



Implementasi Naive Bayes Classifier pada Chatbot Whatsapp untuk Layanan Informasi PPDB

Afrizal Rahman¹, Sholihin²

^{1,2}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang³
Email: ¹afrizalrahman36@gmail.com, ²dosen00404@unpam.ac.id

Abstrak—Perkembangan teknologi informasi yang pesat telah mendorong institusi pendidikan untuk mengadopsi solusi digital guna meningkatkan efisiensi pelayanan informasi kepada masyarakat. *Chatbot* berbasis *Natural Language Processing* (NLP) menjadi salah satu solusi yang efektif dalam merespons pertanyaan pengguna secara otomatis. WhatsApp, sebagai platform komunikasi populer di Indonesia, dipilih sebagai media integrasi *chatbot* dalam layanan informasi Penerimaan Peserta Didik Baru (PPDB). Tantangan utama dalam pengembangan *chatbot* adalah kemampuan mengklasifikasikan pertanyaan pengguna ke kategori yang tepat secara cepat dan akurat. Algoritma Naive Bayes Classifier (NBC) dipilih karena kecepatan dan efektivitasnya dalam menangani data teks berukuran besar, menjadikannya cocok untuk klasifikasi pertanyaan secara *real-time*. Penelitian ini bertujuan mengembangkan dan mengevaluasi implementasi Naive Bayes Classifier dalam *chatbot* WhatsApp untuk meningkatkan efektivitas pelayanan informasi di institusi pendidikan. Berdasarkan studi sebelumnya, penggunaan Naive Bayes telah terbukti unggul dalam klasifikasi pertanyaan administratif dan layanan pelanggan yang repetitif. Implementasi *chatbot* ini diharapkan dapat memberikan kemudahan akses informasi bagi calon peserta didik dan mengurangi beban kerja staf administrasi sekolah.

Kata Kunci: Naive Bayes Classifier, Chatbot WhatsApp, Klasifikasi Pertanyaan, Pendaftaran Sekolah, NLP.

Abstract—Rapid developments in information technology have encouraged educational institutions to adopt digital solutions to improve the efficiency of information services to the public. Chatbots based on Natural Language Processing (NLP) are one effective solution for automatically responding to user queries. WhatsApp, a popular communication platform in Indonesia, was chosen as the medium for integrating chatbots into the New Student Admissions (PPDB) information service. The main challenge in developing chatbots is the ability to classify user questions into the right categories quickly and accurately. The Naive Bayes Classifier (NBC) algorithm was chosen for its speed and effectiveness in handling large amounts of text data, making it suitable for real-time question classification. This study aims to develop and evaluate the implementation of Naive Bayes Classifier in WhatsApp chatbots to improve the effectiveness of information services in educational institutions. Based on previous studies, the use of Naive Bayes has been proven to be superior in classifying repetitive administrative and customer service questions. The implementation of this chatbot is expected to provide easy access to information for prospective students and reduce the workload of school administrative staff.

Keywords: Naive Bayes Classifier, WhatsApp Chatbot, Question Classification, School Registration, NLP.

1. PENDAHULUAN

Perkembangan pesat teknologi informasi mendorong berbagai institusi pendidikan untuk mengadopsi solusi digital guna meningkatkan efisiensi layanan kepada masyarakat, khususnya dalam hal penyampaian informasi terkait Penerimaan Peserta Didik Baru (PPDB). Salah satu inovasi yang sangat menjanjikan adalah pengembangan *chatbot* berbasis *Natural Language Processing* (NLP) yang mampu secara otomatis memahami dan merespon pertanyaan dari calon peserta didik maupun orang tua.

Whatsapp merupakan platform komunikasi populer di Indonesia digunakan sebagai media utama pengembangan *chatbot* ini. Dengan fitur real-time dan kemudahan akses, *chatbot* berbasis Whatsapp dapat menjadi solusi efektif dalam menjawab pertanyaan-pertanyaan umum tentang jadwal pendaftaran, persyaratan, dan biaya secara cepat dan akurat.

