

# Tourist Perceptions Through Sentiment Analysis to Support Tourism Development in Maluku Province

Hennie Tuhuteru <sup>1\*</sup>, Leonardo Petra Refialy <sup>2\*</sup>, Marlisa Laturake <sup>3\*</sup>, Shyrel Gildion Pattirane <sup>4\*</sup>

\* Informatika, Universitas Kristen Indonesia Maluku

[hannytuhuteru@gmail.com](mailto:hannytuhuteru@gmail.com) <sup>1</sup>, [leo.refialy@gmail.com](mailto:leo.refialy@gmail.com) <sup>2</sup>, [marlisalaturake@gmail.com](mailto:marlisalaturake@gmail.com) <sup>3</sup>, [sgpattirane@gmail.com](mailto:sgpattirane@gmail.com) <sup>4</sup>

## Article Info

### Article history:

Received 2023-12-28

Revised 2024-06-25

Accepted 2024-06-26

### Keyword:

Sentiment analysis,  
Tourist perceptions,  
Social media data,  
Maluku Province.

## ABSTRACT

Tourist perceptions obtained through sentiment analysis can provide an overview of tourism development in Maluku Province. This study aims to determine tourists' perceptions of destinations in Maluku based on the results of sentiment analysis. The research uses a quantitative approach by analyzing data scraped and snipped from Facebook, Instagram, TikTok, Google Maps Reviews, and Trip Advisor. Sentiment analysis is conducted by comparing the accuracy levels of the Random Forest, Naïve Bayes, and Support Vector Machine classification models. The comparison results of the three methods show that Random Forest has the highest accuracy rate at 85%. Sentiment analysis results for the entire dataset and analysis per district/city indicate that tourists' perceptions of tourist destinations in Maluku are generally positive, despite being dominated by negative sentiments. The presence of negative and neutral sentiments highlights the need for improvements in the quality of tourist services, particularly in terms of human resources, transportation, accommodation, and infrastructure facilities.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Maluku sebagai daerah kepulauan dan bersejarah memiliki potensi destinasi wisata yang cukup besar. Data [1] dengan keyword “Kepulauan Maluku” menunjukkan popularitas Maluku di seluruh dunia melalui penelusuran Web dengan kategori Pariwisata selama 5 tahun terakhir. Hal ini menunjukkan bahwa pariwisata di Maluku sudah dikenal secara luas bukan hanya di Indonesia, tetapi juga di belahan dunia lain.

Hal ini menjadi perhatian Pemerintah Daerah (PEMDA) melalui penetapan sektor Pariwisata sebagai salah satu sektor unggulan prioritas pada Rencana Pembangunan Jangka Menengah Daerah (RPJMD) Tahun 2019-2024 [2]. Pengembangan pariwisata diprioritaskan karena dapat meningkatkan pertumbuhan ekonomi masyarakat melalui kunjungan wisatawan, terlebih untuk daerah kepulauan karena berhubungan langsung dengan sektor lain seperti transportasi, akomodasi dan ritel [3], [4]. Berdasarkan data [5]–[8], terjadi penurunan jumlah wisatawan yang signifikan

di Maluku saat pandemi Covid-19. Kondisi yang telah membaik namun masih dalam status kedaruratan [9], [10] saat ini turut mempengaruhi adanya peningkatan jumlah wisatawan. Khusus untuk wisatawan mancanegara, terjadi peningkatan yang cukup signifikan pada tahun 2022 dibandingkan 2 tahun sebelumnya, sedangkan untuk wisatawan lokal terjadi penurunan yang cukup signifikan pada tahun 2022 dibandingkan tahun sebelumnya.

Sementara itu, era digital saat ini menjadikan pemanfaatan media sosial (medsos) sebagai salah satu kunci utama dalam pengembangan pariwisata. Medsos seperti Instagram, Facebook, TikTok, YouTube termasuk juga Google Maps Review serta platform khusus untuk wisata seperti TripAdvisor memberikan kemudahan dalam membagikan pengalaman wisatanya atau mencari referensi tentang destinasi wisata sebelum melakukan perjalanan [11]–[14]. Hal tersebut memungkinkan wisatawan membagi pengalaman mereka melalui unggahan pada medsos pribadinya ataupun memberikan ulasan dan/atau rating yang sudah disediakan pada aplikasi. Unggahan dan ulasan tersebut

dapat menjadi bahan pertimbangan dan mempengaruhi keputusan perjalanan pengguna lain yang membacanya [11].

Survei [15] menunjukkan pengguna aktif medsos kini telah mencapai 59,4% sedangkan pengguna internet mencapai 64,4% dari total populasi dunia. Penggunaan yang begitu masif tentu memberikan dampak pada jumlah data unggahan dan ulasan. Data yang besar dan tersedia secara publik tersebut dapat memberikan informasi atau pengetahuan baru, seperti analisis sentimen [16]. Analisis big data terhadap sentimen melalui umpan balik wisatawan menjadi penting karena mereka berharap menikmati liburan atau kunjungannya [13], [17]. Analisis sentimen dapat membantu dalam perencanaan, pengambilan keputusan dan strategi pemasaran bagi para pemangku kepentingan atau pembuat keputusan, termasuk dalam hal ini PEMDA [13]. Analisis sentimen ini *urgent* dilakukan dengan akurasi yang lebih baik agar dapat memperoleh pandangan yang berbeda terkait perkembangan pariwisata di Maluku berdasarkan sentimen wisatawan.

