

ANALISIS SENTIMEN TEMPAT WISATA KABUPATEN GUNUNGKIDUL MENGGUNAKAN METODE NAIVE BAYES BERDASARKAN KRITIK SARAN WISATAWAN

Doni Arya Utama¹

¹Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹*doniautama@gmail.com

Abstrak- Minimnya informasi masih menjadi kendala dalam perencanaan pengelolaan kunjungan wisata ke wilayah Kabupaten Gunung Kidul dan pendataan untuk peningkatan pelayanan. Pendekatan analisis sentimen diusulkan dalam penelitian ini sebagai solusi untuk masalah ini karena media sosial memiliki peran penting dalam penyediaan informasi untuk menilai kunjungan wisatawan. Tujuan dari penelitian ini adalah untuk menemukan model sistem yang dapat memberikan informasi kepada wisatawan dan pengelola objek wisata untuk membantu mereka dalam pengambilan keputusan. Agar dapat diimplementasikan ke dalam sistem dan bermanfaat sebagai sistem pendukung bagi pihak-pihak yang berpartisipasi dalam upaya menaikkan harga jual potensi daerah, khususnya industri pariwisata, pendekatan *Naive Bayes* diterapkan pada penelitian ini. Situs web *TripAdvisor* dan *Google Maps* adalah sumber data untuk komentar pengunjung pada penelitian ini. Berdasarkan penelitian yang telah dilakukan, uji akurasi data split 60% - 40% mendapatkan akurasi 70%, kemudian pada split data 70% - 30% mendapatkan akurasi 80%, dan untuk split data 80% - 20% mendapatkan akurasi mencapai 79%.

Kata Kunci: Wisata, Penelitian, Sistem

SENTIMENT ANALYSIS OF TOURISM PLACES IN GUNUNGKIDUL DISTRICT USING NAIVE BAYES METHOD BASED ON CRITICISM OF TOURISM ADVICE

Abstract- The lack of information is still an obstacle in planning the management of tourist visits to the Gunung Kidul Regency area and data collection for service improvement. The sentiment analysis approach is proposed in this study as a solution to this problem because social media has an important role in providing information to assess tourist visits. The purpose of this research is to find a system model that can provide information to tourists and tourist attraction managers to assist them in making decisions. To be implemented into the system and useful as a support system for parties participating in efforts to increase the selling price of regional potential, especially the tourism industry, the *Naive Bayes* approach was applied to this study. The *TripAdvisor* and *Google Maps* websites are the data sources for visitor comments in this study. Based on the research that has been done, the accuracy test for split data is 60% - 40% to get 70% accuracy, then for split data 70% - 30% to get 80% accuracy, and for split data 80% - 20% to get accuracy reaching 79%.

Keywords: Tour, Research, System

1. PENDAHULUAN

Kemajuan internet dan teknologi memberikan dampak positif yang signifikan bagi industri pariwisata, dengan banyaknya situs *online* yang kini menyediakan layanan pemesanan hotel, restoran, penginapan, dan akomodasi lainnya serta tiket transportasi seperti kereta api dan pesawat terbang. Orang sekarang dapat menggunakan berbagai platform media sosial, seperti *Instagram* dan *Twitter*, untuk menemukan gambar untuk ulasan atau *review* tentang pariwisata dari pengunjung sebelumnya, bahkan jika mereka hanya mencari referensi untuk rencana perjalanan mereka. *TripAdvisor* adalah situs media sosial populer yang digunakan sebagai referensi bagi para pelancong, itu dapat membantu wisatawan domestik dan internasional dalam mengatur perencanaan perjalanan wisata mereka.

Persoalan yang dihadapi oleh Pemerintah Kabupaten Gunung Kidul di bidang pariwisata sebenarnya memiliki pekerjaan yang harus diselesaikan dengan baik, salah satunya adalah pelaksanaan strategi peningkatan objek industri pariwisata yang masih belum ideal. Dinas Pariwisata dan Kebudayaan Kabupaten Gunung Kidul merasa kesulitan untuk menilai kemajuan setiap kunjungan ke kawasan Gunung Kidul karena berbagai alasan, termasuk tingginya jumlah tur. Lokasi wisata harus diperbaiki berdasarkan umpan balik dan pendapat pengunjung. Dengan

adanya kritik dari para wisatawan tersebut, diyakini akan semakin mempermudah Pemerintah Kabupaten Gunung Kidul untuk mengetahui langkah apa yang harus diambil dalam menciptakan objek wisata yang ada.

