P-ISSN : 2549-7480
E-ISSN : 2549-6301

JURNAL ILMIAH INFORMATIKA
E-ISSN : 2549-6301
ISSN : 2549-7480
Faculty of Sains and Technology, Ibrahimy University
S4-Accredited : 105/E/KPT/2022
Journal page is available to : https://journal.ibrahimy.ac.id/index.php/JIMI

KOMPARASI EMPAT ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI INTENT PADA CHATBOT INFORMASI AKADEMIK

Refaya Septianno[1)], Risanto Darmawan[2)], Elmi Devia[3)]

[1] *Teknik Informatika, Universitas Krisnadwipayana*
[2] *Teknik Informatika, Universitas Krisnadwipayana*
[3] *Sistem Informasi, Universitas Krisnadwipayana*

email: [1] refayaseptiano@gmail.com, [2] risantodarmawan@unkris.ac.id, [3] elmidevia@unkris.ac.id

ARTICLE INFO

Article History:
Recieved : 1 November 2026
Accepted : 6 November 2026
Published : 15 November 2026

Keywords:
Chatbot;
Intent Calssification;
Machine Learning;
NLP;
Academic Information.

IEEE style in citing this article:
R. Septianno, R. Darmawan, E. Devia, "Komparasi Empat Algoritma Machine Learning untuk Klasifikasi Intent pada Chatbot Informasi Akademik", *Jurnal.Ilmiah.Informatika*, vol. 11, no. 2, pp. 211-xx, Des. 2026.

Corresponding Author:
Refaya Septianno
Teknik Informatika
Universitas Krisnadwipayana

ABSTRACT

The development of Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies has driven the utilization of chatbots as an automated academic information service medium. A chatbot's success in understanding user queries is heavily influenced by the model's ability to classify intents. This study focuses on comparing the performance of four machine learning algorithms namely Naive Bayes, Logistic Regression, Support Vector Machine (SVM), and XGBoost in the intent classification task for an Indonesian-language academic information chatbot. The methodological approach employed is CRISP-DM, which encompasses the Business Understanding, Data Understanding, Data Preparation, Modelling, and Evaluation phases. The dataset structure is organized in a JSON format containing intent labels, variations of query patterns, and their corresponding academic responses. The data preparation procedure includes case folding, tokenization, stopword removal, and stemming using the Sastrawi library, followed by TF-IDF-based feature extraction. Model evaluation is performed using accuracy, precision, recall, and F1-score metrics to determine which algorithm delivers optimal performance in accurately classifying intents.

© 2026 Jurnal Ilmiah Informatika (Scientific Informatics Journal) with CC BY NC licence

P-ISSN : 2549-7480
R. Septianno dkk / JIMI 11 (2) pp.211-220 E-ISSN : 2549-6301

1. PENDAHLUAN

Disrupsi teknologi informasi dan pemanfaatan kecerdasan buatan telah mentransformasi lanskap layanan akademik di lingkungan perguruan tinggi. Dominasi perkembangan ini terlihat pada adopsi *chatbot* berteknologi *Natural Language Processing* (NLP) yang menawarkan otomatisasi layanan informasi berbasis dialog natural. Melalui penerapan cabang kecerdasan buatan ini, sistem memiliki kapasitas analitis dan generatif terhadap linguistik manusia. Kondisi tersebut memfasilitasi *chatbot* untuk mengidentifikasi pertanyaan sivitas akademika secara presisi dan menyajikan jawaban yang relevan [1].

Kebutuhan terhadap layanan informasi akademik yang cepat dan tersedia sepanjang waktu semakin mendesak seiring bertambahnya jumlah mahasiswa baru pada setiap periode penerimaan. Pada periode tersebut, staf akademik umumnya menerima volume pertanyaan yang tinggi dan bersifat repetitif mengenai jurusan, program studi, biaya pendidikan, dan beasiswa, sementara jam operasional layanan tetap terbatas pada jam kerja dan tidak semua mahasiswa aktif mengakses laman resmi fakultas untuk mencari informasi tersebut. Permasalahan tersebut berimplikasi pada tingginya beban administratif yang harus ditanggung oleh staf akademik. Oleh karena itu, ketepatan respons *chatbot* dalam menjawab berbagai pertanyaan sangat ditentukan oleh efektivitas algoritma klasifikasi teks yang diterapkan untuk mendeteksi maksud (*intent*) dari setiap kueri pengguna .