Namun, pengembangan *chatbot* yang mampu memahami dan mengklasifikasikan pertanyaan pengguna secara tepat memerlukan algoritma yang efisien. Salah satu algoritma yang banyak digunakan dalam klasifikasi teks adalah Naive Bayes Classifier.

Dengan demikian, mengembangkan sistem *chatbot* berbasis Whatsapp dengan implementasi algoritma Naive Bayes Classifier, diharapkan sistem ini dapat memberikan layanan informasi yang efektif, efisien, serta meningkatkan kualitas komunikasi antara institusi pendidikan dan masyarakat.



2. METODE PENELITIAN

Naive Bayes Classifier merupakan metode klasifikasi yang digunakan dalam penelitian ini. Metode ini bekerja berdasarkan teorema Bayes dengan asumsi bahwa setiap fitur pada data masukan bersifat independen atau tidak saling memengaruhi satu sama lain. Asumsi "naif" inilah yang menjadi dasar penamaan metode tersebut, karena pada kenyataannya antarfitur bisa saja memiliki keterkaitan, namun diperlakukan seolah berdiri sendiri untuk menyederhanakan proses perhitungan probabilitas. Dengan pendekatan ini, setiap fitur dihitung kontribusi probabilitasnya secara terpisah terhadap suatu kelas, kemudian digabungkan untuk menentukan hasil klasifikasi akhir. Adapun rumus persamaan dari teorema bayes yaitu:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan:

- $P(C|X)$: Probabilitas posterior, yaitu probabilitas kelas C diberikan data fitur X (yang ingin kita prediksi)
- $P(X|C)$: Probabilitas likelihood, yaitu probabilitas data fitur X muncul jika kelasnya adalah C
- $P(C)$: Probabilitas prior kelas C sebelum melihat data fitur X
- $P(X)$: Probabilitas evidence, yaitu probabilitas data fitur X muncul secara keseluruhan (normalisasi, sering diabaikan dalam klasifikasi karena sama untuk semua kelas)

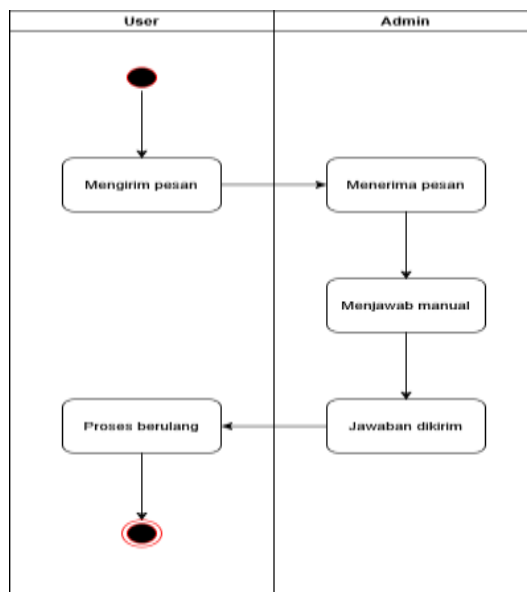
3. ANALISA DAN PEMBAHASAN

3.1 Analisa Sistem

3.1.1 Activity Diagram Sistem Berjalan

Sistem informasi PPDB di banyak sekolah masih manual, mengandalkan admin untuk menjawab pertanyaan calon siswa/orang tua via WhatsApp atau telepon. Cara ini rentan lambat, tidak efisien saat volume pertanyaan tinggi, dan karena tidak ada klasifikasi otomatis, data pertanyaan yang masuk pun tidak terstruktur sehingga sulit dianalisis untuk pengambilan keputusan.