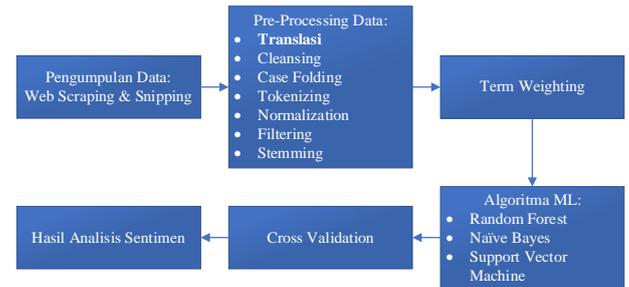
Berdasarkan permasalahan tersebut, maka penelitian ini dilakukan dengan judul “Persepsi Wisatawan Melalui Analisis Sentimen Untuk Mendukung Pengembangan Pariwisata di Provinsi Maluku”. Penelitian ini bertujuan untuk mengetahui sentimen wisatawan terhadap destinasi wisata yang ada di Maluku berdasarkan perbandingan 3 algoritma Machine Learning (ML) dengan tingkat akurasi masing-masing. Hal ini juga dapat dijadikan bahan pertimbangan untuk pengambilan keputusan dalam pengembangan pariwisata sebagai salah satu sektor unggulan di Provinsi Maluku yang berkelanjutan. [18] membandingkan performa analisis sentimen dengan 3 bahasa, yaitu bahasa Turki, bahasa Inggris dan bahasa Arab. Hasil eksperimennya menunjukkan model pre-trained bahasa memiliki kinerja yang baik dalam 3 kelas sentimen. [19] pada penelitiannya tentang analisis sentimen multilingual menjelaskan berbagai teknik untuk meningkatkan penyisipan kata untuk low-resource language dan berfokus pada analisis sentimen multibahasa berbasis deep-learning. Sementara itu, [20] meneliti efek dari kombinasi bahasa Inggris dan bahasa Indonesia untuk dalam membangun teks klasifikasi bahasa Indonesia. Hasil penelitiannya menunjukkan ada peningkatan performa dari kombinasi 2 bahasa tersebut, terutama jika jumlah data bahasa Indonesia lebih sedikit.

Penelitian ini berbeda dari penelitian sebelumnya pada tahap *pre-processing data*. Pada tahap ini, akan ditambahkan fitur translasi seperti yang terlihat pada Gambar 1. Fitur ini bertujuan untuk menerjemahkan teks dari berbagai bahasa ke Bahasa Indonesia menggunakan bahasa pemrograman Python agar seragam dan memudahkan proses analisis sentimen. Selain itu, hingga saat ini belum ada penelitian terkait analisis sentimen terhadap pariwisata di Maluku secara keseluruhan dengan akurasi di atas 60%. Sehingga ini merupakan analisis sentimen pertama yang berkaitan dengan pariwisata di Provinsi Maluku.

## II. METODE

Penelitian ini dilakukan dengan pendekatan kuantitatif memanfaatkan Big Data Analytics untuk sekumpulan data dari berbagai medsos populer seperti Facebook, Instagram, TikTok, Google Maps Review serta seperti TripAdvisor. Kemudian dilakukan pendekatan Data Mining (DM), khususnya text mining dengan memanfaatkan algoritma ML dan teknik Natural Language Processing (NLP) [16], [21]–[24].

Proses Analisis Sentimen pada penelitian ini dilakukan melalui tahapan-tahapan yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Analisis Sentimen

### A. Pengumpulan Data

Pengumpulan data terdiri dari data primer dan data sekunder. Data berkaitan dengan destinasi pariwisata di Maluku dan destinasi super prioritas yang ditetapkan oleh Pemerintah Provinsi Maluku dan Pemerintah masing-masing kabupaten/kota di Maluku melalui penelitian lapangan di Dinas Pariwisata Provinsi Maluku. Penarikan sampel digunakan teknik *purposive sampling*, dengan mengambil 3 destinasi prioritas per kabupaten/kota di Maluku, termasuk destinasi super prioritas Maluku. Sementara itu, data terkait unggahan dan ulasan yang diperoleh dengan teknik *web scraping* dan teknik *snipping* atau pengumpulan data secara langsung. Hal ini perlu dilakukan karena masing-masing platform memiliki aturan kebijakan dan privasi yang berbeda-beda.

### B. Pre-Processing Data

Pre-processing data perlu dilakukan [25]–[27] karena data yang diperoleh melalui komentar dan ulasan memiliki bentuk yang tidak terstruktur sehingga perlu diseragamkan. Pada tahap pertama, dilakukan *translation* dataset yang dikumpulkan ke bahasa Inggris, *cleansing* karakter non-alfabetis, *case-folding* atau mengubah teks menjadi huruf kecil, *Tokenizing* untuk pemisahan kata, *Normalization* untuk menormalkan setiap kata, *Filtering* untuk menghapus kata-kata yang kurang relevan, dan *Stemming* untuk mengubah kata berimbuhan menjadi kata dasar. Kami menambahkan fitur terjemahan menggunakan *library* googletrans sehingga semua teks pada dataset akan diterjemahkan ke bahasa Inggris untuk mempermudah tahapan selanjutnya. Hal ini perlu dilakukan karena komentar atau ulasan yang diperoleh menggunakan berbagai macam bahasa sehingga perlu untuk diseragamkan. Setelah itu, dilakukan pembobotan setiap kata