Penerapan beragam algoritma *machine learning* menunjukkan hasil yang optimal dalam proses klasifikasi *intent*. Berdasarkan kajian Hillary dkk. [2], *Logistic Regression* menawarkan performa yang unggul, kemudahan penerapan, serta efisiensi komputasi yang baik untuk pengolahan teks. Keunggulan dalam mengklasifikasikan teks serta mengenali *intent* percakapan turut ditunjukkan oleh algoritma *Naive Bayes*. Sementara itu, konsistensi tingkat akurasi yang lebih tinggi pada teks berbahasa Indonesia berhasil dicapai oleh *Support Vector Machine* [3]. Performa komparatif juga ditunjukkan oleh *XGBoost* yang memanfaatkan mekanisme *ensemble boosting*, di mana algoritma ini mampu memberikan efisiensi komputasi yang tinggi diiringi tingkat akurasi yang kompetitif [4].

Meskipun keempat algoritma tersebut telah diteliti secara terpisah, belum terdapat penelitian yang secara khusus membandingkan *Naive Bayes*, *Logistic Regression*, SVM, dan *XGBoost* secara bersamaan dalam konteks klasifikasi *intent* pada *chatbot* informasi akademik berbasis Bahasa Indonesia, terlebih pada kondisi *dataset* berskala kecil yang justru paling merepresentasikan situasi riil yang dihadapi banyak perguruan tinggi pada tahap awal pengembangan sistem. Tanpa perbandingan yang sistematis dan terukur pada kondisi tersebut, pengembang sistem tidak memiliki landasan yang kuat untuk menentukan algoritma terbaik, sehingga keputusan yang diambil cenderung bersifat trial-and-error alih-alih berbasis bukti.

Penelitian ini untuk menjawab kebutuhan tersebut. Urgensi penelitian ini terletak pada penyediaan bukti empiris perbandingan performa keempat algoritma secara bersamaan, pada domain informasi akademik, menggunakan *dataset* pertanyaan mahasiswa berbahasa Indonesia berskala kecil yang realistis, dengan menerapkan *pipeline* pra-pemrosesan yang komprehensif serta mengevaluasi performa keempat algoritma secara menyeluruh menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

DOI : https://doi.org/10.35316/jimi.v11i2.211-220
© 2026 Jurnal Ilmiah Informatika with CC BY-NC licence

Penelitian terdahulu yang menjadi dasar pembahasan mencakup pengembangan *chatbot* penerimaan mahasiswa baru, *chatbot* informasi akademik, dan klasifikasi intensi berbahasa Indonesia [5]. Kajian tersebut menunjukkan keberagaman domain dan pendekatan yang digunakan, sedangkan penelitian ini secara khusus membandingkan empat algoritma pada data pertanyaan akademik yang sama.

2. METODE PENELITIAN

Metodologi yang digunakan dalam penelitian ini berbasis kuantitatif dengan penekanan pada uji komparatif kinerja empat *model* pembelajaran mesin, meliputi *Naive Bayes*, *Logistic Regression*, SVM, dan *XGBoost*, dalam mengimplementasikan klasifikasi *intent* untuk sistem tanya jawab akademik otomatis berbahasa Indonesia. Eksperimen ini dilakukan secara objektif dengan menerapkan keempat algoritma tersebut pada satu *dataset* yang seragam di bawah kondisi pengujian yang identik. Selanjutnya, analisis dilakukan secara kuantitatif dengan membandingkan perolehan nilai dari empat metrik evaluasi utama, yakni *accuracy*, *precision*, *recall*, dan *F1-score*. [13]

Permasalahan utama yang diidentifikasi adalah belum tersedianya sistem layanan informasi akademik yang otomatis dan responsif di perguruan tinggi, serta belum adanya perbandingan sistematis antara algoritma *Naive Bayes*, *Logistic Regression*, SVM, dan *XGBoost* dalam konteks klasifikasi *intent* pada *chatbot* informasi akademik berbasis Bahasa Indonesia. Berdasarkan identifikasi tersebut, tujuan penelitian dirumuskan sebagai berikut:

Mengidentifikasi algoritma dengan tingkat efektivitas superior berdasarkan parameter akurasi, presisi, *recall*, dan *F1-score*.

Setelah melakukan tahap understanding maka pada tahapan ini masing-masing dari gambar pada *dataset* dilakukan *preprocessing* agar gambar yang ada dapat digunakan dengan baik pada tahapan selanjutnya, adapun langkah-langkah pada tahap *preprocessing* dapat digambarkan. [12]

Berlanjut dari tahap pembersihan teks, representasi data ditransformasikan ke dalam format vektor angka melalui pemanfaatan metode ekstraksi fitur *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Prosedur ini berperan dalam menentukan bobot atribut yang menggambarkan tingkat kepentingan suatu kata pada dokumen spesifik dibandingkan dengan total korpus secara global. Dibandingkan dengan pendekatan bag-of-words sederhana, *TF-IDF* memiliki keunggulan komparatif dalam memberikan penekanan bobot yang lebih signifikan terhadap kosakata yang unik dan kaya informasi, serta mereduksi dampak dari kata-kata umum yang kurang relevan. [2]

Pada Gambar 1 fase *Modelling* merupakan inti dari penelitian ini, di mana keempat algoritma *machine learning* dilatih menggunakan data latih yang telah disiapkan pada fase sebelumnya.

