

ANALISIS SENTIMEN ULASAN VILLA DI UBUD MENGGUNAKAN METODE NAÏVE BAYES, DECISION TREE, DAN K-NN

Ni Luh Wiwik Sri Rahayu Ginantra¹, C. P. Yanti², G. D. Prasetya³, Ida Bagus Gede Sarasvananda⁴, I Komang Arya Ganda Wiguna⁵

Prodi Teknik Informatika, Institut Bisnis dan Teknologi Indonesia

email: email-penulis¹, christinapy@instiki.ac.id², dhikiprasetya@gmail.com³, sarasvananda@instiki.ac.id⁴, kmaryagw@instiki.ac.id⁵

Abstrak

Penggunaan internet pada sektor pariwisata dapat mempermudah seseorang dalam memperoleh informasi mengenai suatu tempat wisata. Ubud menjadi salah satu destinasi wisata favorit di Kabupaten Gianyar, Provinsi Bali menawarkan berbagai macam jenis wisata yang salah satunya yaitu villa sebagai akomodasi para wisatawan. Informasi mengenai opini wisatawan terhadap Ubud pada *Google Maps*, dapat menjadi bahan evaluasi untuk mempertahankan citra positif pariwisata Ubud. Pada penelitian ini akan dilakukan analisa text mining dengan analisa sentimen villa di Ubud berdasarkan data opini pada *Google Maps* menggunakan metode *Naive Bayes*, *Decision Tree*, dan *k-NN* menggunakan aplikasi RapidMiner. Adapun hasil pengujian dari 2894 data yang dibagi menjadi 2024 data *training* dan 867 data uji. Penggunaan SMOTE *up-sampling* digunakan untuk menyamakan jumlah data yang tidak seimbang. Hasil analisa menunjukkan bahwa metode *k-NN* lebih unggul dalam menganalisis sentimen dengan prediksi sentimen 526 positif, 233 netral, 72 negatif. *Performance confusion matrix* menunjukkan bahwa metode *k-NN* unggul dengan akurasi 91.26%, *precision* 92.97%, *recall* 91.26%, dan *overall performance* 91.83%.

Kata kunci: Analisis sentimen, Villa, Ubud, *Naive Bayes*, *Decision Tree*, *k-NN*

Abstract

The use of the internet in the tourism sector can make it easier for someone who want to obtain information about a tourist places. Ubud is one of the favorite tourist destinations in Gianyar Regency, Bali Province which offers various types of tourism, one of which is villas as accommodation for tourists. Information about tourist opinions on Ubud on Google Maps, can be used as evaluation material to maintain a positive reputation of Ubud tourism. The test results from 2894 data are divided into 2024 training data and 867 test data. The use of SMOTE up-sampling is used to equalize the amount of data that is not balanced. The results of the analysis show that the k-NN method is superior in analyzing sentiment with predictions of 526 positive sentiment, 233 neutral, 72 negative. The performance of confusion matrix shows that the k-NN method excels with 91.26% accuracy, 92.97% precision, 91.26% recall, and 91.83% overall performance

Keywords : *Sentiment Analysis, Villa, Ubud, Naive Bayes, Decision Tree, k-NN*

Diterima Redaksi: 04-07-2022 | Selesai Revisi: 27-10-2022 | Diterbitkan Online: 27-12-2022

DOI: <https://doi.org/10.23887/janapati.v11i3.49450>

PENDAHULUAN

Dalam kehidupan sehari-hari, manusia tidak terlepas dari internet karena penggunaannya yang begitu multifungsi. Seseorang bisa mencari dan mendapatkan informasi yang mereka butuhkan dengan cepat melalui mesin pencarian internet [1]. Pemanfaatan teknologi internet bisa digunakan untuk mengolah data yang sudah ada untuk dijadikan sebagai pengukuran kualitas ataupun perbandingan opini masyarakat di internet [2] terhadap suatu tempat wisata. Ukuran kualitas

ini bisa menjadi bahan pertimbangan bagi pengguna internet dalam mengambil keputusan untuk berkunjung ke tempat wisata [3]. Selain itu juga bermanfaat bagi pihak pengelola wisata yang menjadikannya sebagai bahan evaluasi dalam peningkatan kualitas tempat wisata [4].

Ubud menjadi salah satu destinasi favorit di Bali. Ubud merupakan sebuah kecamatan di Kabupaten Gianyar yang terkenal dengan berbagai jenis wisata, baik itu wisata kuliner, alam, budaya, sejarah, dan religi [5]. Untuk mengakomodasi wisatawan yang sedang

berada di Ubud, mereka bisa memanfaatkan fasilitas hotel, villa, ataupun home stay. Ubud memiliki banyak villa dengan karakteristik dan keunggulan masing-masing untuk memenuhi kebutuhan dan kepuasan wisatawan yang berlibur di daerah Ubud. Walaupun Ubud begitu terkenal dengan keindahan wisatanya, tentu pihak pengelola villa yang ada di Ubud harus tetap menjaga kualitasnya demi mempertahankan citra positif di dunia pariwisata. Hal ini perlu ditelusuri lebih lanjut melalui pendapat/opini wisatawan tentang pengalamannya berlibur di villa daerah Ubud. Salah satu media yang menampung opini para wisatawan mengenai villa di Ubud adalah Google Maps. Hal ini dikarenakan Google Maps merupakan aplikasi peta virtual yang disediakan oleh Google untuk mencari informasi tentang lokasi tertentu [6]. Wisatawan ataupun tour guide sering menggunakan aplikasi ini sebagai acuan ketika mencari lokasi yang diinginkan.

