

KLASIFIKASI TEKS PADA ULASAN OBJEK WISATA FULAN FEHAN MENGUNAKAN METODE K-NEAREST NEIGHBOR (K-NN)

Meisya Maria Alfanita Bhia, Rangga Pahlevi Putra, Aviv Yuniar Rahman

Teknik Informatika, Universitas Widyagama Malang

Jl. Borobudur No.35, Mojolangu, Kec. Lowokwaru, Malang, Indonesia

meisyabhia19@gmail.com

ABSTRAK

Nusa Tenggara Timur (NTT) memiliki banyak potensi wisata yang belum sepenuhnya tergali, salah satunya adalah kawasan wisata Fulan Fehan. Untuk mendukung pengembangan pariwisata di daerah ini, analisis sentiment terhadap ulasan wisatawan dapat memberikan informasi berharga bagi pengelola wisata. Penelitian ini bertujuan untuk melakukan Klasifikasi Tinjauan tentang objek wisata Fulan Fehan yang ada di Google Maps dianalisis menggunakan metode K-Nearest Neighbor (K-NN). Data yang digunakan dalam penelitian ini sebanyak 1.164 ulasan yang diklasifikasi menjadi ulasan positif dan negatif. Proses klasifikasi meliputi tahap pengumpulan data melalui web scraping, preprocessing data yang mencakup cleansing, filtering, tokenizing, transformasi kasus, filter stopword, hingga tahap klasifikasi menggunakan algoritma K-NN. Hasil penelitian menunjukkan bahwa nilai $k=1$ dengan rasio data training dan testing 70:30 memberikan akurasi sebesar 99%, sedangkan $k=10$ memberikan akurasi terendah sebesar 39%. Penelitian ini memberikan kontribusi bagi pengembangan wisata Fulan Fehan dengan menyediakan informasi sentiment pengunjung yang dapat digunakan untuk meningkatkan kualitas layanan dan fasilitas wisata.

Kata kunci : Analisis Sentimen, Fulan Fehan, K-Nearest Neighbor, Klasifikasi Teks, Ulasan Google Maps

1. PENDAHULUAN

Nusa Tenggara Timur (NTT) memiliki potensi besar dalam sektor pariwisata yang bisa dikembangkan. Namun, kurangnya promosi dan terbatasnya akses informasi mengenai kualitas destinasi wisata di daerah tersebut membuat perkembangan pariwisata di NTT belum optimal [1].

Teknologi informasi saat ini telah dimanfaatkan dalam berbagai sektor salah satunya adalah pariwisata. Perkembangan teknologi informasi memiliki peran yang sangat signifikan dalam pengembangan sektor pariwisata dengan menyediakan akses informasi yang lebih mudah, meningkatkan layanan bagi para wisatawan, serta menawarkan berbagai aspek lain yang dapat memberikan keuntungan bagi mereka. [2].

Dari hasil pengamatan yang telah dilakukan, banyak pengunjung objek wisata Nusa Tenggara Timur (NTT) sudah memberikan ulasannya di Google Maps. Melalui ulasan tersebut, masyarakat dapat melakukan analisis guna mengetahui sentimen yang disampaikan oleh para pengunjung dan memberikan wawasan penting untuk meningkatkan mutu pariwisata di NTT. Salah satu kawasan yang memiliki nilai sejarah di NTT adalah Fulan Fehan yang berada di Desa Dirun, Kecamatan Lamaknen, Kabupaten Belu, Provinsi Nusa Tenggara Timur (NTT). Kawasan Fulan Fehan memiliki karakter lanskap berupa hamparan padang rumput yang luas dan berbukit.

Kawasan Fulan Fehan memiliki potensi pengembangan kawasan wisata namun masih terdapat banyak kendala seperti akses jalan menuju kawasan Fulan Fehan yang belum layak untuk digunakan, tidak adanya fasilitas penunjang kawasan wisata yang disediakan seperti tempat parkir, papan penunjuk arah,

toilet, belum tersedia loket masuk kawasan wisata Fulan Fehan [3].

Permasalahan utama yang dihadapi oleh pengelola adalah kesulitan dalam menentukan strategi pemasaran yang tepat karena masih banyak fasilitas pada Fulan Fehan yang masih belum maksimal. Selain itu, dalam proses perbaikan dan peningkatan layanan, pihak pengelola masih kesulitan dalam menentukan sektor mana yang harus diprioritaskan

Masukan yang diperoleh dari wisatawan yang datang ke objek wisata tersebut dapat digunakan sebagai dasar untuk peningkatan layanan serta pengembangan destinasi. Salah satu cara memperoleh informasi tersebut adalah melalui ulasan yang diberikan oleh wisatawan di Google Maps terkait obyek wisata Fulan Fehan. Ulasan di platform ini sangat berharga karena Google Maps merupakan platform global yang populer untuk memberikan ulasan dan rekomendasi terkait dunia pariwisata. [2].