Fase *Evaluation* merupakan tahap terakhir dalam kerangka *CRISP-DM* yang diterapkan dalam penelitian ini. Sebagai tahap penutup dari rangkaian eksperimen, seluruh *model* yang sudah dilatih dievaluasi secara menyeluruh dengan memanfaatkan data uji dan parameter metrik performa. Hasil dari evaluasi ini kemudian dianalisis dan dibandingkan secara komparatif untuk menarik kesimpulan yang menjawab tujuan penelitian. [11]

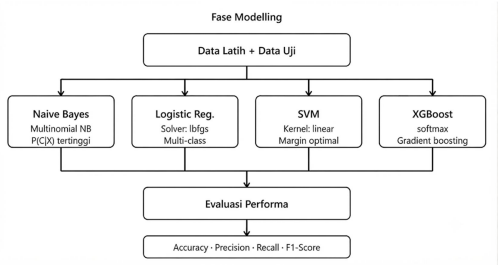
Identifikasi kategori kebutuhan informasi menentukan delapan kategori *intent* yang paling relevan dengan kebutuhan informasi akademik mahasiswa, yaitu sapaan, program studi, jadwal akademik, biaya pendidikan, beasiswa, kontak, pendaftaran, dan rekomendasi jurusan.

DOI : https://doi.org/10.35316/jimi.v11i2.211-220
© 2026 Jurnal Ilmiah Informatika with CC BY-NC licence

P-ISSN : 2549-7480
R. Septianno dkk / JIMI 11 (2) pp.211-220 E-ISSN : 2549-6301

Studi dokumentasi informasi resmi kampus menelusuri informasi resmi yang tersedia pada laman nusamandiri.ac.id untuk memperoleh data faktual mengenai program studi, biaya pendidikan, jalur dan jadwal pendaftaran, kalender akademik, serta kontak layanan, yang selanjutnya menjadi acuan dalam menyusun isi basis pengetahuan.

Pelabelan setiap pola kalimat diberi satu *label intent* yang sesuai, kemudian divalidasi ulang untuk memastikan tidak ada pertanyaan yang memiliki *label* ganda atau ambigu.



Gambar 1. Diagram pemodelan empat algoritma

3. HASIL DAN PEMBAHASAN

Seluruh rangkaian implementasi penelitian dikerjakan dengan memanfaatkan bahasa pemrograman *Python* dalam ekosistem *Google Colab*, serta didukung oleh beberapa pustaka penting, antara lain:

Secara kumulatif, keseluruhan *dataset* terdiri dari 560 sampel data yang terdistribusi dan dikategorikan ke dalam sejumlah kelas *intent* informasi akademik yang relevan. Seluruh data telah melalui proses validasi *label* sehingga setiap pertanyaan hanya memiliki satu kelas *intent*. Keseluruhan *dataset* ini kemudian diaplikasikan sebagai basis data untuk menjalankan siklus pengolahan lanjutan, yang mencakup tahap *preprocessing*, ekstraksi fitur, pelatihan *model* (*training*), hingga pengukuran performa komparatif dari keempat algoritma *machine learning*.

Dalam hasil penelitian ditunjukkan struktur *dataset* yang digunakan dalam penelitian. Data disusun dalam format *JSON* yang terdiri dari komponen *tag*, *patterns*, dan *responses*. Bagian *tag* berfungsi sebagai *label intent*, *patterns* berisi variasi pertanyaan yang dapat diajukan pengguna, sedangkan *responses* berisi jawaban yang akan diberikan *chatbot* ketika *intent* berhasil dikenali. Struktur data ini dipilih karena mampu merepresentasikan data percakapan secara terorganisasi dan memudahkan proses pengolahan pada tahap klasifikasi. [7]

3.1 Persiapan dan Representasi Data

Seluruh teks dalam *dataset* diproses melalui empat tahap pra-pemrosesan yang dilaksanakan secara berurutan. Fase ini berfokus pada pembersihan dan normalisasi struktur teks demi menghasilkan representasi linguistik yang konsisten sebelum memasuki tahap ekstraksi fitur. Proses tersebut dijabarkan ke dalam empat tahapan penting, yakni *case folding*, tokenisasi, penyisihan kata henti (*stopword removal*), dan proses *stemming* menggunakan pustaka *Sastrawi*. [10]

Gambar 2 dibawah menunjukkan pengaruh parameter max_features pada metode TF-IDF terhadap nilai F1-Score algoritma Support Vector Machine (SVM). Parameter max_features digunakan untuk membatasi jumlah kata atau fitur yang dipilih dari hasil ekstraksi TF-IDF sebelum digunakan dalam proses pelatihan model. Tujuannya adalah untuk mengetahui jumlah fitur yang paling optimal dalam merepresentasikan karakteristik setiap intent pada dataset chatbot informasi akademik.

