

Perancangan Chatbot Informasi PPDB Menggunakan Algoritma Multinomial Naive Bayes di TK Delima

Ismail Hasan^{1*}, Ari Nurul Alfian²

^{1,2}Program Studi Teknik Informatika, Universitas Bina Insani, Indonesia

^{1*}ismailhsn266@gmail.com, ²Arin@binainsani.ac.id

Abstrak: TK Delima masih menyampaikan informasi Penerimaan Peserta Didik Baru (PPDB) melalui brosur fisik, komunikasi tatap muka, dan pesan WhatsApp sehingga akses informasi bergantung pada jam operasional dan ketersediaan staf administrasi. Penelitian ini bertujuan merancang chatbot informasi PPDB berbasis website yang mampu mengklasifikasikan intent pertanyaan wali calon peserta didik serta memberikan respons berdasarkan basis pengetahuan yang tersedia. Pengembangan sistem menggunakan Rapid Application Development (RAD) dengan pengumpulan data melalui observasi, wawancara, dan studi pustaka. Masukan pengguna diproses menggunakan Natural Language Processing melalui enam tahap, yaitu case folding, punctuation removal, tokenization, normalisasi kata berbasis kamus, suffix stripping, dan stopword removal. Klasifikasi intent menerapkan Multinomial Naive Bayes yang dikombinasikan dengan Laplace Smoothing untuk mengatasi zero probability dan confidence check untuk menangani masukan di luar domain. Sistem diimplementasikan secara native menggunakan PHP, CodeIgniter 3, dan MySQL tanpa library machine learning eksternal. Data latih terdiri atas 85 kalimat pada tujuh kategori intent. Evaluasi menggunakan 40 kalimat uji menghasilkan accuracy 92,50%, macro-average precision 93,30%, macro-average recall 92,50%, dan macro-average F1-score 92,36%. Seluruh 11 skenario black-box testing berhasil, sedangkan kuesioner terhadap 11 responden memperoleh nilai rata-rata 4,01 atau tingkat persetujuan 80,1% dengan kategori Setuju. Hasil tersebut menunjukkan bahwa chatbot dapat membantu menyediakan informasi PPDB secara terprogram melalui website, meskipun kemampuan mengenali pertanyaan di luar domain masih perlu ditingkatkan.

Kata Kunci: Chatbot; Klasifikasi Intent; Laplace Smoothing; Multinomial Naive Bayes; Natural Language Processing; PPDB

Abstract: TK Delima still disseminates New Student Admission (PPDB) information through printed brochures, face-to-face communication, and WhatsApp messages, making access dependent on school operating hours and administrative staff availability. This study aims to design a website-based PPDB information chatbot capable of classifying the intent of prospective students' parents' questions and returning responses from a predefined knowledge base. The system was developed using Rapid Application Development (RAD), while data were collected through observation, interviews, and literature review. User input is processed using Natural Language Processing through six stages: case folding, punctuation removal, tokenization, dictionary-based normalization,

Ismail Hasan : *Penulis Korespondensi



Copyright © 2026, Ismail Hasan, Ari Nurul Alfian.

suffix stripping, and stopword removal. Intent classification applies Multinomial Naive Bayes combined with Laplace Smoothing to address zero probability and a confidence-check mechanism to handle out-of-domain input. The system was implemented natively using PHP, CodeIgniter 3, and MySQL without external machine-learning libraries. The training dataset contains 85 sentences distributed across seven intent categories. Evaluation on 40 test sentences produced 92.50% accuracy, 93.30% macro-average precision, 92.50% macro-average recall, and 92.36% macro-average F1-score. All 11 black-box testing scenarios were successful, while a questionnaire involving 11 respondents achieved an average score of 4.01 or an 80.1% agreement level, categorized as Agree. The results indicate that the chatbot can support programmed delivery of PPDB information through the website, although its ability to recognize out-of-domain questions still requires improvement.

Keywords: Chatbot; Intent Classification; Laplace Smoothing; Multinomial Naive Bayes; Natural Language Processing; PPDB

1. PENDAHULUAN

Perkembangan teknologi informasi mendorong institusi pendidikan memanfaatkan website sebagai media penyampaian informasi yang dapat diakses tanpa bergantung pada lokasi dan waktu. Website tidak hanya berfungsi sebagai media publikasi, tetapi juga dapat menjadi platform layanan informasi yang mengintegrasikan fitur interaktif. Penelitian mengenai pengembangan sistem informasi berbasis website menunjukkan bahwa pemilihan metode pengembangan dan teknologi implementasi perlu disesuaikan dengan kebutuhan pengguna dan karakteristik organisasi [1], sementara pengembangan aplikasi berbasis PHP dan MySQL tetap relevan untuk membangun layanan web yang terstruktur [13]. Implementasi CodeIgniter 3 juga dapat digunakan untuk membentuk aplikasi berbasis arsitektur Model-View-Controller (MVC) yang modular [14].

Berdasarkan observasi di TK Delima, penyampaian informasi Penerimaan Peserta Didik Baru (PPDB) masih dilakukan melalui brosur fisik, komunikasi tatap muka, dan pesan WhatsApp. Kondisi tersebut menyebabkan jangkauan informasi melalui brosur terbatas pada lingkungan sekitar sekolah, sedangkan komunikasi langsung bergantung pada jam operasional dan ketersediaan staf administrasi. Pertanyaan mengenai biaya, persyaratan, jadwal, kegiatan, seragam, sistem pembelajaran, dan fasilitas juga sering disampaikan dengan topik yang serupa sehingga staf perlu memberikan penjelasan yang sama secara berulang. Oleh karena itu, dibutuhkan media informasi terpusat yang dapat menyediakan informasi PPDB sekaligus membantu menangani pertanyaan berulang dari wali calon peserta didik.

