

Evaluasi Komprehensif Algoritma Naive Bayes Dan Knn Dalam Pengembangan Chatbot Cerdas Untuk Layanan Informasi PMB

Eko Budi Susanto, Mohammad Reza Maulana, Tri Agus Setiawan, Agus Ilyas

Program Studi Teknik Informatika, STMIK Widya Pratama

Jl. Patriot No. 25 Pekalongan, Jawa Tengah

ekobudi.s@stmik-wp.ac.id, reza@stmik-wp.ac.id, triagus@stmik-wp.ac.id,
agusilyas@stmik-wp.ac.id

Abstract – Penerimaan Mahasiswa Baru (PMB) merupakan salah satu kegiatan penting yang dilaksanakan di STMIK Widya Pratama dalam menjaring calon mahasiswa baru. Dari pelaksanaan kegiatan PMB yang dilakukan, pemberian informasi kepada calon mahasiswa menjadi pintu masuk agar calon mahasiswa dapat mendaftar ke perguruan tinggi. Salah satu upaya untuk meningkatkan efisiensi dan kecepatan pemberian informasi maka akan dibangun chatbot. Chatbot pemberian informasi PMB dikembangkan menggunakan metode Naive Bayes. Model Naive Bayes dipilih karena kemampuannya dalam menangani klasifikasi teks, yang sangat berguna dalam merespon berbagai pertanyaan dari calon mahasiswa. Nilai akurasi algoritma Naive Bayes sebesar 88 persen, sedangkan algoritma KNeighborsClassifier (KNN) sebesar 46 persen untuk mengklasifikasi dataset pertanyaan seputar PMB. Manfaat yang diharapkan dari penelitian ini adalah meningkatkan efisiensi dan akurasi dalam penyampaian informasi PMB, serta memberikan pengalaman yang lebih baik bagi calon mahasiswa dalam mengakses informasi yang mereka butuhkan di STMIK Widya Pratama.

Keywords – Chatbot, Penerimaan Mahasiswa Baru, Naive Bayes, KNeighborsClassifier.

Abstrak – New Student Admissions (Penerimaan Mahasiswa Baru - PMB) is one of the important activities carried out at STMIK Widya Pratama in recruiting prospective new students. From the implementation of PMB activities carried out, providing information to prospective students becomes the gateway for prospective students to register for college. One effort to increase the efficiency and speed of providing information is to build a chatbot. The PMB information-providing chatbot was developed using the Naive Bayes method. The Naive Bayes model was chosen because of its ability to handle text classification, which is very useful in responding to various questions from prospective students. The accuracy value of the Naive Bayes algorithm is 88 percent, while the KNeighborsClassifier (KNN) algorithm is 46 percent for classifying datasets of questions about PMB. The expected benefits of this study are to increase efficiency and accuracy in delivering PMB information, as well as providing a better experience for prospective students in accessing the information they need at STMIK Widya Pratama.

Kata Kunci - Chatbot, Penerimaan Mahasiswa Baru, Naive Bayes, KNeighborsClassifier.

I. PENDAHULUAN

STMIK Widya Pratama merupakan salah satu Perguruan Tinggi Swasta (PTS) berada Kota Pekalongan, Jawa Tengah. Berdasarkan informasi pada halaman resmi website pada di <https://stmik-wp.ac.id/> PTS ini memiliki beberapa program studi diantaranya Teknik Informatika dan Sistem Informasi untuk jenjang Strata 1 [1]. Selain itu STMIK Widya Pratama juga memiliki program studi pada jenjang D3 yaitu Komputerisasi Akuntansi. Salah satu hal yang sangat penting bagi PTS dalam keberlangsungan proses pendidikan adalah adanya mahasiswa yang melakukan pendaftaran di perguruan tinggi ini.