Terlihat bahwa nilai *F1-Score* mengalami peningkatan yang cukup signifikan seiring bertambahnya jumlah fitur. Pada penggunaan 50 fitur, *model* menghasilkan *F1-Score* sekitar 0,79, yang menunjukkan bahwa informasi yang digunakan *model*

masih terbatas sehingga kemampuan klasifikasi belum optimal. Ketika jumlah fitur ditingkatkan menjadi 100 dan 200, nilai *F1-Score* meningkat menjadi sekitar 0,90 dan 0,93.

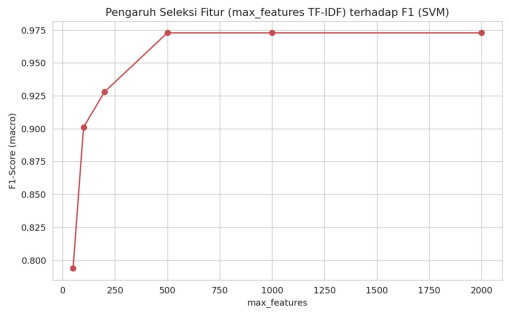
Peningkatan performa yang lebih tinggi terlihat pada penggunaan 500 fitur, di mana nilai *F1-Score* mencapai sekitar 0,97. Setelah melewati 500 fitur, yaitu pada 1000 dan 2000 fitur, nilai *F1-Score* relatif tidak mengalami perubahan yang berarti dan cenderung stabil pada kisaran 0,97.

Oleh karena itu, nilai *max_features* sebesar 500 dapat dianggap sebagai titik optimal karena mampu menghasilkan *F1-Score* yang tinggi dengan kompleksitas komputasi yang lebih efisien dibandingkan penggunaan 1000 atau 2000 fitur.

Mengacu pada Tabel 1, langkah yang dilakukan seusai ekstraksi fitur adalah memecah keseluruhan *dataset* ke dalam dua kelompok menggunakan teknik *train-test split*. Rasio yang diterapkan pada metode ini menetapkan porsi sebesar 80% sebagai data latih, sementara 20% dialokasikan untuk keperluan data uji.

Tabel 1. Pembagian data latih dan data uji

Data Latih	Data Uji	Kelas
448	112	Beasiswa, biaya pendidikan, jadwal akademik, kontak, pendaftaran, program studi, rekomendasi jurusan, sapaan



Gambar 2. Ekstraksi fitur TF-IDF

3.2 Pemodelan Empat Algoritma

Keunggulan utama *Naive Bayes* terletak pada kecepatan dan kesederhanaan komputasinya.

Keunggulan penting yang ditawarkan oleh *Logistic Regression* dan selaras dengan kebutuhan penelitian ini terletak pada kemampuannya menyajikan nilai probabilitas prediksi yang akurat dan terkalibrasi.

Pemilihan kernel linear didasarkan pada karakteristik representasi *TF-IDF* pada ruang fitur berdimensi tinggi yang umumnya telah mampu menghasilkan keterpisahan data secara linear, sehingga penggunaan transformasi kernel yang lebih kompleks seperti RBF atau polinomial menjadi tidak diperlukan lagi. Penerapan parameter regularisasi C berfungsi untuk mengatur trade-off krusial antara memaksimalkan jarak margin dengan menekan seminimal mungkin kesalahan klasifikasi pada data latih, yang pada akhirnya memastikan *model* tidak mengalami *overfitting* terhadap data pelatihannya.

Karena *XGBoost* memerlukan *label* dalam format numerik, *label intent* terlebih dahulu dikodekan menggunakan LabelEncoder sebelum pelatihan. Kelebihan *XGBoost* dibandingkan algoritma *boosting* konvensional adalah tersedianya mekanisme regularisasi bawaan yang membantu mencegah *overfitting*, serta dukungan komputasi paralel yang membuatnya lebih cepat dari implementasi gradient *boosting* tradisional.

