Untuk mendukung pencapaian visi tersebut, dibutuhkan sarana, prasarana, serta layanan akademik yang efektif, efisien, dan responsif.

Pengelolaan layanan akademik yang optimal akan mendukung terciptanya ekosistem pendidikan yang kondusif, berkualitas, dan berkelanjutan (Kemdikbudristek, 2021). Saat ini, Fakultas Hukum Unisba melayani ratusan mahasiswa aktif yang membutuhkan informasi seperti jadwal kuliah, prosedur akademik, beasiswa, hingga informasi non-akademik. Layanan informasi masih banyak bergantung pada interaksi langsung atau melalui media sosial seperti WhatsApp. Namun, sistem ini memiliki keterbatasan, yaitu hanya dikelola oleh satu orang dan tidak tersedia selama 24 jam, sehingga menyebabkan keterlambatan respons dan potensi kehilangan informasi penting. Kondisi ini menunjukkan adanya kesenjangan dalam penyediaan layanan akademik yang belum mampu memberikan akses cepat, responsif, dan berkelanjutan bagi mahasiswa. Sebagai solusi, diperlukan pemanfaatan teknologi informasi yang dapat menyediakan layanan otomatis dan responsif. Salah satu teknologi yang berkembang dalam ranah ini adalah chatbot. Chatbot didefinisikan sebagai program komputer yang meniru percakapan manusia melalui teks atau suara dengan bantuan Natural Language Processing (NLP). Chatbot merupakan salah satu contoh sistem kecerdasan buatan (AI) yang berfungsi sebagai agen interaksi manusia-komputer yang cerdas dan luas digunakan untuk berbagai aplikasi mulai dari edukasi hingga layanan bisnis (An Overview of Chatbot Technology, 2020). Dengan memanfaatkan teknologi Natural Language Processing (NLP), chatbot mampu memahami dan merespons pertanyaan pengguna secara otomatis dan real-time.

Keberhasilan chatbot dalam menjawab pertanyaan secara tepat sangat bergantung pada algoritma klasifikasi teks yang digunakan. Salah satu algoritma yang kerap dipakai adalah Naive Bayes, sebuah metode probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa fitur-fitur dalam data, seperti kata-kata dalam kalimat, merupakan variabel yang saling independen. Meskipun asumsi ini idealistik dan tidak selalu sesuai dengan kenyataan dalam teks, Naive Bayes tetap memberikan performa yang cukup baik dalam berbagai aplikasi klasifikasi teks (Assiroj, dkk 2023). Penelitian lain yang dilakukan oleh Fauzan ,dkk yang mengembangkan chatbot berbasis NLP dengan akurasi 87,5% pada sistem informasi pendaftaran siswa baru, menegaskan efektivitas Naive Bayes dalam konteks chatbot interaktif berbasis teks (Fauzan, dkk 2024)

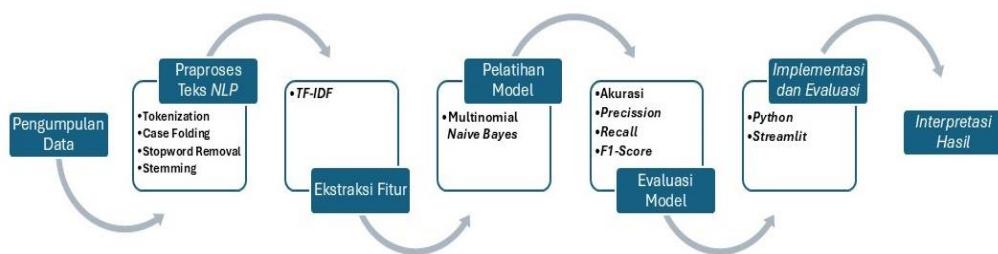
Klasifier Naïve Bayes menunjukkan akurasi dan kecepatan yang tinggi ketika diterapkan pada basis data yang besar. Metode Naïve Bayes sering digunakan dalam menyelesaikan masalah di bidang machine learning karena metode ini dikenal memiliki tingkat akurasi yang tinggi dengan perhitungan yang sederhana (Alimuddin dkk., 2022). Naive Bayes terbukti efektif dalam berbagai tugas NLP seperti klasifikasi teks dan analisis sentimen. Seperti penelitian Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor (KNN) Pada Video Youtube Mengenai Global Warming oleh Himawan, Castaka Sugianto bahwa algoritma Naive Bayes mampu mengklasifikasikan sentimen positif dan negatif dari komentar yang ada dengan akurasi yang lebih tinggi dibandingkan algoritma K-Nearest Neighbor (KNN) (Himawan & Sugianto, 2024). Penelitian lainnya yang dilakukan oleh Ayusari dan Sugianto dalam menganalisis Algoritma Naive Bayes untuk Klasifikasi Ketepatan Waktu Kelulusan