sebelum dilakukan klasifikasi dengan tiga algoritma ML. Kemudian dilakukan validasi menggunakan *k-fold cross validation* untuk memperoleh akurasi yang baik dari masing-masing algoritma tersebut. Hasil analisis sentimen akan diambil berdasarkan akurasi yang paling tinggi dari masing-masing algoritma ML. Selanjutnya, interpretasi data dilakukan berdasarkan hasil analisis yang diperoleh baik secara keseluruhan maupun klasifikasi sentimen per Kabupaten/Kota menggunakan algoritma yang memiliki akurasi paling tinggi.

### C. Term Weighting

Term Weighting atau pembobotan kalimat pada penelitian ini menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Pembobotan kalimat diperlukan untuk mengukur tingkat kepentingan suatu kata (term) dalam sebuah dokumen [28]. TF-IDF dirumuskan sebagai berikut [26], [28]:

$$TF - IDF_{t,d} = TF_{t,d} \times IDF_t$$

Dimana,  $IDF_t = \log \frac{N}{DF_t}$

$t$  = kata-kata yang dihitung;  
 $d$  = bobot kalimat ( $d$ );  
 $TF - IDF_{t,d}$  = kalimat ( $d$ ) terhadap kata ( $t$ );  
 $TF_{t,d}$  = Term Frequency;  
 $IDF_t$  = Inverse Document Frequency;  
 $N$  = jumlah kalimat;  
 $DF_t$  = jumlah kata yang diulang;

### D. Random Forest

Random Forest, adalah algoritma machine learning yang terdiri dari gabungan beberapa algoritma pohon keputusan (Tree). Cara algoritma ini belajar adalah dengan menggunakan teknik pengumpulan data, di mana beberapa pohon keputusan dibangun secara independen dan hasilnya kemudian digabungkan untuk meningkatkan performa dan stabilitas model. Salah satu karakteristik utama dari Random Forest adalah pengenalan elemen keacakan dalam proses pembangunan setiap pohon, baik dalam pemilihan sampel data maupun pemilihan fitur. Dengan pendekatan ini, Random Forest dapat mengatasi masalah overfitting dan menghasilkan model yang lebih kuat dan mampu generalisasi dengan baik [29].

### E. Naïve Bayes

Algoritma Naive Bayes adalah metode klasifikasi berdasarkan teorema probabilitas Bayes dengan asumsi sederhana bahwa setiap fitur yang digunakan adalah independen satu sama lain. Meskipun asumsi ini naif, algoritma ini efektif dan cepat, terutama untuk dataset berdimensi tinggi. Naive Bayes menghitung probabilitas kelas target berdasarkan fitur-fitur yang ada dan memilih kelas dengan probabilitas tertinggi. Cocok untuk tugas

klasifikasi teks dan masalah klasifikasi lainnya, meskipun asumsi independensi tidak selalu terpenuhi [30].

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Secara sederhana, Teorema Bayes memungkinkan penentuan probabilitas dengan menghitung probabilitas terkait lainnya. Setelah diaplikasikan pada algoritma Naive Bayes, rumus ini menghasilkan asumsi dasar, yaitu algoritma selalu menganggap bahwa setiap fitur bersifat mandiri, setara, dan memberikan kontribusi pada hasil ketika memeriksa sebuah fitur.

### F. Support Vector Machine

Algoritma Support Vector Machine (SVM) adalah metode machine learning yang digunakan untuk menggambar garis pemisah terbaik antara dua kelas data. Garis ini disebut hyperplane, dan SVM bertujuan untuk memilih hyperplane dengan margin terbesar, yaitu jarak terdekat antara data dari kedua kelas ke garis tersebut. SVM efektif dalam menangani data berdimensi tinggi dan dapat menggunakan fungsi kernel untuk menangani data yang tidak linier [31].

### G. k-Fold Cross Validation

K-Fold Cross Validation adalah suatu teknik evaluasi kinerja model dalam machine learning. Dalam metode ini, dataset dibagi menjadi  $k$  subset (*k-fold*), dan model diuji  $k$  kali. Pada setiap iterasi, salah satu subset digunakan sebagai data pengujian (*validation set*), sementara  $k-1$  subset lainnya digunakan sebagai data pelatihan (*training set*). Proses ini diulang sebanyak  $k$  kali, sehingga setiap subset menjadi data pengujian satu kali. Dengan demikian, kita dapat mendapatkan perkiraan yang lebih konsisten terkait performa model, dan model tidak hanya bergantung pada satu pembagian khusus dari data [32]. Contoh umum untuk nilai  $k$  dalam *K-Fold Cross Validation* adalah 5 atau 10. Pendekatan ini berguna untuk mencegah terjadinya overfitting atau underfitting yang mungkin terjadi ketika hanya satu pembagian data digunakan untuk mengevaluasi model.

