

## Penerapan Natural Language Processing Pada Sistem Chatbot Sebagai Helpdesk Obyek Wisata Menggunakan Metode Naïve Bayes

Yuhandri<sup>1</sup>, Rini Sovia\*<sup>2</sup>, Afif Syaifullah<sup>3</sup>, Firna Yenila<sup>4</sup>, Randy Permana<sup>5</sup>

<sup>1,2,3,4,5</sup> Universitas Putra Indonesia YPTK Padang  
e-mail: <sup>1</sup>yuyu@upiypk.ac.id, <sup>2</sup>rini\_sovia@upiypk.ac.id, <sup>3</sup>afifsyaifullah1512@gmail.com, <sup>4</sup>firna\_yenila@upiypk.ac.id, <sup>5</sup>randy\_permana@upiypk.ac.id

<sup>2</sup>Universitas Putra Indonesia YPTK Padang  
e-mail: rini\_sovia@upiypk.ac.id

Diterima	Direvisi	Disetujui
01-01-2020	01-02-2020	01-03-2020

**Abstrak** - Keberhasilan pengembangan sektor pariwisata tidak hanya bergantung pada daya tarik wisata semata. Terdapat banyak faktor dalam penghambat perkembangan sektor tersebut. Salah satu faktor tersebut adalah kurangnya perkembangan dalam pola periklanan dan sistem pengelolaan informasi pariwisata. Sebagai akibatnya, sasaran pariwisata tidak terdefinisi dengan baik, dan wisatawan mungkin tidak memilih destinasi tersebut. Bukittinggi sebagai salah satu destinasi wisata yang terdapat di Sumatera Barat juga tidak lepas dari permasalahan tersebut. Kurang tersedianya informasi lengkap tentang potensi pariwisata di Kota Bukittinggi membuat wisatawan mengandalkan sumber-sumber seperti media sosial, internet, dan sumber lainnya untuk mendapatkan informasi. Namun, informasi yang ada belum mencakup seluruh aspek pariwisata, sehingga menyebabkan ketidakpastian bagi calon wisatawan. Untuk mengatasi permasalahan tersebut, dihadirkan sebuah sistem respons obrolan otomatis atau disebut dengan *Chatbot*. Teknologi *Chatbot* merupakan salah satu bentuk sistem *Natural Language Processing* (NLP) dalam kecerdasan buatan. *Chatbot* berperan sebagai agen percakapan yang dapat berfungsi sebagai meja bantuan. Dalam konteks ini, *helpdesk* menjadi elemen penting yang menangani berbagai keluhan dari berbagai pihak dengan menyediakan informasi dan solusi. Dalam penelitian ini, dikembangkan sistem *Chatbot* menggunakan algoritma *Naive Bayes* untuk menjawab pertanyaan umum (FAQ) mengenai informasi pariwisata di Kota Bukittinggi.

Kata Kunci: *Chatbot, Natural Language Processing, Naive Bayes, pariwisata*

**Abstract** - The successful development of the tourism sector does not only depend on tourist attractions. There are many factors that hinder the development of this sector. One of these factors is the slow development of advertising patterns and tourism information management systems. As a result, tourism targets are not well defined, and tourists may not choose the destination. Bukittinggi as one of the tourist destinations in West Sumatra is also not free from these problems. The lack of complete information about tourism potential in Bukittinggi City makes tourists rely on sources such as social media, the internet and other sources to get information. However, existing information does not cover all aspects of tourism, thus causing threats to potential tourists. To overcome this problem, an automatic dialogue response system or what is called a chatbot is presented. Chatbot technology is a form of Natural Language Processing (NLP) system in artificial intelligence. Chatbots act as conversational agents that can function as a help desk. In this context, the helpdesk becomes an important element that handles various complaints from various parties by providing information and solutions. In this research, a Chatbot system was developed using the Naive Bayes algorithm to answer general questions (FAQ) regarding tourism information in Bukittinggi City.

Keywords: *Chatbot, Natural Language Processing, Naive Bayes, tourism*

### PENDAHULUAN

Teknologi informasi terdiri atas dua kata yang memiliki makna berbeda. Kata teknologi berasal dari dua kata, yaitu techno yang berarti seni, dan logia

(logos) yang berarti ilmu, teori. Kata informasi berasal dari kata Perancis kuno pada tahun 1387, yaitu informacion yang diambil dari Bahasa Latin informationem yang berarti garis besar, konsep, ide. Informasi merupakan kata benda dari *informare* yang

berarti aktivitas dalam pengetahuan yang dikomunikasikan. Berdasarkan pendekatan tersebut maka teknologi informasi didefinisikan sebagai seni dan ilmu dalam menyampaikan informasi dari pengirim ke penerima (mukrodin & Mega Sasmita, 2021).