Penerimaan Mahasiswa Baru (PMB) merupakan salah satu proses penting dalam perguruan tinggi yang membutuhkan perhatian khusus [2], [3]. Berdasarkan informasi dari panitia PMB, calon mahasiswa memerlukan informasi yang cepat dan akurat mengenai detail informasi tentang PMB. Namun, keterbatasan sumber daya manusia sering kali menjadi kendala dalam menyediakan layanan informasi yang responsif dan efisien.

Alternatif solusi yang dapat diterapkan untuk memberikan informasi dengan cepat dan responsif kepada calon mahasiswa baru adalah penggunaan chatbot [4], [5].

Dari pembahasan yang telah diuraikan, maka perlu dikembangkan aplikasi chatbot menggunakan algoritma Naive Bayes dan KNN untuk membantu pelaksanaan kegiatan PMB dalam hal pemberian informasi agar respon yang diberikan kepada calon pendaftar menjadi lebih responsif dan cepat. Harapannya dengan adanya respon yang gesit, calon pendaftar dapat terlayani dengan lebih baik. Dengan menerapkan metode ini, diharapkan chatbot yang dikembangkan dapat memberikan informasi yang relevan dan sesuai dengan kebutuhan calon mahasiswa.

A. Chatbot

Chatbot adalah sebuah program komputer yang dirancang untuk mensimulasikan percakapan dengan manusia, biasanya melalui antarmuka teks atau suara [6]. Chatbot memanfaatkan teknologi kecerdasan buatan (AI) untuk memahami, memproses, dan merespons input pengguna dalam bentuk percakapan. Penggunaan chatbot memungkinkan otomatisasi

komunikasi, meningkatkan efisiensi, serta memberikan respons yang cepat dan konsisten kepada pengguna [6], [7], [8], [9].

B. Naive Bayes

Salah satu metode yang dapat digunakan dalam pengembangan chatbot adalah Naive Bayes, yang merupakan metode klasifikasi probabilistik yang sederhana namun efektif [5], [10]. Naive Bayes dapat memprediksi respons yang sesuai berdasarkan pola data yang telah dilatih sebelumnya. Pada penelitian yang dilakukan oleh Diantori dkk menunjukkan bahwa penggunaan metode Naive Bayes, kinerja model menunjukkan nilai akurasi yang baik dengan nilai 88.9%. Algoritma Naive Bayes telah terbukti efektif, dan dengan memperbanyak data serta menyesuaikan model, tingkat akurasi chatbot dapat ditingkatkan secara signifikan [5]. Chatbot ini memiliki peluang besar untuk meningkatkan kualitas interaksi dengan pengguna serta memberikan manfaat dalam menyediakan informasi di lingkungan pendidikan.

C. KNN

KNN merupakan algoritma machine learning yang bekerja dengan cara mengklasifikasikan obyek berdasarkan jarak terdekat dengan obyek yang akan diklasifikasikan, algoritma ini sederhana dan cepat dalam melakukan klasifikasi. Saurina, dkk [11] menggunakan algoritma KNN untuk mengklasifikasi sentimen Pelanggan Batik Ecoprint, algoritma ini kemampuan tinggi dalam mengklasifikasikan sentimen dengan tepat. KNN dapat mengklasifikasikan kata/kalimat dari ulasan pelanggan. Pada penelitian tersebut k-nn classifier menghasilkan presisi, recall dan fi-score masing-masing sebesar 76%, 76% dan 74%. Budianto, dkk, [12] mengklasifikasikan sentimen pengumuman hasil pemilu 2024 dengan algoritma KNN, hasilnya algoritma KNN mampu mengklasifikasikan sentimen dengan tingkat akurasi sebesar 68%. Hutagalung, dkk [13] mengembangkan bot untuk rekomendasi webtoon dengan algoritma KNN, hasil pada penelitiannya bot dapat membantu para pembaca dalam menentukan Webtoon yang ingin dibaca dengan cepat dan akurat.