## III. HASIL DAN PEMBAHASAN

Sebelum melakukan serangkaian proses analisis terhadap dataset, terlebih dahulu dilakukan sedikit eksplorasi terhadap dataset yang digunakan dalam penelitian ini, untuk melihat bagaimana struktur sebaran dari dataset yang digunakan. Dari hasil deskripsi, total data tweet yang ada berjumlah 14.640 dengan jumlah atribut sebanyak 15 atribut. Data tersebut dibagi menjadi 6 maskapai yang masing-masing telah diberi label polaritas sentimen positif, negatif, dan netral. Berikut hasil penjabaran data tweet berdasarkan masing-masing maskapai.

Hasil pengumpulan data yang diperoleh melalui berbagai sumber di internet, yaitu sejumlah 11.752 komentar dan ulasan. Distribusi data komentar sebagaimana terlihat pada Tabel 1 menunjukkan bahwa Kabupaten Maluku Tengah,

Kota Ambon, dan Kabupaten Maluku Tenggara memiliki jumlah data terkumpul lebih banyak dari kabupaten lain di Maluku, yaitu secara berturut-turut 27,5%, 20%, dan 16,62%.

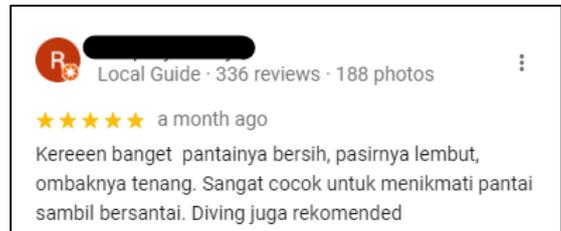
TABEL I  
JUMLAH DATA PER KABUPATEN/KOTA

No.	Kabupaten / Kota	Sampel (Lokasi Wisata)	Jumlah Data	Persentase (%)
1	Kabupaten Buru	Pantai Jikumerasa, Pantai Baikole, Danau Rana	530	4,51%
2	Kabupaten Buru Selatan	Teluk Tifu, Pantai Wamsoba, Pantai Masnana	365	3,11%
3	Kabupaten Kepulauan Aru	Karanguli, Lorang, Wokam	378	3,22%
4	Kabupaten Kepulauan Tanimbar	Pantai Matakus, Pantai Sembunyi, Pantai Weluan	634	5,39%
5	Kabupaten Maluku Barat Daya	Gunung Kerbau, Danau Tihu, Pantai Ilmarang	549	4,67%
6	Kabupaten Maluku Tengah	Pantai Ora, Banda Naira, Benteng Duurstede, Pantai Natsepa, Pantai Liang	3.232	27,50%
7	Kabupaten Maluku Tenggara	Pantai Ngurbloat, Pantai Ngurtafur, Pantai Ngursarnadan	1.953	16,62%
8	Kabupaten Seram Bagian Barat	Alang Asaude, Pulau Osi, Pantai Hatuhuran	739	6,29%
9	Kabupaten Seram Bagian Timur	Pulau Geser, Pantai Gumumae, Tanjung Sesar	491	4,18%
10	Kota Ambon	Pantai Pintu Kota, Gong Perdamaian Dunia, Taman Pattimura	2.350	20,00%
11	Kota Tual	Pulau Bair, Pantai Difur, Pulau Roa	531	4,52%
<b>TOTAL</b>			<b>11.752</b>	<b>100%</b>

Proses analisis sentimen dilakukan menggunakan bahasa pemrograman Python. Sementara itu, untuk pembagian data training dan data testing dilakukan menggunakan k-fold cross

validation dimana teknik ini memiliki akurasi yang lebih baik jika proses pembagian dilakukan secara manual atau menggunakan holdout method [33], [34].

Data yang dikumpulkan tentu memiliki karakter yang unik dan perlu untuk dilakukan pre-processing data. Hal ini diperlukan untuk mengubah data yang tidak terstruktur menjadi terstruktur agar dapat mempermudah proses analisis data. Sebagai contoh, salah satu komentar yang diperoleh melalui Google Map Review yang ditunjukkan Gambar 2.



Gambar 2. Ulasan Pengunjung pada Salah Satu Destinasi Wisata di Maluku

Ulasan tersebut kemudian dilakukan pre-processing data seperti yang terlihat pada Tabel 2.

TABEL II  
TAHAPAN PRE-PROCESSING DATA

Pre-Processing	Hasil
<b>Translation</b>	It's really cool, the beach is clean, the sand is soft, the waves are calm. Very suitable for enjoying the beach while relaxing. Diving is also recommended
<b>Cleansing</b>	Its really cool the beach is clean the sand is soft the waves are calm Very suitable for enjoying the beach while relaxing Diving is also recommended
<b>Case Folding</b>	its really cool the beach is clean the sand is soft the waves are calm very suitable for enjoying the beach while relaxing diving is also recommended
<b>Tokenizing</b>	['its', 'really', 'cool', 'the', 'beach', 'is', 'clean', 'the', 'sand', 'is', 'soft', ',', 'the', 'waves', 'are', 'calm', ',', 'very', 'suitable', 'for', 'enjoying', 'the', 'beach', 'while', 'relaxing', ',', 'diving', 'is', 'also', 'recommended']
<b>Normalization</b>	[[['it'], ['s'], ['really'], ['cool'], [','], ['the'], ['beach'], ['is'], ['clean'], [','], ['the'], ['sand'], ['is'], ['soft'], [','], ['the'], ['waves'], ['are'], ['calm'], [','], ['very'], ['suitable'], ['for'], ['enjoying'], ['the'], ['beach'], ['while'], ['relaxing'], [','], ['diving'], ['is'], ['also'], ['recommended']]]
<b>Filtering</b>	['really', 'cool', ',', 'beach', 'clean', ',', 'sand', 'soft', ',', 'waves', 'calm', ',', 'suitable', 'enjoying', 'beach', 'relaxing', ',', 'diving', 'also', 'recommended']
<b>Stemming</b>	['really', 'cool', ',', 'beach', 'clean', ',', 'sand', 'soft', ',', 'wave', 'calm', ',', 'suitabl', 'enjoy', 'beach', 'relax', ',', 'dive', 'also', 'recommend']