Perkembangan teknologi informasi berbasis computer yang sangat cepat telah membuat banyak perubahan dalam sendi kehidupan manusia. Salah satu perkembangan teknologi terkini adalah kecerdasan buatan (*artificial intelligence*). Adanya kecerdasan buatan, computer dapat melakukan tugas tertentu seperti yang dilakukan oleh manusia robot *chatting* (*Chatbot*). Secara garis besar, *Chatbot* dapat melakukan tugas-tugas seperti melayani percakapan (*chatting*) dengan siapapun, seperti menyapa, menjawab, dan juga melayani permintaan dari orang lain (mukrodin & Mega Sasmita, 2021).

*Chatbot* berperan sebagai agen percakapan yang dapat digunakan sebagai helpdesk. *Helpdesk* merupakan struktur atau program yang menangani segala bentuk keluhan dari berbagai pihak dengan menyediakan layanan untuk memberi informasi dan memberikan solusi kepada *user* (Prasetyo et al., 2021). Keberhasilan pengembangan pariwisata tidaklah hanya tergantung pada berapa banyak obyek wisata yang dimiliki, keindahan, kealamian dan keunikan budaya dan tradisi masyarakat disekitar obyek atau kawasan wisata namun yang jauh lebih penting adalah sumber daya manusia sebagai pengelola, sistem manajemen pengelolaan pariwisata dan informasi pariwisata itu sendiri. Kota Bukittinggi merupakan salah satu kota di Sumatera Barat yang memiliki berbagai obyek wisata yang menarik.

Salah satu faktor yang menghambat perkembangan suatu pariwisata adalah pola promosi dan sistem pengelolaan informasi pariwisata yang belum baik sehingga terkadang objek wisata menjadi tidak dikenal dan tentunya tidak menjadi objek tujuan para wisatawan untuk berwisata. Informasi mengenai obyek wisata yang ada di kota Bukittinggi di dapat wisatawan berasal dari berbagai macam media sosial seperti facebook, instagram, twitter, dan mesin pencari seperti Google. Namun, informasi yang didapat oleh wisatawan tidak mampu mencakup seluruh wisata yang ada di kota Bukittinggi, sehingga wisatawan hanya mampu memperoleh informasi hanya berdasarkan yang sudah ada di media social, padahal masih banyak obyek wisata yang menarik dan belum terdapat di internet atau *social media* manapun.

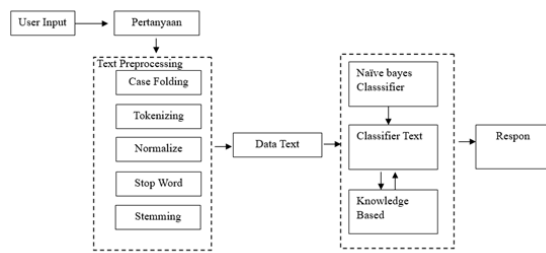
Teknologi *Chatbot* merupakan salah satu bentuk sistem *Natural Language Processing* (NLP), dan merupakan salah satu bidang ilmu Kecerdasan Buatan (*Artificial Intelligence*) yang mempelajari tentang interaksi antara computer dan bahasa alami manusia. Pemrosesan bahasa alami (NLP) baru-baru ini menarik banyak perhatian dalam representasi dan analisis komputasi bahasa manusia, dimana penerapan dari konsep ini dapat diterapkan pada

bidang yang berbeda, termasuk terjemahan mesin, deteksi spam email, ekstraksi informasi, ringkasan, kedokteran, dan menjawab pertanyaan (Khurana et al., 2023). Salah satu penerapan dari NLP yaitu mesin penjawab pertanyaan atau *chatbot* bertujuan untuk berkomunikasi dengan manusia dan dengan bantuan itu membuat banyak pekerjaan yang berlebihan menjadi lebih mudah bagi manusia (Maitri & Sutopo, 2019). *Chatbot* berbasis AI, juga dikenal sebagai agen percakapan atau asisten virtual, adalah perangkat lunak yang mengandalkan pembelajaran mesin (ML) dan pemrosesan bahasa alami (NLP) untuk mensimulasikan percakapan alami manusia dan mengotomatiskan tugas melalui interaksi teks atau suara (Zhou et al., 2023). Teknologi ini memungkinkan chatbots untuk mendukung banyak pelanggan secara bersamaan, memberikan dukungan multibahasa, tersedia 24 jam tanpa biaya tambahan, membebaskan manusia terhadap tugas yang berulang, dan mendukung berbagai layanan (William et al., 2023). *Chatbot* menyediakan layanan pelayanan informasi yang efisien dan dengan demikian bermanfaat bagi para wisatawan. Algoritma Naive Bayes dikembangkan berdasarkan dari teorema Bayes dan termasuk kedalam kategori algoritma supervised learning. Chatbots dapat dilatih sebelumnya untuk memahami pertanyaan pengguna dan menghasilkan jawaban secara instan dengan NLP (Dongbo et al., 2023). Algoritma ini menggunakan *training dataset* (kumpulan data untuk dilatih) yang akan memandu komputer dalam menghasilkan keluaran yang sesuai dengan harapan (Hutama et al., 2021).