3.3 Evaluasi *Model* dan *Confusion Matrix*

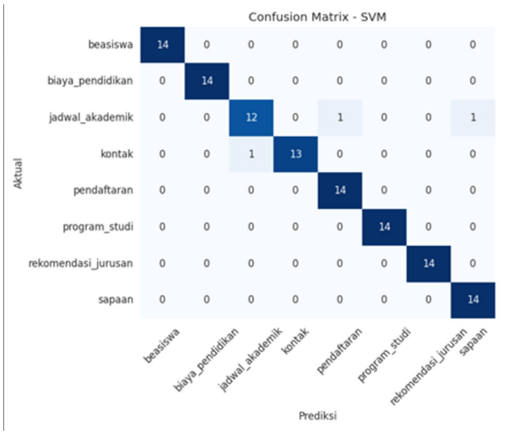
Pada Gambar 3 sebagian besar kategori *intent* berhasil diprediksi dengan benar. Kategori beasiswa, biaya_pendidikan, pendaftaran, program_studi, rekomendasi_jurusan, dan sapaan masing-masing memiliki 14 data uji yang

P-ISSN : 2549-7480
R. Septianno dkk / JIMI 11 (2) pp.211-220 E-ISSN : 2549-6301

seluruhnya diklasifikasikan dengan tepat. Lebih lanjut, pada kategori jadwal_akademik, *model* sukses mengenali 12 data dengan benar, sementara terdapat 2 data yang keliru dikategorikan masing-masing 1 data sebagai pendaftaran dan 1 data sebagai sapaan. Sementara itu, pada kategori kontak, sebanyak 13 data berhasil diprediksi secara akurat, disertai 1 data yang meleset dan diklasifikasikan sebagai jadwal_akademik.

Meninjau matriks konfusi (*confusion matrix*) untuk algoritma *XGBoost* Dalam hasil penelitian, tampak bahwa *model* berkinerja sangat optimal dalam mengelompokkan sebagian besar data uji. *Intent* rekomendasi_jurusan dan sapaan memperoleh hasil terbaik dengan masing-masing 14 data terklasifikasi benar tanpa kesalahan, sedangkan *intent* kontak dan pendaftaran memiliki 13 prediksi benar. Selain itu, *intent* jadwal_akademik berhasil dikenali sebanyak 12 data, *intent* program_studi sebanyak 12 data, *intent* biaya_pendidikan sebanyak 11 data, dan *intent* beasiswa sebanyak 10 data.

Meninjau *classification report* Dalam hasil penelitian, SVM terbukti unggul sebagai *model* terbaik dengan mencatatkan tingkat akurasi mencapai 97% pada data uji. Lebih lanjut, nilai rata-rata makro (*macro average*) dan tertimbang (*weighted average*) untuk parameter *precision*, *recall*, serta *F1-score* turut menyentuh angka 0,97, yang menegaskan kestabilan performa *model* secara merata pada setiap kategori *intent*.



Gambar 3. Confusion matrix SVM

3.4 Perbandingan Performa dan Analisis *Model* Terbaik

Lebih lanjut, Tabel 2 merangkum perbandingan kinerja empat algoritma *machine learning*, yakni *Naive Bayes*, *Logistic Regression*, *Support Vector Machine* (SVM), dan *XGBoost*. Berdasarkan evaluasi menggunakan parameter *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, algoritma SVM mencatatkan performa terbaik dengan skor akurasi 0,97, serta nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,97. Hasil tersebut memperlihatkan bahwa SVM tidak hanya handal dalam mengelompokkan *intent* pertanyaan mahasiswa secara akurat, tetapi juga memiliki stabilitas yang baik dalam menyeimbangkan aspek ketepatan prediksi (*precision*) dan tingkat kelengkapan pengenalan kelas (*recall*).

Di posisi berikutnya, algoritma *Logistic Regression* berada di urutan kedua dengan mencatatkan tingkat akurasi 0,96, serta nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,96.

Di sisi lain, algoritma *XGBoost* berada di posisi terbawah dengan perolehan skor akurasi 0,88, *precision* 0,89, *recall* 0,88, dan *F1-score* 0,88. Temuan tersebut memperlihatkan bahwa *model XGBoost* belum optimal dalam menangani *dataset* pada penelitian ini. Salah satu kendala utamanya disinyalir dari ukuran *dataset*

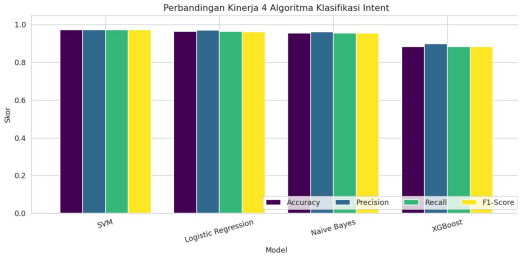
yang relatif kecil, yaitu sekitar 560 data sampel, sehingga mekanisme *boosting* yang biasanya memerlukan skala data yang masif belum mampu bekerja secara optimal.