Penelitian sebelumnya mengenai sentiment analisis untuk objek wisata di Kalimantan Barat [4] menghasilkan akurasi yang diperoleh bervariasi untuk setiap lokasi wisata. Akurasi tertinggi mencapai 0,76, sedangkan yang terendah adalah 0,38. Analisis sentimen terhadap destinasi wisata di Kalimantan Barat menunjukkan hasil yang positif. Berdasarkan kinerja metode Naive Bayes, nilai rata-rata f1-score untuk kelas positif mencapai 0,73, lebih tinggi dibandingkan kelas netral yang hanya 0,53, dan kelas negatif yang hanya memperoleh 0,14. Keunggulan k-NN terletak pada prinsip yang sederhana, di mana metode ini beroperasi berdasarkan jarak terpendek antara sampel uji dan sampel pelatihan. Dalam penelitian ini, preprocessing yang diterapkan meliputi tokenisasi, penyaringan stopwords, dan penggunaan bi-

gram. Dataset yang digunakan mencakup 50 ulasan positif dan 50 ulasan negatif. Eksperimen telah dilakukan dengan variasi nilai k , dan hasil menunjukkan bahwa akurasi tertinggi dicapai pada $k=7$, dengan akurasi 77,01%, presisi 92,38%, recall 61,56%, dan nilai AUC sebesar 0,894 [5].

2. TINJAUAN PUSTAKA

2.1. Klasifikasi Teks

Klasifikasi teks merujuk pada proses pengelompokan teks ke dalam kategori yang telah ditentukan sebelumnya. Proses ini berguna untuk mengorganisasi, menyusun, dan mengkategorikan berbagai jenis informasi. Contohnya, artikel berita bisa dikelompokkan berdasarkan topik tertentu, percakapan dalam obrolan bisa diurutkan menurut bahasa yang digunakan, sementara penyebutan merek dapat dikategorikan berdasarkan sentimen yang ditimbulkan, dan lain-lain [6].

2.2. K-Nearest Neighbor

Metode *K-Nearest Neighbors* (K-NN) adalah salah satu metode nonparametrik yang paling populer diperkenalkan oleh Fix dan Hodges pada tahun 1951. Karena hanya ada satu parameter K (jumlah nearest neighbors) yang perlu ditentukan, mudah untuk menerapkan metode K-NN [7].

K-Nearest Neighbors (K-NN) adalah metode klasifikasi yang metode voting untuk memprediksi kelas output dari data masukan. Metode ini berdasarkan pada suatu algoritma yang menghitung jarak antara suatu titik data dan titik data lainnya yang berdekatan.

Algoritma ini menggunakan jarak sebagai parameter untuk memprediksi kelas output dari data masukan K-NN akan menentukan K titik data terdekat dengan titik data masukan dan kemudian menghitung jumlah kelas yang berbeda akan menjadi kelas output yang diprediksi [8].

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Dimana:

$d_{(x,y)}$ = jarak antara objek a dan b

n = banyaknya variabel

x_i = nilai objek data training a pada variabel ke i

y_i = nilai objek data training b pada variabel ke i

2.3. Web Scrapping

Web scraping merujuk pada proses pengumpulan data dengan cara yang tidak melibatkan penggunaan program yang berkomunikasi dengan API, atau dilakukan oleh manusia melalui browser web. Umumnya, ini dilakukan dengan membuat program otomatis yang mengirim permintaan ke server web untuk mendapatkan data, biasanya dalam bentuk HTML dan file lain yang membentuk halaman web, lalu mem-parsing data tersebut untuk mendapatkan informasi yang diinginkan. [6].

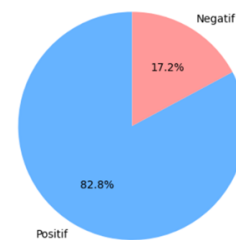
2.4. Google Colaboratory

Google Colaboratory adalah sebuah platform yang mirip dengan Jupyter Notebooks. Jupyter sendiri adalah alat open source yang berbasis web, mengintegrasikan berbagai bahasa pemrograman, pustaka, dan alat untuk visualisasi data [9].