Penelitian ini menggunakan algoritma Naive Bayes untuk mengembangkan sistem *Chatbot* yang mampu mencari informasi yang dibutuhkan mengenai objek wisata dengan menggabungkan algoritma tersebut dengan *Natural Language Processing* (NLP). Model penjawab pertanyaan (QA) yang digunakan pada NLP memastikan bahwa pertanyaan dan jawaban yang diberikan sesuai dengan keakuratan yang baik, dimana salah satu konsep yang dapat menjamin mutu dari jawaban tersebut menerapkan klasifikasi pertanyaan (Thananukhun et al., 2023). NLP digunakan untuk membantu *Chatbot* dalam menganalisa dan memahami bahasa alami manusia sementara algoritma Naive Bayes digunakan pada tahap pengklasifikasian teks yang dilatih sehingga mampu menangani permintaan pengguna (Ratiasadara et al., 2023).

## METODE PENELITIAN

Berikut merupakan metodologi penelitian yang diusulkan dari penelitian *Chatbot* berbasis Naive bayes Classifier.



Sumber:(Yuhandri dkk, 2023)

Gambar 1. Metode Penelitian

Berdasarkan gambar 1 sistem akan mengolah bahasa dari input teks pengguna, input text merupakan rangkaian text yang diinputkan kedalam sistem dalam bentuk bahasa alami manusia oleh pengguna. Input text kemudian dilakukan Preprocessing dengan menggunakan konsep Natural Language Processing. Terdapat beberapa tahapan dalam preprocessing data didalam penelitian ini yaitu : (a) Case Folding yaitu mengubah seluruh teks menjadi huruf kecil untuk menghindari perbedaan besar-kecil huruf, (b) Tokenizing yaitu dengan membagi teks menjadi token-token, seperti kata-kata atau frasa, (c) Normalisasi mengatasi variasi dalam teks seperti mengganti singkatan, ejaan alternatif, atau sinonim dengan bentuk standar, (d) Stop Word Removal menghilangkan kata-kata umum (stop words) yang tidak memiliki makna signifikan dalam analisis, (e) Stemming atau Lemmatization Mengubah kata-kata ke bentuk dasarnya (stemming) atau ke bentuk kata dasar (lemmatization).

Model klasifikasi yang digunakan pada penelitian ini menggunakan algoritma klasifikasi Naive Bayes. Model ini akan digunakan untuk mengklasifikasikan teks ke dalam kategori atau label yang sesuai berdasarkan pengetahuan berbasis aturan.

Berdasarkan hasil klasifikasi dan pengetahuan berbasis aturan, sistem dapat menghasilkan respon yang sesuai kepada pengguna akhir. Respon ini dapat berupa informasi, rekomendasi, atau tindakan yang diambil berdasarkan masukan pengguna.

**Hasil Penelitian dan Pembahasan**

Natural Language Processing berbasis Chatbot dapat memproses pertanyaan yang diajukan oleh pengguna kedalam bahasa alami, dan mencari kata kunci yang relevan dengan informasi yang diinginkan pengguna. Untuk mencapai hal tersebut perlu dikaitkan beberapa kata kunci yang telah didefinisikan sebelumnya dengan rule yang tersedia pada basis pengetahuan seperti yang terlihat pada tabel 1.