Penelitian ini bertujuan untuk menganalisis sentimen terhadap ulasan wisatawan yang sudah berkunjung ke tempat wisata pada Kabupaten Gunung Kidul dengan menggunakan metode *Naive Bayes*, karena tingkat kesalahan yang lebih rendah yang diperoleh ketika dataset besar, bersama dengan akurasi dan kecepatan yang lebih tinggi. *Review* objek wisata Kabupaten Gunung Kidul dikumpulkan dari kolom *review* di *website TripAdvisor* dan *Google Maps*. Penelitian ini memberikan informasi yang akan berguna untuk penelitian selanjutnya, khususnya dalam pengembangan ilmu pemerintahan terkait peningkatan potensi pariwisata dalam pelaksanaan otonomi daerah. Informasi ini nantinya akan bermanfaat bagi Dinas Kebudayaan dan Pariwisata Kota Wonosari, pengelola tempat wisata di Kabupaten Gunung Kidul, dan pihak lain yang membutuhkan. melihat banyaknya opini positif dan negatif, yang mungkin bisa ditelaah untuk dianalisis lebih lanjut.

Rizaldi et al. (2021) menyatakan bahwa *Naive Bayes* adalah teknik yang diterapkan untuk menentukan kelas dari setiap masalah, selain itu pada algoritma ini menggunakan probabilitas yang tertinggi untuk dilakukan klasifikasi data yang telah dibagi berdasarkan tiap - tiap masalah.

Sunardi et al. (2018) menyebutkan bahwa *Naive Bayes Classifier* dapat digunakan untuk melakukan klasifikasi berdasarkan probabilitas dan teorema *Bayes*, dengan asumsi setiap variabel memiliki karakteristik independen X, atau karakteristik independen yang tidak terkait dengan variabel lain. Itu sebuah cara. *Taksonomi Naive Bayes* melibatkan dua fase klasifikasi teksnya: fase pelatihan dan fase klasifikasi. Selama fase pelatihan, analisis sampel dilakukan pada data seperti dokumen yang representatif.

Kelebihan dari metode *Naive Bayes Classifier* yaitu sederhana namun memiliki akurasi yang tinggi. Selain itu metode ini terbukti memiliki akurasi dan kecepatan yang tinggi ketika diaplikasikan ke dalam *database* meski dengan data yang besar (Sevsa dan Wahyudi, 2019).

Hendra dan Fitriyani (2021) menyebutkan bahwa *Naive Bayes Classifier* memiliki kekurangan yaitu probabilitas tidak dapat mengukur seberapa besar tingkat dari keakuratan prediksi. Selain itu NBC juga memiliki kelemahan pada proses seleksi atribut yang dapat mempengaruhi nilai dari akurasi.

Akurasi sebesar 65% ditemukan pada studi analisis sentimen sebelumnya, yang menggunakan *Naive Bayes* untuk menganalisis sentimen destinasi wisata Jakarta pasca covid-19[1], pada analisis sentimen destinasi wisata di Kota Tegal, teknik *Naive Bayes* mencapai nilai akurasi sebesar 77,50% [2], sedangkan akurasi uji *Naive Bayes* pada analisis sentimen tempat wisata di Bali sebesar 90,71% [3].

Penelitian ini dilatar belakangi oleh *research GAP* pada penelitian-penelitian terdahulu. Berdasarkan penelitian dilakukan Soemantri & Dairoh (2019). Fanissa et al. (2018). Khairunissa et al. (2021). Mereka memperoleh kesimpulan bahwa kinerja sistem terbaik dihasilkan ketika menggunakan metode *Naive Bayes*. Namun berbeda dengan penelitian Pati & Umar (2022), tingkat akurasi yang diperoleh dengan menggunakan metode K-NN lebih baik dibanding dengan metode *Naive Bayes Classifier*.

Perbedaan penelitian ini dengan yang sebelumnya yaitu penelitian ini pengoperasian program berbasis *website* sehingga dapat diakses melalui PC, laptop, Maupun *Smartphone* sehingga penggunaan nya lebih efisien.