Chatbot dapat digunakan sebagai antarmuka percakapan untuk menyampaikan informasi berdasarkan basis pengetahuan yang tersedia. Penerapan chatbot pada layanan informasi telah digunakan untuk membantu menjawab pertanyaan pelanggan [2] dan pada layanan PPDB melalui platform komunikasi [19]. Tantangan penting pada chatbot berbasis teks adalah menentukan maksud atau intent pengguna ketika pertanyaan ditulis dengan susunan kalimat, singkatan, tanda baca, dan bentuk kata yang berbeda. Klasifikasi intent menjadi bagian penting karena hasil klasifikasi menentukan respons yang akan diberikan oleh sistem [3].

Beberapa penelitian terdahulu telah mengembangkan chatbot pada konteks layanan informasi dan pendidikan. Apriliani dkk. menerapkan Natural Language Processing untuk membangun aplikasi chatbot pada SMK YPE Nusantara Slawi [7], sedangkan Mamuriyah dkk. membangun chatbot interaktif menggunakan algoritma Naive Bayes untuk menjawab pertanyaan pengguna secara otomatis [8]. Pada konteks PPDB, Santoso dkk. merancang program chatbot berbasis website untuk sistem informasi dan pelayanan PPDB di SMK Al-Bahri Bekasi [9], sementara Rafi dkk. mengimplementasikan chatbot penerimaan peserta

didik baru di SMK Intensif Baitussalam menggunakan pendekatan NLP [11]. Pada jenjang pendidikan anak usia dini, Gardhitatama dkk. mengembangkan chatbot untuk mendukung pembelajaran bahasa Inggris di TK PKK Kartini Sidodadi [10], dan Huberta dan Wijaya merancang chatbot website untuk menyampaikan informasi program studi [18]. Penelitian-penelitian tersebut menunjukkan bahwa chatbot secara umum efektif diterapkan pada layanan informasi maupun konteks PPDB, namun implementasi pada jenjang Taman Kanak-kanak yang mengintegrasikan klasifikasi intent berbasis Multinomial Naive Bayes, Laplace Smoothing, dan confidence check untuk menangani pertanyaan di luar domain masih belum banyak dibahas. Penelitian ini melengkapi celah tersebut dengan berfokus pada chatbot informasi PPDB di TK Delima yang sebelumnya belum memiliki media informasi digital.

Penelitian ini menerapkan Natural Language Processing (NLP) untuk menstrukturkan masukan bahasa Indonesia tidak baku dan Multinomial Naive Bayes sebagai algoritma klasifikasi intent. Multinomial Naive Bayes sesuai untuk data teks yang direpresentasikan melalui frekuensi kemunculan token dan telah digunakan dalam berbagai tugas klasifikasi teks [4], [12], [16], [17]. Laplace Smoothing diterapkan untuk mencegah zero probability ketika token pada data uji tidak ditemukan pada suatu kelas [5]. Sistem dikembangkan menggunakan Rapid Application Development (RAD) karena pendekatan ini mendukung pengembangan iteratif dan keterlibatan pengguna selama proses perancangan [6], [15]. Pendekatan tersebut dipilih untuk menangani variasi pertanyaan PPDB berbahasa Indonesia secara terstruktur dan menghasilkan klasifikasi intent yang dapat langsung digunakan untuk menentukan respons chatbot.

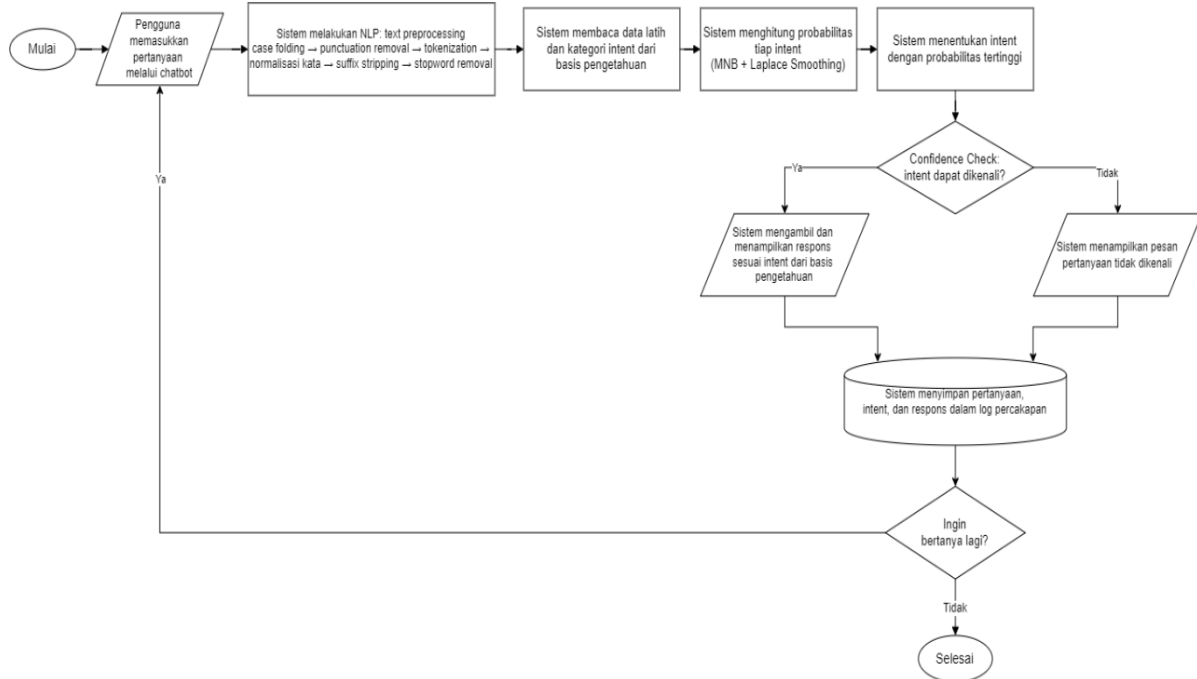
Tujuan penelitian adalah menganalisis kebutuhan informasi PPDB dan jenis pertanyaan pengguna, merancang serta mengimplementasikan chatbot PPDB berbasis website, menerapkan Multinomial Naive Bayes untuk mengklasifikasikan intent, dan menguji kinerja model menggunakan Confusion Matrix. Kontribusi penelitian ini bersifat aplikatif, yaitu menerapkan klasifikasi intent berbasis Multinomial Naive Bayes secara terintegrasi pada chatbot layanan informasi PPDB TK Delima. Implementasi dilakukan secara native menggunakan PHP dan CodeIgniter 3 tanpa library machine learning eksternal, dengan mengintegrasikan enam tahap text preprocessing, Laplace Smoothing, dan confidence check untuk menangani masukan di luar domain. Pendekatan tersebut diterapkan pada konteks TK Delima yang sebelumnya belum memiliki website resmi dan chatbot untuk layanan informasi PPDB. Website berfungsi sebagai media informasi dan platform implementasi chatbot; penelitian tidak mencakup proses pendaftaran daring.