Tabel 1. Kata Kunci dan Rule

Kategori Pertanyaan	Kata Kunci	Rule
Destinasi	Wisata	C1R1
Destinasi	Destinasi	C1R2
Lokasi	Lokasi	C2R1
Lokasi	Letak	C2R2
Deskripsi	Sejarah	C3R1
Deskripsi	Tentang	C3R2
Tarif	Tarif	C4R1
Tarif	Ongkos	C4R2
Tarif	Biaya	C4R3
Waktu	Kapan	C5R1
Waktu	Jam	C5R2
Transportasi	Transportasi	C6R1
Transportasi	Kendaraan	C6R2

Kemudian, dengan menggunakan metode Naive Bayes, sistem Chatbot mengklasifikasikan pertanyaan pengguna ke dalam kategori-kategori tertentu. Kategori tersebut meliputi nama obyek wisata, jenis obyek wisata, atau kriteria obyek wisata. Selanjutnya, sistem Chatbot dapat mengakses basis data yang berisi informasi tentang obyek wisata, dan memberikan jawaban yang sesuai dengan pertanyaan pengguna pada Tabel 2.

Tabel 2. Jawaban Berdasarkan Rule

Rule	Jawaban
C1R1	Ada begitu banyak wisata yang terdapat di daerah Bukittinggi yang dapat memanjakan mata, [Destinasi]
C1R2	Terdapat 14 destinasi wisata yang tersebar di berbagai wilayah kota bukittinggi, yaitu [destinasi]
C2R1	Wisata [destinasi] berlokasi di [lokasi]
C2R2	Wisata [destinasi] merupakan salah satu wisata yang sering dikunjungi oleh wisatawan, terletak di [jarak] kilometer dari [destinasi]
C3R1	[destinasi] [sejarah]
C3R2	[destinasi] [deskripsi]
C4R1	Tarif yang dipatok untuk memasuki [destinasi] adalah sebesar [tarif]
C4R2	Untuk transportasi umum seperti angkot biasanya mematok biaya sebesar Rp. 4000, dan untuk transportasi umum berbasis online lainnya membutuhkan biaya yang bervariasi tergantung dari jarak tempuh yang dibutuhkan.
C4R3	Biaya yang dibutuhkan untuk menikmati wisata {destinasi} sebesarRp. [tarif]
C5R1	Wisada [destinasi] tersebut dibuka pada pukul 8 pagi hingga jam 5 sore
C5R2	Waktu yang dibutuhkan dari [destinasi] menuju [destinasi] yaitusekitar [waktu] jam

C6R1	Kendaraan pribadi roda 2 dan 4 apat digunakan untuk menuju destinasi wisata [destinasi]
C6R2	Anda dapat menggunakan kendaraan umum seperti Angkot, Grab, Gojek ataupun Maxim untuk menuju ke destinasi wisata yang anda inginkan.

Sebagai contoh, terdapat beberapa pertanyaan yang masuk kedalam sistem *Chatbot* seperti tabel 3.

Tabel 3. Contoh Pertanyaan

Kategori Pertanyaan	Pertanyaan
Destinasi	Apa Saja destinasi Wisata yg terdapat di Bukittinggi?
Lokasi	Terletak didaerah mana wisata jam gadang?
Deskripsi	Jelaskan tentang jembatan limpapeh?
Tarif	Berapa biaya masuk ke wisata Lobang Jepang?
Waktu	Jam berapa wisata Benteng Fort de Kock buka?
Transportasi	Apa transportasi/kendaraan umum yg digunakan ke wisata Taman Ngarai Maaram?

Pertanyaan yang masuk kedalam sistem akan dilakukan preprocessing sesuai tahapan yang sudah dijelaskan pada bab sebelumnya. Langkah pertama adalah melakukan Case Folding terhadap data dimana teks dirubah menjadi huruf kecil seperti yang terlihat pada tabel 4.

Tabel 4. Hasil Case Folding

Kategori Pertanyaan	Pertanyaan
Destinasi	apa saja destinasi wisata yg terdapat di bukittinggi
Lokasi	terletak didaerah mana wisata jam gadang
Deskripsi	jelaskan tentang jembatan limpapeh
Tarif	berapa biaya masuk ke wisata lobang jepang
Waktu	jam berapa wisata benteng fort de kock buka
Transportasi	apa transportasi kendaraan umum yg digunakan ke wisata tamanngarai maaram

Tahapan tokenizing bertujuan untuk memecah kalimat menjadi sebuah kata atau token dengan cara membelah kata dan mendefinisikan struktur atau token dengan cara menentukan unsur sintaksis per kata seperti yang terlihat pada tabel 5.