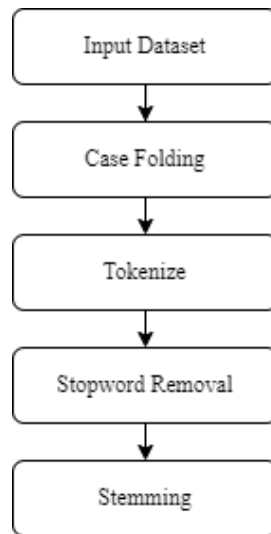
2. METODE PENELITIAN

2.1 Data Penelitian

Data yang digunakan untuk melakukan penelitian ini adalah 500 ulasan atau *review* yang didapatkan dari hasil *scraping* data wisata Kabupaten Gunung Kidul pada *TripAdvisor* dan *Google Maps*.

2.2 Persiapan Data

Data yang telah diberi label sentimen akan diproses terlebih dahulu, diikuti dengan data untuk kritik dan rekomendasi. Tindakan yang diambil selama *preprocessing* data dapat dilihat pada gambar 1.



Gambar 1. Preprocessing

- Case folding* meliputi proses mengubah semua teks kapital menjadi huruf kecil.
- Tokenizing* meliputi proses memecah urutan teks menjadi kelompok kata. Proses ini juga dapat menghapus pembatas, angka, tanda baca, dan karakter non-huruf.
- Stopword removal* untuk menghapus istilah yang sering digunakan yang kurang berhubungan dengan sentimen. Kata henti pada dasarnya tidak berarti, atau bisa juga berupa preposisi, konjungsi, dll. Pustaka dari *Natural Language Toolkit* (NLTK) digunakan untuk menghapus kata henti.
- Stemming* adalah proses menemukan akar kata. Peneliti mengembangkan *file stemming* menggunakan pendekatan ini, yang dapat digunakan untuk menghilangkan imbuhan kata. Pembentukan batang dilakukan dengan bantuan pustaka literatur.
- Term weighting* (TF-IDF) adalah proses pembobotan kata. Hal ini dilakukan dengan mencari nilai *Term Frequency* (TF), kemudian nilai *Document Frequency* (DF), kemudian nilai *Inverse Document Frequency* (IDF) dan menghitung bobotnya.

2.3 Pemodelan

Pada titik ini, bahan *preprocessing* yang sudah dimodelkan oleh peneliti. Selanjutnya data uji dan data latih dipisahkan dari dataset. 80%-20%, 60%-40%, dan 70%-30% dari jumlah total data yang dianalisis kemudian dinilai menggunakan split data. Metode *Naive Bayes* digunakan oleh peneliti.

2.4 Evaluasi

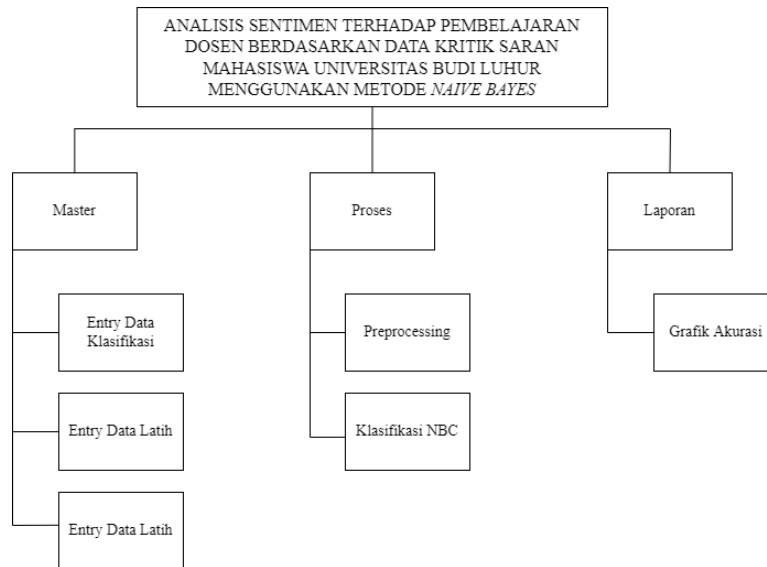
Pada titik ini, peneliti menilai keefektifan metode klasifikasi dengan memeriksa seberapa besar akurasi, presisi, dan daya ingat diperoleh saat menggunakan data umum dibandingkan dengan algoritma *Naive Bayes*.