Notebook Jupyter dapat digunakan baik secara lokal maupun melalui cloud. Setiap dokumen terdiri dari beberapa sel, di mana setiap sel dapat berisi kode atau markdown, dan hasilnya akan terintegrasi dalam dokumen. Jenis output yang dihasilkan meliputi teks, tabel, grafik, dan diagram. Google Colaboratory (Colab) merupakan inisiatif yang bertujuan untuk memperluas akses pendidikan dan penelitian dalam bidang pembelajaran mesin. Notebook Colab dibangun di atas Jupyter dan berfungsi sebagai objek dalam Google Documents, memungkinkan pengguna untuk berbagi dan berkolaborasi pada notebook yang sama [10].

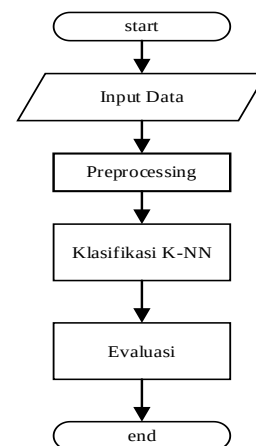
3. METODE PENELITIAN

Data yang digunakan pada penelitian ini diambil dari kolom ulasan pengunjung objek wisata Fulan Fehan pada Google Maps menggunakan teknik web scrapping dengan memanfaatkan aplikasi Instant Data Scapper.



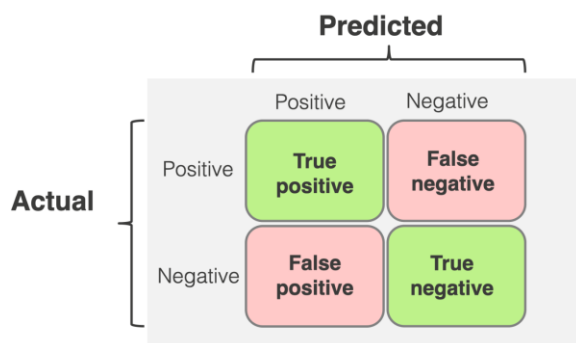
Gambar 1. Persentase perbandingan respon positif dan negatif

Data yang digunakan adalah data hasil scrapping sebanyak 1164 yang terdiri dari variabel yaitu ulasan dan dibagi menjadi dua kategori yaitu positif dan negatif. Ulasan yang mempunyai respon positif berjumlah 965 ulasan, dan respon negatif sebesar 200 ulasan.



Gambar 2. Flowchart Analisis Sentimen Menggunakan K-NN

Hasil data ulasan yang telah mengalami preprocessing akan diklasifikasikan menggunakan algoritma *K-Nearest Neighbors* (K-NN). Pada proses klasifikasi KNN, langkah awal adalah input data ulasan yang telah diproses, selanjutnya akan memasukkan nilai k dan perbandingan data training dan data testing yang akan digunakan, pada penelitian ini menggunakan $k = 1$ hingga $k=10$, dan perbandingan data 90:10 hingga 10:90. Proses selanjutnya akan dihitung jarak antar ecludian dan mengidentifikasi jarak terdekat. Dari K tetangga terdekat yang sudah ditemukan, lakukan voting berdasarkan kelas (positif dan negatif). Data baru akan diklasifikasikan ke dalam kelas yang paling banyak muncul di antara tetangga terdekat tersebut. Setelah mengklasifikasikan data uji, bandingkan hasil klasifikasi dengan label yang benar menggunakan matiks confusion.



Gambar 3. Matrix confusion

Berdasarkan mtrix confusion akan dihitung presisi, recall, F-1 score, dan akurasi.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (5)$$

Dimana:

TP = True positive

TN = True negative

FP = False Positive

FN = False Negative

4. HASIL DAN PEMBAHASAN

Klasifikasi KNN pada penelitian ini menggunakan lima perbandingan untuk data training dan data testing yaitu 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80 dan 10:90 sedangkan untuk nilai k menggunakan 1, 2, 3, 4, 5, 6, 7, 8, 9, dan 10.