Tabel 5. Tokenisasi Teks

D1	D2	D3	D4	D5	D6
apa	terletak	jelaskan	berapa	jam	apa
saja	didaerah	tentang	biaya	berapa	transportasi
destinasi	mana	jembatan	masuk	benteng	kendaraan
wisata	wisata	limpapeh	ke	fort	umum
yg	jam		wisata	de	yg
terdapat	gadang		lobang	kock	digunakan
di			jepang	buka	ke
bukittinggi					wisata
					taman
					ngarai
					maaram

Normalizaton bertujuan untuk mengembalikan kata menjadi ejaan yang benar menggunakan tabel normalisasi untuk mengganti dengan kata baku dari kata yang tidak baku dan laveinsthein distance untuk pengkoreksian ejaan dengan mengambil jarak kesalahan 1.

Tabel 6. Normalisasi Teks

D1	D2	D3	D4	D5	D6
apa	Terletak	jelaskan	berapa	jam	apa
saja	Didaerah	tentang	biaya	berapa	transportasi
destinasi	Mana	jembatan	masuk	benteng	kendaraan
wisata	Wisata	limpapeh	ke	fort	umum
yang	Jam		wisata	de	yang
terdapat	Gadang		lobang	kock	digunakan
di			jepang	buka	ke
bukittinggi					wisata
					taman
					ngarai
					maaram

Tahap stopword removal bertujuan untuk menghapus kata yang tidak dibutuhkan untuk mengurangi dimensi ruang pemrosesan. Kata tersebut merupakan kata preposisi, kata tanya dan lainnya yang bukan kata kunci yang penting untuk diproses.

Tabel 7. Stopword Removal

D1	D2	D3	D4	D5	D6
apa	letak	<b>jelas</b>	<b>berapa</b>	jam	apa
<b>saja</b>	didaerah	tentang	biaya	<b>berapa</b>	transportasi
destinasi	<b>mana</b>	jembatan	masuk	benteng	kendaraan
wisata	wisata	limpapeh	ke	fort	umum
<b>yang</b>	jam		wisata	de	<b>yg</b>
<b>terdapat</b>	gadang		lobang	kock	<b>digunakan</b>
<b>di</b>			jepang	buka	<b>ke</b>
bukittinggi					wisata
					taman
					ngarai
					maaram

Pada tabel 7 teks yang ditulis bold akan dihilangkan dari data input yang terdapat setelah melalui tahapan Normalisasi teks.

Tahapan Stemming untuk mengubah kata dengan suffix menjadi kata dasar. Proses stemming dimulai dengan menentukan apakah input berupa kata dasar atau tidak dengan membandingkan kata dengan kamus kata dasar. Proses akan menghilangkan

imbuhan kemudian mengganti kata tersebut dengan kata dasar jika kata yang dicari belum berbentuk kata dasar. Proses akan mengembalikan kata awal tersebut jika kata yang dicari sudah berbentuk kata dasar.

Tabel 8. Stemming

D1	D2	D3	D4	D5	D6
destinasi	letak	tentang	biaya	jam	transportasi
wisata	daerah	jembatan	masuk	benteng	kendaraan
bukittinggi	wisata	limpapeh	wisata	fort	umum
	jam		lobang	de	wisata
	gadang		jepang	kock	taman
				buka	ngarai
					maaram

Pada tabel 8, semua data teks sudah berubah menjadi kata dasar. Hasil setelah stemming merupakan hasil akhir text preprocessing dan digunakan untuk tahapan pembobotan kata.

Pembobotan kata bertujuan untuk mengubah kata menjadi vektor dimana setiap dianggap sebagai satu fitur. Pembobotan TF-IDF digunakan dalam penelitian ini yang merupakan gabungan dari TF (Term Frequency) untuk mengkalkulasi bobot kata dalam dokumen kumulatif dan IDF untuk menghitung bobot kemunculan kata pada kumpulan dokumen.

Tabel 9. Pembobotan Dokumen

Term	TF						DF	D/DF	IDF	IDF+1	TF.IDF					
	D1	D2	D3	D4	D5	D6					D1	D2	D3	D4	D5	D6
benteng	0	0	0	0	1	0	1	6	1,791759 469	2,791759 469	0	0	0	0	2,791759 469	0
biaya	0	0	0	1	0	0	1	6	1,791759 469	2,791759 469	0	0	0	2,791759 469	0	0
buka	0	0	0	0	1	0	1	6	1,791759 469	2,791759 469	0	0	0	0	2,791759 469	0
bukittinggi	1	0	0	0	0	0	1	6	1,791759 469	2,791759 469	2,791759 59469	0	0	0	0	0
daerah	0	1	0	0	0	0	1	6	1,791759 469	2,791759 469	0	2,791759 469	0	0	0	0
de	0	0	0	0	1	0	1	6	1,791759 469	2,791759 469	0	0	0	0	2,791759 469	0
destinasi	1	0	0	0	0	0	1	6	1,791759 469	2,791759 469	2,791759 59469	0	0	0	0	0
fort	0	0	0	0	1	0	1	6	1,791759 469	2,791759 469	0	0	0	0	2,791759 469	0
gadang	0	1	0	0	0	0	1	6	1,791759 469	2,791759 469	0	2,791759 469	0	0	0	0
jam	0	1	0	0	1	0	2	3	1,098612 289	2,098612 289	0	2,098612 289	0	0	2,098612 289	0