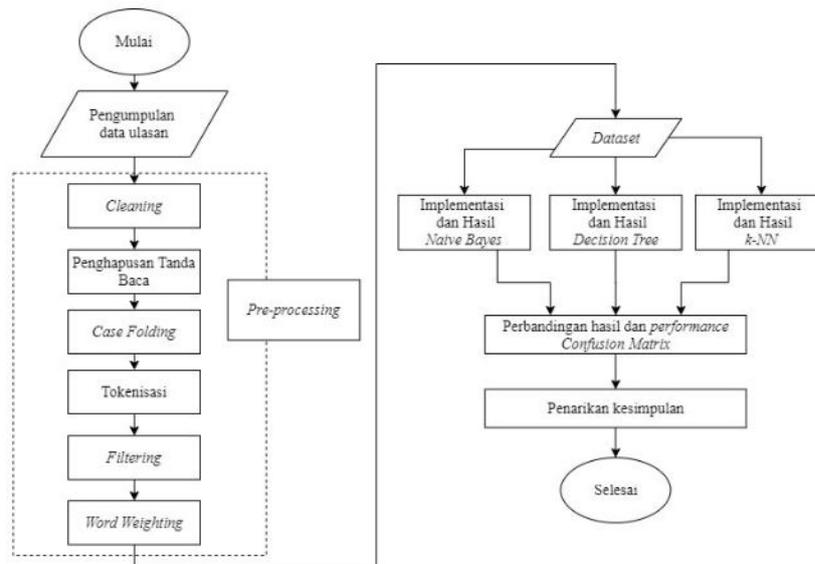
Untuk menghasilkan informasi dari data opini yang ada pada Google Maps, perlu dilakukan analisa *text mining* dengan analisa sentimen untuk villa di Ubud dengan memisahkan opini ke dalam sentimen positif, negatif dan netral. *Text Mining* adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dokumen, *clustering*, *information extraction*, analisis sentimen dan *information retrieval* dimana *text mining* merupakan variasi dari *data mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar [7]. Analisis sentimen ini menggunakan data ulasan dari 40 villa di daerah Ubud pada situs Google Maps. Data 40 villa di Ubud ini diambil dari masing-masing lima villa di delapan desa yang ada di Kecamatan Ubud, yaitu Desa Loddunduh, Desa Peliatan, Desa Mas, Desa

Petulu, Desa Sayan, Desa Singakerta, Desa Kedewatan, dan Kelurahan Ubud. Adapun rentang waktu yang diambil adalah 3 tahun, yaitu dari Januari 2019 hingga Januari 2022. Penentuan rentang waktu ini dikarenakan dalam tiga tahun bisa memberikan ulasan yang lebih relevan dibanding rentang waktu lebih lama, dimana terdapat kemungkinan bahwa villa tersebut sudah berubah atau tidak beroperasi lagi. Data yang dikumpulkan adalah 4585 data ulasan yang kemudian setelah melalui proses cleaning menjadi 2894. Kemudian data tersebut dibagi menjadi data training dengan kelas positif sebanyak 1647, 155 netral, dan 222 negatif. Sedangkan data uji pada kelas positif sebanyak 706, 66 netral, dan 95 negatif.

Penelitian ini penulis akan menganalisis sentimen ulasan villa di Ubud dengan membandingkan metode Naive Bayes, Decision Tree dan k-NN. Untuk mengetahui proses analisis sentimen akan diolah menggunakan RapidMiner [8]. Kemudian untuk menyeimbangkan data ulasan digunakan metode SMOTE up-sampling [9]. Berdasarkan uraian di atas maka penelitian ini fokus dilakukan dengan judul "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree dan k-NN".

METODE

Pada tahap penelitian diuraikan tentang alur penelitian yang dimulai dari pengumpulan data hingga analisis. Adapun penjelasan dari tahapan dalam penelitian analisa sentimen pada Villa di daerah Ubud menggunakan *text mining* metode Naive Bayes, Decision Tree, dan k-NN, akan dijabarkan melalui flowchart pada gambar 1 dibawah ini:



Gambar 1. Alur Penelitian

Pengumpulan Data

Pengumpulan data dimaksudkan untuk mendapatkan data yang akan digunakan dalam penelitian. Data yang digunakan untuk penelitian text mining adalah ulasan dari 40 villa yang diambil masing-masing sebanyak lima villa di delapan desa yang ada di Kecamatan Ubud. Dalam proses pengumpulannya menggunakan Instant Data Scraper. Berikut merupakan tahapan pengumpulan data dari penelitian ini.