2. METODE PENELITIAN

Data penelitian diperoleh melalui observasi, wawancara, dan studi pustaka. Observasi dilakukan terhadap proses penyampaian informasi PPDB di TK Delima, sedangkan wawancara dilakukan dengan kepala sekolah untuk mengetahui alur kerja, topik pertanyaan yang sering disampaikan, dan kebutuhan sistem. Hasil wawancara menjadi dasar penetapan tujuh kategori intent, yaitu sapaan, biaya, syarat, seragam, pembelajaran, jadwal, dan kegiatan. Studi pustaka digunakan untuk mendukung konsep NLP, text preprocessing, Multinomial Naive Bayes, Laplace Smoothing, RAD, dan evaluasi model.

Pengembangan perangkat lunak menggunakan RAD yang terdiri atas Requirements Planning, User Design, Construction, dan Cutover. Pada Requirements Planning dilakukan identifikasi kebutuhan fungsional, nonfungsional, informasi PPDB, serta kebutuhan chatbot. User Design mencakup perancangan alur website dan chatbot, flowchart, use case, activity diagram, basis data, dan prototipe antarmuka. Pada Construction, website dan chatbot dibangun menggunakan PHP, CodeIgniter 3, dan MySQL, sedangkan model

klasifikasi dikembangkan secara native. Tahap Cutover mencakup penerapan sistem, black-box testing, evaluasi model menggunakan Confusion Matrix, dan uji coba pengguna.



Gambar 1. Flowchart Sistem Usulan Chatbot PPDB TK Delima

A. Text Preprocessing

Pipeline pemrosesan pertanyaan terdiri atas enam tahap text preprocessing sebagaimana disajikan pada Tabel 1.

Tabel 1. Tahapan Text Preprocessing dan Contoh

Tahap	Fungsi	Contoh (Input → Output)
Case folding	Mengubah seluruh huruf menjadi huruf kecil	"Kapan PPDB Dibuka?" → "kapan ppdb dibuka?"
Punctuation removal	Menghapus tanda baca dan simbol	"kapan ppdb dibuka?" → "kapan ppdb dibuka"
Tokenization	Memecah kalimat menjadi token	"kapan ppdb dibuka" → [kapan, ppdb, dibuka]
Normalisasi kata	Memetakan kata tidak baku ke bentuk baku menggunakan dictionary mapping	[kpn, ppdb, dibuka] → [kapan, ppdb, dibuka]
Suffix stripping	Menghapus imbuhan akhir tertentu (-nya, -kan, -an, -lah, -ku, -mu) dengan daftar pengecualian	[biayanya, daftarkan] → [biaya, daftar]
Stopword removal	Menghapus kata umum yang tidak memberikan informasi penting terhadap proses klasifikasi	[apa, saja, syarat, pendaftaran] → [syarat, pendaftaran]

Keenam tahap tersebut dijalankan secara berurutan pada setiap pertanyaan yang masuk, dan token hasil preprocessing digunakan sebagai fitur pada proses klasifikasi Multinomial Naive Bayes.

B. Klasifikasi Intent dengan Multinomial Naive Bayes

Pada tahap pelatihan, sistem membaca korpus data latih dari MySQL dan menghitung prior probability setiap kelas serta conditional probability setiap token. Pada tahap

inferensi, pertanyaan pengguna diproses dengan pipeline yang sama, kemudian sistem menghitung skor untuk seluruh kelas dan memilih kelas dengan nilai Maximum A Posteriori.

Prior Probability dan Posterior Probability (Skor Naive Bayes):

$$P(c | d) \propto P(c) \times \prod_i P(t_i | c) \quad (1)$$

Keterangan: c = kelas atau intent yang dievaluasi; d = pertanyaan (dokumen) yang diklasifikasikan; $P(c)$ = prior probability kelas c , dihitung dari proporsi jumlah kalimat kelas c terhadap seluruh data latih; $\prod_i P(t_i | c)$ = perkalian conditional probability seluruh token pada pertanyaan, mewakili peluang gabungan seluruh kata dalam dokumen; $P(c | d)$ = skor posterior kelas c yang digunakan untuk memilih kelas dengan nilai Maximum A Posteriori.

Untuk mengatasi token yang belum pernah muncul pada suatu kelas, conditional probability dihitung menggunakan Laplace Smoothing, dengan $|V|$ menyatakan jumlah token unik pada vocabulary. Perhitungan posterior selanjutnya dilakukan pada ruang logaritmik untuk menghindari underflow.

