

Analisis Sentimen Taman Wisata Karang Resik Dengan Metode Naïve Bayes

Salma Natasya Azzahra¹, Ruuhwan², Missi Hikmatyar³

^{1,2,3} Universitas Perjuangan; Jl. Peta No.177, (0265) 326058

^{1,2,3} Jurusan Teknik Informatika, FT UNPER, Tasikmalaya

e-mail: ¹1903010053@unper.ac.id, ²ruuhwan@unper.ac.id, ³missi@unper.ac.id

ABSTRAK

Taman Wisata Karang Resik merupakan salah satu tempat wisata yang langsung dikelola oleh Pemkot Tasikmalaya yang berada di perbatasan antara Kota Tasikmalaya dan Kabupaten Ciamis. Pada perkembangan teknologi sekarang ini, pengunjung Taman Wisata Karang Resik dapat memberikan sebuah komentar terhadap kondisi dan fasilitas yang tersedia di Taman Wisata karang Resik melalui Google Maps. Analisis sentimen dilakukan bertujuan untuk mengetahui sentimen pengunjung terhadap Taman Wisata Karang Resik. Metode yang akan digunakan pada penelitian ini adalah Naïve Bayes dengan pembobotan Term Frequency Inverse Document Frequency (TF-IDF). Pada penelitian ini akan dibagi menjadi 7 tahap, diantaranya pengambilan data, data exploration, visualisasi data, preprocessing data, penerapan naïve bayes, evaluation, dan prediction. Data akan dikelompokkan menjadi 2 kelas yaitu positif dan negatif dengan data yang diambil sebanyak 1500 data. Hasil akurasi dari pengujian yang dilakukan dengan perbandingan data training, dan data testing 70:30 adalah 90% dan melakukan prediksi ulasan baru sebanyak 8 kali dan hasil prediksi akurat.

Kata Kunci : Analisis Sentimen, Naïve Bayes, Taman Wisata Karang Resik, TF-IDF

ABSTRACT

Karang Resik Tourism Park is one of the tourist attractions directly managed by the Tasikmalaya city government which is located on the border between Tasikmalaya City and Ciamis Regency. In today's technological development, visitors to Karang Resik Tourist Park can provide a comment on the conditions and facilities available at Karang Resik Tourist Park through Google Maps. Sentiment analysis is carried out to determine the sentiment of visitors to Karang Resik Tourism Park. The method that will be used in this research is Naïve Bayes with Term Frequency Inverse Document Frequency (TF-IDF) weighting. This research will be divided into 7 stages, including data collection, data exploration, data visualization, data preprocessing, application of naïve bayes, evaluation, and prediction. The data will be grouped into 2 classes, namely positive and negative with 1500 data taken. The results of the tests carried out with the ratio of training data, and 70:30 testing data are 90% and predict new reviews 8 times and the prediction results are accurate.

Keywords: Karang Resik Tourist Park, Naïve Bayes, Sentiment Analysis, TF-IDF

1. PENDAHULUAN

Taman wisata karang resik salah satu taman kota yang dikelola oleh Pemkot Tasikmalaya yang berlokasi di Jl. Mochammad Hatta, kota Tasikmalaya. Taman rekreasi ini sangat cocok untuk bersantai atau *healing* bersama keluarga karena di taman wisata tersebut terdapat banyak tempat yang cukup untuk memanjakan mata. Taman ini memiliki beberapa konsep yang berbeda diantaranya tempat untuk edukasi, tempat bermain, tempat untuk swafoto serta ada tempat untuk wisata kuliner. Ada keunikan di taman wisata ini yaitu adanya tempat yang bernuansa berbagai negara seperti India, Korea, Jepang, Yunani dan Belanda.[1]

Pada era teknologi informasi saat ini banyak manfaat yang diperoleh, salah satunya pengunjung taman wisata dapat memberikan komentar terhadap kondisi dan fasilitas yang tersedia di Taman Wisata Karang Resik. Pengunjung dapat memberikan komentar sekaligus rating di *website* yang sudah disediakan oleh Google dengan cara pengunjung hanya perlu memasukkan *keyword* 'Taman Wisata Karang Resik Tasikmalaya' pada halaman Google Maps.