Mahasiswa Politeknik TEDC, Algoritma Naive Bayes berhasil mengklasifikasikan ketepatan waktu kelulusan mahasiswa dengan akurat. Dari 251 data, 202 mahasiswa dikategorikan lulus tepat waktu dan 49 tidak tepat waktu menggunakan F1-Score (Ayusari & Sugianto, 2025). Algoritma Naive Bayes dalam konteks layanan akademik, dapat digunakan untuk mengklasifikasikan pertanyaan mahasiswa ke dalam kategori intent yang spesifik seperti jadwal kuliah, beasiswa, administrasi akademik ataupun kategori lainnya seputar akademik. Penelitian sebelumnya yang membahas chatbot untuk layanan akademik masih terbatas, terutama yang secara spesifik mengkaji layanan akademik di Fakultas Hukum Unisba sebagai domain. Selain itu, belum banyak penelitian yang mengembangkan klasifikasi intent sesuai kebutuhan informasi akademik, serta merancang protokol praproses data teks berbahasa Indonesia yang sistematis.

Tujuan penelitian ini adalah untuk menganalisis efektivitas dan akurasi algoritma Naive Bayes dalam klasifikasi teks pada chatbot layanan akademik Fakultas Hukum Unisba, sehingga dapat memberikan kontribusi dalam pengembangan sistem layanan pendidikan berbasis kecerdasan buatan di perguruan tinggi.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen dan pengujian system klasifikasi berbasis teks. Metodologi penelitian kuantitatif merupakan jenis penelitian berdasarkan paradigma yang menekankan penggunaan data dalam bentuk angka untuk menjawab pertanyaan penelitian secara objektif. Metodologi adalah cara sistematis yang digunakan untuk mencapai tujuan penelitian, mencakup bahan, alat, prosedur, dan analisis data. Penelitian ilmiah harus dilakukan secara sistematis, terkontrol, dan kritis untuk menghasilkan kebenaran ilmiah yang bersifat menerangkan dan memprediksi (Aiman, 2025). Untuk mencapai hasil dari tujuan penelitian maka penulis membuat alur proses penelitian seperti pada gambar berikut:



Gambar 1. Alur Proses Penelitian

Berdasarkan gambar 1 alur proses penelitian, penelitian diawali dengan melakukan pengumpulan data dimana dataset yang digunakan merupakan kumpulan pertanyaan mahasiswa terkait layanan akademik seperti informasi jadwal, beasiswa, pendaftaran sidang, dan informasi umum fakultas. Data dikumpulkan dari dokumentasi pertanyaan yang sering diajukan mahasiswa melalui pesan whatsapp, serta sintesis dan observasi langsung berdasarkan wawancara dengan staf akademik terkait hal apa saja yang sering ditanyakan oleh mahasiswa.

Tahap kedua adalah praproses teks NLP, yang berfungsi untuk membersihkan dan menormalisasi data teks sehingga siap digunakan dalam model pembelajaran mesin. Proses ini meliputi tokenization untuk memecah kalimat menjadi kata-kata, case folding untuk mengubah semua huruf menjadi huruf kecil, stopword removal untuk menghapus kata-kata umum yang tidak bermakna penting, seperti “dan”, “atau”, “yang”, serta stemming untuk mengembalikan kata ke bentuk kata dasarnya tanpa imbuhan. Tahapan ini memastikan bahwa data teks memiliki konsistensi dan kualitas yang baik untuk pemodelan.

Setelah melalui praproses teks, selanjutnya adalah ekstraksi fitur yang bertujuan mengubah teks menjadi representasi numerik agar dapat dipahami oleh algoritma. Dalam penelitian ini, metode Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk menghasilkan vektor fitur yang merepresentasikan bobot kata dalam teks.