Meninjau keseluruhan hasil perbandingan, algoritma *Support Vector Machine* (SVM) terbukti keluar sebagai *model* dengan performa terbaik. Hal ini diperkuat oleh pencapaian metrik evaluasi mencakup akurasi, *precision*, *recall*, serta *F1-score* yang mencatatkan skor paling unggul melampaui algoritma-algoritma pembanding lainnya. Dengan demikian, SVM dipilih sebagai *model* terbaik untuk diterapkan pada *chatbot* informasi akademik dalam penelitian ini.

Berdasarkan visualisasi pada Gambar 4, disajikan perbandingan nilai akurasi, *precision*, *recall*, dan *F1-score* dari keempat algoritma *machine learning* yang diuji dalam klasifikasi *intent chatbot* informasi akademik.

Tabel 2. Perbandingan hasil evaluasi keempat model

Model	Accuracy	Precision	Recall	F1-score
SVM	0,97	0,97	0,97	0,97
Logistic Regression	0,96	0,96	0,96	0,96
Naive Bayes	0,95	0,96	0,95	0,95
XGBoost	0,88	0,89	0,88	0,88



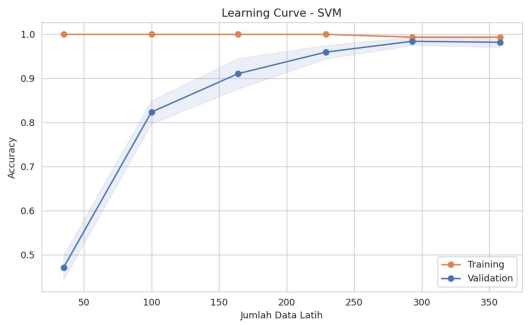
Gambar 4. Perbandingan kinerja empat algoritma

3.5 *Overfitting*, *Underfitting*, dan *Learning Curve*

Berdasarkan hasil evaluasi, SVM terbukti sebagai *model* dengan tingkat kestabilan tertinggi, ditunjukkan oleh perolehan train *accuracy* di kisaran 0,99, *cross-validation* 0,98, serta test *accuracy* 0,97. Kesenjangan yang minim di antara ketiga metrik ini membuktikan bahwa *model* sanggup menangkap pola informasi dengan sangat baik sekaligus terhindar dari permasalahan *overfitting* yang signifikan. Kondisi serupa juga terlihat pada *Logistic Regression* dan *Naive Bayes*, yang memiliki perbedaan nilai relatif kecil sehingga kemampuan generalisasinya masih tergolong baik.

Sebaliknya, *XGBoost* menunjukkan selisih yang lebih besar antara train *accuracy* (1,00) dengan *cross-validation* (0,90) dan test *accuracy* (0,88). Gejala ini menandakan adanya potensi *overfitting*, di mana *model* menunjukkan performa yang sangat tinggi saat mempelajari data latih, tetapi gagal mempertahankan tingkat akurasi yang sama ketika dihadapkan pada data pengujian baru.

Pada Gambar 5 bahwa nilai *training accuracy* pada algoritma SVM berada pada kisaran 0,99–1,00 untuk seluruh jumlah data latih. Sementara, validation *accuracy* mengalami peningkatan seiring bertambahnya jumlah data latih. Pada saat jumlah data latih masih sedikit, nilai validation *accuracy* relatif rendah, yaitu sekitar 0,47, namun meningkat menjadi 0,82, 0,91, dan 0,96 ketika jumlah data latih bertambah.



Gambar 5. Learning curve SVM

P-ISSN : 2549-7480
R. Septianno dkk / JIMI 11 (2) pp.211-220 E-ISSN : 2549-6301

3.6 Implementasi *Chatbot* Informasi Akademik

Sebagai tahap penutup dalam rangkaian metodologi Cross Industry Standard Process for *Data Mining* (*CRISP-DM*), fase *deployment* berfokus pada penerapan *model* paling optimal hasil pengujian komparatif ke lingkungan sistem *chatbot akademik*. *Model* terbaik tersebut diintegrasikan langsung ke dalam arsitektur aplikasi *chatbot*, sehingga memungkinkan sistem untuk mengenali dan mengklasifikasikan *intent* dari ragam pertanyaan pengguna secara otomatis dan adaptif.

Begitu *intent* berhasil diidentifikasi secara akurat, langkah berikutnya adalah pencarian respons yang sesuai pada basis pengetahuan (knowledge base) sistem yang dikelola dan disimpan menggunakan format berkas *JSON*. Jawaban yang memiliki *intent* yang sama dengan hasil prediksi kemudian dikirimkan kembali ke antarmuka *chatbot* sebagai respons kepada pengguna. Dengan mekanisme tersebut, *chatbot* mampu memberikan informasi akademik secara otomatis, cepat, dan konsisten tanpa memerlukan intervensi langsung dari staf administrasi.

