

Klasifikasi Pola Bahasa pada Chatbot Perpustakaan Menggunakan Metode TF-IDF dan Support Vector Machine (SVM)

Eka Andy Yulianta¹, Muhammad Ghufon²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *¹andy68153@gmail.com , ²ghufon532@gmail.com

Abstrak – Perkembangan teknologi informasi mendorong perpustakaan untuk mendigitalkan layanannya dengan menggunakan chatbot yang didukung oleh kecerdasan buatan. Masalah utama dalam membuat chatbot adalah kemampuan sistem untuk memahami dan mengelompokkan ucapan pengguna secara tepat. Penelitian ini mencoba menggunakan dua metode, yaitu *Term Frequency–Inverse Document Frequency (TF-IDF)* dan *Support Vector Machine (SVM)*, untuk mengklasifikasikan jenis pertanyaan pengguna pada chatbot perpustakaan. Data yang digunakan adalah pertanyaan yang dibagi ke dalam dua kategori, yaitu *faq* dan *cari_buku*. Proses penelitian mencakup beberapa tahap, seperti mengolah teks, memberi bobot dengan TF-IDF, melatih model SVM, dan mengevaluasi hasilnya menggunakan *Confusion Matrix*. Hasil uji coba menunjukkan akurasi model mencapai 87,5%, yang menunjukkan metode ini efektif dalam membantu pengembangan chatbot perpustakaan berbasis teks.

Kata Kunci — Chatbot, Perpustakaan, SVM, TF-IDF

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong digitalisasi layanan pendidikan, termasuk sistem perpustakaan yang kini mengalami transformasi menuju platform digital berbasis kecerdasan buatan (AI)[1]. Kemajuan teknologi serta automasi sangat memengaruhi cara kerja perpustakaan, terutama di era saat ini yang ditandai dengan internet dan kecerdasan buatan. Hal ini mendorong pergeseran pengelolaan perpustakaan ke arah yang lebih berbasis teknologi informasi [2]. Salah satu solusi yang menjanjikan adalah dengan menggunakan aplikasi chatbot

Chatbot dirancang dengan kemampuan merekomendasikan buku atau referensi lainnya berdasarkan riwayat pencarian dan kebutuhan akademik siswa, sehingga dapat membantu mempercepat dan mempermudah proses pencarian informasi. Teknologi ini memanfaatkan pemrosesan bahasa alami untuk memungkinkan komputer memahami dan merespons bahasa sehari-hari, sehingga interaksi terasa seperti percakapan dengan manusia [3]. Hasil penelitian sebelumnya oleh [4] juga menunjukkan bahwa sistem chatbot dapat bekerja dengan baik, mampu berinteraksi dengan pengguna, dan menjawab pertanyaan mereka.

Meskipun demikian, pemanfaatan chatbot AI dalam pelayanan publik masih menghadapi sejumlah tantangan, seperti akurasi pemahaman bahasa, keandalan sistem dalam menjawab pertanyaan kompleks, serta isu keamanan dan privasi data [5]. Oleh karena itu, diperlukan sebuah mekanisme yang cerdas untuk mengklasifikasikan pola bahasa pengguna (*natural language processing/NLP*) secara akurat. Metode ini memungkinkan komunikasi antara manusia dan komputer dalam bahasa alami. Salah satu algoritma yang digunakan untuk menganalisis teks dan mengekstrak informasi penting dari dokumen adalah TF-IDF. *Term Frequency-Inverse Document Frequency (TF-IDF)* adalah teknik dalam text mining yang memberikan bobot pada kata-kata di dalam dokumen. Algoritma ini dikenal karena efisiensinya, kemudahannya penggunaannya, dan akurasi hasilnya. Metode ini menghitung nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* untuk setiap kata dalam setiap dokumen dalam korpus [6]. Selanjutnya pada penelitian Riyadh [7] membuktikan bahwa algoritma *Support Vector Machine (SVM)* efektif dalam meningkatkan akurasi evaluasi kualitas jawaban pada sistem chatbot.

Penelitian ini bertujuan menghasilkan model klasifikasi pola bahasa yang akurat sehingga chatbot perpustakaan dapat memberikan respons otomatis yang tepat dan relevan. Integrasi model ke dalam chatbot diharapkan meningkatkan kualitas layanan, mempercepat pencarian informasi, serta mendukung digitalisasi layanan perpustakaan.