jambatan	0	0	1	0	0	0	1	6	1,791759469	2,791759469	0	0	2,791759469	0	0	0
jepang	0	0	0	1	0	0	1	6	1,791759469	2,791759469	0	0	0	2,791759469	0	0
kendaraan	0	0	0	0	0	1	1	6	1,791759469	2,791759469	0	0	0	0	0	2,791759469
kock	0	0	0	0	1	0	1	6	1,791759469	2,791759469	0	0	0	0	2,791759469	0
letak	0	1	0	0	0	0	1	6	1,791759469	2,791759469	0	2,791759469	0	0	0	0
limpapeh	0	0	1	0	0	0	1	6	1,791759469	2,791759469	0	0	2,791759469	0	0	0
lobang	0	0	0	1	0	0		6	1,791759469	2,791759469	0	0	0	2,791759469	0	0
maaram	0	0	0	0	0	1	1	6	1,791759469	2,791759469	0	0	0	0	0	2,791759469
masuk	0	0	0	1	0	0	1	6	1,791759469	2,791759469	0	0	0	2,791759469	0	0
ngarai	0	0	0	0	0	1	1	6	1,791759469	2,791759469	0	0	0	0	0	2,791759469
taman	0	0	0	0	0	1	1	6	1,791759469	2,791759469	0	0	0	0	0	2,791759469
tentang	0	0	1	0	0	0	1	6	1,791759469	2,791759469	0	0	2,791759469	0	0	0
transportasi	0	0	0	0	0	1	1	6	1,791759469	2,791759469	0	0	0	0	0	2,791759469
umum	0	0	0	0	0	1	1	6	1,791759469	2,791759469	0	0	0	0	0	2,791759469
wisata	1	1	0	1	0	1	4	1,5	0,405465108	1,405465108	1,405465108	1,405465108	0	1,405465108	0	1,405465108

Setelah didapatkan bobot dari setiap kata, selanjutnya penjumlahan bobot kata berdasarkan kategori pertanyaan.

Tabel 10. Penjumlahan Bobot

Term	Σ W Kata t						IDF
	Destinasi	Lokasi	Deskripsi	Tarif	Waktu	Akses	
benteng	0	0	0	0	2,791759469	0	1,791759469
biaya	0	0	0	2,791759469	0	0	1,791759469
buka	0	0	0	0	2,791759469	0	1,791759469
bukittinggi	2,791759469	0	0	0	0	0	1,791759469
daerah	0	2,791759469	0	0	0	0	1,791759469
de	0	0	0	0	2,791759469	0	1,791759469

destinasi	2,791759469	0	0	0	0	0	1,791759469
fort	0	0	0	0	2,791759469	0	1,791759469
gadang	0	2,791759469	0	0	0	0	1,791759469
jam	0	2,098612289	0	0	2,098612289	0	1,098612289
jambatan	0	0	2,791759469	0	0	0	1,791759469
jepang	0	0	0	2,791759469	0	0	1,791759469
kendaraan	0	0	0	0	0	2,791759469	1,791759469
kock	0	0	0	0	2,791759469	0	1,791759469
letak	0	2,791759469	0	0	0	0	1,791759469

limpa peh	0	0	2,7917 59469	0	0	0	1,7917 59469
loban g	0	0	0	2,7917 59469	0	0	1,7917 59469
maara m	0	0	0	0	0	2,7917 59469	1,7917 59469
masu k	0	0	0	2,7917 59469	0	0	1,7917 59469
ngarai	0	0	0	0	0	2,7917 59469	1,7917 59469
taman	0	0	0	0	0	2,7917 59469	1,7917 59469
tentan g	0	0	2,7917 59469	0	0	0	1,7917 59469
transp ortasi	0	0	0	0	0	2,7917 59469	1,7917 59469
umum	0	0	0	0	0	2,7917 59469	1,7917 59469
wisata	1,4054 65108	1,4054 65108	0	1,4054 65108	0	1,4054 65108	0,4054 65108
Σ	6,9889 84047	11,879 3558	8,3752 78408	12,572 50299	16,057 40963	18,156 02192	42,714 54519

Training dilakukan untuk menentukan model klasifikasi yang dijadikan acuan untuk pengujian. Langkah pertama dalam fase training adalah menggunakan maximum likelihood untuk menghitung nilai prior probabilities P(c) dan dilanjutkan dengan menggunakan laplace smoothing (ls) untuk mencari nilai conditional probabilities.