Conditional Probability (Likelihood) dengan Laplace Smoothing:

$$P(t | c) = \frac{\text{count}(t, c) + 1}{\sum_{w \in V} \text{count}(w, c) + |V|} \quad (2)$$

Keterangan: $\text{count}(t, c)$ = jumlah kemunculan token t pada kelas c ; $\sum \text{count}(w, c)$ = total frekuensi seluruh token pada kelas c ; $|V|$ = jumlah token unik pada vocabulary; penambahan angka 1 merupakan Laplace Smoothing untuk mencegah zero probability ketika suatu token tidak ditemukan pada kelas tersebut.

Log Probability (Log-Posterior):

$$\log P(c | d) \propto \log P(c) + \sum_i \log P(t_i | c) \quad (3)$$

Keterangan: c = kelas atau intent yang dievaluasi; d = pertanyaan (dokumen) yang diklasifikasikan; $P(c)$ = prior probability kelas c ; $P(t_i | c)$ = conditional probability token ke- i pada kelas c hasil Persamaan (2); $\sum_i$ = penjumlahan log-probability atas seluruh token pada pertanyaan. Persamaan (3) merupakan bentuk logaritmik dari Persamaan (1) yang digunakan untuk menghindari underflow saat mengalikan banyak probabilitas kecil.

C. Confidence Check untuk Masukan di Luar Domain

Setelah skor setiap kelas diperoleh, sistem menerapkan confidence check dengan membandingkan selisih skor dua kelas tertinggi. Selisih skor dihitung menggunakan persamaan berikut:

$$\Delta = |\text{Score}_1 - \text{Score}_2| \quad (4)$$

Keterangan: Δ = selisih skor antara dua kelas dengan peringkat tertinggi; Score_1 = skor (log-posterior) kelas dengan peringkat pertama; Score_2 = skor kelas dengan peringkat kedua, sebagaimana dicontohkan pada perhitungan log-score di subbagian B.

Threshold awal confidence check sebesar 0,3 digunakan sebagai nilai kandidat dan selanjutnya diuji melalui analisis sensitivitas terhadap beberapa nilai threshold, yaitu 0,1, 0,2, 0,3, 0,4, dan 0,5. Pengaruh masing-masing threshold terhadap accuracy, precision, recall, F1-score, serta kemampuan sistem mengenali pertanyaan di luar domain dianalisis pada tahap evaluasi. Keputusan klasifikasi kemudian ditentukan sebagai berikut:

$$\begin{aligned} \text{Prediksi} &= \text{intent dengan } \text{Score}_1, \text{ jika } \Delta \geq 0,3; \\ &\text{atau tidak_dikenali, jika } \Delta < 0,3 \end{aligned} \quad (5)$$

Keterangan: jika $\Delta \geq 0,3$, hasil klasifikasi diterima dan chatbot menampilkan jawaban sesuai intent dari basis pengetahuan; jika $\Delta < 0,3$, pertanyaan ditetapkan sebagai tidak_dikenali dan chatbot menampilkan pesan fallback. Seluruh pertanyaan, intent hasil klasifikasi, dan respons disimpan dalam log percakapan.

D. Dataset dan Skema Evaluasi

Data latih terdiri atas 85 kalimat yang dilabeli secara manual pada tujuh kategori intent. Evaluasi model menggunakan 40 kalimat uji yang tidak terdapat pada data latih, terdiri atas masing-masing lima kalimat untuk tujuh kategori intent dan lima kalimat di luar domain dengan label tidak_dikenali. Kinerja model diukur menggunakan accuracy, precision, recall, dan F1-score. Pengujian fungsional dilakukan melalui 11 skenario black-box testing. Uji coba pengguna dilakukan terhadap 11 responden wali calon peserta didik menggunakan 15 pernyataan skala Likert 1–5 yang dikelompokkan ke dalam aspek website, chatbot, dan perbandingan layanan informasi.

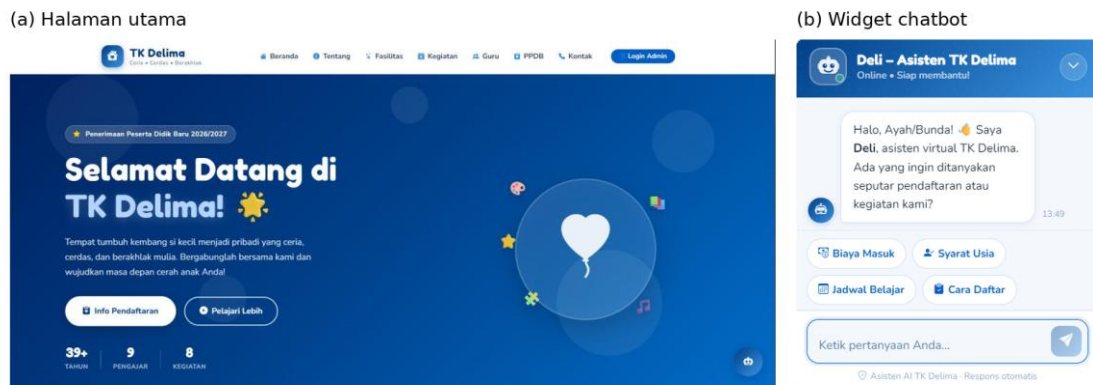
Tabel 2. Distribusi Data Latih Chatbot

Intent	Jumlah Kalimat Latih	Persentase
Sapaan	11	12,94%
Biaya	15	17,65%
Syarat	15	17,65%
Seragam	11	12,94%
Pembelajaran	11	12,94%
Jadwal	13	15,29%
Kegiatan	9	10,59%
Total	85	100%

3. HASIL DAN PEMBAHASAN

A. Implementasi Sistem

Implementasi menghasilkan website informasi PPDB TK Delima yang terintegrasi dengan chatbot Deli. Sistem menggunakan arsitektur MVC pada CodeIgniter 3 dan basis data MySQL. Wali calon peserta didik dapat mengakses informasi sekolah dan menggunakan chatbot melalui widget pada halaman website. Admin dapat mengelola basis pengetahuan yang berisi nama intent, contoh pertanyaan sebagai data latih, jawaban chatbot, log percakapan, serta konten website. Ketika pertanyaan dikirim, sistem melakukan preprocessing, menghitung skor Multinomial Naive Bayes dengan Laplace Smoothing, menjalankan confidence check, kemudian menampilkan jawaban sesuai intent atau pesan tidak dikenali.