D. Confusion matrix

Confusion matrix adalah alat yang digunakan dalam pembelajaran mesin (machine learning) dan statistik untuk mengevaluasi kinerja model klasifikasi [14]. Confusion matrix menyajikan informasi dalam bentuk tabel yang membandingkan prediksi model dengan hasil yang sebenarnya. Matriks ini memungkinkan kita untuk melihat bagaimana model klasifikasi bekerja pada setiap kelas target dan untuk mengidentifikasi di mana model melakukan kesalahan.

Berikut adalah penjelasan komponen utama dalam confusion matrix:

True Positive (TP): Jumlah prediksi positif yang benar. Ini adalah kasus di mana model memprediksi kelas positif, dan hasil sebenarnya juga positif.

True Negative (TN): Jumlah prediksi negatif yang benar. Ini adalah kasus di mana model memprediksi kelas negatif, dan hasil sebenarnya juga negatif.

False Positive (FP): Jumlah prediksi positif yang salah (juga dikenal sebagai "Type I Error"). Ini adalah kasus di mana model memprediksi kelas positif, tetapi hasil sebenarnya negatif.

False Negative (FN): Jumlah prediksi negatif yang salah (juga dikenal sebagai "Type II Error"). Ini adalah kasus di mana model memprediksi kelas negatif, tetapi hasil sebenarnya positif.

Dengan menggunakan confusion matrix, kita dapat menghitung berbagai metrik kinerja model, seperti:

Accuracy (Akurasi): Persentase prediksi yang benar di antara semua prediksi yang dibuat.

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

Precision (Presisi): Proporsi prediksi positif yang benar di antara semua prediksi positif.

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

Recall (Sensitivitas atau True Positive Rate): Proporsi kasus positif yang benar-benar diidentifikasi oleh model.

$$Recall = \frac{TP}{TP + FN}$$

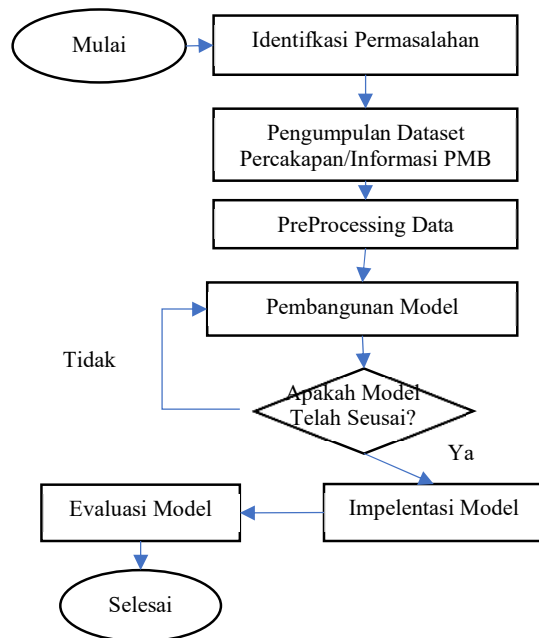
F1 Score: Rata-rata harmonik dari precision dan recall, memberikan keseimbangan antara keduanya.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Confusion matrix adalah alat yang sangat berguna untuk memahami bagaimana model klasifikasi bekerja pada setiap kelas dan untuk mengidentifikasi area di mana model mungkin memerlukan perbaikan.

II. METODE PENELITIAN

Dalam pengembangan chatbot untuk pemberian informasi terkait penerimaan mahasiswa baru (PMB) di STMIK Widya Pratama, kerangka pemikiran yang digunakan melibatkan beberapa langkah utama seperti yang terlihat pada gambar 1. Gambar 1 menunjukkan alur penelitian yang akan dilakukan dimulai dari proses identifikasi permasalahan yang ada. Setelah itu pengumpulan dataset dilakukan dilanjutkan dengan pre-processing data. Kemudian pembangunan model menggunakan Naive Bayes dan KNeighborsClassifier dilakukan. Jika model sudah sesuai maka dilakukan tahap berikutnya yaitu implementasi dan dilanjutkan dengan evaluasi model. Bahasa pemrograman yang digunakan untuk membangun model pada penelitian ini adalah python.