Selanjutnya pembobotan kata dilakukan terhadap dataset dengan menggunakan TF-IDF. Gambaran hasil pembobotan menggunakan teknik tersebut dapat dilihat pada Tabel 3.

TABEL III  
PEMBOBOTAN DENGAN TF-IDF PADA DATASET

Komentar / Ulasan	Kata					
	1	2	3	4	5	...
1	0.000	0.545	1.023	0.031	0.000	...
2	0.253	0.000	0.000	0.254	0.000	...
3	0.311	0.104	0.302	0.000	0.000	...
4	0.000	0.000	0.870	0.000	0.000	...
...	...	...	...	...	...	...
11.752	0.000	0.202	0.000	0.000	0.000	...

A. Hasil Klasifikasi Sentimen

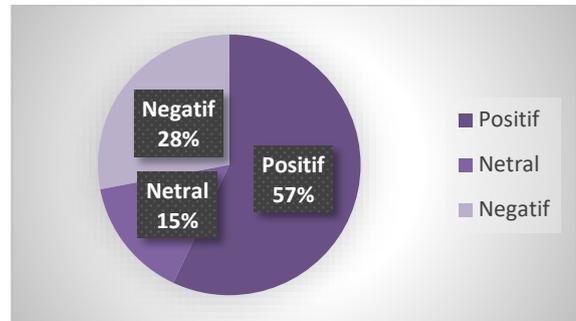
Proses pembentukan model klasifikasi terhadap dataset yang sudah melalui tahap pembobotan sebelumnya dilakukan dengan 3 algoritma Machine Learning, yaitu Random Forest (RF), Naïve Bayes Classifier (NBC) dan SVM. Hasil pengujian 3 model tersebut di atas menggunakan *k-fold cross validation* dapat dilihat pada Tabel 4. Nilai k akan menentukan berapa kali proses *training* dan validasi dilakukan. Pengujian dilakukan dengan parameter akurasi, dimana nilai yang paling merupakan tingkat akurasi yang terbaik. Akurasi sudah tentu sangat berpengaruh terhadap performance model yang dihasilkan.

TABEL IV  
VALIDASI MODEL RANDOM FOREST

k-Fold	Accuracy		
	Random Forest	Naïve Bayes	SVM
2	0.82	0.77	<b>0.83</b>
3	0.83	0.78	0.81
4	0.84	<b>0.79</b>	0.79
5	<b>0.85</b>	0.77	0.80
6	0.83	0.78	0.78
7	0.83	0.75	0.77
8	0.84	0.73	0.77
9	0.82	0.72	0.76
10	0.82	0.68	0.72

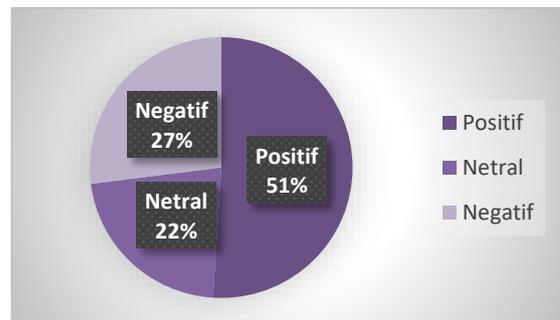
Berdasarkan hasil yang diperoleh pada Tabel 4, dapat dilihat bahwa masing-masing model memiliki tingkat akurasi yang berbeda berdasarkan nilai k. Pada metode Random Forest, akurasi tertinggi terdapat pada 5-fold dengan nilai 0.85 atau 85%. Model Naïve Bayes memiliki tingkat akurasi tertinggi pada 4-fold, yaitu sebesar 0.79 atau 79%. Sedangkan untuk SVM, akurasi tertinggi terdapat 2-fold sebesar 0.83 atau 83%. Hal ini juga membuktikan bahwa penggunaan cross-validation turut mempengaruhi tingkat akurasi dari masing-masing model. Berdasarkan Tabel 4 juga dapat diketahui bahwa Algoritma Random Forest memiliki tingkat akurasi yang lebih baik dari Naïve Bayes dan SVM.

Klasifikasi sentimen dilakukan dengan menggunakan jumlah fold yang menghasilkan akurasi tertinggi pada masing-masing model algoritma. Hasil klasifikasi sentimen dengan algoritma Random Forest dapat dilihat pada Gambar 3. Terlihat bahwa terdapat 52% sentimen positif, 21% sentimen netral, dan 27% sentimen negatif.