Pada Gambar 6 halaman *chatbot* terdiri atas beberapa komponen utama yang saling mendukung proses interaksi pengguna dengan sistem. Pada bagian sisi kiri halaman, disediakan menu navigasi utama yang berfungsi untuk memudahkan pengguna berpindah antarhalaman aplikasi secara fleksibel, yang mencakup menu *Chatbot*, *Admin Dashboard*, dan Live Compare. Adanya fitur navigasi ini memberikan kemudahan bagi pengguna maupun administrator untuk menjangkau berbagai fasilitas yang disediakan secara fleksibel sesuai keperluan.

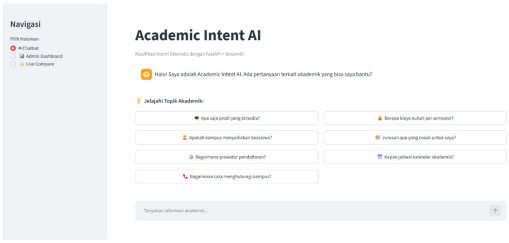
Dalam hasil penelitian pengguna memasukkan pertanyaan "Jurusan apa yang cocok untuk saya?" melalui kolom percakapan pada halaman *chatbot*. Hasil klasifikasi menunjukkan bahwa sistem berhasil mengidentifikasi pertanyaan tersebut sebagai *intent* rekomendasi_jurusan dengan tingkat *confidence* sebesar 96,3%. Besaran nilai *confidence* ini berfungsi untuk mengindikasikan seberapa tinggi tingkat kepercayaan *model* terhadap keputusan klasifikasi atau prediksi yang diberikan kepada pengguna.

Dalam hasil penelitian pengguna mengajukan pertanyaan "Bagaimana prosedur pendaftaran?" melalui halaman *chatbot*. Hasil klasifikasi menunjukkan bahwa sistem berhasil mengidentifikasi pertanyaan tersebut sebagai *intent* pendaftaran dengan nilai *confidence* sebesar 98,2%. Tingginya nilai *confidence* tersebut mengindikasikan bahwa *model* mempunyai tingkat kepercayaan yang kuat terhadap ketepatan keputusan klasifikasi yang dihasilkan.

Dalam hasil penelitian hasil klasifikasi menunjukkan bahwa sistem berhasil mengenali pertanyaan tersebut sebagai *intent* jadwal_akademik dengan nilai *confidence* sebesar 99,7%. Lonjakan nilai *confidence* ke tingkat yang sangat tinggi mengindikasikan bahwa *model* memiliki derajat kepercayaan yang kuat dan optimal terhadap ketepatan keputusan klasifikasi yang diberikan.

Hasil implementasi pada beberapa skenario pengujian, yaitu rekomendasi jurusan, pendaftaran, dan jadwal akademik, menunjukkan bahwa sistem mampu mengidentifikasi *intent* dengan baik dan menghasilkan respons yang sesuai dengan kebutuhan pengguna. Penerapan ini secara konkret mengonfirmasi bahwa *model* unggulan yang diperoleh dari proses perbandingan tidak hanya unggul pada metrik evaluasi pengujian, melainkan benar-benar aplikatif dan berfungsi secara optimal di dalam ekosistem layanan *chatbot* informasi akademik.

DOI : https://doi.org/10.35316/jimi.v11i2.211-220
© 2026 Jurnal Ilmiah Informatika with CC BY-NC licence



Gambar 6. Halaman chatbot Academic Intent AI

4. KESIMPULAN

Penelitian ini berhasil menjawab permasalahan mengenai algoritma *machine learning* yang paling tepat untuk mengklasifikasikan *intent* pada *chatbot* informasi akademik. Berdasarkan serangkaian uji coba, algoritma *Support Vector Machine* (SVM) terkonfirmasi unggul dalam memberikan klasifikasi *intent* yang sangat akurat dan stabil, yang pada gilirannya memungkinkan sistem *chatbot* untuk menyajikan jawaban relevan sesuai dengan esensi pertanyaan mahasiswa.

Evaluasi yang mengacu pada empat parameter metrik *accuracy*, *precision*, *recall*, dan *F1-score* mengonfirmasi bahwa *Support Vector Machine* (SVM) keluar sebagai *model* terbaik dengan masing-masing nilai metrik mencapai 0,97. Angka ini melampaui performa *Logistic Regression* di angka 0,96, *Naive Bayes* sebesar 0,95, dan *XGBoost* pada level 0,88. Adapun *XGBoost* memperlihatkan hasil terendah akibat rentan mengalami *overfitting* ketika memproses *dataset* berskala kecil. Dengan demikian, SVM direkomendasikan sebagai algoritma utama untuk diterapkan sebagai inti mesin klasifikasi *intent* pada sistem *chatbot* informasi akademik di perguruan tinggi.