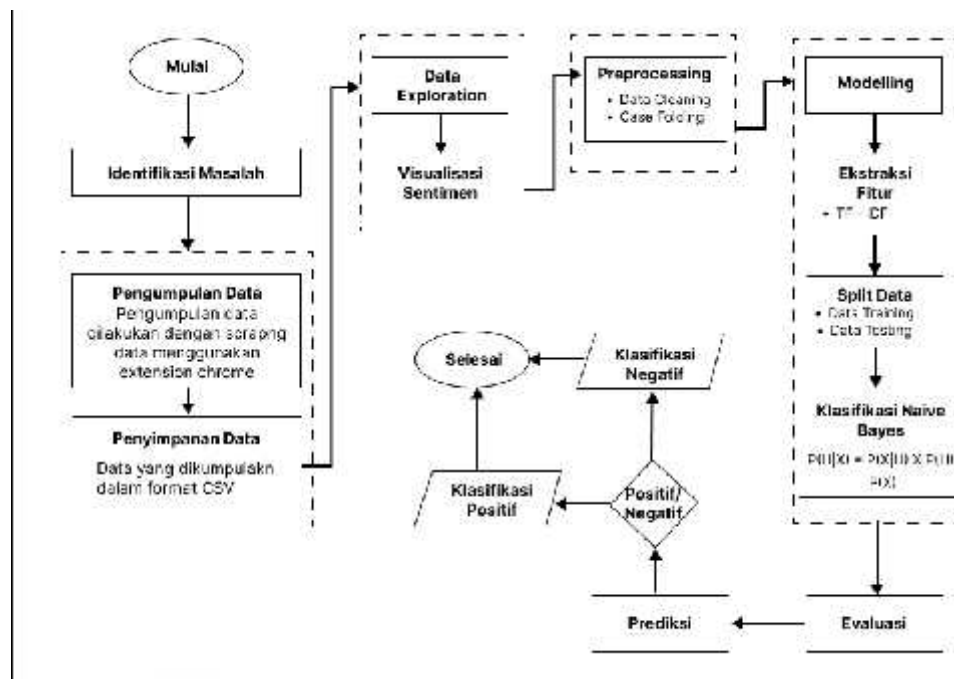
Masalah yang dihadapi adalah terkadang rating yang diberikan tidak sinkron dengan komentar yang diberikan. Hal itu menimbulkan kebingungan karena adanya keambiguan. Sedangkan *feedback* tersebut sangat penting untuk menjadi bahan evaluasi Pemkot untuk menjadi lebih baik di waktu yang akan datang. Oleh karena itu dilakukan analisis sentimen untuk mengetahui sentimen pengunjung Taman Wisata Karang Resik. Metode yang digunakan adalah Naïve Bayes dan TF-IDF sebagai fitur ekstraksinya.

Penelitian yang telah dilakukan sebelumnya mengenai analisis sentimen diantaranya, penggunaan naïve bayes dalam penilaian tempat wisata di kota Tegal[2], penelitian tersebut menyatakan bahwa metode Naïve Bayes menghasilkan nilai akurasi lebih besar dibandingkan metode Decision Tree yaitu sebesar 77,50 %. Penelitian lain menggunakan Naïve Bayes untuk klasifikasi komentar terhadap pembelajaran Jarak Jauh [3] menghasilkan nilai akurasi 68%. Penelitian menggunakan Naïve Bayes Multinomial pada berita hoax di twitter [4] menghasilkan akurasi sebesar 72%. Penggunaan Naive Bayes terhadap opini film pada twitter[5] menghasilkan akurasi sebesar 90% dengan precision 92%, recall 90%, dan f-measure 90%..

2. METODE PENELITIAN

2.1 Metodologi Penelitian

Metode penelitian yang digunakan pada tahap ini dimulai dari identifikasi masalah kemudian dilanjutkan dengan pembuatan dataset. Setelah dataset sudah dibuat selanjutnya mulai mengolah dataset tersebut, dimulai dari *preprocessing*, pelatihan model sehingga menghasilkan nilai akurasi dari pelatihan model tersebut kemudian pada proses terakhir yaitu menguji model tersebut dengan sebuah kalimat dan model tersebut akan mengklasifikasi kalimat tersebut kedalam klasifikasi positif atau negatif seperti pada Gambar 1.