4.1. Preprocessing Data

Data ulasan yang telah didapatkan akan dilakukan preprocessing terlebih dahulu.

a. *Cleansing Data*, pada proses ini akan menghapus tanda baca, angka dan karakter khusus yang tidak diperlukan.

	ulasan	kategori	ulasan_clean
0	Hamparan padang rumput hijau, sapi dan kuda ya...	positif	Hamparan padang rumput hijau sapi dan kuda yan...
1	Udara nya sangat menusuk ketika berkunjung di ...	negatif	Udara nya sangat menusuk ketika berkunjung di ...
2	pemandangan bagus sepertinya pagi menjelang si...	positif	pemandangan bagus sepertinya pagi menjelang si...
3	5/5 Salah satu lokasi wisata terbaik di Nusa T...	positif	Salah satu lokasi wisata terbaik di Nusa Tengg...
4	Fulan Fehan, masuk kabupaten belu. Padang rump...	positif	Fulan Fehan masuk kabupaten belu Padang rumput...
...	...	...	...
1160	keren sekali	positif	keren sekali
1161	Very cold	negatif	Very cold
1162	land above the clouds	positif	land above the clouds
1163	Wooooow	positif	Wooooow
1164	It's cool	positif	Its cool

Gambar 4. Hasil cleansing data

b. *Translate*, hasil *cleansing* akan diterjemahkan ke dalam Bahasa Indonesia karena ada beberapa ulasan yang masih menggunakan bahasa asing.

ulasan_clean	ulasan_terjemahan
Hamparan padang rumput hijau sapi dan kuda yan...	Hamparan padang rumput hijau sapi dan kuda yan...
Udara nya sangat menusuk ketika berkunjung di ...	Udara nya sangat menusuk ketika berkunjung di ...
...	...
pemandangan bagus sepertinya pagi menjelang si...	pemandangan bagus sepertinya pagi menjelang si...
Salah satu lokasi wisata terbaik di Nusa Tengg...	Salah satu lokasi wisata terbaik di Nusa Tengg...
Fulan Fehan masuk kabupaten belu Padang rumput...	Fulan Fehan masuk kabupaten belu Padang rumput...
...	...
keren sekali	keren sekali
Very cold	Sangat dingin
land above the clouds	mendarat di atas awan
Wooooow	Wooooow
Its cool	Itu keren

Gambar 5. Hasil terjemahan

c. *Transform Cases*, proses ini bertujuan untuk merubah bentuk tulisan pada data ulasan ke dalam *lower case*. Hal ini dilakukan agar mengurangi redudansi data, missal kata “Bagus” dan “bagus” akan dianggap model yag berbeda, dengan mengubah ke dalam huruf kecil kedua kata tersebut akan menjadi model yang sama.

ulasan_terjemahan	ulasan_terjemahan_lower
Hamparan padang rumput hijau sapi dan kuda yan...	hamparan padang rumput hijau sapi dan kuda yan...
Udara nya sangat menusuk ketika berkunjung di ...	udara nya sangat menusuk ketika berkunjung di ...
...	...
pemandangan bagus sepertinya pagi menjelang si...	pemandangan bagus sepertinya pagi menjelang si...
Salah satu lokasi wisata terbaik di Nusa Tengg...	salah satu lokasi wisata terbaik di nusa tengg...
Fulan Fehan masuk kabupaten belu Padang rumput...	fulan fehan masuk kabupaten belu padang rumput...
...	...
keren sekali	keren sekali
Sangat dingin	sangat dingin
mendarat di atas awan	mendarat di atas awan
Wooooow	woooooow
Itu keren	itu keren

Gambar 6. Hasil transform cases ke huruf kecil

d. *Tokenize*, proses tokenisasi bertujuan untuk membagi teks mentah menjadi unit-unit kecil yang disebut token, yang biasanya berupa kata, frasa,

atau bahkan karakter individu. Missal kalimat “Wisata Fulan Fehan bagus sekali” maka akan dibagi menjadi lima kata menjadi [“Wisata”, “Fulan”, “Fehan”, “bagus”, “sekali”].

	ulasan_terjemahan_lower	ulasan_token
0	hamparan padang rumput hijau sapi dan kuda yan...	[hamparan, padang, rumput, hijau, sapi, dan, k...
1	udara nya sangat menusuk ketika berkunjung di ...	[udara, nya, sangat, menusuk, ketika, berkunjung...
2	pemandangan bagus seperti pagi menjelang si...	[pemandangan, bagus, seperti, pagi, menela...
3	salah satu lokasi wisata terbaik di nusa tengg...	[salah, satu, lokasi, wisata, terbaik, di, nus...
4	fulan fehan masuk kabupaten belu padang rumput...	[fulan, fehan, masuk, kabupaten, belu, padang,...
...	...	...
1160	keren sekali	[keren, sekali]
1161	sangat dingin	[sangat, dingin]
1162	mendarat di atas awan	[mendarat, di, atas, awan]
1163	wooooo	[wooooo]
1164	itu keren	[itu, keren]