Setelah itu, dilakukan pelatihan model menggunakan algoritma Multinomial Naive Bayes. Algoritma ini dipilih karena memiliki kinerja yang baik dalam klasifikasi teks, sederhana dalam implementasi, dan efektif untuk dataset berbahasa Indonesia dengan jumlah data yang terbatas. Pada tahap ini, model dilatih untuk mengenali pola dan mengklasifikasikan pertanyaan berdasarkan intent yang telah ditentukan. Tahap berikutnya adalah evaluasi model, yang dilakukan untuk menilai performa klasifikasi dari algoritma. Evaluasi dilakukan dengan menghitung metrik Akurasi untuk mengukur persentase jumlah prediksi yang benar dari total prediksi, Precision, Recall untuk mengukur seberapa tepat model dalam memprediksi kelas tertentu, dan F1-score untuk menghitung rata-rata harmonis dari precision dan recall untuk memberikan gambaran menyeluruh terhadap kinerja model. Model yang telah dievaluasi kemudian dilanjutkan pada tahap implementasi dan simulasi, yaitu integrasi model ke dalam antarmuka chatbot. Tahap terakhir interpretasi hasil yaitu menganalisis performa algoritma Naive Bayes sebagai algoritma pada chatbot layanan akademik berdasarkan nilai dari proses evaluasi model dan menguji aplikasi chatbot melalui simulasi percakapan dengan pengguna, di mana chatbot menerima input berupa pertanyaan, melakukan klasifikasi intent dengan model Naive Bayes, dan memberikan jawaban yang sesuai.

3. HASIL DAN PEMBAHASAN

Proses penelitian dimulai dengan membuat dataset berisi pertanyaan, jawaban dan intent seputar informasi akademik. Dataset ini disimpan dalam format file .csv (*Comma-Separated Values*) dan diberinama “dataset_pertanyaan.csv”, berisi 250 data pertanyaan, jawaban dan intent untuk selanjutnya dilakukan praproses teks *NLP* terhadap dataset, ekstraksi fitur, pelatihan model menggunakan algoritma *Naive Bayes*, evaluasi model untuk menghitung tingkat akurasi serta dilakukan implementasi dan simulasi aplikasi *chatbot*. Distribusi dataset yang digunakan pada penelitian ini dapat dilihat pada tabel berikut:

Tabel 1. Distribusi dataset

No.	Intent	Jumlah Data	Presentase(%)
1	Ujian	30	12%
2	KRS	26	10%
3	Pendaftaran	12	5%
4	Skripsi	39	16%
5	Administrasi	32	13%
6	Lainnya	29	12%
7	Wisuda	22	9%
8	Jadwal Kuliah	14	6%
9	Nilai	30	12%
10	Beasiswa	16	6%
	Total	250	100%

Berdasarkan Tabel 1, distribusi dataset menunjukkan adanya variasi jumlah data pada setiap kelas intent. Kelas *Skripsi* (16%), *Administrasi* (13%), serta *Ujian* dan *Nilai* (12%) merupakan kelas dengan jumlah data relatif besar. Sebaliknya, kelas seperti *Pendaftaran* (5%), *Beasiswa* (6%), dan *Jadwal Kuliah* (6%) memiliki jumlah data yang lebih kecil. Kondisi ini memperlihatkan adanya ketidakseimbangan kelas (class imbalance) pada dataset.

Penelitian ini tetap menggunakan dataset apa adanya tanpa dilakukan proses penyeimbangan atau augmentasi data. Hal ini bertujuan untuk mencerminkan kondisi riil pertanyaan mahasiswa yang memang tidak terdistribusi secara merata pada tiap kategori. Dengan demikian, hasil evaluasi model

dapat memberikan gambaran yang lebih objektif terhadap kinerja algoritma Naive Bayes dalam menangani data dengan distribusi alami, meskipun terdapat risiko penurunan performa pada kelas minoritas.

3.1 Hasil Praproses Teks NLP

Setelah dataset dibuat, selanjutnya dilakukan praproses teks *NLP* menggunakan metode *Tokenization*, *case Folding*, *Stopword Removal*, dan *Stemming*. Hasil dari proses tersebut dapat dilihat pada gambar berikut:

	Teks Asli	Case Folding	Tokenisasi	Stopword Removal	Stemming
0	Bagaimana cara daftar mahasiswa baru?	bagaimana cara daftar mahasiswa baru?	[berapa, biaya, pendaftaran, kuliah, di, fakul...]	[biaya, pendaftaran, kuliah, fakultas, hukum]	[biaya, pendaftaran, kuliah, fakulta, hukum]
1	Syarat pendaftaran mahasiswa baru apa saja?	syarat pendaftaran mahasiswa baru apa saja?	[berapa, biaya, pendaftaran, kuliah, di, fakul...]	[biaya, pendaftaran, kuliah, fakultas, hukum]	[biaya, pendaftaran, kuliah, fakulta, hukum]
2	Kapan jadwal pendaftaran mahasiswa baru?	kapan jadwal pendaftaran mahasiswa baru?	[berapa, biaya, pendaftaran, kuliah, di, fakul...]	[biaya, pendaftaran, kuliah, fakultas, hukum]	[biaya, pendaftaran, kuliah, fakulta, hukum]
3	Tanggal berapa jadwal pendaftaran mahasiswa baru?	tanggal berapa jadwal pendaftaran mahasiswa baru?	[berapa, biaya, pendaftaran, kuliah, di, fakul...]	[biaya, pendaftaran, kuliah, fakultas, hukum]	[biaya, pendaftaran, kuliah, fakulta, hukum]
4	Berapa biaya pendaftaran kuliah di Fakultas Hu...	berapa biaya pendaftaran kuliah di fakultas hu...	[berapa, biaya, pendaftaran, kuliah, di, fakul...]	[biaya, pendaftaran, kuliah, fakultas, hukum]	[biaya, pendaftaran, kuliah, fakulta, hukum]

Gambar 2. Hasil Praproses Teks

3.2 Hasil Ekstraksi Fitur

Setelah data teks dibersihkan melalui praposes teks, langkah selanjutnya adalah mengubah teks menjadi bentuk numerik agar bisa diproses oleh algoritma pembelajaran. Proses ekstraksi fitur dilakukan menggunakan metode *TF-IDF* untuk merepresentasikan data teks ke dalam bentuk numerik. Hasil ekstraksi menghasilkan matriks dengan ukuran n dokumen × m fitur. Gambar 3 menunjukkan cuplikan dari hasil ekstraksi 5 dokumen dengan 10 fitur pertama. Terlihat bahwa kata biaya memiliki nilai TF-IDF yang cukup tinggi pada dokumen ke-5 (0,460281), menunjukkan kata tersebut cukup representatif dalam membedakan dokumen tersebut dari yang lain. hasil dari ekstraksi fitur dapat dilihat pada gambar berikut:

	administrasi	akademik	aktif	akun	bata	batal	beasiswa	berkelompok	berprestasi	biaya
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.460281

Gambar 3. Hasil Ekstraksi Fitur *TF-IDF*

Konfigurasi parameter yang digunakan dalam proses ekstraksi fitur *TF-IDF* pada penelitian ini menggunakan nilai bawaan (default) dari library *scikit-learn* karena tidak dilakukan penyesuaian khusus. Rincian parameter dapat dilihat pada Tabel 2 berikut:

Tabel 2. Parameter *TF-IDF*

No.	Parameter	Nilai yang Digunakan	Keterangan
1	n-gram range	(1,1)	Hanya menggunakan unigram (kata tunggal)
2	min_df	1	Semua kata yang muncul minimal sekali tetap dipertahankan
3	max_df	1	Semua kata tetap digunakan tanpa ada yang difilter karena terlalu sering
4	use_idf	TRUE	IDF diterapkan untuk menurunkan bobot kata yang sering muncul
5	norm	L2	Normalisasi vektor menggunakan L2 norm

Berdasarkan Tabel 2 dapat disimpulkan bahwa penelitian ini menggunakan konfigurasi standar TF-IDF, yaitu unigram dengan mempertahankan semua kata yang muncul minimal sekali. Penerapan IDF membantu menurunkan bobot kata yang terlalu sering muncul, sementara normalisasi L2 menjaga skala fitur agar tetap seimbang. Dengan setelan ini, representasi teks dapat digunakan secara optimal dalam proses pembelajaran mesin menggunakan algoritma Naive Bayes.

3.3 Pelatihan Model

Tahap selanjutnya adalah melakukan pelatihan model yaitu menerapkan algoritma Multinomial *Naive Bayes* untuk melakukan klasifikasi intent pada dataset pertanyaan. Dalam model multinomial *Naive Bayes*, algoritma ini menghitung probabilitas berdasarkan jumlah kemunculan kata (fitur) dalam dokumen tertentu. Proses pelatihan dilakukan menggunakan data latih, sebagaimana ditunjukkan pada potongan kode berikut: `model = MultinomialNB(); model.fit(X_train, y_train)`. Tahap ini menghasilkan model Multinomial *Naive Bayes* yang siap digunakan untuk melakukan prediksi terhadap pertanyaan mahasiswa. Dengan demikian, sistem dapat memetakan input teks ke dalam kelas intent yang sesuai berdasarkan probabilitas tertinggi yang dihitung oleh algoritma.