Pre-processing

Pre-processing merupakan proses awal persiapan text mining untuk mempersiapkan teks menjadi data yang bisa diolah lebih lanjut. Manfaat dari pre-processing data ini adalah agar mendapat kumpulan kata yang penting dan berguna dalam analisa sentimen. Berikut merupakan tahap-tahap *pre-preprocessing* :

- a. **Cleaning**
Proses cleaning pada RapidMiner menggunakan operator replace dan filter. Fungsinya adalah untuk menghilangkan kata-kata yang tidak diperlukan seperti template default, kalimat terduplikat dan kolom tabel yang kosong.
- b. **Penghapusan Tanda Baca**
Proses penghapusan tanda baca menggunakan operator replace yang parameternya berisi tanda baca
- c. **Case Folding**
Case folding menyeragamkan kata-kata menjadi lower-case menggunakan operator *Transform Case* [10]. Proses ini dilakukan di dalam proses yang dinamakan Process Document from Data bersama dengan

tokenisasi, filtering stopwords dan kalimat yang tidak bermanfaat.

- d. **Filtering**
Filtering pada proses ini untuk menghilangkan stopwords dan kata yang tidak bermanfaat. Proses filter stopwords dilakukan dengan operator Filter Stopwords (dictionary) yang menggunakan stopwords dataset buatan Oswin Rahadiyan Hartono yang diperoleh dari situs kaggle.com [11]. Sedangkan kata yang tidak bermanfaat menggunakan operator Filter Tokens (by Length) dengan parameter minimal karakter 4 dan maksimal adalah 25.
- e. **Tokenisasi**
Tokenisasi memisahkan kata per kata dengan operator *tokenize* pada *text processing*. Tujuannya untuk dikelompokkan kemunculan suatu kata pada proses selanjutnya.
- f. **Word Weighting(TF-IDF)**
Word Weighting memberikan skor pada setiap kemunculan kata. Pada proses ini menggunakan operator Process Documents from Data yang di dalamnya menggunakan metode *frequency-inverse document frequency (TF-IDF)* [12]. Untuk menghubungkan data tabel dengan operator ini menggunakan operator Nominal to Text, karena data pada tabel bersifat polinomial sedangkan Process Documents from Data harus bersifat teks.
- g. **Imbalance Data (SMOTE Up Sampling)**

Untuk menyeimbangkan data sentimen yang tidak seimbang, maka menggunakan operator SMOTE untuk membuat data sintesis. Operator ini otomatis mendeteksi kelas mana yang menjadi kelas minoritas, kemudian dilakukan sampling untuk membuat random data dari kelas tersebut hingga memiliki data yang seimbang. Jika kelas minoritas memiliki jumlah lebih dari satu bisa menggunakan dua operator SMOTE up sampling.

Proses Klasifikasi Ulasan

Proses klasifikasi teks dengan metode Naive Bayes dan Decision Tree dari hasil pre-processing yang menghasilkan dataset, dilakukan untuk mendapatkan hasil kategori, apakah itu ulasan yang bersifat positif, negatif atau netral. Untuk mendapatkan hasil tersebut penulis menggunakan aplikasi bantuan Rapidminer dalam menghitung dan menggabungkan hasil klasifikasi yang didapatkan.

a. Labeling Dataset

Data yang sudah melalui pre-processing akan diberikan label secara manual oleh penulis dan dibantu oleh teman penulis. Label yang diberikan dipertimbangkan berdasarkan rating dan isi ulasan. Ulasan dengan rating lebih atau sama dengan empat akan dikategorikan positif, kemudian rating tiga akan dikategorikan netral, dan rating kurang dari dua akan dikategorikan negatif. Selanjutnya pada bagian ulasan, setiap kalimat yang mengandung kata-kata positif akan diberikan poin 1, kemudian setiap kalimat yang mengandung kata-kata negatif akan diberikan poin 0, dan kalimat yang mengandung kata-kata negatif akan diberikan -1. Jumlah poin rating dijumlahkan dengan poin ulasan akan diakumulasikan dan akan menjadi nilai yang menentukan label sentimen.

b. Implementasi Model

Implementasi model yaitu memasukan data uji ke dalam model mesin untuk dipelajari menggunakan *Cross Validation*. Data akan diuji yang sudah dilatih untuk menentukan sentimen keseluruhan data. Dalam hal ini adalah metode Naive Bayes, Decision Tree, dan k-NN.

Skenario Pengujian

Berdasarkan rencana pengujian yang dijelaskan, maka dilakukan skenario pengujian. Tujuannya adalah untuk mengetahui hasil pengujian metode yang diharapkan. Pada skenario kali ini menggunakan skenario *Confusion Matrix* untuk mendapatkan performance *accuracy*, *recall*, dan *precision* [13]. Berikut adalah penjelasan untuk masing-masing skenario.

a. Skenario Pengujian Klasifikasi Sentimen

Berdasarkan rencana pengujian yang dijelaskan. Tujuan dilakukannya pengujian ini adalah untuk mengetahui hasil pengujian yang diharapkan. Pada skenario kali ini dilakukan dengan perhitungan dari aplikasi RapidMiner yang dilakukan pada masing-masing metode.