Gambar 7. Hasil tokenize

- e. *Stopword*, *stopword* dalam penelitian ini menggunakan sastrawi, Stopword Sastrawi adalah daftar kata umum yang sering muncul dalam teks tetapi dianggap tidak memiliki makna signifikan dalam analisis teks, seperti kata sambung, kata depan, atau kata ganti. Selain penggunaan *stopword* sastrawi, akan ditambahkan juga beberapa kata yang harus dihapus seperti kata nya, rp, satu, dan yang lainnya.

	ulasan_token	ulasan_stopword
0	[hamparan, padang, rumput, hijau, sapi, dan, k...	[hamparan, padang, rumput, hijau, sapi, kuda, ...
1	[udara, nya, sangat, menusuk, ketika, berkunjung...	[udara, menusuk, berkunjung, pagi, hari, asri]
2	[pemandangan, bagus, seperti, pagi, menela...	[pemandangan, bagus, pagi, menjelang, siang, j...
3	[salah, satu, lokasi, wisata, terbaik, di, nus...	[lokasi, wisata, terbaik, timur, meski, jauh, ...
4	[fulan, fehan, masuk, kabupaten, belu, padang...	[padang, rumput, luas, banyak, sapi, kuda, sin...
...	...	...
1160	[keren, sekali]	[keren]
1161	[sangat, dingin]	[dingin]
1162	[mendarat, di, atas, awan]	[mendarat, atas, awan]
1163	[wooooo]	[wooooo]
1164	[itu, keren]	[keren]

Gambar 8. Hasil *stopword*

4.2. Perbandingan 90:10

Tabel 1. Hasil perbandingan 90:10

k	Presisi	Recall	F-1 Score	Akurasi
1	92%	100%	96%	98%
2	92%	100%	96%	98%
3	92%	100%	96%	98%
4	81%	100%	90%	96%
5	81%	100%	90%	96%
6	43%	100%	60%	75%
7	63%	100%	77%	89%
8	32%	100%	49%	61%
9	37%	100%	64%	68%
10	24%	100%	39%	42%

Nilai akurasi terbesar pada $k = [1,2,3]$ sebesar 98% dan nilai akurasi terendah pada $k=10$ sebesar 42%. Berdasarkan grafik tersebut, dapat disimpulkan bahwa semakin tinggi nilai k pada pengujian KNN untuk data 90:10 nilai akurasi semakin menurun.

4.3. Perbandingan 80:20

Tabel 2. Hasil perbandingan 80:20

k	Presisi	Recall	F-1 Score	Akurasi
1	85%	100%	92%	97%
2	78%	100%	88%	95%
3	85%	100%	92%	97%
4	70%	100%	82%	93%
5	78%	100%	88%	95%
6	51%	100%	67%	84%
7	53%	100%	70%	85%
8	31%	100%	48%	63%
9	31%	100%	48%	64%
10	22%	100%	36%	42%

Nilai akurasi terbesar pada $k=[1,3]$ sebesar 97% dan nilai akurasi terendah pada $k=10$ sebesar 42%. Berdasarkan grafik tersebut, dapat disimpulkan bahwa semakin tinggi nilai k pada pengujian KNN untuk data 80:20 nilai akurasi semakin menurun.

4.4. Perbandingan 70:30

Tabel 3. Hasil perbandingan 70:30

k	Presisi	Recall	F-1 Score	Akurasi
1	100%	95%	97%	99%
2	89%	100%	94%	98%
3	89%	100%	94%	98%
4	64%	100%	78%	91%
5	64%	100%	78%	91%
6	44%	100%	62%	80%
7	44%	100%	62%	80%
8	26%	100%	41%	54%
9	26%	100%	41%	55%
10	22%	100%	35%	42%

Nilai akurasi terbesar pada $k=1$ sebesar 99% dan nilai akurasi terendah pada $k=10$ sebesar 42%. Berdasarkan grafik tersebut, dapat disimpulkan bahwa semakin tinggi nilai k pada pengujian KNN untuk data 70:30 nilai akurasi semakin menurun.

4.5. Perbandingan 60:40

Tabel 4. Hasil perbandingan 60:40

k	Presisi	Recall	F-1 Score	Akurasi
1	76%	100%	86%	95%
2	61%	100%	75%	90%
3	61%	100%	75%	90%
4	35%	100%	52%	71%
5	35%	100%	52%	71%
6	23%	100%	38%	49%
7	23%	100%	38%	50%
8	21%	100%	35%	43%
9	22%	100%	36%	44%
10	20%	100%	34%	39%

Nilai akurasi terbesar pada $k=1$ sebesar 95% dan nilai akurasi terendah pada $k=10$ sebesar 39%. Berdasarkan grafik tersebut, dapat disimpulkan bahwa semakin tinggi nilai k pada pengujian KNN untuk data 70:30 nilai akurasi semakin menurun.