Gambar 1. Alur Penelitian

1. Identifikasi Permasalahan

Proses PMB melibatkan berbagai tahapan mulai dari pengumpulan informasi, pendaftaran, seleksi, hingga pengumuman hasil. Berdasarkan informasi dari panitia PMB, calon mahasiswa memerlukan informasi yang cepat dan akurat mengenai detail informasi terkait kegiatan akademik maupun non-akademik, prosedur penerimaan, persyaratan, serta batas waktu yang ditetapkan. Namun, keterbatasan sumber daya manusia sering kali menjadi kendala dalam menyediakan layanan informasi yang responsif dan efisien. Alternatif solusi yang dapat diterapkan untuk memberikan informasi dengan cepat dan responsif kepada calon mahasiswa baru adalah penggunaan chatbot.

2. Pengumpulan Data

Dataset merupakan hasil informasi percakapan antara petugas penerimaan mahasiswa baru dengan calon pendaftar atau calon mahasiswa, acuan jawaban dari pertanyaan calon pendaftar adalah informasi yang ada di brosur penerimaan mahasiswa baru tahun akademik 2024-2025 STMIK Widya Pratama. Data dikumpulkan menjadi dataset yang terdiri dari variable/atribut/field yaitu: patterns, tag, dan respon.

Tabel 1. Atribut Dataset

No	Nama Field	Keterangan
1	Patterns	Berupa pertanyaan seputar pmb
2	Tag	Kelas/Label dari Patterns
3	Respon	Respon yang akan dipilih

Atribut dataset terdiri dari patterns, tag, dan respon. Pattern berisi tentang pertanyaan-

pertanyaan seputar PMB. Tag merupakan kelas/label dari pertanyaan. Dan Respon merupakan jawaban yang nantinya akan dipilih secara random oleh algoritma. Dataset tersebut disimpan dalam file *.json.

	text_input	intents
0	Hallo Kak	salam
1	Hi	salam
2	Hai	salam
3	Halo	salam
4	Apa Kabar	salam
...
122	Daah	bye
123	Semoga harimu menyenangkan	bye
124	Ok makasih	bye
125	Sampai Jumpa Lagi	bye
126	Ok bye	bye

Gambar 2. Isi Penggalan Dataset

Gambar 2 merupakan tampilan isi dataset yang ditampilkan dengan library pandas dataframe pada python. File dataset dipanggil dengan menggunakan library python JSONParser, kemudian di ubah menjadi format dataframe dengan menggunakan library get_dataframe(). Dataset terdiri dari 126 record dan 21 kelas.

3. PreProcessing Data

Pada tahapan preprocessing data dilakukan untuk mengubah huruf besar menjadi huruf kecil dan menghilangkan tanda baca (cleansing). Teknik yang digunakan yaitu case folding. Berikut contoh proses preprocessing data dengan menggunakan case folding:

Tabel 2. Contoh Implementasi Case Folding

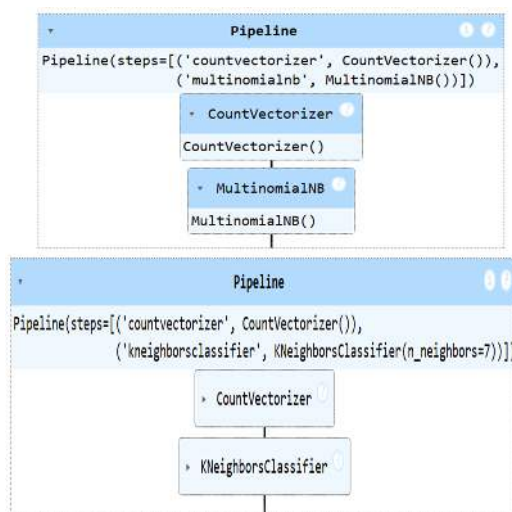
Kalimat	Hasil Case Foling dan Cleansing
Apa saja Program Studi yang dimiliki?	apa saja program studi yang dimiliki
"Selamat Siang Kak ?"	selamat siang kak