b. Skenario Pengujian Performance

Skenario pengujian performance dilakukan untuk mengetahui tingkat *accuracy*, *recall* dan *precision* dari penggunaan metode, yang saat ini digunakan adalah Naive Bayes, Decision Tree dan k-NN. Pengukuran kinerja sistem menggunakan Confusion Matrix. Data ulasan villa di Ubud pada Google Maps diuji menggunakan pembagian 70% data latih dan 30% data uji. Dari Confusion matrix binary diperoleh nilai akurasi, presisi dan *recall*. Ukuran penilaian biasanya diberikan dengan persentasi dari 1 sampai 100 persen. Sebuah sistem bisa dianggap baik jika nilai akurasi, *precision*, dan *recall* memiliki tingkat persentasi tinggi. Nilai akurasi merupakan penilaian seberapa akurat pengklasifikasian data pada suatu model algoritma yang digunakan. Akurasi bisa didapat dari perbandingan antara data positif dan negatif yang terklasifikasi benar dengan keseluruhan data.

$$\text{akurasi} = \frac{TP}{\text{jumlah data}} \times 100 \% \quad (1)$$

Precision adalah kecocokan antara bagian data yang diambil dengan informasi yang dibutuhkan. Atau *precision* adalah ukuran kinerja seberapa banyak hasil relevan dengan informasi yang dicari. Nilai presisi didapat dari nilai jumlah data positif yang diklasifikasikan benar dibagi dengan total data yang diklasifikasi positif.

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \% \quad (2)$$

Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Nilai recall merupakan ukuran kinerja berupa nilai persentase data kategori positif yang diklasifikasikan negatif.

$$Recall = \frac{TP}{TP+FN} \times 100 \% \quad (3)$$

Naive Bayes Classifier

Naive Bayes Classifier merupakan metode pengklasifikasian yang berakar pada Teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris, Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Kegunaan metode Naive Bayes adalah untuk klasifikasi dokumen teks, metode *machine learning* yang menghasilkan probabilitas, untuk

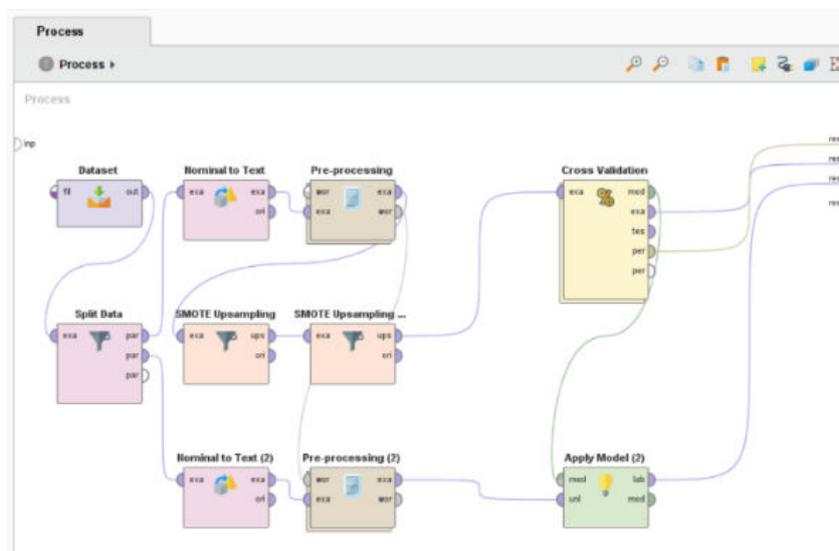
diagnosa medis secara otomatis, mendeteksi *spam* atau *filtering spam* [14].

Analisis Sentimen

Analisis sentimen merupakan sebuah proses memahami, mengekstraksi, dan mengolah data teks untuk menemukan jenis sentimen yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. Analisis sentimen bisa digunakan untuk memudahkan pengguna pada proses sentimen/ pandangan hingga penentuan keputusan pada suatu objek [15]. Analisis sentimen biasa digunakan untuk menilai pandangan suka dan tidak-suka publik terhadap suatu barang atau jasa.

HASIL DAN PEMBAHASAN

Berdasarkan tabel dataset ulasan atau sentimen di atas, untuk melakukan pengujian analisa sentimen dari masing-masing metode akan diimplementasikan ke dalam skema RapidMiner dengan skema sebagai berikut.



Gambar 2. Skema Rapidminer

- a. Input data
Pada proses input data, yang dilakukan adalah memasukkan data dan memilih kolom

mana saja yang akan digunakan. Kolom yang digunakan adalah kolom *rating*, *review*, dan sentimen.

Format your columns.