2.5 Rancangan Pengujian

Program *Naive Bayes Classifier* dilakukan pengujian untuk menilai ketepatan penghitungan sentimen positif dan negatif dari data. pengujian dijalankan menggunakan data sampel yang dibagi menjadi 40% data uji dan 60% data latih, 30% data uji dan 70% data latih, 20% data uji dan 80% data latih.

2.6 Rancangan Menu

Berikut merupakan rancangan menu dari *Naive Bayes Classifier* dapat dilihat pada gambar 2.



Gambar 2. Rancangan menu

3. HASIL DAN PEMBAHASAN

3.1 Data Collection

Pada tahapan ini peneliti memiliki data sebanyak 500 yang akan digunakan. Peneliti menggunakan data kritik saran wisatawan Kabupaten Gunung Kidul pada rentang waktu 2022/2023 dapat dilihat pada tabel 1.

Tabel 1. Data Collection

Topik	Tanggal	Jumlah
Kritik saran wisatawan 2022-2023	26 Juni 2023	500 Data
Total		500 Data

Berikut ini merupakan sampel dari data kritik saran mahasiswa Universitas Budi Luhur pada tabel 2.

Tabel 2. Sampel data komentar

ID	Text
1	Pengalaman naik Jeep di pinggir pantai, seru apalagi konvoi, jgn lupa foto2 di pantai parangtritis
2	Banyak spot menarik sebenarnya. Cuma gak ditata aja.
3	Berkesan banget, gak kan bosan
4	Bentuk gumuk pasirnya sudah tidak terjaga
5	Pantai selatan jogja yg cukup terkenal di masanya, pantai pertamakali yg saya kunjungi waktu baru di jogja, penuh kenangan, cuman kebersihannya perlu diperhatikan
82	Mantab seru jos pokoknya
94	Sekarang ramai, banyak jeep, sandboard, warung
111	Menyenangkan pengalaman naik jeep di parangtritis
114	Pemandangan yang indah

3.2 Pemberian label sentimen

Data yang telah melalui tahap reduksi (*cleaning data*) harus diberi label untuk memastikan sentimen dari masing-masing data sebelum memulai tahap pembersihan data. Setiap potongan data memiliki label sentimen positif dan negatif. Di bawah ini adalah contoh data sentimen positif dan negatif dapat dilihat pada tabel 3 dan 4.

Tabel 3. Contoh Sentimen Positif

ID	Text	Sentimen
131	Indah dan nyaman serta bersih	Positif

Tabel 4. Contoh Sentimen Negatif

ID	Text	Sentimen
312	Seperti biasa selalu ramai, disayangkan banyak sampah popok bayi di pinggir pantai	Negatif

3.3 Tahapan Preprocessing

Preprocessing, yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, merupakan prosedur untuk membersihkan data. Peneliti menggunakan kamus *stopwords* untuk metode *stopword removal* dan leksikon bahasa Indonesia yang cukup besar untuk bagian *stemming*. Hasil dari tahap *preprocessing* dapat di lihat pada tabel 5.

Tabel 5. Tahapan Preprocessing

Preprocessing	Sebelum	Sesudah
<i>Case folding</i>	Cave tubing dengan track yang menantang namun sayangnya masih kurang populer padahal sangat mengasyikkan	cave tubing dengan track yang menantang namun sayangnya masih kurang populer padahal sangat mengasyikkan
<i>Tokenize</i>	cave tubing dengan track yang menantang namun sayangnya masih kurang populer padahal sangat mengasyikkan	"cave", "tubing", "dengan", "track", "yang", "menantang", "namun", "sayangnya", "masih", "kurang", "populer", "padahal", "sangat", "mengasyikkan"
<i>Stopword removal</i>	"cave", "tubing", "dengan", "track", "yang", "menantang", "namun", "sayangnya", "masih", "kurang", "populer", "padahal", "sangat", "mengasyikkan"	"cave", "tubing", "track", "menantang", "namun", "sayangnya", "masih", "kurang", "populer", "padahal", "mengasyikkan"
<i>Stemming</i>	"cave", "tubing", "track", "menantang", "namun", "sayangnya", "masih", "kurang", "populer", "padahal", "mengasyikkan"	"cave", "tubing", "track", "menantang", "namun", "sayang", "masih", "kurang", "populer", "padahal", "asyik"

3.4 TF-IDF

Ini adalah salah satu teknik pembobotan, yang disebut TF-IDF, dan perhitungannya terlihat pada tabel 6.