Gambar 1 Metodologi Penelitian

2.2 Identifikasi Masalah

Identifikasi masalah pada penelitian ini diantaranya:

- Pengunjung taman wisata Karang Resik ketika memberi ulasan pada google maps terkadang apa yang diutrakan tidak sesuai dengan rating yang diberikan.
- Pengelola taman wisata Karang Resik yang tidak terlalu aktif membaca ulasan pengunjung pada google maps.

2.3 Pengumpulan Data

Pada tahap ini peneliti mengumpulkan data dari *review* pengunjung Taman Wisata Karang Resik di *website* Google Maps dengan cara *scraping* untuk mempercepat proses pengumpulan data. Tools yang digunakan pada *scraping* yaitu *extension* yang disediakan oleh Google Chrome yaitu *Data Scraper*.

2.4 Data Exploration

Tahapan eksplorasi data yang bertujuan untuk lebih memahami karakteristik dari data. Agar dapat memproses data yang telah dibuat, harus membuat query terlebih dahulu.

2.5 Data Preprocessing

Preprocessing adalah proses untuk membersihkan kata-kata yang tidak bermakna, penggandaan kata, dan *missing value*. Tahapan *preprocessing* pada penelitian ini diantaranya *cleaning data*, dan *case folding*

2.6 Modelling

Pada tahap penerapan model ini, ada beberapa proses yang harus dilakukan diantaranya split data, menghitung kemunculan atribut dengan TF-IDF, melakukan klasifikasi menggunakan Naïve Bayes.

2.7 Evaluasi

Tahap ini adalah evaluasi model yang bertujuan untuk mengetahui akurasi yang diperoleh dari model yang digunakan yaitu *Naïve Bayes*. Evaluasi model dilakukan dengan cara melihat tabel *confusion matrix* yang digunakan sebagai acuan tahap evaluasi.

2.8 Prediksi

Tahap ini merupakan tahap terakhir dari penelitian ini yang bertujuan untuk menguji apakah model yang diterapkan pada analisis sentimen ini bekerja dengan baik atau ada *error*.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan *extensions* dari Google Chrome yaitu *Data Scaper*. *Extensions* ini merupakan salah satu *extensions* Google Chrome yang berfungsi sebagai *web scraper* untuk mengambil data pada *website* secara *online*. Data yang diambil sebanyak 1500 ulasan pada Google Maps dengan *keyword* “Taman Wisata Karang Resik”. Contoh ulasan yang sudah diperoleh dan diberi label dapat dilihat pada tabel 1.

Tabel 1. Pengumpulan Data

No	Label	Reviews
1	Positif	Tempat wisata yg cocok buat santai bersama keluarga
2	Positif	Tempat luas dan nyaman...cocok untuk liburan keluarga..
3	Negatif	Wisata nya bagus cuma untuk tiket saya rasa terlalu mahal
4	Positif	Tempat yang nyaman, bnyak spot foto, harga yang terjangkau, pokonya cocok deh dengan saku kita... hihi
5	Negatif	Minim petugas patroli, tolong perbanyak lagi tempat sampah

3.2. Data Exploration

Tahap ekplorasi ini adalah tahap awal untuk mengetahui suatu pola/*pattern* pada data. Untuk mengetahui pola pada penelitian ini dilakukan menggunakan teknik visualisasi data. Pada gambar 2 terlihat bahwa dari data yang terkumpul, ulasan positif lebih banyak dibanding ulasan negatif seperti yang diperlihatkan oleh Gambar 2.



Gambar 2. Visualisasi Data

3.3. Data Preprocessing

Tahap ini merupakan proses manipulasi data mentah menjadi data yang dapat dipahami serta meningkatkan kinerja data. Pada penelitian ini ada 2 tahap yang dilakukan diantaranya:

1. *Cleaning Data*

Sebelumnya dataset berisi simbol khusus dan ruang kosong maka diperlukan tahap ini untuk membersihkan simbol tersebut, seperti penghapusan karakter khusus, dan angka seperti yang terlihat di Tabel 2.