Replace errors with missing values

	Nama polynominal	Role polynominal	Rating integer	Waktu polynominal	Review polynominal	Sentimen polynominal/label
1	BAJU SENILAU	Local Guide	15	?	Akses jalan susah	negatif
2	Ludewico Marchetto	Local Guide	15	3 tahun lalu	Aku tinggal di ini dari 1...	negatif
3	Dian Trihasanti	Local Guide	15	sekitaran lalu	bad experience and ba...	negatif
4	Nisa Indra	Local Guide	15	2 tahun lalu	Bagus dan bersih Saya...	negatif
5	EIPhena Tanjung	Local Guide	15	sekitaran lalu	Bagus sech Tp Ada Be...	negatif
6	Ahmad Dermachikah	Local Guide	15	4 bulan lalu	Banyak kebisingan di ...	negatif
7	Jauhan Khoirul Kowad	?	15	2 tahun lalu	Banyak kutu di kasumy...	negatif
8	bagus sigit	Local Guide	15	?	Baru pertama kali niem...	negatif
9	nahq de	Local Guide	15	2 tahun lalu	Berikut adalah seperan...	negatif
10	CC	Local Guide	15	?	Berjalanlah 10 menit di...	negatif
11	CC	Local Guide	15	?	Berjalanlah 10 menit di...	negatif
12	Jore Carol	Local Guide	15	3 tahun lalu	breakfastnya parah	negatif
13	Meghna Jain	Local Guide	15	2 tahun lalu	Bukan tempat yang bag...	negatif
14	Kebut Laani	Local Guide	15	4 tahun lalu	Dilihat dr segi spiritual...	negatif
15	Ladex jul	Local Guide	15	?	Diluar ekspektasi	negatif
16	Nick	Local Guide	15	3 tahun lalu	Fotofoto itu membuat k...	negatif
17	Leigh Downing	Local Guide	15	2 tahun lalu	handuk bau tidak ada b...	negatif
18	Francesco F De Troya	Local Guide	15	3 bulan lalu	Hathati dengan tempat...	negatif
19	Laura Pardo	Local Guide	15	7 tahun lalu	Hotel ini rugin rantik N...	netanar

no problems.

Previous Finish Cancel

Gambar 3. Input Data Ulasan

b. *Split* data

Data yang sudah diinputkan kemudian akan dibagi menjadi data training dan data uji menggunakan operator *Split* Data. Di dalam operator kita memberikan pembagian data 70:30. Data yang dibagi akan membuat himpunan data yang kemudian didistribusikan dengan proporsi yang tepat (*stratified sampling*) dari label kelas yang ditentukan.

c. *Pre-processing*

Pada *pre-processing* data melalui tranformasi data agar bisa diolah oleh algoritma nantinya. Pertama mengubah seluruh kalimat menjadi lower case, membagi kata per kata dengan tokenize, dan memfilter kata-kata dengan filter tokens.



Gambar 4. Skema *Pre-processing*

Selanjutnya memetakan kata-kata dengan TF-IDF. Hasil yang didapat setelah mengeksekusi

proses *pre-processing* dengan TF-IDF adalah sebagai berikut.

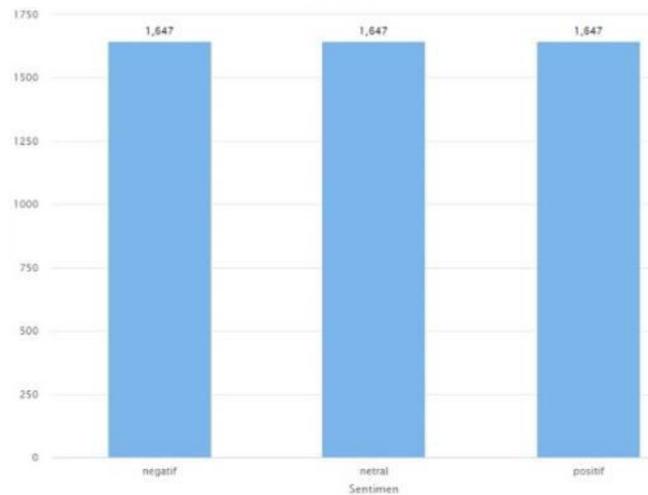
Sentimen	text	adalah	adanya	adapun	adat	adem	adil	adivana	admin
negatif	akses jalan s...	0	0	0	0	0	0	0	0
negatif	tinggal dari hi...	0	0	0	0	0	0	0	0
negatif	bagus bersih...	0	0	0	0	0	0	0	0
negatif	banyak kebisi...	0.145	0	0	0	0	0	0	0
negatif	berjalanlah ...	0	0	0	0	0	0	0	0
negatif	breakfastnya ...	0	0	0	0	0	0	0	0
negatif	hotel memilik...	0	0	0	0	0	0	0	0
negatif	hotel tidak ra...	0	0	0	0	0	0	0	0
negatif	tidak hubung...	0	0	0	0	0	0	0	0
negatif	kalo tamu ch...	0	0	0	0	0	0	0	0
negatif	kamar delux...	0	0	0	0	0	0	0	0
negatif	kamar tidak c...	0	0	0	0	0	0	0	0
negatif	kami memes...	0	0	0	0	0	0	0	0
negatif	kami tiba res...	0.072	0	0	0	0	0	0	0
negatif	kerusakan lis...	0	0	0	0	0	0	0	0
negatif	lariyah hotel s...	0	0	0	0	0	0	0	0
negatif	layanan pela...	0	0	0	0	0	0	0	0
negatif	layanan yang ...	0	0	0	0	0	0	0	0

Gambar 5. Hasil TF-IDF

d. SMOTE *up-sampling*

Pada proses *up-sampling* ini kelas data minor kelas netral dan kelas negatif dibuatkan data sintesis agar bisa menyamai jumlah data di

kelas mayoritas yaitu kelas positif. Setelah dieksekusi data yang sebelumnya tidak seimbang sekarang menjadi setara jumlah data 1647.