Tabel 6. Tabel Dokumen TF-IDF

ID Dokumen	Sebelum Preprocessing	Isi Dokumen Sesudah Preprocessing
d1	Cocok untu menikmati hangatnya mentari pagi	['cocok','menikmati','hangat', 'mentari','pagi']
d2	Wisata jeep nya recomended buat dicoba	['wisata','jeep','recomended', 'buat', 'coba']
d3	Serunya gak ada duanya	['seru', 'gak', 'ada', 'dua']

Setelah itu, dihitung kemunculan *term* atau kata pada dokumen dapat dilihat pada tabel 7.

Tabel 7. Hitungan Kemunculan Term

TERM	TF		
	d1	d2	d3
cocok	1	0	0
menikmati	1	0	0
hangat	1	0	0
mentari	1	0	0
pagi	1	0	0
wisata	0	1	0

jeep	0	1	0
recommended	0	1	0
buat	0	1	0
coba	0	1	0
seru	0	0	1
gak	0	0	1
ada	0	0	1
dua	0	0	1

Setelah itu dilakukan TF Normalisasi untuk 3 dokumen dengan cara membagi nilai TF dengan panjang dokumen, sebagaimana contoh nilai d1 pada kata ‘sudah’ adalah 1 dibagi panjang dokumen yaitu 5, didapatkan hasil sebesar 0.2 dapat dilihat pada tabel 8.

Tabel 8. TF Normalisasi

TF NORMALISASI		
d1	d2	d3
0.2	0	0
0.2	0	0
0.2	0	0
0.2	0	0
0.2	0	0
0	0.2	0
0	0.2	0
0	0.2	0
0	0.2	0
0	0.2	0
0	0	0.25
0	0	0.25
0	0	0.25
0	0	0.2

Selanjutnya menghitung DF dengan menghitung jumlah dokumen yang mengandung kata tersebut. Misalnya, frase ‘cocok’ hanya muncul di dokumen 1, bukan dokumen 2 atau 3. Akibatnya, dapat dikatakan bahwa DF adalah 1. LOG (total dokumen dibagi DF), misalnya LOG (total dokumen 3 dibagi dengan DF), kemudian dihitung sebagai bagian dari prosedur perhitungan IDF dapat dilihat pada tabel 9.

Tabel 9. Perhitungan IDF

DF	IDF
1	$\text{LOG}(3/1) = 0.477$
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477
1	0.477

Selanjutnya proses ke perhitungan TF. Untuk setiap frase di setiap dokumen, d1, d2, dan d3, nilai TF dan nilai IDF dikalikan dalam prosedur perhitungan IDF dilihat pada tabel 10.

Tabel 10. Perhitungan TF.IDF

TF IDF		
d1	d2	d3
0.095	0	0

0.095	0	0
0.095	0	0
0.095	0	0
0.095	0	0
0	0.095	0
0	0.095	0
0	0.095	0
0	0.095	0
0	0.095	0
0	0	0.119
0	0	0.119
0	0	0.119
0	0	0.119

3.5 Naive Bayes

Setiap masalah dapat dikategorikan menggunakan teori *Bayes* dan probabilitas menggunakan pendekatan *Naive Bayes*. Berikut merupakan contoh perhitungan *Naive Bayes* dapat dilihat pada tabel 11.

Tabel 11. Data Sample Naive Bayes

ID	Text	Sentimen
227	pantai masih bersih ombak besar pantai selatan	Positif
307	pantai bersih dagang ramah wisata pantai recomended banget	Positif
384	pantai luas bersih main anak luas cukup datar pantai	Positif
385	ada kayu2 sama sampah plastik	Positif
399	agak kotor karena ujung buang sampah sembarangan harga makan variasi	Negatif

Karena terdapat 3 data dokumen positif dari 5 data maka probabilitas dokumen positif adalah $P(\text{positif}) = \frac{3}{5} = 0.6$ dan probabilitas untuk dokumen negatif adalah $P(\text{negatif}) = \frac{2}{5} = 0.4$