Berikut adalah gambar *activity diagram* sistem berjalan:



Gambar 1. Activity Diagram Sistem Berjalan

3.1.2 Activity Diagram Sistem Usulan

Sistem baru yang diusulkan adalah chatbot otomatis berbasis WhatsApp yang mampu menerima pertanyaan dari pengguna, melakukan pra-pemrosesan teks, mengklasifikasikan pertanyaan menggunakan Naive Bayes Classifier, kemudian menjawab pertanyaan secara otomatis sesuai kategorinya, serta menyimpan data percakapan untuk keperluan pelatihan ulang atau evaluasi model. Sistem ini dapat berjalan selama 24 jam penuh setiap hari, memberikan respons yang cepat dan konsisten, sehingga membantu admin sekolah dalam menyederhanakan alur layanan informasi.

Berikut adalah gambar *activity diagram* sistem usulan:



Gambar 2. Activity Diagram Sistem Usulan

3.2 Perhitungan Teorema Bayes

Pada penelitian ini, total data pelatihan yang digunakan berjumlah 300 data. Data pelatihan ini terdiri dari beberapa kategori atau kelas yang mewakili berbagai jenis pertanyaan dalam layanan informasi PPDB pada chatbot WhatsApp. Pengelompokan data berdasarkan kategori sangat penting sebagai dasar bagi algoritma Naive Bayes untuk membangun model klasifikasi yang akurat.

Jumlah data pada setiap kategori ini mencerminkan variasi pertanyaan yang masuk ke sistem dan digunakan untuk menghitung probabilitas prior masing-masing kategori dalam proses klasifikasi. Distribusi jumlah data per kategori adalah sebagai berikut:

Tabel 1 Total Data

Kode	Nama Kategori	Total Data
A	Jadwal	70
B	Biaya	60
C	Persyaratan	55
D	Salam	50
E	Umum	75

Selain itu, fitur kata kunci seperti kata "PPDB" yang muncul dalam data diteliti frekuensinya untuk masing-masing kategori guna menentukan probabilitas likelihood pada tiap kelas. Frekuensi kemunculan kata "PPDB" adalah sebagai berikut:



Tabel 2 Total Kata Kunci

Kode	Nama Kategori	Kata Kunci	Total Data
A	Jadwal	PPDB	15
B	Biaya	PPDB	13
C	Persyaratan	PPDB	14
D	Salam	PPDB	12
E	Umum	PPDB	48

Besaran total kata pada masing-masing kategori (jumlah kata total yang terhitung dari keseluruhan data dalam kategori tersebut) adalah:

Tabel 3 Total Kata

Kode	Nama Kategori	Total Kata
A	Jadwal	390
B	Biaya	380
C	Persyaratan	350
D	Salam	360
E	Umum	400

Selain itu, ukuran vocabulary, yaitu total kata unik yang muncul di seluruh dataset adalah sebanyak 500 kata.

Informasi jumlah data ini digunakan dalam perhitungan Teorema Bayes dengan metode *Laplace smoothing* yang memastikan probabilitas kata yang jarang muncul tetap terhitung secara matematis dan mencegah kemunculan nilai nol. Metode ini membantu model Naive Bayes dalam mengklasifikasikan pertanyaan secara akurat, berdasarkan probabilitas posterior yang dihitung dari prior dan likelihood pada tiap kategori.

3.3 Probabilitas Prior

Probabilitas prior menggambarkan seberapa besar kemungkinan sebuah kategori muncul dalam data pelatihan. Probabilitas ini dihitung dengan membagi jumlah data pada kategori tertentu dengan total keseluruhan data pelatihan. Berikut adalah rumus dari probabilitas prior:

$$P(C) = \frac{\text{jumlah data di kategori}}{\text{total data}}$$

Berikut ini adalah perhitungan probabilitas prior :

1. $P(A) = \frac{70}{300} = 0,23$
2. $P(B) = \frac{60}{300} = 0,2$
3. $P(C) = \frac{55}{300} = 0,18$
4. $P(D) = \frac{50}{300} = 0,16$
5. $P(E) = \frac{75}{300} = 0,25$