2. METODE PENELITIAN

2.1 Pengumpulan data

Dataset diperoleh dari himpunan percakapan pengguna dengan chatbot perpustakaan yang telah diberi label sesuai kategori yaitu cari buku, dan lihat kebijakan Perustakaan. Jumlah dataset 30 kalimat sesuai standar klasifikasi pola bahasa

2.2 Pra-pemrosesan Teks

Tahapan prapemrosesan bertujuan untuk menyederhanakan teks agar lebih mudah diolah oleh algoritma. Pada penelitian ini digunakan prapemrosesan dasar agar tetap kompatibel dengan lingkungan eksekusi program, yaitu:

1. Konversi huruf menjadi lowercase
Semua teks diubah ke huruf kecil untuk konsistensi.
2. Pembersihan karakter dasar
Menghilangkan karakter yang tidak digunakan dalam pemrosesan (opsional).

Pendekatan ini digunakan agar model dapat tetap dilatih tanpa ketergantungan library eksternal seperti Sastrawi, mengingat pembatasan instalasi pada lingkungan eksekusi.

2.3 TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode untuk menghitung hubungan suatu kata (term) terhadap dokumen. Dari namanya metode ini terdiri dari 2 proses, *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). TF digunakan untuk mengukur seberapa sering suatu kata muncul dalam satu kalimat. Semakin tinggi frekuensi kemunculan suatu kata maka semakin tinggi nilai TF. IDF digunakan untuk mengukur seberapa sering suatu kata atau frekuensi suatu kata muncul dalam satu korpus [8]. Korpus adalah kumpulan kata yang disusun secara sistematis. Semakin rendah frekuensi kemunculan suatu kata pada korpus maka semakin tinggi nilai IDF

2.4 Train–Test Split

Dataset dibagi menjadi dua bagian:

1. 80% data latih (training)
2. 20% data uji (testing)

Pembagian ini bertujuan menguji kemampuan generalisasi model terhadap data baru yang belum pernah dipelajari.

2.5 SVM

Support Vector Machine (SVM) merupakan algoritma supervised learning yang banyak digunakan dalam klasifikasi teks karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan akurasi yang baik [9]. SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan data ke dalam kelas tertentu, sehingga cocok untuk klasifikasi pertanyaan pada chatbot [10].

2.6 Evaluasi

Tahap evaluasi bertujuan mengukur performa model chatbot setelah dilakukan pelatihan. Evaluasi dilakukan menggunakan *Confusion Matrix* dan *Cross-Validation*, yang selanjutnya digunakan untuk menghitung metrik seperti nilai akurasi, nilai presisi, dan nilai Recall.

$$\text{Akurasi} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots\dots\dots(1)$$

$$\text{Presisi} = \frac{TP}{(TP+FP)} \dots\dots\dots(2)$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \dots\dots\dots(3)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan data

Selama fase pengumpulan data, dataset dikumpulkan dari sekelompok pertanyaan yang biasanya diajukan pengguna kepada layanan chatbot perpustakaan. Data tersebut terdiri dari kalimat teks yang telah diberi label sesuai dengan kebutuhan pengguna, yaitu pertanyaan umum (FAQ) dan pencarian buku (cari_buku). Jumlah total kalimat data yang digunakan adalah 50, memenuhi standar minimal untuk penelitian klasifikasi pola bahasa skala kecil.

Dataset ini dirancang untuk mewakili pola bahasa alami yang digunakan pengguna saat berinteraksi dengan sistem chatbot perpustakaan. Setiap entri data mencakup pertanyaan yang dipasangkan dengan label kelas yang sesuai, sehingga cocok untuk digunakan sebagai data pelatihan dan pengujian dalam proses pembelajaran terawasi.

Tabel 3.1. Contoh data pengumpulan

pertanyaan	jawaban	label
jam berapa perpustakaan buka	Perpustakaan buka setiap hari Senin sampai Jumat pukul 08.00 sampai 16.00 WIB	faq
jam berapa perpustakaan tutup	Perpustakaan tutup setiap hari Sabtu Minggu dan hari libur nasional	faq
apakah tersedia buku kecerdasan buatan	Silakan cek ketersediaan buku kecerdasan buatan di katalog online	cari_buku

3.2 Pra-pemrosesan Teks

Hasil pra-pemrosesan menunjukkan bahwa semua teks pertanyaan berhasil diubah menjadi bentuk yang lebih konsisten. Semua kalimat diubah menjadi huruf kecil dan tidak ada karakter yang tidak penting, sehingga tidak ada perbedaan dalam penulisan kata. Teks yang dihasilkan menjadi lebih rapi dan siap digunakan pada tahap ekstraksi fitur dengan metode TF-IDF.