Kemudian dapat dilakukan perhitungan probabilitas untuk menentukan kelas negatif atau positif sebagai berikut :

$$\begin{aligned}
 P(\text{bersih}|\text{positif}) &= \frac{\frac{3}{24} + \frac{1}{46}}{\frac{3}{24} + \frac{1}{46}} = \frac{27}{184} = 0.1467 \\
 P(\text{ramah}|\text{positif}) &= \frac{\frac{1}{24} + \frac{1}{46}}{\frac{1}{24} + \frac{1}{46}} = \frac{35}{552} = 0.0634 \\
 P(\text{recomended}|\text{positif}) &= \frac{\frac{1}{24} + \frac{1}{46}}{\frac{1}{24} + \frac{1}{46}} = \frac{35}{552} = 0.0634 \\
 P(\text{pantai}|\text{positif}) &= \frac{\frac{6}{24} + \frac{1}{46}}{\frac{6}{24} + \frac{1}{46}} = \frac{25}{92} = 0.2717 \\
 P(\text{sampah}|\text{negatif}) &= \frac{\frac{2}{22} + \frac{1}{46}}{\frac{2}{22} + \frac{1}{46}} = \frac{57}{506} = 0.1126 \\
 P(\text{kotor}|\text{negatif}) &= \frac{\frac{1}{22} + \frac{1}{46}}{\frac{1}{22} + \frac{1}{46}} = \frac{17}{253} = 0.0671
 \end{aligned}$$

Maka nilai probabilitas setiap kata pada kelas dapat dilihat pada tabel 12.

Tabel 12. Probabilitas Kata

Word	nk	N	Vocabulary	P
Pantai	6	24	46	0.2717
masih	1	24	46	0.0634
bersih	3	24	46	0.1467
ombak	1	24	46	0.0634
besar	1	24	46	0.0634

selatan	1	24	46	0.0634
dagang	1	24	46	0.0634
ramah	1	24	46	0.0634
wisata	1	24	46	0.0634
recomended	1	24	46	0.0634
banget	1	24	46	0.0634
luas	1	24	46	0.0634
main	1	24	46	0.0634
anak	1	24	46	0.0634
cukup	1	24	46	0.0634
datar	1	24	46	0.0634
ada	1	22	46	0.0671
kayu	1	22	46	0.0671
sama	1	22	46	0.0671
sampah	2	22	46	0.1126
plastik	1	22	46	0.0671
ke bawa	1	22	46	0.0671
agak	1	22	46	0.0671
kotor	1	22	46	0.0671
karena	1	22	46	0.0671
ujung	1	22	46	0.0671
buang	1	22	46	0.0671
sembarangan	1	22	46	0.0671
harga	1	22	46	0.0671
makan	1	22	46	0.0671
variasi	1	22	46	0.0671

Metode pemeriksaan dokumen pada setiap kelas ditunjukkan pada tabel di atas, dan dapat dilanjutkan ke proses klasifikasi untuk membentuk kelas pada dokumen baru. Dengan menerapkan rumus ini, nilai tertinggi dari perhitungan *Naive Bayes* akan digunakan sebagai kelas positif, dan nilai terendah akan digunakan sebagai kelas negatif. Sebagai contoh, menentukan kelas data baru dapat dilihat pada tabel 13.

Tabel 13. Menentukan Sentimen

Data	Sentimen
salah satu spot sunset terbaik di jogja	???

Berdasarkan data tersebut, dapat menghitung nilai *naive bayes* untuk mendapatkan kelas positif dan negatif.

Positif :

$$p(\text{Positif}) \cdot p(\text{salah}|\text{positif}) \cdot p(\text{satu}|\text{positif}) \cdot p(\text{spot}|\text{positif}) \cdot p(\text{sunset}|\text{positif}) \cdot p(\text{terbaik}|\text{positif}) \cdot p(\text{di}|\text{positif}) \cdot p(\text{jogja}|\text{positif}) = 0.6 \times 0.0634 \times 0.0634 \times 0.0634 \times 0.0634 \times 0.0634 \times 0.0634 \times 0.0634 = 2.470 \times 10^{-2}$$

Negatif :

$$p(\text{Negatif}) \cdot p(\text{salah}|\text{Negatif}) \cdot p(\text{satu}|\text{Negatif}) \cdot p(\text{spot}|\text{Negatif}) \cdot p(\text{sunset}|\text{Negatif}) \cdot p(\text{terbaik}|\text{Negatif}) \cdot p(\text{di}|\text{Negatif}) \cdot p(\text{jogja}|\text{Negatif}) = 0.4 \times 0.0671 \times 0.0671 \times 0.0671 \times 0.0671 \times 0.0671 \times 0.0671 \times 0.0671 = 2.450 \times 10^{-2}$$

Dapat disimpulkan bahwa nilai penjumlahan memiliki nilai positif lebih besar yaitu 2.470×10^{-2} sedangkan nilai negatif lebih kecil yaitu 2.450×10^{-2} sehingga dapat disimpulkan kalimat tersebut mendapat sentimen positif dilihat pada tabel 14.