Gambar 6. Hasil data training dengan SMOTE

Pengujian dan Perbandingan

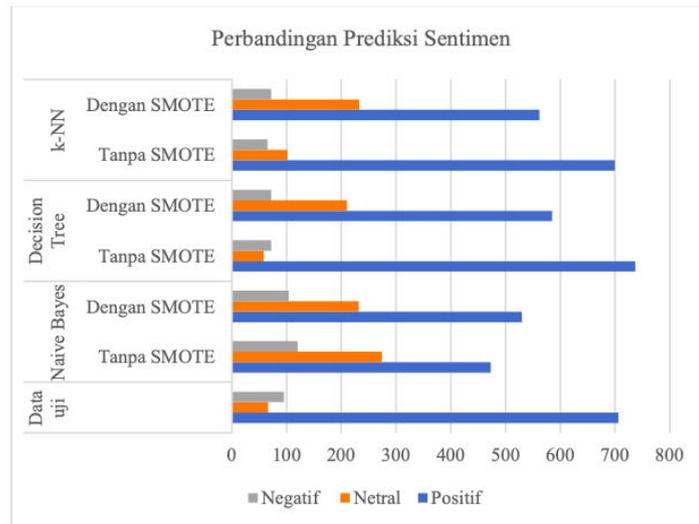
Pada tahap pengujian dan perbandingan penulis memaparkan hasil prediksi sentimen dari masing-masing metode, hasil performance dengan metode *Confusion Matrix*, dan perbandingannya.

a. Perbandingan Prediksi sentimen

Pada prediksi sentimen membandingkan ketiga metode yaitu metode Naive Bayes, Decision Tree, dan k-NN yang menggunakan dan tanpa menggunakan SMOTE *up-sampling*. Berikut adalah perbandingan dari ketiga metode.

Tabel 1. Perbandingan prediksi

Data Uji	Metode	Naive Bayes	Decision Tree	k-NN
pred. Positif	Tanpa SMOTE	473	737	700
	Dengan SMOTE	530	585	562
pred. Netral	Tanpa SMOTE	274	58	102
	Dengan SMOTE	232	102	233
pred. Negatif	Tanpa SMOTE	120	72	65
	Dengan SMOTE	104	65	72



Gambar 7. Grafik Perbandingan Prediksi Setiap Metode

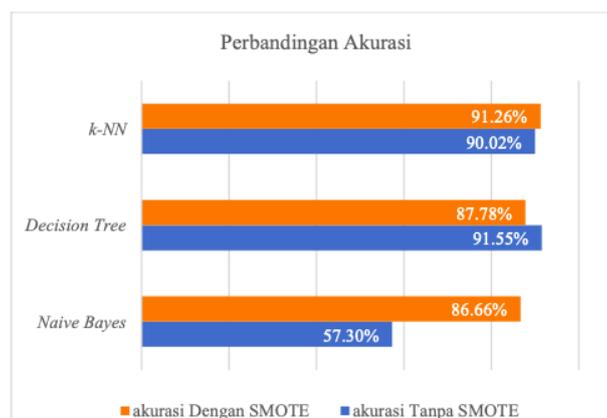
b. Perbandingan akurasi

Pada hasil yang disajikan diketahui bahwa tanpa SMOTE up-sampling metode Decision Tree memiliki akurasi yang paling tinggi, kemudian metode k-NN nomor dua dan metode Naive Bayes dengan akurasi yang paling rendah. Akan tetapi setelah menggunakan SMOTE up-sampling metode k-NN memiliki

persentase akurasi yang paling tinggi, lalu Decision Tree kedua dan Naive Bayes paling rendah. Kemudian diantara ketiga metode tersebut hanya metode Decision Tree yang mengalami penurunan akurasi dibanding dua metode lain yang mengalami peningkatan akurasi.

Tabel 2. Perbandingan Akurasi

Metode	Hasil Akurasi	
	Tanpa SMOTE	Dengan SMOTE
Naive Bayes	57.30%	86.66%
Decision Tree	91.55%	87.78%
k-NN	90.02%	91.26%



Gambar 8. Grafik Perbandingan Akurasi Setiap Metode

c. Perbandingan Precision

Hasil perhitungan confusion matrix untuk precision tanpa menggunakan SMOTE up-

sampling menunjukkan bahwa pred. Positif metode k-NN lebih unggul dengan persentase 94.96%. Lalu pada pred. Netral metode

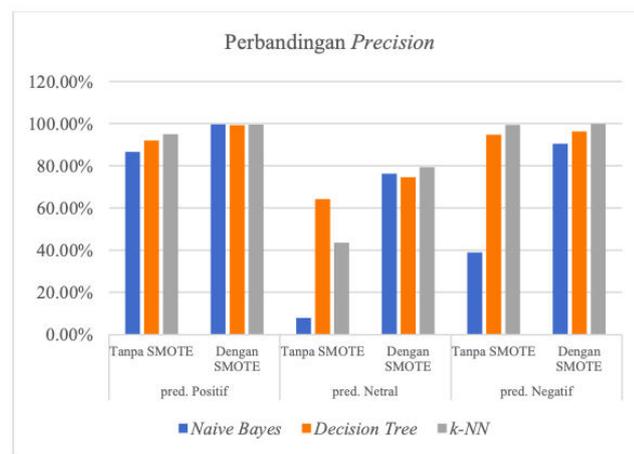
Decision Tree yang lebih unggul dengan persentase 64.29%. Kemudian untuk pred. Negatif metode yang lebih unggul adalah k-NN dengan persentase 99.41%. Bisa diketahui jika persentase yang dihasilkan setiap metode perbandingannya cukup jauh atau tidak setara.