3.4 Probabilitas Likelihood

Probabilitas likelihood adalah peluang munculnya kata tertentu "PPDB" pada kategori tertentu. Untuk menghindari probabilitas nol pada kata yang mungkin jarang muncul, digunakan teknik *Laplace smoothing*. Probabilitas ini dihitung dengan rumus:

$$P(X | C) = \frac{\text{frekuensi kata "PPDB" di C} + 1}{\text{total kata} + \text{ukuran vocabulary}}$$



Berikut ini adalah perhitungan probabilitas likelihood:

1. $P("PPDB" | A) = \frac{15+1}{390+500} = \frac{16}{890} = 0,0179$
2. $P("PPDB" | B) = \frac{13+1}{380+500} = \frac{14}{880} = 0,0159$
3. $P("PPDB" | C) = \frac{14+1}{350+500} = \frac{15}{850} = 0,0176$
4. $P("PPDB" | D) = \frac{12+1}{360+500} = \frac{13}{860} = 0,0151$
5. $P("PPDB" | E) = \frac{48+1}{400+500} = \frac{49}{900} = 0,0544$

3.5 Probabilitas Posterior

Probabilitas posterior menunjukkan kemungkinan kategori tertentu diberikan kemunculan kata "PPDB". Nilai ini proporsional terhadap hasil perkalian probabilitas prior dan likelihood:

$$P(C|"PPDB") = P("PPDB"|C) \times P(C)$$

Berikut ini adalah perhitungan probabilitas posterior:

1. $P(A|"PPDB") = 0,0179 \times 0,23 = 0,00411$
2. $P(B|"PPDB") = 0,0159 \times 0,2 = 0,00318$
3. $P(C|"PPDB") = 0,0176 \times 0,18 = 0,00316$
4. $P(D|"PPDB") = 0,0151 \times 0,16 = 0,00241$
5. $P(E|"PPDB") = 0,0544 \times 0,25 = 0,0136$

Kategori yang memiliki nilai posterior terbesar adalah kategori E dengan nilai 0,0136. Oleh karena itu, sistem klasifikasi Naive Bayes akan menetapkan bahwa pertanyaan yang mengandung kata "PPDB" termasuk dalam kategori E.

3.6 Validasi Data

Melakukan evaluasi performa mode klasifikasi, seperti *Naive Bayes* pada sistem *chatbot* diperlukan data hasil prediksi yang telah diklasifikasikan ke dalam kategori TP, FP, FN, dan TN. Berikut adalah data untuk menghitung *confusion matrix*:

Tabel 4 Data Confusion Matrix

No	Nama	Total
1	True Positive	60
2	False Negative	5
3	False Positive	10
4	True Negative	100

1. Accuracy

Mengukur proporsi prediksi yang benar, baik untuk kelas positif maupun negatif, dari keseluruhan data pengujian. Nilai akurasi yang tinggi menandakan bahwa model mampu melakukan klasifikasi dengan baik. Berikut rumus untuk menghitung akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Berikut ini adalah perhitungan akurasi:

$$Akurasi = \frac{60 + 100}{60 + 100 + 10 + 5} = \frac{160}{175} = 0,914 \text{ atau } 91.4 \%$$



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 7 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1913-1920

Setelah menghitung nilai akurasi, diperoleh nilai 91.4 % dengan ini model klasifikasi mampu memprediksi dengan benar sejumlah data dari keseluruhan data. Nilai akurasi yang tinggi menunjukkan bahwa model berperforma baik dalam mengklasifikasikan pertanyaan pada chatbot.

2. Precision

Precision menunjukkan seberapa akurat model dalam memprediksi data positif. Nilai *precision* yang tinggi berarti sebagian besar prediksi positif yang dilakukan oleh model adalah benar-benar positif, sehingga tingkat kesalahan FP rendah. Berikut rumus untuk menghitung *precision*:

$$Precision = \frac{TP}{TP + FP}$$

Berikut ini adalah perhitungan *precision*:

$$Precision = \frac{60}{60 + 10} = \frac{60}{70} = 0,857 \text{ atau } 85.7\%$$

Nilai *precision* sebesar 0,857 atau 85.7 menunjukkan bahwa prediksi positif adalah benar, bahwa model cukup handal dalam mengidentifikasi data positif dengan tingkat kesalahan FP yang relatif kecil.