Gambar 2. Implementasi Halaman Utama Website dan Widget Chatbot Deli

B. Contoh Perhitungan Multinomial Naive Bayes

Untuk memperjelas proses klasifikasi, digunakan contoh data latih yang terdiri atas tujuh kalimat dari tiga kategori intent, yaitu biaya, syarat, dan sapaan. Data tersebut digunakan sebagai ilustrasi perhitungan, sedangkan implementasi model menggunakan seluruh 85 kalimat data latih sebagaimana ditunjukkan pada **Tabel 2**. Pertanyaan uji yang digunakan adalah "Biaya pendaftaran?". Setelah melalui tahap text preprocessing, diperoleh token [biaya, pendaftaran].

Menentukan prior probability

Pada sampel terdapat 3 kalimat untuk intent biaya, 2 kalimat untuk intent syarat, dan 2 kalimat untuk intent sapaan, sehingga:

$$P(\text{Biaya}) = 3/7 = 0,4286$$

$$P(\text{Syarat}) = 2/7 = 0,2857$$

$$P(\text{Sapaan}) = 2/7 = 0,2857$$

Menghitung conditional probability dengan Laplace Smoothing

Perhitungan dilakukan menggunakan Persamaan (2). Pada contoh sampel, ukuran vocabulary yang digunakan adalah $|V| = 12$.

Untuk intent biaya, total frekuensi token adalah 9. Karena token "biaya" muncul 2 kali dan "pendaftaran" muncul 2 kali:

$$P(\text{biaya}|\text{Biaya}) = (2+1)/(9+12) = 3/21 = 0,1429$$

$$P(\text{pendaftaran}|\text{Biaya}) = (2+1)/(9+12) = 3/21 = 0,1429$$

Untuk intent syarat, total frekuensi token adalah 6. Token "biaya" tidak muncul, sedangkan "pendaftaran" muncul 1 kali:

$$P(\text{biaya}|\text{Syarat}) = (0+1)/(6+12) = 1/18 = 0,0556$$

$$P(\text{pendaftaran}|\text{Syarat}) = (1+1)/(6+12) = 2/18 = 0,1111$$

Untuk intent sapaan, total frekuensi token adalah 4 dan kedua token tidak muncul pada kelas tersebut:

$$P(\text{biaya}|\text{Sapaan}) = (0+1)/(4+12) = 1/16 = 0,0625$$

$$P(\text{pendaftaran}|\text{Sapaan}) = (0+1)/(4+12) = 1/16 = 0,0625$$

Menghitung log-score menggunakan Persamaan (3). Perhitungan dilakukan pada ruang logaritmik untuk menghindari underflow:

$$\text{Score}(\text{Biaya}) = \ln(0,4286) + \ln(0,1429) + \ln(0,1429) = -4,7391$$

$$\text{Score}(\text{Syarat}) = \ln(0,2857) + \ln(0,0556) + \ln(0,1111) = -6,3404$$

$$\text{Score}(\text{Sapaan}) = \ln(0,2857) + \ln(0,0625) + \ln(0,0625) = -6,7979$$

Tabel 3. Hasil Perhitungan Log-Score dan Peringkat Intent

Intent	Log-Score	Peringkat
Biaya	-4,7391	1
Syarat	-6,3404	2
Sapaan	-6,7979	3

Karena nilai log-score yang paling mendekati nol adalah -4,7391, maka pertanyaan "Biaya pendaftaran?" diklasifikasikan sebagai intent biaya.

C. Confidence check. Sistem membandingkan dua skor tertinggi menggunakan Persamaan (4) dan (5):

$$\Delta = |-4,7391 - (-6,3404)| = 1,6013$$

Dengan ambang confidence check sebesar 0,3, nilai $\Delta = 1,6013 > 0,3$ sehingga hasil klasifikasi diterima dan respons intent biaya ditampilkan. Apabila $\Delta < 0,3$, sistem menetapkan hasil sebagai tidak_dikenali dan menampilkan respons fallback.

D. Evaluasi Akurasi Model

Evaluasi pada 40 kalimat uji menghasilkan 37 prediksi benar dan tiga prediksi salah sehingga accuracy mencapai 92,50%. Enam kelas, yaitu biaya, syarat, sapaan, jadwal, seragam, dan pembelajaran, memiliki recall 100%. Kelas kegiatan memiliki recall 80% karena satu pertanyaan "ada pramuka tidak di TK Delima" diprediksi sebagai tidak_dikenali. Pada kelas tidak_dikenali, tiga dari lima pertanyaan di luar domain berhasil ditolak, sedangkan "besok cuaca cerah tidak" dan "rekomenadasi film bagus apa" diprediksi sebagai sapaan. Kondisi tersebut menyebabkan precision kelas sapaan turun menjadi 71,43% dan recall kelas tidak_dikenali menjadi 60%. Mengingat ukuran sampel uji yang relatif kecil (40 kalimat), estimasi accuracy, precision, recall, dan F1-score pada penelitian ini memiliki margin kesalahan yang cukup besar dan belum divalidasi melalui cross-validation, sehingga generalisasi hasil ke variasi pertanyaan yang lebih luas masih perlu diverifikasi lebih lanjut. Keterkaitan pola kesalahan klasifikasi ini dengan persepsi pengguna dibahas lebih lanjut pada Sub-bab G.