3.3 TF-IDF

Hasil ekstraksi fitur dengan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) menunjukkan bahwa setiap kalimat pertanyaan berhasil diubah menjadi vektor numerik berdimensi tinggi. Teknik ini memberikan bobot pada kata-kata yang penting dalam membedakan jenis pertanyaan dari pengguna.

Kata-kata yang muncul sering pada satu kategori tetapi tidak sering muncul di kategori lain mendapatkan bobot TF-IDF yang lebih besar. Hal ini membantu proses klasifikasi dengan memberikan kontribusi yang signifikan. Hasil ini menunjukkan bahwa metode TF-IDF mampu menangkap pola bahasa dalam dataset chatbot perpustakaan secara efektif, kata-kata penting diantaranya.

Tabel 3.2. Contoh Kata_kata penting

Kata penting	Nilai IDF
Yang bisa	3.614959778036198
Saya ingin	3.614959778036198
Ruang	3.614959778036198
Perpustakaan menyediakan	3.614959778036198
Layanan	3.614959778036198
Peminjaman buku	3.614959778036198
peminjaman	3.614959778036198

Vektor yang dihasilkan bersifat konsisten dan informatif, sehingga bisa digunakan secara optimal oleh algoritma *Support Vector Machine* (SVM) pada tahap klasifikasi. Dengan demikian, hasil dari TF-IDF memberikan dasar fitur yang kuat untuk meningkatkan kemampuan model dalam memahami maksud pertanyaan pengguna.

3.4 Train–Test Split

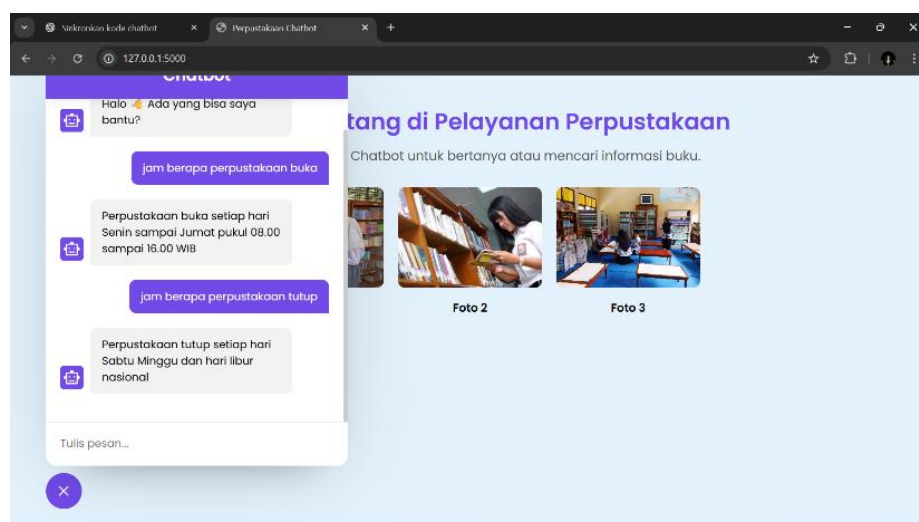
Hasil membagi data dengan metode train–test split menunjukkan bahwa dataset terbagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20% dari seluruh data. Data latih digunakan untuk membuat dan melatih model klasifikasi, sedangkan data uji digunakan untuk mengetahui seberapa baik model bisa mengklasifikasikan data yang belum pernah diajarkan sebelumnya.

Cara membagi data ini menjaga distribusi kelas tetap seimbang di data latih dan data uji, sehingga tidak ada bias terhadap salah satu kategori. Dengan cara pembagian ini, evaluasi model bisa dilakukan secara adil dan mampu menunjukkan performa model ketika digunakan di dunia nyata.

3.5 SVM

Hasil klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) menunjukkan bahwa model berhasil mempelajari pola cara orang berbicara berdasarkan fitur TF-IDF. Model mampu mengelompokkan pertanyaan pengguna ke dalam kategori yang tepat, seperti pertanyaan umum (FAQ).

Sebagian besar data uji berhasil diprediksi dengan benar oleh model, yang menunjukkan bahwa SVM efektif dalam memisahkan antar kategori pada data teks yang kompleks. Hasil ini membuktikan bahwa model mampu membedakan pola bahasa di setiap kategori pertanyaan, meskipun ada beberapa kalimat yang menggunakan kata serupa. Kemampuan model dalam melakukan klasifikasi secara akurat menjadi dasar keberhasilan sistem chatbot dalam memberikan jawaban yang tepat sesuai dengan pertanyaan pengguna.