4.6. Perbandingan 50:50

Tabel 5. Hasil perbandingan 50:50

k	Presisi	Recall	F-1 Score	Akurasi
1	74%	100%	85%	94%
2	45%	100%	62%	81%
3	45%	100%	62%	81%
4	29%	100%	45%	51%
5	29%	100%	45%	51%
6	22%	100%	36%	45%
7	22%	100%	36%	45%
8	21%	100%	35%	41%
9	21%	100%	35%	42%
10	21%	100%	34%	39%

Nilai akurasi terbesar pada $k=1$ sebesar 95% dan nilai akurasi terendah pada $k=10$ sebesar 39%. Berdasarkan grafik tersebut, dapat disimpulkan bahwa semakin tinggi nilai k pada pengujian KNN untuk data 50:50 nilai akurasi semakin menurun. Penurunan terjadi karena KNN mulai mempertimbangkan lebih banyak tetangga yang mungkin berada cukup jauh dari titik yang diuji..

4.7. Perbandingan 40:60

Tabel 6. Hasil perbandingan 40:60

k	Presisi	Recall	F-1 Score	Akurasi
1	58%	100%	73%	89%
2	33%	100%	50%	69%
3	33%	100%	50%	69%
4	24%	100%	38%	50%
5	24%	100%	38%	50%
6	21%	100%	34%	41%
7	21%	100%	35%	42%
8	20%	100%	33%	38%
9	60%	90%	72%	89%
10	57%	90%	70%	88%

Nilai akurasi terbesar pada $k=1$ dan $k=9$ sebesar 89% dan nilai akurasi terendah pada $k=8$ sebesar 38%. Nilai k pada pengujian KNN untuk data 60:40 pada $k=2$ hingga $k=8$ nilai akurasi semakin menurun. Penurunan terjadi karena KNN mulai mempertimbangkan lebih banyak tetangga yang mungkin berada cukup jauh dari titik yang diuji. Hal ini dapat menyebabkan pengaruh dari tetangga yang secara geografis atau dalam ruang fitur tidak relevan dengan kelas sebenarnya, sehingga mempengaruhi prediksi. Pada $k=9$ nilai akurasi naik menjadi 89% hal ini dikarenakan terdapat keseimbangan yang baik antara menggunakan cukup banyak tetangga untuk stabilitas tanpa terlalu banyak memediasi data yang tidak relevan. Ini menunjukkan bahwa model pada nilai k ini lebih generalis dengan tingkat stabilitas yang lebih tinggi, tanpa overfitting seperti pada $k=1$

4.8. Perbandingan 30:70

Tabel 7. Hasil perbandingan 30:70

k	Presisi	Recall	F-1 Score	Akurasi
1	48%	100%	65%	82%
2	29%	100%	44%	60%
3	29%	100%	44%	60%
4	23%	100%	37%	46%
5	23%	100%	37%	46%
6	21%	100%	34%	38%
7	22%	100%	36%	42%
8	20%	100%	33%	36%
9	54%	76%	63%	86%
10	51%	88%	64%	84%

Nilai akurasi terbesar pada $k=9$ sebesar 86% dan nilai akurasi terendah pada $k=8$ sebesar 36%. Nilai k pada pengujian KNN untuk data 30:70 dari $k=1$ ke $k=8$ nilai akurasi semakin menurun. Penurunan terjadi karena KNN mulai mempertimbangkan lebih banyak tetangga yang mungkin berada cukup jauh dari titik yang diuji. Hal ini dapat menyebabkan pengaruh dari tetangga yang secara geografis atau dalam ruang fitur tidak relevan dengan kelas sebenarnya, sehingga mempengaruhi prediksi. Pada $k=9$ nilai akurasi naik menjadi 89% hal ini dikarenakan terdapat keseimbangan yang baik antara menggunakan cukup banyak tetangga untuk stabilitas tanpa terlalu banyak memediasi data yang tidak relevan. Ini menunjukkan bahwa model pada nilai k ini lebih generalis dengan tingkat stabilitas yang lebih tinggi..