Tabel 11. laplace smoothing

Term	LS					
benten g	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,06451 6477	0,01642 8301
biaya	0,02011 9296	0,01831 7064	0,0195 7337	0,06858 3142	0,01701 4918	0,01642 8301
buka	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,06451 6477	0,01642 8301
bukitti nggi	0,07628 753	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,01642 8301
daerah	0,02011 9296	0,06945 3902	0,0195 7337	0,01808 7419	0,01701 4918	0,01642 8301
de	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,06451 6477	0,01642 8301

destina si	0,07628 753	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,01642 8301
fort	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,06451 6477	0,01642 8301
gadang	0,02011 9296	0,06945 3902	0,0195 7337	0,01808 7419	0,01701 4918	0,01642 8301
jam	0,02011 9296	0,05675 7481	0,0195 7337	0,01808 7419	0,05272 2634	0,01642 8301
jambatan	0,02011 9296	0,01831 7064	0,0742 1751	0,01808 7419	0,01701 4918	0,01642 8301
jepang	0,02011 9296	0,01831 7064	0,0195 7337	0,06858 3142	0,01701 4918	0,01642 8301
kendar aan	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,06229 2166
kock	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,06451 6477	0,01642 8301
letak	0,02011 9296	0,06945 3902	0,0195 7337	0,01808 7419	0,01701 4918	0,01642 8301
limpap eh	0,02011 9296	0,01831 7064	0,0742 1751	0,01808 7419	0,01701 4918	0,01642 8301
lobang	0,02011 9296	0,01831 7064	0,0195 7337	0,06858 3142	0,01701 4918	0,01642 8301
maara m	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,06229 2166
masuk	0,02011 9296	0,01831 7064	0,0195 7337	0,06858 3142	0,01701 4918	0,01642 8301
ngarai	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,06229 2166
taman	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,06229 2166
tentang	0,02011 9296	0,01831 7064	0,0742 1751	0,01808 7419	0,01701 4918	0,01642 8301
transpo rtasi	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,06229 2166
umum	0,02011 9296	0,01831 7064	0,0195 7337	0,01808 7419	0,01701 4918	0,06229 2166
wisata	0,04839 6264	0,04406 1059	0,0195 7337	0,04350 8655	0,01701 4918	0,03951 7705

Pada tahapan testing, data yang telah dilatih akan diujikan untuk klasifikasi.

Tabel 12. laplace smoothing

Jenis Pertanyaan	Isi Pertanyaan
?	Kendaraan umum apa yang biasa digunakan untuk ke daerah wisata bukitinggi?

Tahap *ekstraksi entitas* ini dilakukan ketika *Chatbot* akan mengidentifikasi frasa yang terkait dengan kelompok entitas (NER). Metode yang digunakan untuk melakukan identifikasi entitas (NER) terdiri dari penggunaan aturan yang telah ditentukan berdasarkan analisis.

Aturan untuk pengenalan entitas (NER) adalah sebagai berikut :

1. Jika token terdapat di dalam field *relative\_name* pada tabel destinasi yang ada di dalam database, maka pengenalan entitas sebagai [destinasi].
2. Jika token terdapat di dalam field *relative\_name* pada tabel transportasi yang ada di dalam database, maka pengenalan entitas sebagai [transportasi].

Kata “transportasi” dapat dikenali sebagai entitas diskon karena token “transportasi” terdapat dalam *field relative\_name* pada tabel diskon yang ada di dalam *database*, sehingga dikenali sebagai entitas [transportasi]. Setelah transportasi dikenali maka *rule* akan dikenali melalui basis aturan yang sudah tersimpan didalam database. Pada contoh ini C6R2 dikenali sebagai *rule* yang dijalankan ketika transportasi dipilih.