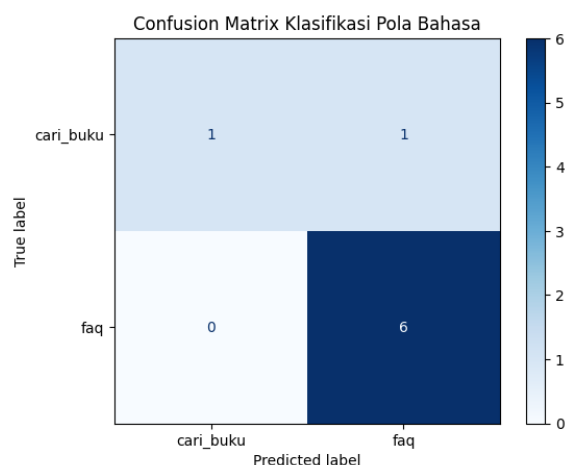
Gambar 3.1. Hasil chatbot dalam mencari jawaban

Gambar 3.1 adalah interaksi chatbot dengan pengguna untuk mencari informasi tentang perpustakaan. Pertama masuk halaman layanan online, selanjutnya jika ingin menanyakan informasi perpustakaan bisa langsung masuk halaman chatbot.

3.6 Evaluasi

Hasil evaluasi menunjukkan bahwa sistem chatbot yang menggunakan teknik text mining dengan kombinasi TF-IDF dan *Support Vector Machine* (SVM) berhasil mengklasifikasikan pola bahasa pengguna secara baik. Dari gambar 2 *Confusion Matrix*, terlihat bahwa semua data pada kelas *faq* terklasifikasi dengan benar, sedangkan pada kelas *cari_buku* ada satu data yang salah dikategorikan sebagai *faq*. Dari hasil tersebut, sistem mencapai akurasi sebesar 0,875 (87,5%), artinya sebagian besar data uji bisa diprediksi secara tepat. Tingkat keberhasilan yang tinggi pada kelas *faq* menunjukkan bahwa model mampu mengenali pola bahasa yang informatif dan berulang, serta mampu menangkapnya dengan baik melalui metode pembobotan TF-IDF.

Dari *Classification Report*, kelas *cari_buku* memiliki nilai *precision* 1,00, *recall* 0,50, dan *f1-score* 0,67. Hal ini menunjukkan bahwa meskipun prediksi cukup tepat, model masih kurang mampu mengenali semua variasi pola pertanyaan pencarian buku. Sebaliknya, kelas *faq* menunjukkan performa yang sangat baik dengan *precision* 0,86, *recall* 1,00, dan *f1-score* 0,92. Nilai *macro average f1-score* sebesar 0,79 serta *weighted average f1-score* sebesar 0,86 menunjukkan bahwa model memiliki kinerja yang stabil dan seimbang pada kedua kelas. Secara keseluruhan, hasil evaluasi ini membuktikan bahwa sistem klasifikasi pola bahasa yang dikembangkan cukup akurat, stabil, dan layak digunakan sebagai chatbot berbasis teks, serta bisa dipakai sebagai objek penelitian dalam kajian klasifikasi pola bahasa.



Gambar 3.2. Hasil evaluasi

Gambar 3.2 adalah *Confusion Matrix* menunjukkan bahwa model mengklasifikasikan data dengan baik, dengan performa sangat optimal pada kelas *faq* dan satu kesalahan klasifikasi pada kelas *cari_buku* akibat kemiripan pola bahasa.

4 SIMPULAN

Penelitian ini berhasil menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *Support Vector Machine* (SVM) untuk mengelompokkan pola ucapan dalam chatbot perpustakaan ke dalam dua kategori, yaitu *faq* dan *cari_buku*. Dari hasil pengevaluasian dengan *Confusion Matrix* dan *Classification Report*, sistem mencapai akurasi sebesar 87,5%. Performa sistem sangat bagus pada kategori *faq* dan cukup baik pada kategori *cari_buku*. Nilai *precision*, *recall*, dan *f1-score* menunjukkan bahwa model mampu mengenali maksud pengguna secara stabil, meskipun masih ada beberapa kesalahan klasifikasi karena adanya kesamaan pola ucapan antar kategori. Secara keseluruhan, hasil penelitian menunjukkan bahwa kombinasi TF-IDF dan SVM efektif dan layak digunakan dalam pengembangan chatbot perpustakaan berbasis kecerdasan buatan.