Tabel 14. Hasil prediksi *Naive Bayes*

Data	Sentimen
salah satu spot sunset terbaik di jogja	Positif

3.6 Hasil Pengujian Program

Dengan *dataset* kritik saran wisatawan pengunjung Kabupaten Gunung Kidul sebanyak 500 baris data dilakukan pengujian dan didapatkan hasil dapat dilihat pada tabel 15.

Tabel 15. Hasil Pengujian Program

Data Latih	Data Uji	Split Data	Akurasi	Presisi	Recall
400	100	80 – 20	79%	79%	100%
350	150	70 – 30	80%	80%	100%
300	200	60 – 40	79%	79%	100%

Berdasarkan analisis yang dilakukan didapatkan akurasi pada pengujian split data 60% – 40% mendapatkan hasil akurasi sebesar 79%, kemudian pada split data 70% – 30% mendapatkan hasil akurasi sebesar 80%, lalu untuk hasil split data 80% – 20% mendapatkan hasil akurasi sebesar 79%.

3.7 Hasil Analisis Sentimen

Berdasarkan pengujian yang dilakukan, terdapat 350 sentimen positif dan 150 sentimen negatif. Berdasarkan hasil sentimen disimpulkan bahwa kinerja pengelola pariwisata Kabupaten Gunung Kidul sudah baik, namun tidak menutup kemungkinan peningkatan kualitas pelayanan dan fasilitas kebersihan di kawasan wisata.

4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa metode *Naive Bayes* berhasil atau mampu dalam menganalisis sentimen terhadap tempat wisata di Kabupaten Gunung Kidul. Untuk melakukan analisis sentimen tersebut, dibagi menjadi dua kategori sentimen positif dan negatif. Data yang digunakan untuk melakukan penelitian ini adalah 500 ulasan atau *review* yang didapatkan dari hasil *scraping* data wisata Kabupaten Gunung Kidul pada *TripAdvisor* dan *Google Maps*. Tahap dalam analisis sentimen penelitian ini adalah *preprocessing* (*case folding, tokenize, stopword removal, stemming*). Berdasarkan pengujian metode algoritma *Naive Bayes* dengan split data 60 – 40 menghasilkan nilai akurasi 79%, Pada split data 70 – 30 menghasilkan nilai akurasi 80%, kemudian pada split data 80 – 20 mendapatkan nilai akurasi 79%.

DAFTAR PUSTAKA

- [1] N. A. Rahma, Garno, and N. Sulistiyowati, "Analisis Sentimen Tempat Wisata Di Jakarta Pasca Covid-19 Dengan Algoritma Naive Bayes Universitas Pahlawan Tuanku Tambusai," *J. Pendidik. dan Konseling*, vol. 4, no. 6, pp. 5894–5908, 2022.
- [2] O. Somantri and D. Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 2, p. 191, 2019, doi: 10.26418/jp.v5i2.32661.
- [3] A. Sentimen et al., "Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 418–427, 2022.
- [4] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier," *Pros. SNST*, vol. 7, pp. 37–42, 2016, [Online]. Available: https://www.publikasiilmiah.unwahas.ac.id/index.php/PROSIDING_SNST_FT/article/view/1506/1589
- [5] A. Fahreza, "Analisis Sentimen Opini Publik Terhadap Pariwisata Di Masa Pandemi Covid-19 Menggunakan Metode Naive Bayes," vol. 17, no. 1, pp. 49–54, 2022.
- [6] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [7] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.
- [8] S. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] G. K. Pati and E. Umar, "Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K-Nearest Neighbor," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2309, 2022, doi: 10.30865/mib.v6i4.4635.
- [10] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 205–215, 2022, doi: 10.23887/janapati.v11i3.49450.