Tabel 2. Cleaning Data

Reviews	Cleaning
Keren...Areanya luas, segala ada..Dan kalo mau keliling semua lumayan cape banget. ...	Keren Areanya luas segala ada Dan kalo mau keliling semua lumayan cape banget

2. *Case Folding*

Tahap ini yaitu mengubah kalimat menjadi *lowercase*. Sehingga huruf sebelumnya 'A' - 'Z' diubah menjadi 'a' - 'z' sehingga data menjadi lebih efisien dan efektif. Bisa dilihat di Tabel 3.

Tabel 3. Case Folding

Reviews	Case Folding
Keren...Areanya luas, segala ada..Dan kalo mau keliling semua lumayan cape banget. ...	keren areanya luas segala ada dan kalo mau keliling semua lumayan cape banget

3.4. Modelling

Tahap ini adalah penerapan metode yang akan digunakan. Pada penelitian ini menggunakan metode Naïve Bayes. Sebelumnya harus melakukan pembobotan kata dan pembagian data *training* dan data *testing*.

3.4.1. Term Frequency – Inverse Document Frequency

Untuk melakukan pembobotan TF-IDF, maka diperlukan 3 tahap yang dilakukan, pertama menghitung TF, kedua menghitung IDF, dan ketiga hasil dari TF-IDF. Berikut contoh kalimat pada Tabel 4, Tabel 5 hasil TF, Tabel 6 hasil IDF, dan Tabel 7 Hasil TF-IDF.

Tabel 4. Contoh Kalimat

Reviews
tempat wisata cocok buat santai bersama keluarga
tempat luas dan cocok untuk liburan keluarga

Menghitung TF menggunakan persamaan:

$$TF_{td} = tf_{td} \quad (1)$$

Keterangan:

d = dokumen

t = term

TF_{td} : Nilai Term Frequency pada term (t) dalam dokumen (d)

tf_{td} : Frekuensi kemunculan term (t) dalam dokumen (d)

Tabel 5. TF

R1	TF	R2	TF
tempat	1	tempat	1
wisata	1	luas	1
cocok	1	dan	1
buat	1	cocok	1
santai	1	untuk	1
bersama	1	liburan	1
keluarga	1	keluarga	1

Tahap selanjutnya menghitung idf dengan persamaan:

$$idf = \log \left(\frac{N}{df_t} \right) \quad (2)$$

Keterangan:

N = total dokumen keseluruhan

df_t = jumlah kata pada *term* (t) yang ada pada seluruh dokumen (N)

idf = hasil dari *Inverse Document Frequency* (IDF)

Tabel 6. IDF

R1	IDF	R2	IDF
tempat	$\log\left(\frac{2}{2}\right) = 0$	tempat	$\log\left(\frac{2}{2}\right) = 0$
wisata	$\log\left(\frac{2}{1}\right) = 1$	luas	$\log\left(\frac{2}{1}\right) = 1$
cocok	$\log\left(\frac{2}{2}\right) = 0$	dan	$\log\left(\frac{2}{1}\right) = 1$
buat	$\log\left(\frac{2}{1}\right) = 1$	cocok	$\log\left(\frac{2}{2}\right) = 0$
santai	$\log\left(\frac{2}{1}\right) = 1$	untuk	$\log\left(\frac{2}{1}\right) = 1$
bersama	$\log\left(\frac{2}{1}\right) = 1$	liburan	$\log\left(\frac{2}{1}\right) = 1$
keluarga	$\log\left(\frac{2}{2}\right) = 0$	keluarga	$\log\left(\frac{2}{2}\right) = 0$

Tahap terakhir menghitung TF-IDF dengan persamaan:

$$TF - IDF = (TF) \times (IDF + 1) \quad (3)$$

Tabel 7. TF - IDF

R1	TF-IDF 1	TF-IDF 2
tempat	$1 \times (0 + 1) = 1$	$1 \times (0 + 1) = 1$
wisata	$1 \times (1 + 1) = 2$	$0 \times 2 = 0$
cocok	$1 \times (0 + 1) = 1$	$1 \times (0 + 1) = 1$
buat	$1 \times (1 + 1) = 2$	$0 \times 2 = 0$
santai	$1 \times (1 + 1) = 2$	$0 \times 2 = 0$
bersama	$1 \times (1 + 1) = 2$	$0 \times 2 = 0$
keluarga	$1 \times (0 + 1) = 1$	$1 \times (0 + 1) = 1$
luas	$0 \times 2 = 0$	$1 \times (1 + 1) = 2$
dan	$0 \times 2 = 0$	$1 \times (1 + 1) = 2$
untuk	$0 \times 2 = 0$	$1 \times (1 + 1) = 2$
liburan	$0 \times 2 = 0$	$1 \times (1 + 1) = 2$