Sedangkan setelah menggunakan SMOTE up-sampling menunjukkan bahwa precision untuk pred. Positif yang lebih unggul adalah

metode Naive Bayes dengan persentase 99.63% akan tetapi tidak jauh berbeda dengan persentase metode lain yang sama-sama mendapat 99%. Lalu untuk pred. Netral yang lebih unggul adalah metode k-NN dengan persentase 79.43%. Dan untuk pred. Negatif yang lebih unggul adalah metode k-NN dengan persentase 99.93%.

Tabel 3. Perbandingan *Precision*

Metode	Naive Bayes	Decision Tree	k-NN	
pred. Positif	Tanpa SMOTE	86.63%	92.07%	94.96%
	Dengan SMOTE	99.63%	99.29%	99.55%
pred. Netral	Tanpa SMOTE	7.87%	64.29%	43.54%
	Dengan SMOTE	76.30%	74.53%	79.43%
pred. Negatif	Tanpa SMOTE	38.94%	94.76%	99.41%
	Dengan SMOTE	90.50%	96.32%	99.93%



Gambar 9. Perbandingan *Precision* Setiap Metode

d. Perbandingan *Recall*

Hasil tanpa menggunakan SMOTE up-sampling menunjukkan kelas positif metode yang lebih unggul adalah Decision Tree dengan persentase 99.33%. Lalu pada kelas netral yang lebih unggul adalah metode k-NN dengan persentase 58.71%. Kemudian pada kelas negatif yang lebih unggul adalah metode Decision Tree. Pada bagian kelas netral ketiga metode mendapat hasil yang sangat rendah.

Sedangkan setelah menggunakan SMOTE up-sampling menunjukkan bahwa nilai *Recall* untuk kelas positif yang lebih unggul adalah Decision Tree dengan persentase 84.34%. Lalu

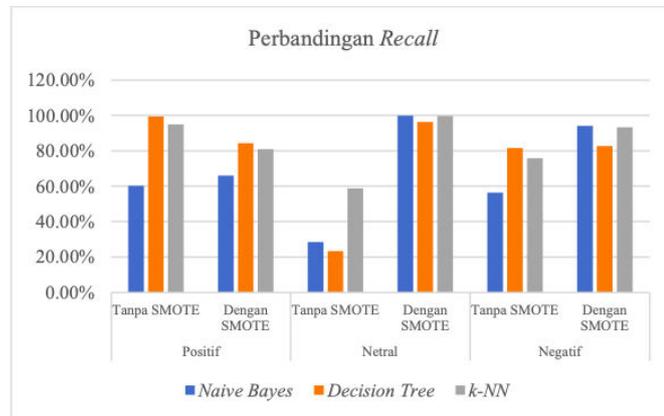
pada kelas netral yang lebih unggul adalah metode Naive Bayes dengan persentase 99.88%. Kemudian pada kelas negatif yang lebih unggul adalah metode Naive Bayes dengan persentase 94.11%.

Dari ketiga kelas yang dibandingkan ketika menggunakan dan tidak menggunakan SMOTE up-sampling diketahui bahwa hanya kelas positif yang mengalami penurunan persentase. Sedangkan untuk kedua kelas yang lainnya mengalami kenaikan persentase yang berarti penggunaan up-sampling ini memberikan manfaat yang baik untuk hasil *Recall*.

Tabel 5. Perbandingan *Recall*

Metode	Naive Bayes	Decision Tree	k-NN	
Positif	Tanpa SMOTE	60.17%	99.33%	94.90%
	Dengan SMOTE	66.00%	84.34%	80.94%

Netral	Tanpa SMOTE	28.39%	23.23%	58.71%
	Dengan SMOTE	99.88%	96.30%	99.64%
Negatif	Tanpa SMOTE	56.31%	81.53%	75.68%
	Dengan SMOTE	94.11%	82.70%	93.20%



Gambar 10. Perbandingan Recall Setiap Metode

e. Perbandingan Keseluruhan

Perbandingan keseluruhan merupakan nilai overall dari masing-masing aspek yang diperoleh pada setiap metode dengan cara menghitung persentase kelas positif, netral, dan negatif kemudian dibagi tiga. Kemudian untuk mendapatkan hasil overall performance diperoleh dengan cara menjumlahkan persentase akurasi, overall precision dan overall recall lalu dibagi tiga. Dari keseluruhan perbandingan performance menggunakan

Confusion Matrix tanpa SMOTE up-sampling dari ketiga metode yang digunakan bisa kita ketahui bahwa metode k-NN menjadi paling unggul dengan persentase overall performance 81.92%. Hasil ini tidak terpaut jauh dengan metode Decision Tree yang mendapat *overall performance* 81.10%. Kemudian metode Naive Bayes mendapat *overall performance* terendah dengan persentase 50.02%.