3.4 Hasil Evaluasi Model

Secara keseluruhan, model menghasilkan akurasi sebesar 96%, yang menunjukkan bahwa algoritma *Naive Bayes* mampu melakukan klasifikasi intent pertanyaan mahasiswa dengan cukup baik. Nilai *macro average* untuk *precision* mencapai 83%, *recall* 82%, dan *f1-score* 82%. Sementara itu, *weighted average precision* sebesar 95%, *recall* 96%, dan *f1-score* 95%, sehingga kinerja model relatif stabil pada berbagai kelas. Beberapa kelas seperti *administrasi*, *KRS*, *skripsi*, dan *ujian* berhasil diklasifikasikan sempurna (*precision*, *recall*, dan *f1-score* = 1.00). Sementara itu, kelas *pendaftaran* dan *lainnya* masih menunjukkan sedikit kesalahan dengan *f1-score* 0.88 dan 0.86. Namun, kelas *beasiswa* tidak terprediksi sama sekali (*f1-score* = 0.00) karena jumlah data yang sangat terbatas. Dengan demikian, meskipun performa model secara keseluruhan sangat baik, ketidakseimbangan distribusi data masih menjadi faktor utama yang memengaruhi hasil pada kelas tertentu. Hasil keseluruhan evaluasi model dapat dilihat pada gambar berikut :

```
Akurasi: 0.96
Precision: 0.9488888888888889
Recall: 0.96
F1-Score: 0.9510714285714286

== Classification Report ==
      precision    recall   f1-score  support
administrasi       1.00     1.00     1.00      5
beasiswa          0.00     0.00     0.00      1
      krs           1.00     1.00     1.00     13
lainnya            1.00     0.75     0.86      4
pendaftaran        0.78     1.00     0.88      7
      skripsi        1.00     1.00     1.00     13
      ujian           1.00     1.00     1.00      7

accuracy           -         -       0.96     50
macro avg          0.83     0.82     0.82     50
weighted avg       0.95     0.96     0.95     50
```

Gambar 4. Hasil Evaluasi Model

3.5 Implementasi dan Evaluasi

Setelah melakukan pengujian dan pelatihan pada algoritma *naive bayes*, selanjutnya adalah membuat aplikasi *chatbot* sederhana untuk simulasi dengan menggunakan algoritma *naive bayes*. Pembuatan aplikasi sederhana ini menggunakan bahasa pemrograman *python* dan aplikasi *Visual Studio Code* untuk menulis dan menjalankan kode *python*. Aplikasi *chatbot* ini dibuat dalam versi web dengan menggunakan *streamlit* sebagai *framework* terbuka atau *open source* yang akan secara otomatis mengubah kode *python* menjadi aplikasi web yang dapat berinteraksi dengan pengguna. Setelah aplikasi dibuat dan dijalankan, aplikasi *chatbot* dapat menjawab pertanyaan seputar layanan akademik secara akurat sesuai dataset yang telah diberikan dan dilatihkan pada algoritma *naive bayes*. Berikut tampilan dari aplikasi *chatbot* layanan akademik Fakultas Hukum Unisba dengan memnggunakan algoritma *naive bayes*.



Gambar 5. Aplikasi *Chatbot* Layanan Akademik Fakultas Hukum Unisba

3.6 Interpretasi Hasil

Hasil pengujian menunjukkan bahwa algoritma *Naive Bayes* mampu mencapai tingkat akurasi sebesar 85% dalam mengklasifikasikan pertanyaan mahasiswa ke dalam kategori layanan akademik. Nilai akurasi tersebut dapat dikatakan cukup baik, mengingat karakteristik bahasa alami yang digunakan mahasiswa sering kali bervariasi, baik dari segi tata bahasa, penggunaan sinonim, maupun struktur kalimat. Hal ini membuktikan bahwa *Naive Bayes* efektif dalam menangani teks dengan variasi yang tinggi karena algoritma ini bekerja berdasarkan probabilitas kemunculan kata-kata pada setiap kelas. Temuan ini sejalan dengan teori bahwa *Naive Bayes* cocok digunakan dalam permasalahan klasifikasi teks dan sistem berbasis Natural Language Processing (NLP), terutama pada dataset berukuran sedang dengan fitur yang relatif sederhana. Akurasi 85% menunjukkan bahwa sebagian besar pertanyaan mahasiswa dapat dijawab dengan benar oleh chatbot, meskipun masih terdapat beberapa kasus kesalahan klasifikasi akibat keterbatasan dataset dan overlap antar kategori pertanyaan.