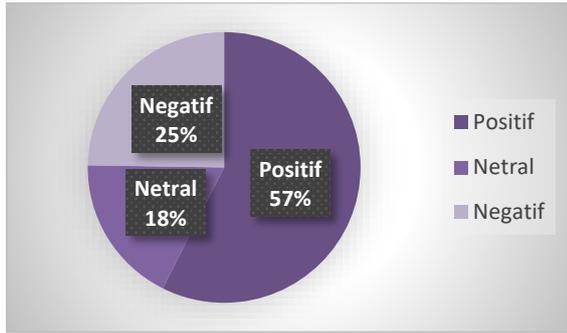
Gambar 3. Klasifikasi Sentimen Menggunakan Algoritma Random Forest

Klasifikasi sentimen dengan algoritma Naïve Bayes dilakukan dengan nilai fold yang memberikan akurasi paling tinggi, yaitu 4. Hasil klasifikasi dengan metode ini seperti yang terlihat pada Gambar 4 menunjukkan bahwa terdapat 51% sentimen positif dari total keseluruhan dataset, 22% termasuk sentimen netral, dan 27% sisanya dikategorikan sebagai sentimen negatif.



Gambar 4. Klasifikasi Sentimen Menggunakan Algoritma Naïve Bayes

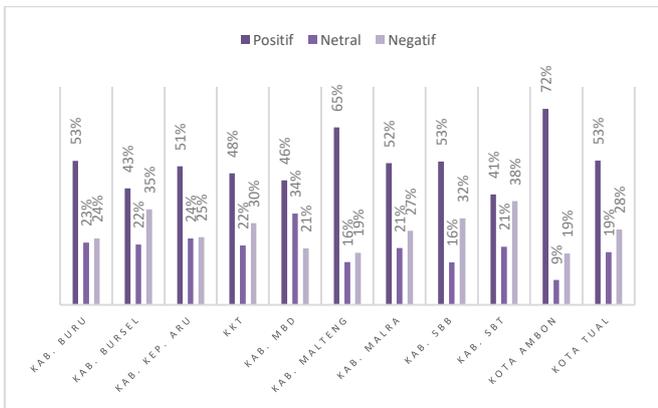
Selanjutnya, SVM memiliki akurasi yang tinggi ketika fold-nya bernilai 2. Hasil klasifikasi sentimen dengan metode ini terlihat pada Gambar 5. Berdasarkan hasil yang diperoleh, diketahui terdapat 57% sentimen positif, 18% termasuk sentimen netral, dan 25% sisanya dikategorikan sebagai sentimen negatif.



Gambar 5. Klasifikasi Sentimen Menggunakan Algoritma SVM

**B. Pembahasan**

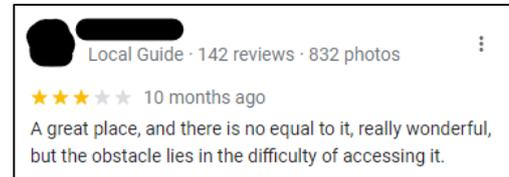
Secara keseluruhan, hasil klasifikasi sentimen dengan menggunakan 3 algoritma ini didominasi oleh sentimen positif. Hal ini menunjukkan persepsi wisatawan terhadap destinasi pariwisata di Provinsi Maluku dapat dikatakan baik. Walaupun demikian, terdapat sentimen negatif dan netral yang dapat dijadikan untuk perbaikan dan peningkatan kualitas pariwisata. Untuk membahas hal ini secara komprehensif, kami melakukan klasifikasi sentimen dengan algoritma Random Forest untuk masing-masing lokasi wisata di 11 Kabupaten/Kota di Maluku. Algoritma ini dipilih karena terbukti memiliki akurasi yang tinggi jika dibandingkan dengan Naïve Bayes dan SVM. Hasil klasifikasi sentimen per Kabupaten/Kota di Provinsi Maluku ditunjukkan pada Gambar 6.



Gambar 6. Distribusi Sentimen per Kabupaten/Kota di Maluku Menggunakan Algoritma Random Forest

Berdasarkan hasil yang diperoleh pada Gambar 6, dapat terlihat bahwa tingkat sentimen untuk masing-masing Kabupaten/Kota di Maluku memiliki persentase yang cukup bervariasi. Namun secara keseluruhan, klasifikasi sentimen yang dihasilkan didominasi oleh sentimen positif. Hal ini menunjukkan bahwa persepsi wisatawan terhadap pariwisata di seluruh Kabupaten/Kota dapat dikatakan baik karena distribusi sentimen yang lebih banyak pada sentimen positif. Perlu juga diingat bahwa jumlah ulasan dan komentar yang diperoleh dalam dataset untuk masing-masing Kabupaten/Kota cukup berbeda, terutama untuk daerah yang

berada cukup jauh dari Kota Ambon, kurang populer serta kabupaten yang masih terhitung baru terbentuk seperti Kabupaten Buru Selatan. Kota Ambon, Kabupaten Maluku Tengah dan Maluku Tenggara memiliki ulasan yang cukup banyak karena terdapat beberapa lokasi wisata yang sudah terkenal secara nasional seperti Pantai Ora, Banda Neira, Pantai Ngurbloat dan Pantai Pintu Kota.