Proses mengubah dari huruf besar ke huruf kecil dengan library "lower()". Sedangkan untuk menghilangkan tanda baca (cleansing) menggunakan "string.punctuation". Pada tahap ini juga dilakukan proses label encoding. Label encoding merupakan teknik untuk mengolah data, mengubah nilai-nilai kolom, kategori menjadi nilai numerik atau label. Proses ini mengubah data kategorikal menjadi angka 0 atau 1. Proses label encoding dilakukan secara otomatis bersamaan dengan proses vektorisasi. Vektorisasi yaitu proses mengubah teks menjadi numerik (label encoding) dengan menghitung kata atau word yang memiliki frekuensi yang sering muncul. Proses vektorisasi, pada bagian ini menggunakan

library “CountVectorizer” yang diimport dari library scikit-learn yaitu “sklearn.feature_extraction.text”.

4. Pembangunan Model

Modeling menggunakan algoritma machine learning yaitu algoritma klasifikasi probabilitas naïve bayes, algoritma ini sederhana namun efektif, dapat memprediksi respons yang sesuai pola data yang telah dilatih. Algoritma klasifikasi lainnya yaitu: KNeighborsClassifier (KNN), algoritma ini kemampuan tinggi dalam mengklasifikasikan sentimen dengan tepat [12]. Pada eksperimen ini menggunakan library python MultinomialNB [15] dan KNeighborsClassifier [16].



Gambar 3. Model Yang Diusulkan

5. Implementasi Model

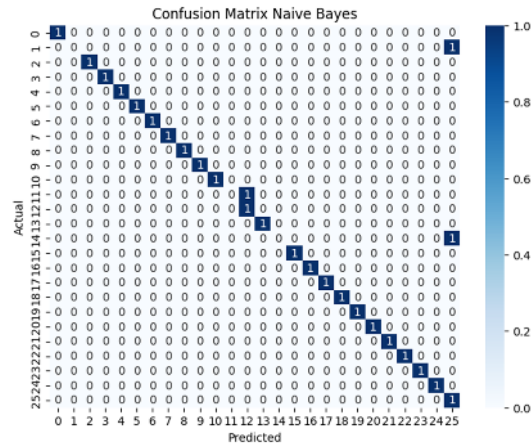
Model yang diusulkan selanjutnya diimplementasikan untuk melakukan klasifikasi pada dataset yang digunakan. Kedua model yang diusulkan dilatih menggunakan dataset yang dibagi menjadi dua bagian yaitu: 20 peren untuk data testing, 80 persen untuk data training. Dalam membangun dan melatih masing-masing model menggunakan teknik pipeline. Pipeline merupakan serangkaian proses yang dilalui suatu pertanyaan sejak pengguna menanyakannya hingga chatbot memberikan respons.

6. Evaluasi Model

Masing-masing model akan diuji dengan data testing yang telah disiapkan sebelumnya. Teknik evaluasi yang digunakan yaitu confusion matrix klasifikasi [14], untuk mengetahui tingkat akurasi, f1-score, recall, dan precision. Selanjutnya hasil evaluasi ini akan dicari model dengan tingkat kinerja yang terbaik, untuk dipilih sebagai model yang akan diimplementasikan pada aplikasi chatbot.

III. HASIL DAN PEMBAHASAN

Model dilatih dengan dataset yang telah dibagi menjadi data training dan data testing, kemudian dilakukan pengujian akurasi menggunakan confusion matrix untuk mengetahui tingkat akurasi, precision, recall, dan f1-score. Berdasarkan hasil dari gambar 4 dan tabel 3 didapatkan tingkat akurasi dari naïve bayes yaitu: 88 persen.