3.4.2. Split Data

Split data dilakukan untuk membagi data *training* dan data *testing* sebelum menggunakan metode yang akan diterapkan. Pada penelitian ini dilakukan split data dengan perbandingan data *training* sebanyak 80% dan data *testing* sebanyak 20%.

3.4.3. Klasifikasi Naïve Bayes

Setelah data diubah menjadi numerik dengan TF-IDF, maka tahap ini adalah penerapan naïve bayes. Pada tahap ini ada 3 langkah yang harus dilalui dimulai mencari peluang hipotesis $P(H_i)$, peluang banyaknya hipotesis dalam kejadian $P(H|X)$, dan mengalikan hasil perhitungan $P(H_i)$ dan $P(H|X)$ seperti yang diperlihatkan oleh Tabel 8.

Tabel 8. Klasifikasi Naïve Bayes

ulasan	Fitur				
	tempat	wisata	...	luas	Label
ulasan 1	1	2	...	0	Positif
ulasan 2	1	0	...	1	Positif

Ulasan terbaru yang akan diinputkan berupa kalimat “tempat bagus tapi harga tiket mahal”. Untuk memprediksi ulasan tersebut maka dilakukan langkah pertama menggunakan persamaan:

$$P(H_i) = \frac{N_i}{N} \quad (1)$$

$$P(H_{positif}) = \frac{2}{2} = 1$$

Keterangan:

N_i : jumlah data pada data *testing*

N : jumlah semua data *testing*

Peluang hipotesis positif sudah ditemukan yaitu bernilai 1 apabila hasil akhir < 1 maka hasil tersebut termasuk ke dalam kelas negatif. Langkah selanjutnya mencari peluang hipotesis pada kejadian $P(X|H)$ dengan persamaan:

$$P(X|H_i) = \prod_{k=1}^n P(x_k|H_{positif}) \quad (2)$$

$$= \prod_{k=1}^n P(x_k|H_{positif})$$

Keterangan:

$P(x_k|H_i)$: peluang munculnya term x_k dalam kelas H_i

Contoh perhitungan menggunakan data *testing* “tempat bagus tapi harga tiket mahal” dengan kelas positif. untuk mencari nilai $P(X|H_i)$ menggunakan persamaan:

$$\prod_{k=1}^n P(x_k|H_{positif}) = \frac{T_{kpositif} + 1}{T_{positif} + \beta} \quad (3)$$

Keterangan:

β : total kata pada data *testing*

T_{ki} : jumlah term x_k dalam kelas H_i pada data *testing*

T_i : jumlah seluruh *term* dalam kelas H_i pada data *testing*

$$\text{tempat} = \frac{2+1}{1+2+\dots+22} = \frac{3}{44}$$

$$\text{bagus} = \frac{0+1}{1+2+\dots+22} = \frac{1}{44}$$

$$\text{tapi} = \frac{0+1}{1+2+\dots+22} = \frac{1}{44}$$

$$\text{harga} = \frac{0+1}{1+2+\dots+22} = \frac{1}{44}$$

$$\text{tiket} = \frac{0+1}{1+2+\dots+22} = \frac{1}{44}$$

$$\text{mahal} = \frac{0+1}{1+2+\dots+22} = \frac{1}{44}$$

$$\text{Hasil: } \frac{3}{44} \times \frac{1}{44} \times \frac{1}{44} \times \frac{1}{44} \times \frac{1}{44} \times \frac{1}{44} = 4.13433 \times 10^{-10}$$

Hasil akhir dari perhitungan naïve bayes kurang dari 1 berarti ulasan “tempat bagus tapi harga tiket mahal” termasuk ke dalam kelas **negatif**.