Tabel 4. Hasil Evaluasi Model per Kelas Intent Berdasarkan Confusion Matrix

Intent	TP	FP	FN	Precision	Recall	F1-Score
biaya	5	0	0	100%	100%	100%
syarat	5	0	0	100%	100%	100%
sapaan	5	2	0	71,43%	100%	83,33%
jadwal	5	0	0	100%	100%	100%
seragam	5	0	0	100%	100%	100%
pembelajaran	5	0	0	100%	100%	100%
kegiatan	4	0	1	100%	80%	88,89%

Intent	TP	FP	FN	Precision	Recall	F1-Score
tidak_dikenali	3	1	2	75%	60%	66,67%
Total	37	3	3	—	—	—
Macro-average	—	—	—	93,30%	92,50%	92,36%

Secara keseluruhan model macro-average precision 93,30%, macro-average recall 92,50%, dan macro-average F1-score 92,36%. Nilai ini menunjukkan bahwa Multinomial Naive Bayes mampu mengklasifikasikan sebagian besar variasi pertanyaan pada domain PPDB yang disediakan. Hasil tersebut sejalan dengan penelitian Muhidin dkk. yang menunjukkan bahwa Multinomial Naive Bayes dapat memberikan performa tinggi pada chatbot informasi berbasis teks [12]. Namun, angka penelitian yang berbeda tidak dapat dibandingkan secara langsung karena domain, jumlah data, distribusi kelas, dan skenario evaluasinya tidak sama. Temuan kesalahan pada pertanyaan di luar domain juga menguatkan perlunya penambahan variasi data latih dan penyempurnaan mekanisme penolakan masukan.

E. Analisis Sensitivitas Threshold Confidence Check

Penentuan threshold 0,3 dilakukan secara sistematis melalui pengujian sensitivitas terhadap lima nilai threshold, yaitu 0,1, 0,2, 0,3, 0,4, dan 0,5. Hasil pengujian menunjukkan bahwa performa klasifikasi pada data uji meningkat dari threshold 0,1 hingga mencapai performa tertinggi pada threshold 0,3 dan 0,4, dengan accuracy sebesar 92,50%, precision sebesar 93,30%, recall sebesar 92,50%, dan F1-score sebesar 92,36%. Ketika threshold dinaikkan menjadi 0,5, performa mengalami penurunan menjadi accuracy 82,50%, precision 88,62%, recall 82,50%, dan F1-score 84,24%. Pada pengujian out-of-domain (OOD), kemampuan deteksi meningkat dari 40% (2 dari 5 sampel) pada threshold 0,1 dan 0,2 menjadi 60% (3 dari 5 sampel) pada threshold 0,3, 0,4, dan 0,5. Berdasarkan hasil tersebut, threshold 0,3 dipilih sebagai nilai terendah pada rentang threshold dengan performa klasifikasi optimum. Threshold 0,3 dipilih dibandingkan 0,4 karena memberikan performa yang sama pada data uji dengan nilai threshold yang lebih rendah, sehingga bersifat kurang restriktif dalam menerima hasil klasifikasi. Pengujian OOD menggunakan lima sampel sehingga hasil deteksinya bersifat indikatif dan masih memerlukan validasi dengan jumlah sampel yang lebih besar.

Tabel 5. Hasil Pengujian Sensitivitas Threshold Confidence Check

Threshold	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	OOD Benar	OOD Detection
0,1	87,50%	89,06%	87,50%	87,39%	2/5	40%
0,2	90,00%	91,15%	90,00%	89,48%	2/5	40%
0,3	92,50%	93,30%	92,50%	92,36%	3/5	60%
0,4	92,50%	93,30%	92,50%	92,36%	3/5	60%
0,5	82,50%	88,62%	82,50%	84,24%	3/5	60%

F. Pengujian Fungsional

Pengujian fungsional menunjukkan seluruh 11 skenario black-box testing berhasil, meliputi akses halaman utama, penggunaan widget chatbot, respons terhadap pertanyaan relevan dan di luar konteks, login admin, pengelolaan intent, log percakapan, konten PPDB, serta logout. Hasil tersebut menunjukkan bahwa fungsi utama pada sisi pengguna dan admin berjalan sesuai spesifikasi yang ditetapkan.

G. Uji Coba Pengguna

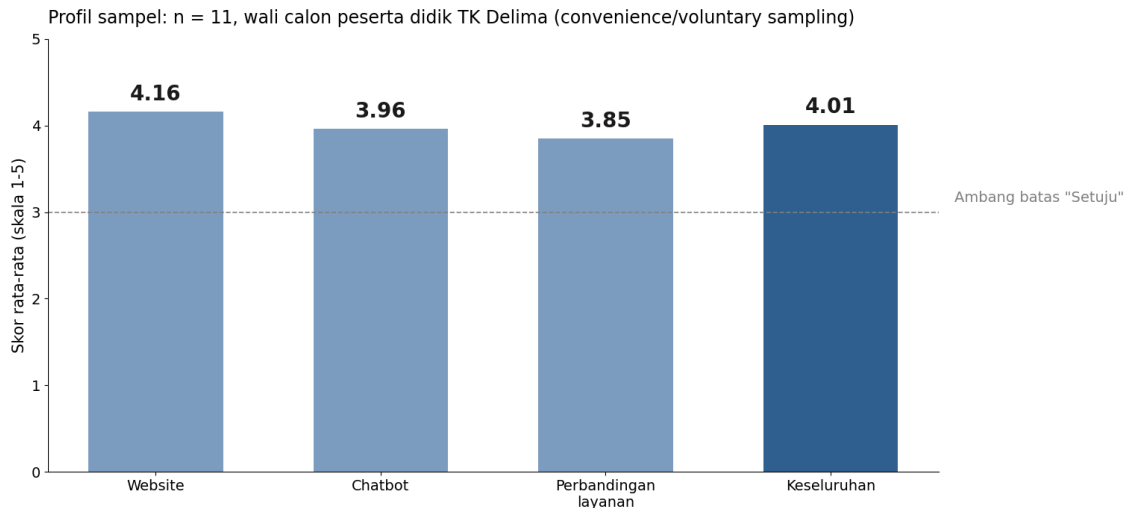
Uji coba pengguna menghasilkan nilai rata-rata keseluruhan 4,01 atau tingkat persetujuan 80,1% dengan kategori Setuju. Aspek website memperoleh rata-rata 4,16,

Ismail Hasan : *Penulis Korespondensi



Copyright © 2026, Ismail Hasan, Ari Nurul Alfian.

aspek chatbot 3,96, dan aspek perbandingan layanan informasi 3,85; seluruhnya berada pada kategori Setuju. Responden terdiri atas 11 wali calon peserta didik TK Delima yang berpartisipasi secara sukarela setelah kuesioner dibagikan oleh pihak sekolah melalui grup WhatsApp wali murid (*convenience/voluntary sampling*), sebagaimana diilustrasikan pada Gambar 3.