3.7 Implementasi Chatbot

1. Halaman Chat



Gambar 3. Halaman Chat



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 7 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1913-1920

2. Halaman Layanan Chat



Gambar 4. Halaman Layanan Chat

4. KESIMPULAN

1. Perancangan dan implementasi *Naive Bayes Classifier* pada *chatbot* Whatsapp untuk mengklasifikasikan pertanyaan terkait PPDB berhasil dilakukan dengan baik. Algoritma *Naive Bayes* mampu mengklasifikasikan pertanyaan secara efektif dan mendukung sistem *chatbot* dalam memberikan jawaban yang relevan dan akurat kepada pengguna.
2. Aplikasi *chatbot* yang dikembangkan dapat memudahkan para orang tua dalam mencari informasi tentang PPDB secara cepat, efisien dan akurat melalui platform Whatsapp. *Chatbot* ini memberikan kemudahan akses informasi yang interaktif, sehingga para orang tua tidak perlu lagi menunggu admin sekolah untuk memberikan informasi secara manual.

REFERENCES

- Amirulloh, I., Pertiwi, M. W., & Wibisono, T. (2024). RANCANG BANGUN CHATBOT WHATSAPP MENGGUNAKAN NODE JS DAN MODEL NATURAL LANGUAGE PROCESSING UNTUK LAYANAN PPDB SMK YPC TASEMALAYA. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(1). <https://doi.org/10.23960/jitet.v12i1.3846>.
- Diantoni, C., Mufidah, R., Triana Informatika, H., Karawang, S., Ronggo Waluyo, J. H., Timur, T., Karawang, J., & Barat, I. (2024). MEMBANGUN CHATBOT UNTUK INFORMASI MAGANG DAN STUDI INDEPENDEN KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 2). <https://kampusmerdeka.kemdikbud.go.id/>
- Lestari, D., & Subekti, L. (2024). Implementasi Chatbot pada Telegram sebagai Monitoring Assistant dengan Analisis Teks Klasifikasi Menggunakan Metode Support Vector Machine. *Journal of Internet and Software Engineering*, 5(2).
- Harahap, A. K., Djunaedi, F. L., Syastra, T., & Hasian, I. (2024). PERANCANGAN APLIKASI WHATSAPP CHATBOT UNTUK PARIWISATA DI KABUPATEN PURWAKARTA. 5(1), 582–593.
- Habsy, B. A., Septiani, L., Kurniawan, T. H., Khofifah, D., Shakila, D. N., Mubarak, A. K., & Fajariyah, S. A. N. (2024). Penerapan Manajemen Layanan Informasi dalam Bimbingan dan Konseling. *WISSEN: Jurnal Ilmu Sosial dan Humaniora*, 2(4), 247-259.



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 7 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1913-1920

- Amanda, Y., & Ujianti, M. H. (2025). Perancangan Sistem Informasi Penerimaan Peserta Didik Baru Pada Daycare Dan Pre School Ananda Mandiri Slawi Berbasis Web. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 177-184.
- Mamuriyah, N., Ladi, B.-S., Gajah Mada, J., Indah, T., Sekupang, K., Batam, K., & Riau, K. (2024). PEMBANGUNAN CHATBOT INTERAKTIF DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES. 4(2), 82–94. <https://doi.org/10.51903/informatika.v4i2.864>.
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 1).
- Narendra, D., Setyo Utomo, S., Dwi Bhakti, H., Aisyiyah, P., & Devi, R. (2025). PENERAPAN ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI PENILAIAN KINERJA PEGAWAI DI KEDAI XYZ. <https://doi.org/10.8734/Kohesi.v1i2.36>.