4.9. Perbandingan 20:80

Tabel 8. Hasil perbandingan 20:80

k	Presisi	Recall	F-1 Score	Akurasi
1	42%	100%	59%	77%
2	25%	100%	41%	50%
3	100%	47%	64%	91%
4	78%	57%	66%	90%
5	80%	37%	50%	88%
6	63%	42%	50%	86%
7	55%	17%	26%	83%
8	50%	24%	33%	83%
9	62%	8%	14%	83%
10	22%	8%	12%	79%

Nilai akurasi terbesar pada $k=3$ sebesar 91% dan nilai akurasi terendah pada $k=2$ sebesar 50%. Nilai akurasi dari $k=1$ ke $k=2$ mengalami penurunan hal ini dikarenakan model tidak menemukan keseimbangan antar tetangga yang terdekat, sedangkan pada $k=3$ terjadi kenaikan nilai akurasi yang signifikan yang berarti bahwa model pada $k=3$ menemukan keseimbangan yang baik antar tetangga terdekat dan pada $k=5$ hingga $k=10$ mengalami penurunan akurasi yang disebabkan tidak menemukan keseimbangan yang sesuai.

4.10. Perbandingan 10:90

Tabel 9. Hasil perbandingan 10:90

k	Presisi	Recall	F-1 Score	Akurasi
1	93%	38%	54%	89%
2	61%	52%	56%	86%
3	67%	17%	27%	84%
4	51%	24%	33%	83%
5	44%	12%	19%	82%
6	27%	14%	18%	79%
7	9%	3%	4%	79%
8	15%	6%	8%	79%
9	36%	3%	5%	83%
10	26%	3%	5%	82%

Nilai akurasi terbesar pada $k=1$ sebesar 93% dan nilai akurasi terendah pada $k=6$, $k=7$, dan $k=8$ sebesar 79%. Nilai k pada pengujian KNN untuk data 10:90 nilai akurasi semakin menurun. Penurunan terjadi karena KNN mulai mempertimbangkan lebih banyak tetangga yang mungkin berada cukup jauh dari titik yang diuji. Hal ini dapat menyebabkan pengaruh dari tetangga yang secara geografis atau dalam ruang fitur tidak relevan dengan kelas sebenarnya, sehingga mempengaruhi prediksi. Pada $k=9$ terjadi peningkatan karena data menemukan sedikit keseimbangan antar tetangga, akan tetapi pada $k=10$ mengalami penurunan nilai akurasi lagi.

4.11. Pembahasan

Tabel 10. Hasil perbandingan antar rasio

k	Rasio Data Training								
	90	80	70	60	50	40	30	20	10
1	98%	97%	99%	95%	94%	89%	82%	77%	89%
2	98%	95%	98%	90%	81%	69%	60%	50%	86%
3	98%	97%	98%	90%	81%	69%	60%	91%	84%
4	96%	93%	91%	71%	51%	50%	46%	90%	83%
5	96%	95%	91%	71%	51%	50%	46%	88%	82%
6	75%	84%	80%	49%	45%	41%	38%	86%	79%
7	89%	85%	80%	50%	45%	42%	42%	83%	79%
8	61%	63%	54%	43%	41%	38%	36%	83%	79%
9	68%	64%	55%	44%	42%	89%	86%	83%	83%
10	42%	42%	42%	39%	39%	88%	84%	79%	82%

Berdasarkan hasil pengujian, pada rasio 90:10 menghasilkan nilai akurasi tertinggi pada $k=[1,2,3]$ sebesar 98% dan akurasi terendah pada $k=10$ dengan nilai akurasi sebesar 42%. Pada rasio 80:20 menghasilkan nilai akurasi tertinggi pada $k=[1,3]$ sebesar 97% dan akurasi terendah pada $k=10$ dengan nilai akurasi sebesar 42%. Pada rasio 70:30 menghasilkan nilai akurasi tertinggi pada $k=1$ sebesar 99% dan akurasi terendah pada $k=10$ dengan nilai akurasi sebesar 42%. Pada rasio 60:40 menghasilkan nilai akurasi tertinggi pada $k=1$ sebesar 95% dan akurasi terendah pada $k=10$ dengan nilai akurasi sebesar 39%. Pada rasio 50:50 menghasilkan nilai akurasi tertinggi pada $k=1$ sebesar 94% dan akurasi terendah pada $k=10$ dengan nilai akurasi sebesar 39%.