Gambar 3. Profil Sampel dan Ringkasan Skor Uji Coba Pengguna

Tabel 6. Ringkasan Hasil Uji Coba Pengguna

Aspek	Nilai Rata-rata	Interpretasi
Website	4,16	Setuju
Chatbot	3,96	Setuju
Perbandingan layanan informasi	3,85	Setuju
Keseluruhan	4,01 (80,1%)	Setuju

Tabel 7. Hasil Uji Coba Pengguna per Indikator (n=11)

No	Aspek	Indikator Penilaian	Skor	Persentase
1	Website	Tampilan website mudah dipahami dan digunakan	46	83,6%
2	Website	Informasi yang tersedia lengkap dan jelas	44	80,0%
3	Website	Menu navigasi mudah ditemukan dan diakses	48	87,3%
4	Website	Dapat mengakses informasi kapan saja tanpa terikat jam operasional sekolah	47	85,5%
5	Website	Membantu mendapatkan informasi pendaftaran dengan lebih mudah	44	80,0%
6	Chatbot	Mampu menjawab pertanyaan dengan jelas	40	72,7%
7	Chatbot	Jawaban sesuai dengan pertanyaan yang diajukan	39	70,9%
8	Chatbot	Memberikan respons dengan cepat saat digunakan	45	81,8%
9	Chatbot	Bahasa yang digunakan mudah dipahami	45	81,8%
10	Chatbot	Tampilan widget mudah digunakan	45	81,8%
11	Chatbot	Dapat digunakan tanpa mengalami kesulitan	48	87,3%
12	Chatbot	Membantu memperoleh informasi tanpa menghubungi pihak sekolah langsung	43	78,2%
13	Perbandingan	Lebih memudahkan dibanding cara sebelumnya (datang langsung/WhatsApp)	42	76,4%
14	Perbandingan	Lebih nyaman dibanding menghubungi admin secara langsung	42	76,4%
15	Perbandingan	Secara keseluruhan membantu meningkatkan layanan informasi pendaftaran	43	78,2%

Ismail Hasan : *Penulis Korespondensi



Copyright © 2026, Ismail Hasan, Ari Nurul Alfian.

Berdasarkan Tabel 7, indikator dengan persentase tertinggi adalah kemudahan menemukan menu navigasi website dan kemudahan menggunakan chatbot, masing-masing sebesar 87,3%. Sementara itu, persentase terendah terdapat pada indikator kesesuaian jawaban chatbot dengan pertanyaan pengguna, yaitu sebesar 70,9%. Hasil tersebut menunjukkan bahwa aspek kemudahan penggunaan telah dinilai positif, sedangkan kesesuaian respons chatbot masih menjadi aspek yang perlu ditingkatkan.

Temuan pada Tabel 4 menunjukkan bahwa kesalahan klasifikasi pada intent kegiatan dan tidak dikenali berkaitan dengan rendahnya indikator kesesuaian respons pada Tabel 7 (70,9%, skor 39 dari 55). Kesalahan klasifikasi pada intent kegiatan menyebabkan pertanyaan yang relevan menghasilkan respons fallback, sedangkan kesalahan klasifikasi pertanyaan di luar domain sebagai sapaan menghasilkan respons yang tidak sesuai dengan maksud pengguna. Dengan demikian, hasil analisis Confusion Matrix dan kuesioner menunjukkan adanya hubungan antara ketepatan klasifikasi intent dan persepsi pengguna terhadap kesesuaian respons chatbot. Perbaikan dapat dilakukan dengan memperluas variasi data latih pada intent kegiatan, menambah data uji di luar domain, serta mengevaluasi kembali basis pengetahuan dan pasangan intent-jawaban.

4. KESIMPULAN

Penelitian ini menghasilkan chatbot informasi PPDB TK Delima berbasis website yang dibangun menggunakan PHP, CodeIgniter 3, dan MySQL serta mengimplementasikan Multinomial Naive Bayes secara native. Masukan pengguna diproses melalui enam tahap text preprocessing, kemudian diklasifikasikan menggunakan Multinomial Naive Bayes dengan Laplace Smoothing dan confidence check. Model menggunakan 85 kalimat latih pada tujuh kategori intent dan dievaluasi menggunakan 40 kalimat uji. Hasil pengujian memperoleh accuracy 92,50%, macro-average precision 93,30%, macro-average recall 92,50%, dan macro-average F1-score 92,36%. Seluruh 11 skenario black-box testing berhasil dan uji coba terhadap 11 responden memperoleh tingkat persetujuan 80,1% dengan kategori Setuju. Dengan demikian, website dan chatbot dapat membantu penyampaian informasi PPDB dan menangani pertanyaan berulang berdasarkan basis pengetahuan yang tersedia. Keterbatasan utama terdapat pada kemampuan mengenali sebagian masukan di luar domain. Keterbatasan lain terletak pada ukuran dataset yang relatif kecil, yaitu 85 kalimat latih dan 40 kalimat uji, serta belum dilakukannya validasi silang (cross-validation). Kondisi tersebut membatasi kemampuan penelitian dalam menilai konsistensi dan generalisasi model, sehingga hasil evaluasi perlu diverifikasi lebih lanjut menggunakan dataset yang lebih besar dan beragam. Pengembangan selanjutnya disarankan menambah kategori intent dan variasi data latih, memperluas dataset, mengevaluasi mekanisme confidence check, membandingkan Multinomial Naive Bayes dengan metode klasifikasi lain, serta mempertimbangkan integrasi chatbot dengan platform komunikasi seperti WhatsApp.