Gambar 7. Sampel Ulasan Terkait Destinasi Wisata di Maluku

Walaupun demikian, adanya sentimen negatif dapat menjadi tolak ukur untuk melihat adanya beberapa masalah yang perlu diperbaiki. Demikian juga dengan sentimen netral yang dapat menunjukkan perlu adanya upaya peningkatan pelayanan maupun kepuasan wisatawan. Secara umum, beberapa keluhan yang dialami oleh wisatawan seperti kesulitan mengakses lokasi wisata karena keterbatasan alat transportasi dan tingginya biaya perjalanan dari Kota Ambon ke daerah-daerah lain, cuaca yang tidak menentu, guesthouse yang kurang memadai, kurangnya promosi untuk lokasi destinasi wisata di beberapa Kabupaten/Kota, serta ketersediaan jaringan seluler. Salah satu ulasan ini dapat dilihat pada Gambar 7. Dengan demikian, Pemerintah daerah, baik Pemerintah Provinsi Maluku maupun Pemerintah Kabupaten/Kota masih perlu melakukan upaya untuk menarik wisatawan untuk berkunjung, terutama untuk daerah-daerah yang belum terlalu populer. Upaya yang dapat dilakukan berupa peningkatan promosi pariwisata daerah, meningkatkan kualitas fasilitas dan layanan pariwisata, keamanan dan kenyamanan wisatawan, serta yang paling penting dalam bidang transportasi ke daerah-daerah yang cukup jauh dari pusat kota kabupaten atau kota madya.

**IV. KESIMPULAN**

Penelitian yang dilakukan dapat memberikan gambaran bagaimana persepsi wisatawan terhadap pariwisata di Maluku melalui analisis sentimen terhadap komentar dan ulasan yang diperoleh melalui media sosial seperti Facebook, Instagram, TikTok, serta ulasan pada Google Maps dan TripAdvisor. Dalam pemodelan untuk klasifikasi sentimen, digunakan 3 model, yaitu algoritma Random Forest, Naïve Bayes dan Support Vector Machine (SVM). Hasil pengujian terhadap ketiga algoritma tersebut menunjukkan Random Forest memiliki akurasi yang paling baik daripada Naïve Bayes dan SVM, yaitu 85% dengan 5-fold.

Secara keseluruhan, persepsi wisatawan terhadap destinasi pariwisata di Provinsi Maluku dapat dikatakan baik berdasarkan hasil analisis sentimen yang didominasi oleh sentimen positif. Namun, adanya sentimen negatif dan netral

dapat mengindikasikan bahwa Pemerintah Daerah, baik Pemerintah Provinsi maupun Pemerintah Kabupaten/Kota masih perlu melakukan upaya peningkatan kualitas pelayanan terhadap wisatawan baik dari segi sumber daya manusia, transportasi, akomodasi, maupun sarana dan prasarana. Penelitian selanjutnya masih perlu dilakukan dengan meningkatkan jumlah data teks yang digunakan dan penggunaan metode analisis sentimen lainnya.

#### UCAPAN TERIMA KASIH

Terima kasih kepada Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat yang telah mendanai penelitian ini dengan Nomor Kontrak Induk 188/E5/PG.02.00.PL/2023 dan Nomor Kontrak Turunan 328/LL12/PG/2023, 06/UKIM.H6/N/KP.PKN/2023.