5. REFERENSI

[1] Munawirah, M., Mardjani, E., & Kristian, K, "Aplikasi Chatbot Berbasis Dialogflow dan NLP untuk Pelayanan Informasi Akademik pada Fakultas Ilmu Komputer Universitas Tomakaka", 99–115, 2025.

[2] Hillary, R., Wijaya, A. C., & Susanto, M. W, "Pengembangan Chatbot Analisis Data Mahasiswa dengan Term Frequency - Inverse Document Frequency dan Logistic Regression [Development of Student Data Analysis Chatbot Using Term Frequency - Inverse Document Frequency and Logistic Regression]", 01, 875–885, 2026.

[3] Hidayat, R., Mahdiana, D., & Fergina, A, "Comparative Analysis of Logistic Regression , SVM , Xgboost , and Random Forest Algorithms for Diabetes Classification", 7(1), 281–291. https://doi.org/10.32493/jtsi.v7i1.38258, 2024.

[4] Armiady, D, "Analisis Algoritma Logistic Regression Dan Support Vector Machine Pada Kasus Klasifikasi Citra Hewan Rawa Dengan Dataset Yang Tidak Seimbang", Data Sciences Indonesia (DSI), 4(1), 69-77, 2024.

[5] Al Farisi, F. A., Perdana, R. S., & Adikara, P. P, "Klasifikasi Intensi dengan Metode Long Short-Term Memory pada Chatbot Bahasa Indonesia", Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 11(5), 1051–1058. https://doi.org/10.25126/jtiik.2024118000, 2024.

[6] Fauzan, M. F., Imanda, R., & Hasbi, M. A, "Designing an Chatbot with NLP Technology in a Website-Based New Student Admission Information System", 8(2), 358–366, 2024.

[7] Jtik, J., Teknologi, J., Qalimaturrahmah, M., & Santoso, D. B, "Aplikasi Layanan dan Informasi Akademik Berbasis Chatbot Telegram

P-ISSN : 2549-7480
R. Septianno dkk / JIMI 11 (2) pp.211-220 E-ISSN : 2549-6301

Menggunakan Natural Language Processing", 8(2), 2024.

[8] Setiyorini, A, "Pengembangan Aplikasi Chatbot Untuk Layanan Penerimaan Mahasiswa Baru Berbasis Natural Language Processing", Jurnal Informatika: Jurnal Pengembangan IT, 10(4), pp.1115-1126, 2025.

[9] Unisba, F. H., Teks, K., & Akademik, L, "Aplikasi chatbot berbasis nlp untuk layanan akademik studi kasus fakultas hukum unisba", 20, 89–97, 2025.

[10] Angelina, S. J., Negara, A. B. P., & Muhardi, H, "Analisis pengaruh penerapan stopword removal pada performa klasifikasi sentimen tweet bahasa Indonesia", Jurnal Aplikasi dan Riset Informatika, 1(2), 165-173, 2023.

[11] Bumigora Information Technology (BITe), "Evaluasi metrik klasifikasi: Accuracy, precision, recall, dan F1-score pada model machine learning", Jurnal Bumigora Information Technology, 6(1). https://journal.universitasbumigora.ac.id/bite/article/view/4909, 2024.

[12] Carolina, V. P., Utami, E., & Yaqin, A, "Riset Jurnal Literatur: Penggunaan Metode Stemming Pada Bahasa Daerah Melayu-Ambon", Journal of Electrical Engineering and Computer (JEECOM), 6(1), 186-196, 2024.

[13] Fahmi, M. N, "Implementasi mechine learning menggunakan Python Library: Scikit-Learn (supervised dan unsupervised learning)", Sains Data Jurnal Studi Matematika dan Teknologi, 1(2), 87-96, 2023.

[14] Gumelar, A., Moeslim, S., Firmansyah, E., & Sutara, B, "Perbandingan Kinerja XGBoost dan IndoBERT untuk Klasifikasi Teks Kesehatan Bahasa Indonesia", 5(2), 62–72, 2025.

[15] Hakim, L., Sobri, A., Sunardi, L., & Nurdiansyah, D, "Prediksi penyakit jantung berbasis mesin learning dengan menggunakan metode k-nn", 07(02), 14–20, 2025.

[16] Wardhana, R. G., Wang, G., & Sibuea, F, "Penerapan machine learning dalam prediksi tingkat kasus penyakit di Indonesia", Journal of Information System Management (JOISM), 5(1), 40-45, 2023.

DOI : https://doi.org/10.35316/jimi.v11i2.211-220 © 2026 Jurnal Ilmiah Informatika with CC BY-NC licence