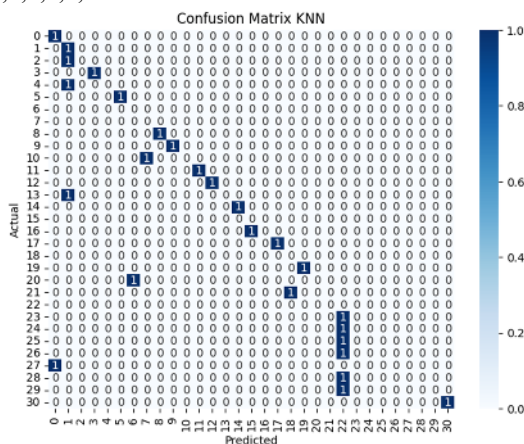
Gambar 4 Hasil Confusion Matrix Naive Bayes

Tabel 3. Tabel Precision, Recall, F1-Score Naive Bayes

	precision	recall	f1-score
apa itu ti	1.00	1.00	1.00
apa kabar	0.00	0.00	0.00
apa kemampuan lo	1.00	1.00	1.00
apa sih sistem informasi	1.00	1.00	1.00
apa yang dipelajari	1.00	1.00	1.00
bagaimana cara daftarnya	1.00	1.00	1.00
hallo kak	1.00	1.00	1.00
hallo selamat malam	1.00	1.00	1.00
hallo selamat siang	1.00	1.00	1.00
hubungi siapa ya	1.00	1.00	1.00
info fakultas	1.00	1.00	1.00
jadi apa lulusannya	0.00	0.00	0.00
jadi apa setelah lulus	0.50	1.00	0.67
jelaskan tentang program studi ti	1.00	1.00	1.00
kamu bisa apa	0.00	0.00	0.00
malam	1.00	1.00	1.00
mendaftarnya caranya bagaimana	1.00	1.00	1.00
nama kamu siapa	1.00	1.00	1.00
pendaftaran	1.00	1.00	1.00
profesi pekerjaan	1.00	1.00	1.00

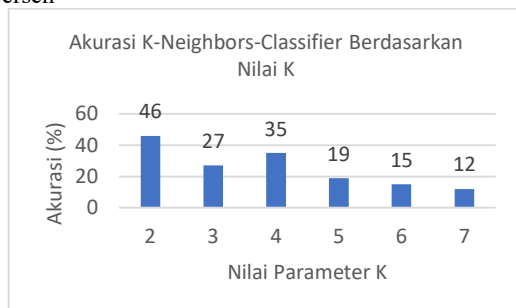
	precision	recall	f1-score
sampai jumpa lagi	1.00	1.00	1.00
selamat tinggal	1.00	1.00	1.00
seperti apa ti itu	1.00	1.00	1.00
siang	1.00	1.00	1.00
telepon	1.00	1.00	1.00
tugas kamu apa	0.33	1.00	0.50

Model K-Neighbors-Classifier dilatih dan ditraining dengan dataset yang sama, diukur tingkat akurasi dengan metode yang sama yaitu confusion matrix. Pada eksperimen ini dilakukan pengujian untuk nilai parameter k pada algoritma K-Neighbors-Classifer. Nilai parameter yang diujikan adalah 2,3,4,5,6,7.



Gambar 5. Hasil Confusion Matrix KNN dengan Nilai K=2

Dari hasil eksperimen didapatkan semakin bertambah nilai k nya semakin kecil tingkat akurasi. Hasilnya didapatkan bahwa nilai parameter k yang terbaik adalah 2. Hasil akurasi yang didapat sebesar 46 persen