Tabel 6. Perbandingan Keseluruhan

Metode		Naive Bayes	Decision Tree	k-NN
akurasi	Tanpa SMOTE	57.30%	91.55%	90.02%
	Dengan SMOTE	86.66%	87.78%	91.26%
Overall Precision	Tanpa SMOTE	44.48%	83.71%	79.30%
	Dengan SMOTE	88.81%	90.05%	92.97%
overall recall	Tanpa SMOTE	47.99%	59.78%	60.43%
	Dengan SMOTE	86.66%	87.78%	91.26%
Overall performance	Tanpa SMOTE	50.02%	81.10%	81.92%
	Dengan SMOTE	87.38%	88.54%	91.83%

Sedangkan perbandingan performance setelah menggunakan SMOTE *up-sampling* menunjukkan bahwa metode k-NN tetap lebih unggul diantara ketiga metode dengan persentase 91.83%. Walaupun tetap lebih tinggi, persentasenya mengalami kenaikan yang menunjukkan bahwa setelah

menggunakan SMOTE *up-sampling* kualitas hasil yang didapatkan menjadi lebih baik.

KESIMPULAN

Berdasarkan hasil penelitian ini, metode Naive Bayes, Decision Tree, dan k-NN berhasil atau mampu dalam menganalisis

sentimen villa di Ubud dengan baik. Penggunaan SMOTE up-sampling terbukti bisa meningkatkan hasil performance dari ketiga metode yaitu Naive Bayes, Decision Tree, dan k-NN. Metode Naive Bayes setelah menggunakan SMOTE *up-sampling* mendapat prediksi sentimen 530 positif, 232 netral, dan 104 negatif. Performance confusion matrix yang diperoleh yaitu, akurasi 86.66%, precision 88.81%, recall 86.66%, dan overall performance 87.38%. Metode Decision Tree setelah menggunakan SMOTE up-sampling mendapat prediksi sentimen 585 positif, 210 netral, dan 72 negatif. Performance confusion matrix yang diperoleh yaitu, akurasi 87.78%, precision 90.05%, recall 87.78%, dan overall performance 88.54%. Metode k-NN setelah menggunakan SMOTE up-sampling mendapat prediksi sentimen 562 positif, 233 netral, dan 72 negatif. Performance confusion matrix yang diperoleh yaitu, akurasi 91.26%, precision 92.97%, recall 91.26%, dan overall performance 91.83%. Dari ketiga metode yang digunakan untuk analisis sentimen villa di Kecamatan Ubud, diperoleh bahwa metode k-NN lebih baik dan optimal dibanding metode Naive Bayes dan Decision Tree. Nilai akurasi pada k-NN 91.26% menunjukkan *excellent classification*.

REFERENSI

- [1] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [2] T. Hendrawati and C. P. Yanti, "Analysis of Twitter Users Sentiment against the Covid-19 Outbreak Using the Backpropagation Method with Adam Optimization," *Journal of Electrical, Electronics and Informatics*, vol. 5, no. 1, 2021.
- [3] O. Somantri and Dairoh, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," *Jurnal Edukasi dan Penelitian Informatika*, vol. 5, no. 2, 2019, [Online]. Available: www.google.com/maps
- [4] A. Rifa, H. Sujaini, D. Prawira, and J. H. Hadari Nawawi, "Sentiment Analysis Objek Wisata Kalimantan Barat Pada Google Maps Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika*, vol. 7, no. 3, 2021.
- [5] A. Thalia and S. Nugroho, "Strategi Pengembangan Desa Sayan Ubud, Kabupaten Gianyar, Bali, Sebagai Desa Wisata Berbasis Wisata Alam Bija," *Jurnal Destinasi Pariwisata*, vol. 7, no. 2, 2019.
- [6] D. Retno Utari and A. Wibowo, *Pemanfaatan Google Maps dalam Pembuatan Aplikasi Pemantau Kondisi Jalan dan Lalu lintas*. [Online]. Available: <http://maps.google.com>
- [7] Sudiantoro et al., "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier," *Din. Inform.*, vol. 10, no. 2, pp. 398–401, 2018.
- [8] H. S. Utama, D. Rosiyadi, D. Aridarma, and B. S. Prakoso, "Sentimen Analisis Kebijakan Ganjil Genap Di Tol Bekasi Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 247–254, Sep. 2019, doi: 10.33480/pilar.v15i2.705.
- [9] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote Dan K-Nearest Neighbor," 2018.
- [10] A. H. Tri Jaka, "Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining."
- [11] Oswin and R. Hartono, "Indonesian Stoplist Most Common Words (Stop Words) in Bahasa Indonesia." <https://www.kaggle.com/oswinrh/indonesian-stoplist>
- [12] A. T. Jaka, "Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining," *Inform. UPGRIS*, vol. 1, pp. 1–9, 2015.
- [13] I. Widhi Saputro and B. Wulan Sari, "Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa Naive Bayes Algorithm Performance Test for Student Study Prediction," *Citec Journal*, vol. 6, no. 1, 2019.
- [14] Widiyanto and M. Haldi, "Algoritma Naive Bayes," 2019. <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/>
- [15] Cahyani, Riza, and P. Pandu Adikara, "Analisis Sentimen terhadap Ulasan Hotel menggunakan Boosting Weighted Extreme Learning Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 8, pp. 2548–964, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>