Dengan demikian, hasil penelitian ini menegaskan bahwa implementasi algoritma *Naive Bayes* dalam sistem chatbot layanan akademik Fakultas Hukum Unisba dapat membantu meningkatkan efisiensi layanan, khususnya dalam memberikan respon cepat terhadap pertanyaan yang sering diajukan mahasiswa.

4. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma *Naive Bayes* dapat digunakan secara efektif untuk mengklasifikasikan pertanyaan pengguna ke dalam kategori tertentu dalam konteks layanan akademik Fakultas Hukum Unisba. Hasil pengujian memperlihatkan bahwa algoritma ini mampu memberikan tingkat akurasi yang cukup baik pada data uji yaitu 96%, sehingga dapat diimplementasikan dalam sistem chatbot berbasis Natural Language Processing (NLP) untuk membantu menjawab pertanyaan mahasiswa secara otomatis. Begitupun setelah algoritma diterapkan pada aplikasi chatbot, chatbot dapat menjawab pertanyaan yang diberikan seputar informasi akademik secara akurat. Saran penulis dataset yang digunakan dalam penelitian ini masih terbatas dalam jumlah dan variasi pertanyaan. Untuk meningkatkan performa sistem, diperlukan perluasan dataset baik dari segi kuantitas maupun keragaman jenis pertanyaan yang akan membantu model mengenali pola bahasa mahasiswa yang lebih kompleks dan bervariasi. Meskipun *Naive Bayes* menunjukkan hasil yang cukup baik, penelitian selanjutnya disarankan membandingkan performa dengan algoritma lain seperti Support Vector Machine (SVM), Random Forest, atau metode berbasis deep learning (misalnya LSTM atau BERT) untuk melihat peningkatan akurasi dan ketahanan model.

5. REFERENSI

- Alimuddin, A., Putra, R., & Yuliani, S. (2022). Implementasi algoritma Naïve Bayes untuk klasifikasi teks berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 9(3), 451–460.
- An Overview of Chatbot Technology. (2020). *International Journal of Advanced Computer Science and Applications (IJACSA)*, 11(12), 1–7.
- Assiroj, P., Kurnia, A., & Alam, S. (2023). The performance of Naïve Bayes, Support Vector Machine, and Logistic Regression on Indonesia immigration sentiment analysis. *Bulletin of Electrical Engineering and Informatics*, 12(6), 3843–3852.
- Ayusari, N., & Sugianto, C. A. (2025). Algoritma Naive Bayes untuk klasifikasi ketepatan waktu kelulusan mahasiswa Politeknik TEDC Bandung. *Journal of Applied Information Technology and Innovation*, 1(1), 1–9.
- BAN-PT. (2023). *Keputusan akreditasi BAN-PT tahun 2023*. Badan Akreditasi Nasional Perguruan Tinggi.
- Fauzan, M. F., Imand, R., & Hasbi, M. A. (2024). Designing a chatbot with NLP technology in a website-based new student admission information system. *Jurnal JAIC*, 8(2), 358–366.
- Himawan, A., & Sugianto, C. A. (2024). Perbandingan algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) pada video YouTube mengenai global warming. *Jurnal Informatika*, 15(2), 112–120.

- Kemdikbudristek. (2021). *Transformasi digital layanan akademik perguruan tinggi*. Jakarta: Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.
- Kementerian Pendidikan dan Kebudayaan. (2020). *Statistik pendidikan tinggi 2020*. Jakarta: Kemdikbud.
- Ummul Aiman. (2025). Metodologi penelitian kuantitatif untuk ilmu komputer. *Jurnal Ilmiah Teknologi dan Sistem Informasi*, 12(1), 55–64.
- Universitas Islam Bandung (UNISBA). (2023). *Visi dan misi Fakultas Hukum*. Bandung: UNISBA.