Gambar 6. Perbandingan Akurasi Berdasarkan Nilai K

Tabel 4. Tabel Precision, Recall, F1-Score KNN dengan Nilai K=2

	precision	recall	f1-score
apa itu ti	0.50	1.00	0.67
apa kabar	0.25	1.00	0.40
apa kemampuan lo	0.00	0.00	0.00
apa sih sistem informasi	1.00	1.00	1.00
apa yang dipelajari	0.00	0.00	0.00
bagaiman cara daftarnya	1.00	1.00	1.00
bagaimana caranya	0.00	0.00	0.00
hai selamat siang	0.00	0.00	0.00
hallo kak	1.00	1.00	1.00
hallo selamat malam	1.00	1.00	1.00
hallo selamat siang	0.00	0.00	0.00
hubungi siapa ya	1.00	1.00	1.00
info fakultas	1.00	1.00	1.00
jadi apa lulusannya	0.00	0.00	0.00
jadi apa setelah lulus	1.00	1.00	1.00
jelaskan tentang program studi si	0.00	0.00	0.00
jelaskan tentang program studi ti	0.00	0.00	0.00
kamu bisa apa	1.00	1.00	1.00
kamu siapa	0.00	0.00	0.00
malam	1.00	1.00	1.00
mendaftarnya caranya bagaimana	0.00	0.00	0.00
nama kamu siapa	0.00	0.00	0.00
p	0.00	0.00	0.00
pendaftaran	0.00	0.00	0.00
profesi pekerjaan	0.00	0.00	0.00
sampai jumpa lagi	0.00	0.00	0.00
selamat tinggal	0.00	0.00	0.00
seperti apa ti itu	0.00	0.00	0.00
siang	0.00	0.00	0.00
telepon	0.00	0.00	0.00
tugas kamu apa	1.00	1.00	1.00

Tabel 5. Perbandingan Hasil Akurasi

Algoritma	Akurasi (%)
K-Neighbors-Classifier	46
MultinomialNB	88

IV. KESIMPULAN

Berdasarkan hasil pembahasan yang telah dijabarkan, maka penelitian ini dapat disimpulkan bahwa tingkat akurasi (tabel 5) algoritma naïve bayes

pada pengembangan chatbot penerimaan mahasiswa baru di Institut Widya Pratama sebesar 88 persen, yang berarti bahwa dari 10 pertanyaan yang disampaikan terdapat 2 jawaban yang salah. Dibandingkan dengan algoritma K-Neighbors-Classifer, algoritma Naïve Bayes memiliki performa yang lebih baik pada dataset percakapan penerimaan mahasiswa baru. Hal ini sesuai dengan karakter dari algoritma Naïve Bayes yang mempunyai kinerja yang bagus di bidang natural language processing. Nilai parameter k yang menghasilkan tingkat akurasi yang baik pada algoritma K-Neighbors-Classifer adalah 2, dengan tingkat akurasi 46 persen.

Diperlukan implementasi lebih lanjut untuk pengembangan aplikasi chatbot mengingat performa dari algoritma Naïve Bayes yang baik dan menambahkan data training yang lebih variatif dan spesifik untuk meningkatkan akurasinya. Meskipun K-NN tidak sebaik algoritma naïve bayes pada penelitian ini, penyesuaian parameter seperti nilai k dan penggunaan algoritma pengukuran jarak yang berbeda dapat meningkatkan performa K-NN.

V. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Yayasan Pendidikan Widya Pratama yang telah memberikan dukungan financial terhadap penelitian ini, dan kepada akun youtube NgodingPython [17] yang telah memberikan referensi dan pengetahuan yang mendukung penelitian ini.