Dari semua rasio yang digunakan pada pengujian, nilai $k = 1$ menghasilkan nilai akurasi pada tiap-tiap rasio dengan nilai akurasi tertinggi pada rasio 70:30 sebesar 99%, sedangkan $k=10$ menghasilkan nilai akurasi terendah untuk tiap-tiap pengujian dengan nilai akurasi terendah pada rasio 50:50 sebesar 39%. Maka untuk klasifikasi text pada ulasan objek wisata fulan fehan menggunakan KNN dengan $k=1$ dan rasio 70:30.

Ketika nilai k kecil (misalnya 1 atau 3), model KNN hanya mempertimbangkan sejumlah kecil tetangga terdekat untuk menentukan kelas dari titik yang diuji. Ini membuat prediksi lebih sensitif terhadap pola lokal dari data. Namun, saat nilai k semakin besar, KNN mulai mempertimbangkan lebih banyak tetangga yang mungkin berada cukup jauh dari titik yang diuji. Hal ini dapat menyebabkan pengaruh dari tetangga yang secara geografis atau dalam ruang fitur tidak relevan dengan kelas sebenarnya, sehingga mempengaruhi prediksi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian dan pembahasan, proses sentiment analisis pada ulasan objek wisata fulan fehan menggunakan KNN mendapatkan nilai $k = 1$ menghasilkan nilai presisi 100%, recall 95%, f-1 score 97%, akurasi 99% pada rasio 70:30, sedangkan $k=10$ menghasilkan nilai presisi 21%, recall 100%, f-1 score 34% akurasi 39%, pada rasio 50:50. Maka untuk klasifikasi text pada ulasan objek wisata fulan fehan menggunakan metode KNN dengan $k = 1$ dan rasio 70:30. Untuk penelitian selanjutnya dapat mengambil data ulasan dari platform wisata yang lain agar data yang didapatkan semakin banyak dan akurasi atau ketepatan model diharapkan semakin baik, serta diharapkan menggunakan variabel yang lain tidak hanya ulasan seperti rating yang diberikan oleh pemberi ulasan terhadap tempat wisata.

DAFTAR PUSTAKA

- [1] S. B. Tandilino and M. Kabu, "Design Village Tourism Package Model to Support the Application of Estate Tourism Program in East Nusa Tenggara," in *Proceedings of the International Conference on Applied Science and Technology on Social Science (ICAST-SS 2020)*, 2021. doi: 10.2991/assehr.k.210424.058.
- [2] A. A. Arifiyanti, M. F. Pandji, and B. Utomo, "Analisis Sentimen Ulasan Pengunjung Objek Wisata Gunung Bromo pada Situs Tripadvisor," *Explor. J. Sist. Inf. dan Telemat.*, vol. 13, no. 1, 2022, doi: 10.36448/jsit.v13i1.2539.
- [3] D. Budiyo, H. Kurniawan, A. Sumiati, and Y. Assa, "Analisis Kesesuaian Lanskap Fulan Fenan Sebagai Objek Wisata Sejarah Yang Berkelanjutan," *J. Buana Sains*, vol. 22, no. 2, 2022.
- [4] A. Rifa'i, H. Sujaini, and D. Prawira, "Sentiment Analysis Objek Wisata Kalimantan Barat Pada Google Maps Menggunakan Metode Naive

- Bayes,” *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, 2021, doi: 10.26418/jp.v7i3.48132.
- [5] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [6] I. Wirawan and A. Zubaidi, “Klasifikasi Teks Ulasan Pada Web Tripadvisor Tentang Wisata Alam Pulau Lombok Menggunakan Metode Naive Bayes Classifier,” *J. Teknol. Inf. ...*, 2022.
- [7] S. Sahara and R. Agung Permana, “Metode Knn Pada Sentiment Analisis Review Produk Game Android,” *Ijns.org Indones. J. Netw. Secur.*, vol. 11, no. 2, pp. 2302–5700, 2022, [Online]. Available: www.zdnet.com
- [8] R. Amilia and E. Prasetyo, “Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Pada Anak Menggunakan Metode K-Nearest Neighbor Studi Kasus Rumah Sakit Pku Muhammadiyah Ujung Pangkah Gresik,” *Indexia*, vol. 2, no. 2, 2021, doi: 10.30587/indexia.v2i2.2557.
- [9] D. S. R. Sukhdeve and S. S. Sukhdeve, “Google Colaboratory,” in *Google Cloud Platform for Data Science*, 2023. doi: 10.1007/978-1-4842-9688-2_2.
- [10] F. Wilyani, Q. Nuryan Arif, and F. Aslimar, “Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory,” *J. Pengabd. Pada Masy. Indones.*, vol. 3, no. 1, 2024.