5 SARAN

Untuk penelitian selanjutnya, disarankan menggunakan dataset yang lebih besar dan berbagai jenis agar kemampuan model dalam mengenali berbagai cara orang berbicara menjadi lebih baik. Selain itu, dengan menerapkan langkah-langkah persiapan teks yang lebih lengkap, seperti pengurangan kata tidak penting dan pengubahan kata ke bentuk dasarnya dalam bahasa Indonesia, maka kualitas data dan tingkat keakuratan pengelompokan hasil akan meningkat. Penelitian berikutnya juga bisa mencoba menggunakan atau

membandingkan berbagai metode klasifikasi lain, seperti *Naïve Bayes*, *Random Forest*, atau pendekatan *deep learning*. Selain itu, sistem chatbot bisa dihubungkan langsung dengan layanan perpustakaan agar hasil pengelompokan bisa digunakan secara langsung dan cepat.

DAFTAR PUSTAKA

- [1] A. Zulfahrizan, M. I. Drilanang, F. B. Hutagalung, dan A. Perdana, “PENGEMBANGAN APLIKASI LIBRARY INTEGRATED MANAGEMENT (LIM) SEBAGAI ALAT MANAJEMEN PEMINJAMAN BUKU DENGAN INTEGRASI AI CHATBOT BERBASIS WEB,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, hlm. 6014–6020, 2025.
- [2] O. Alvansyah, N. A. Yolandari, M. F. Zulfi, A. N. Nasution, dan A. Perdana, “Pembuatan Website Peminjaman Buku Perpustakaan Digital dengan Integrasi Fitur Kecerdasan Buatan untuk Meningkatkan Literasi,” *Jurnal Bisnis Digital dan Entrepreneur (BISENTER)*, vol. 3, no. 2, hlm. 341–347, 2025.
- [3] S. H. Bariah, W. Pratiwi, dan K. A. N. Imania, “Pengembangan Virtual Assistant Chatbot Berbasis Whatsapp Pada Pusat Layanan Informasi Mahasiswa Institut Pendidikan Indonesia-Garut,” *Petik: Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, vol. 8, no. 1, hlm. 66–79, 2022.
- [4] D. Kristanto, R. A. Ramadhani, dan A. B. Setiawan, “Pengembangan Chatbot Layanan Informasi Kampus Menggunakan TF-IDF”, doi: 10.33364/algoritma/v.22-2.2350.
- [5] F. Hulu, “Pemanfaatan Chatbot Ai Berbasis Bahasa Indonesia Untuk Pelayanan Publik Digital,” *Jurnal Ilmu Artificial Intelligence Nusantara*, vol. 1, no. 1, hlm. 25–32, 2025.
- [6] M. S. Maulana, Y. Anshori, R. Azhar, R. Laila, dan N. T. Lapatta, “IMPLEMENTASI PEMBOBOTAN TF-IDF PADA CHATBOT TELEGRAM UNTUK SISTEM LAYANAN INFORMASI,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 3, hlm. 1869–1877, Agu 2025, doi: 10.29100/jupi.v10i3.6314.
- [7] M. Riyadh dan R. Artikel, “PENGUNAAN SUPPORT VECTOR MACHINE (SVM) DALAM SISTEM QUALITY ASSURANCE (QA) CHAT BOT DAN VOICE BOT UNTUK MENINGKATKAN AKURASI RESPON INFORMASI ARTIKEL ABSTRAK.”
- [8] M. R. A. F. Rifqi dan M. R. Al Fauzan, “Kombinasi TF-IDF dan Neural Network Untuk Pelayanan Informasi Al-Qur’an Dalam Bentuk Chatbot,” *JURNAL FASILKOM*, vol. 14, no. 2, hlm. 318–324, 2024.
- [9] F. Abdusyukur, “Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter,” *Komputa: Jurnal Ilmiah Komputer Dan Informatika*, vol. 12, no. 1, hlm. 73–82, 2023.
- [10] N. Setiawan, M. Akbar, A. Anto Tri Susilo, J. K. Jendral Besar Moh Soeharto, K. Lubuk Linggau, dan S. Selatan, “Pengembangan Chatbot Untuk Layanan Informasi Keanggotaan Guru Metode Support Vector Machine.”