Pengujian akurasi jawaban untuk menentukan tingkat akurasi chatbot dalam memberikan jawaban berdasarkan entitas sebagai kata kunci. Terdapat tiga bagian dalam pemberian jawaban oleh chatbot yaitu: chatbot mampu memberikan jawaban yang sesuai dengan yang diharapkan berdasarkan pertanyaan yang diajukan (jawaban benar), *chatbot* memberikan jawaban yang tidak sesuai dengan yang diharapkan dalam menjawab pertanyaan (jawaban salah), dan jawaban yang tidak bisa dijawab oleh *chatbot* (jawaban ditolak). Setelah melakukan pengujian diperoleh hasil akurasi jawaban yang dapat dilihat pada tabel 13.

Tabel 13. Laplace Smoothing

Jawaban yang sebenarnya	Jawaban yang diberikan		
	Benar	Salah	Ditolak
Benar	35	2	1
Salah	0	0	0
Ditolak	0	0	0

Berdasarkan tabel diatas, pengujian akurasi jawaban dengan menggunakan 40 dataset memiliki tingkat akurasi yaitu dengan menggunakan rumus akurasi.

$$\text{Akurasi} = \frac{33}{38} \times 100\% = 92,1\%$$

## KESIMPULAN

*Natural Language Processing* dan Naïve Bayes Classifier dapat membantu *Chatbot* memahami arti dari pertanyaan berdasarkan bahasa alami manusia dengan nilai akurasi klasifikasi kategori pertanyaan sebesar 94.24%. Nilai akurasi ini dipengaruhi text preprocessing dan pembobotan kata. Pada penerapan *Natural Language Processing* pada sistem *Chatbot* ini dapat memudahkan pengguna dalam melakukan pertanyaan dengan cepat dan tepat, karena *Chatbot* ini akan berinteraksi langsung ke pengguna melalui tanya jawab.

## REFERENSI

- Dongbo, m., miniaoui, s., fen, l., althubiti, s. A., & alsenani, t. R. (2023). Intelligent chatbot interaction system capable for sentimental analysis using hybrid machine learning algorithms. *Information processing & management*, 60(5), 103440. <https://doi.org/https://doi.org/10.1016/j.ipm.2023.103440>
- Hutama, r. C., fauziah, f., & komalasari, r. T. (2021). Aplikasi chatbot berbasis teks menggunakan algoritma naive bayes classifier faq grabads. *String (satuan tulisan riset dan inovasi teknologi)*, 6(1), 90. <https://doi.org/10.30998/string.v6i1.9919>
- Khurana, d., koli, a., khatler, k., & singh, s. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications*, 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Maitri, a. L., & sutopo, j. (2019). Rancangan bangun chatbot sebagai pusat informasi lembaga kursus dan pelatihan menggunakan pendekatan natural language processing. *Eprints.uty.ac.id*, 1–9. <http://eprints.uty.ac.id/>
- Mukrodin, mukrodin, & mega sasmita, n. (2021). Rtificial inteligence dalam apilkasi chatbot sebagai helpdesk obyek wisata dengan permodelan natural language processing (studi kasus: kabupaten cilacap). *Smart comp :jurnalnya orang pintar komputer*, 10(1), 7–14. <https://doi.org/10.30591/smartcomp.v10i1.2135>
- Prasetyo, v. R., benarkah, n., & chrisintha, v. J. (2021). Implementasi natural language processing dalam pembuatan chatbot pada program information technology universitas surabaya. *Teknika*, 10(2), 114–121. <https://doi.org/10.34148/teknika.v10i2.370>
- Ratiasasadara, p. W., sudarno, s., & tarno, t. (2023). Analisis sentimen penerapan ppkm pada twitter menggunakan naïve bayes classifier dengan



- seleksi fitur chi-square. *Jurnal gaussian; vol 11, no 4 (2022): jurnal gaussiando - 10.14710/j.gauss.11.4.580-590*.  
<https://ejournal3.undip.ac.id/index.php/gaussian/article/view/35636>
- Thananukhun, k., jaiyen, s., jitkajornwanich, k., & hanskunatai, a. (2023). Question classification for thai conversational chatbots using artificial neural networks and multilingual bert models. *2023 15th international conference on knowledge and smart technology (kst)*, 1–4. <https://doi.org/10.1109/kst57286.2023.10086784>
- William, p., lanke, g. R., inukollu, v. N. R., singh, p., shrivastava, a., & ro. (2023). *Framework for design and implementation of chat support system using natural language processing*. Ieee xplore.
- Zhou, s., silvasstar, j., clark, c., salyers, a. J., chavez, c., & bull, s. S. (2023). An artificially intelligent, natural language processing chatbot designed to promote covid-19 vaccination: a proof-of-concept pilot study. *Digital health, 9*, 20552076231155680. <https://doi.org/10.1177/20552076231155679>