3.5. Evaluasi

Untuk mengetahui hasil performa model Naïve Bayes, maka diperlukan dengan adanya tahapan evaluasi. Pada tahapan ini akan menghasilkan sebuah *confusion matrix* ukuran 2x2 dan *classification report* yang terdiri dari *accuracy*, *precision*, *recall* dan *f1-score*. Pada tahap evaluasi model yang diterapkan, dapat dilihat bahwa model tersebut menghasilkan *accuracy* sebesar **84%**. Adapun hasil *precision*, *recall*, *f1-score* pada tiap kelas dapat dikatakan bahwa keakuratan antara data yang diminta dengan data prediksi oleh pengguna untuk kelas positif sebesar **62%** dan kelas negatif **95%**. Sedangkan keberhasilan model dalam menemukan kembali informasi untuk kelas positif sebesar **86%** dan kelas negatif sebesar **83%** seperti yang diperlihatkan oleh Gambar 3.

Naïve Bayes				
[[63 10]				
[38 190]]				
Naïve Bayes				
	precision	recall	f1-score	support
0	0.62	0.86	0.72	73
1	0.95	0.83	0.89	228
accuracy			0.84	301
macro avg	0.79	0.85	0.81	301
weighted avg	0.87	0.84	0.85	301

Gambar 3. Confusion Matrix

3.6. Prediksi

Pada tahap ini melakukan uji coba model yang telah diterapkan untuk melakukan prediksi ulasan yang diinputkan oleh *user* dengan menghitung tiap kata menggunakan perhitungan tf-idf.

Setiap kata memiliki bobot yang berbeda sehingga dapat ditentukan apakah ulasan yang diinputkan termasuk kelas positif atau kelas negatif. Pada penelitian ini, contoh ulasan pertama “Dalam kebersihan tidak terjaga karena suka muncul bau tidak sedap” oleh sistem, ulasan tersebut diubah kebentuk numerik menggunakan tf-idf kemudian disimpan pada *array* dan dihitung berdasarkan rumus Naïve Bayes kemudian ulasan tersebut akan diklasifikasikan berdasarkan hasil perhitungan ke dalam kelas negatif. Pada penelitian ini dilakukan pengujian sebanyak 8 kali dan hasilnya sesuai dengan nada kalimat yang diinputkan.

4. KESIMPULAN

Berdasarkan hasil yang diperoleh pada penelitian ini menggunakan metode Naïve Bayes dan pembobotan Term Frequency Inverse Document Frequency diantaranya:

- a. Telah melakukan pengumpulan data ulasan pengunjung tempat wisata Karang Resik dari google maps sebanyak 1500 data menggunakan Data Scraper.
- b. Telah melakukan analisis sentimen untuk mengetahui nada komentar positif atau negatif dari pengunjung terhadap tempat wisata Krang Resik.
- c. Melakukan pembobotan kata menggunakan Term Frequency Inverse Document Frequency.
- d. Melakukan analisis sentimen dengan metode Naïve Bayes dengan menghasilkan nilai akurasi 84%.

5. SARAN

Pada penelitian ini masih bisa dikembangkan, sehingga peneliti mengusulkan saran agar penelitian ini menjadi lebih baik kedepannya, diantaranya memperbanyak dataset, melakukan tahapan *preprocessing* lainnya seperti tokenization, normalization, dan stemming dan bereksperimen menggunakan pembobotan yang lain.

DAFTAR PUSTAKA

- [1] B. D. Kusuma, "Berwisata Asyik ke Karang Resik Tasikmalaya," Pinhome, 31 09 2021. [Online]. Available: <https://www.pinhome.id/info-area/karang-resik-tasikmalaya/>. [Accessed 20 09 2022].
- [2] O. Somantri and Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," JEPIN (Jurnal Edukasi dan Penelitian Informatika), vol. 5, p. 191, 2019.
- [3] N. C. Siregar, "Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar," Jurnal Teknologia, vol. III, 2020.
- [4] C. S. Sriyano, "Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada," e-Proceeding of Engineering, vol. VIII, p. 3396, 2021.
- [5] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," Jurnal Inovtec Polbeng, vol. 3, 2018.