5. REFERENCES

- [1] E. Rahmi, E. Yumami, and N. Hidayasari, "Analisis Metode Pengembangan Sistem Informasi Berbasis Website: Systematic Literature Review," REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer, vol. 7, pp. 821–834, 2023.
- [2] N. R. Fahmi and D. F. Viandhana, "Pusat Pelayanan Wifi Berbasis Chatbot Untuk Layanan Pelanggan," in Prosiding Seminar Nasional Teknologi dan Sains, vol. 5, pp. 379–386, 2026.
- [3] A. Dwiyo, Abdiansah, and M. Fachrurrozi, "Analisis Perbandingan Klasifikasi Intent Chatbot Menggunakan Deep Learning BERT, RoBERTa, dan IndoBERT," Journal of Information System Research (JOSH), vol. 6, pp. 595–606, 2024.

Ismail Hasan : *Penulis Korespondensi



Copyright © 2026, Ismail Hasan, Ari Nurul Alfian.

- [4] N. P. A. Noviyanti and I. G. N. A. C. Putra, "Penerapan Multinomial Naïve Bayes dan Chi-Square pada Analisis Sentimen Makan Bergizi Gratis," *Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA)*, vol. 3, pp. 785–794, 2025.
- [5] Y. A. Rizky, A. Aziz, and W. Harianto, "Implementasi Naive Bayes dengan Menggunakan Metode Laplace Smoothing," *RAINSTEK: Jurnal Terapan Sains dan Teknologi*, vol. 6, no. 3, pp. 164–172, 2024.
- [6] Sondang, "Penerapan Metode RAD dalam Pengembangan Sistem Informasi Pemesanan Jasa Percetakan Berbasis Web pada Percetakan Karya Sehati Jaya," *E-Jurnal Manaj. Inform. Komput.*, vol. 8, pp. 871–881, 2024.
- [7] D. Apriliani, S. F. Handayani, and I. T. Saputra, "Implementasi Natural Language Processing (NLP) dalam Pengembangan Aplikasi Chatbot pada SMK YPE Nusantara Slawi," *Techno.COM*, vol. 22, pp. 1037–1047, 2023.
- [8] N. Mamuriyah, Haeruddin, and Hero, "Pembangunan Chatbot Interaktif dengan Menggunakan Algoritma Naive Bayes," *Informatika: Jurnal Teknik Informatika dan Multimedia*, vol. 4, pp. 82–94, 2024.
- [9] B. Santoso, Jefi, B. A. Wahid, F. E. Schadu, and D. R. Zain, "Perancangan Program Chatbot Berbasis Website dalam Sistem Informasi dan Pelayanan PPDB pada SMK Al-Bahri Bekasi," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, pp. 11260–11268, 2024.
- [10] D. H. Gardhitatama, H. Yuana, and F. Febrinita, "Rancang Bangun Aplikasi Pembelajaran Bahasa Inggris Berbasis Chatbot (Studi Kasus: TK PKK Kartini Sidodadi)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, pp. 829–834, 2022.
- [11] R. R. Rafi, A. Setiawan, A. A. Rosadi, and A. Sanjaya, "Chatbot Penerimaan Peserta Didik Baru di SMK Intensif Baitussalam Menggunakan NLP," in *Prosiding Seminar Nasional Teknologi dan Sains*, vol. 4, pp. 130–135, 2025.
- [12] A. Muhidin, M. Danny, and E. Rilvani, "Algoritme Multinomial Naïve Bayes pada Aplikasi Chatbot Layanan Informasi Berbasis Teks," *J. Ilm. Komput.*, vol. 19, pp. 71–80, 2023.
- [13] F. Sinlae, I. Maulana, F. Setiyansyah, and M. Ihsan, "Pengenalan Pemrograman Web: Pembuatan Aplikasi Web Sederhana dengan PHP dan MySQL," *Jurnal Siber Multi Disiplin (JSMD)*, vol. 2, pp. 68–82, 2024.
- [14] P. Kustianto, S. Auliana, B. Rakhim, S. Permana, A. Rohman, and A. Munawir, "Penerapan Framework CodeIgniter 3 pada Aplikasi Penjualan di Toko Kue Kering Miya Jaya Serang," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, pp. 831–836, 2025.
- [15] N. W. Rhomadhona, M. A. Muhammad, P. B. Wintoro, and Y. Mulyani, "Rapid Application Development untuk Sistem Informasi Event," *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, vol. 13, no. 1, 2025.
- [16] R. Chandra and E. M. Sipayung, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Menggunakan Algoritma Naive Bayes Classifier," *TEKNOSI: Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 3, pp. 156–164, 2024.
- [17] R. Indransyah, Y. H. Chrisnanto, and P. N. Sabrina, "Klasifikasi Sentimen Pergelaran MotoGP di Indonesia Menggunakan Algoritma Correlated Naïve Bayes Classifier," *INFOTECH Journal*, vol. 8, no. 2, pp. 60–66, 2022.
- [18] B. Huberta and A. B. Wijaya, "Perancangan Chatbot Website Program Studi Framework," *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, vol. 11, 2023.
- [19] I. Amirulloh, M. W. Pertiwi, and T. Wibisono, "Whatsapp Menggunakan Node JS dan Model Natural Language Processing untuk Layanan PPDB," *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, vol. 12, 2024.