DAFTAR PUSTAKA

- [1] Unit Humas & Kerjasama STMIK Widya Pratama, "Website STMIK Widya Pratama." Accessed: Aug. 26, 2024. [Online]. Available: <https://stmik-wp.ac.id/>
- [2] D. Heryati, I. Zulkifli, R. M. Fajri, and S. Kom, "Aplikasi Chatbot untuk Penerimaan Mahasiswa Baru Universitas Indo Global Mandiri Menggunakan Deep Learning," *The Journal of Intelligent Networks and IoT Global*, vol. 1, no. 1, 2023.
- [3] M. Nugraha, L. Sakinah, R. A. Setiawan, and H. Mulyani, "Rancang Bangun Sistem Informasi Penerimaan Mahasiswa Baru Berbasis Web dengan Menggunakan Framework Laravel," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4179.
- [4] A. Nurul Puteri, F. Tamrin, K. Rahman Nasir, D. Widya Anggraeni, and M. Arafah, "Aplikasi Chatbot untuk Layanan Informasi Penerimaan Mahasiswa Baru," in *Prosiding Seminar Nasional Teknik Elektro dan Informatika (SNTEI) 2022-Teknik Informatika*, 2022, pp. 168–174.
- [5] C. Diantoni *et al.*, "Membangun Chatbot untuk Informasi Magang dan Studi Independen Kampus Merdeka Dengan Algoritma Naive Bayes," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, 2024, [Online]. Available: <https://kampusmerdeka.kemdikbud.go.id/>
- [6] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," in *IFIP Advances in Information and Communication Technology*, Springer, 2020, pp. 373–383. doi: 10.1007/978-3-030-49186-4_31.
- [7] C. Kooli, "Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions," *Sustainability (Switzerland)*, vol. 15, no. 7, Apr. 2023, doi: 10.3390/su15075614.
- [8] M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, "Interacting with educational chatbots: A systematic review," *Educ Inf Technol (Dordr)*, vol. 28, no. 1, pp. 973–1018, Jan. 2023, doi: 10.1007/s10639-022-11177-3.
- [9] C.-C. Lin, A. Y. Q. Huang, and S. J. H. Yang, "A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022)," *Sustainability*, vol. 15, no. 5, p. 4012, Feb. 2023, doi: 10.3390/su15054012.
- [10] E. B. Susanto, Paminto Agung Christianto, Mohammad Reza Maulana, and Satriedi Wahyu Binabar, "Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLE Samsat Jawa Tengah," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 234–241, Dec. 2022, doi: 10.37859/coscitech.v3i3.4343.
- [11] N. Saurina, T. Rahayuningsih, L. Retnawati, F. Teknik, U. Wijaya, and K. Surabaya, "Analisis Sentimen Ulasan Pelanggan Batik Ecoprint Menggunakan Naïve Bayes Dan KNN Classifier," vol. 9, no. 2, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [12] T. Aurelly Claudia Budianto, H. Fatoni, M. Ayu Syaharani, C. Rozikin Informatika, U. Singaperbangsa Karawang Jl HSRonggo Waluyo, and T. Timur, "ANALISIS SENTIMEN PENGUMUMAN HASIL PEMILU 2024 DI SOSIAL MEDIA X MENGGUNAKAN KNN DAN NAÏVE BAYES CLASSIFIER," 2024.
- [13] D. Durbin Hutagalung, C. Hanifurohman, and D. R. Baskhara, "Bot Sistem Rekomendasi Webtoon Dengan Metode User-Collaborative Filtering Menggunakan Algoritma K-Nearest Neighbor (KNN)." [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- [14] M. M. Mijwil and M. Aljanabi, "A Comparative Analysis of Machine Learning Algorithms for Classification of Diabetes Utilizing Confusion Matrix Analysis," *Baghdad Science Journal*, vol. 21, no. 5, pp. 1712–1728, 2024, doi: 10.21123/BSJ.2023.9010.
- [15] "1.9. Naive Bayes — scikit-learn 1.1.3 documentation." Accessed: Nov. 30, 2022. [Online]. Available: https://scikit-learn.org/stable/modules/naive_bayes.html
- [16] scikit-learn.org, "Nearest Neighbors Classification," <https://scikit-learn.org/stable/modules/neighbors.html#>.
- [17] NgodingPython, "[LIVE] Weekend Project : Membuat Chatbot Sederhana," <https://www.youtube.com/watch?v=HihKQ8F0k4c&t=929s>.