#### DAFTAR PUSTAKA

- [1] Google, "Data Google Trends," *Google Trends*, 2023. <https://trends.google.co.id/trends/explore?cat=67&date=today-5-y&q=%2Fm%2F0jdmd&hl=id> (diakses Apr 09, 2023).
- [2] Pemerintah Daerah Provinsi Maluku, "RPJMD: Rencana Pembangunan Jangka Menengah Daerah Provinsi Maluku Tahun 2019-2024," Ambon, 2019. [Daring]. Tersedia pada: [https://malukuprov.go.id/pdf\\_import/RPJMD\\_2019-2024.pdf](https://malukuprov.go.id/pdf_import/RPJMD_2019-2024.pdf).
- [3] Bidang Neraca Wilayah & Analisis Statistik, *Analisis Sektor Unggulan Provinsi Maluku 2022*. Ambon: Badan Pusat Statistik Provinsi Maluku, 2022.
- [4] A. Liu, Y. R. Kim, dan H. Song, "Toward an accurate assessment of tourism economic impact: A systematic literature review," *Ann. Tour. Res. Empir. Insights*, vol. 3, no. 2, hal. 100054, 2022, doi: 10.1016/j.annale.2022.100054.
- [5] F. Hendriks, "Provinsi Maluku Dalam Angka 2023," BPS Provinsi Maluku, Ambon, 2023. [Daring]. Tersedia pada: <https://maluku.bps.go.id/publication/2023/02/28/5e8944e1ca42a5199c4c577e/provinsi-maluku-dalam-angka-2023.html>.
- [6] F. Hendriks, "Provinsi Maluku Dalam Angka 2022," BPS Provinsi Maluku, Ambon, 2022. [Daring]. Tersedia pada: <https://maluku.bps.go.id/publication/2022/02/25/2a70c4b4b14c5150791bad4c/provinsi-maluku-dalam-angka-2022.html>.
- [7] A. J. Leatemia dan F. Hendriks, "Provinsi Maluku Dalam Angka 2021," BPS Provinsi Maluku, Ambon, 2021. [Daring]. Tersedia pada: <https://maluku.bps.go.id/publication/2021/02/26/972e698a6e9a5506eabfd7d/provinsi-maluku-dalam-angka-2021.html>.
- [8] A. J. Leatemia dan F. Hendriks, "Provinsi Maluku Dalam Angka 2020," BPS Provinsi Maluku, Ambon, 2020. [Daring]. Tersedia pada: <https://maluku.bps.go.id/publication/2020/05/19/afc99b2033e3746f260dbb1/provinsi-maluku-dalam-angka-2020.html>.
- [9] Rokom, "Indonesia Bersiap Menuju Endemi," *Sehat Negeriku*, Jakarta, 2022.
- [10] D. Shabrina, "Status Kedaruratan Covid-19 Selama Mudik Masih Berlanjut," *Media Indonesia*, Apr 03, 2023.
- [11] C. Liu dan H. T. Chong, "Social media engagement and impacts on post-COVID-19 travel intention for adventure tourism in New Zealand," *J. Outdoor Recreat. Tour.*, no. February, hal. 100612, 2023, doi: 10.1016/j.jort.2023.100612.
- [12] Y. Yuan, C. S. Chan, S. Eichelberger, H. Ma, dan B. Pikkemaat, "The effect of social media on travel planning process by Chinese tourists: the way forward to tourism futures," *J. Tour. Futur.*, hal. 1–20, 2022, doi: 10.1108/JTF-04-2021-0094.
- [13] K. Puh dan M. Bagić Babac, "Predicting sentiment and rating of tourist reviews using machine learning," *J. Hosp. Tour. Insights*, 2022, doi: 10.1108/JHTI-02-2022-0078.
- [14] S. Mukhopadhyay, T. Jain, S. Modgil, dan R. K. Singh, "Social media analytics in tourism: a review and agenda for future research," in *Benchmarking An International Journal*, vol. ahead-of-p, no. ahead-of-print, Singapore: Springer Nature Singapore Pte Ltd, 2022, hal. 193–236.
- [15] S. Kemp, "Digital 2023 Global Overview Report," Singapore, 2023. [Daring]. Tersedia pada: <https://www.slideshare.net/DataReportal/digital-2023-global-overview-report-summary-version-january-2023-v02>.
- [16] Y. Li, Z. Lin, dan S. Xiao, "Using social media big data for tourist demand forecasting: A new machine learning analytical approach," *J. Digit. Econ.*, vol. 1, no. 1, hal. 32–43, 2022, doi: 10.1016/j.jdec.2022.08.006.
- [17] M. Thelwall, "Sentiment Analysis for Tourism," in *Big Data and Innovation in Tourism, Travel, and Hospitality: Managerial Approaches, Techniques, and Applications*, Springer Singapore, 2019, hal. 1–223.
- [18] P. Savci dan B. Das, "Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of Arabic, English, and Turkish languages," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 3, hal. 227–237, 2023, doi: 10.1016/j.jksuci.2023.02.017.
- [19] H. Nankani, H. Dutta, H. Shrivastava, P. V. N. . R. Krishna, D. Mahata, dan R. R. Shah, "Multilingual Sentiment Analysis," *Deep Learn. Approaches Sentim. Anal.*, hal. 193–236, 2020, doi: <https://doi.org/10.1007/978-981-15-1216-2>.
- [20] I. F. Putra dan A. Purwarianti, "Improving Indonesian Text Classification Using Multilingual Language Model," hal. 0–4, 2020.
- [21] E. Fernandes, S. Moro, dan P. Cortez, "Data Science, Machine learning and big data in Digital Journalism: A survey of state-of-the-art, challenges and opportunities," *Expert Syst. Appl.*, vol. 221, no. September 2021, hal. 119795, 2023, doi: 10.1016/j.eswa.2023.119795.
- [22] X. Shu dan Y. Ye, "Knowledge Discovery: Methods from data mining and machine learning," *Soc. Sci. Res.*, vol. 110, no. October 2022, hal. 102817, 2023, doi: 10.1016/j.ssresearch.2022.102817.
- [23] S. García-Méndez, F. de Arriba-Pérez, A. Barros-Vila, dan F. J. González-Castaño, "Detection of temporality at discourse level on financial news by combining Natural Language Processing and Machine Learning," *Expert Syst. Appl.*, vol. 197, no. September 2021, hal. 116648, 2022, doi: 10.1016/j.eswa.2022.116648.
- [24] J. Atkinson dan A. Escudero, "Evolutionary natural-language coreference resolution for sentiment analysis," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, hal. 100115, 2022, doi: 10.1016/j.jjime.2022.100115.
- [25] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine," *Inf. Syst. Dev.*, vol. 5, no. 2, hal. 7–13, 2020.
- [26] H. Tuhuteru dan A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J. Pengemb. IT*, vol. 3, no. 3, hal. 394–401, 2018, doi: 10.30591/jpit.v3i3.977.
- [27] M. J. Denny dan A. Spirling, "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It," *Polit. Anal.*, vol. 26, no. 2, hal. 168–189, 2018, doi: 10.1017/pan.2017.44.
- [28] E. Prayitno, T. Suprawoto, dan ..., "Optimasi Hasil Pencarian Pada Web Scrapping Menggunakan Pembobotan Kata Tf-Idf," *J. Innov. Res. Knowl.*, vol. 1, no. 7, hal. 241–246, 2021, [Daring]. Tersedia pada: <https://bajangjournal.com/index.php/JIRK/article/view/822>.
- [29] A. B. Shaik dan S. Srinivasan, *A brief survey on random forest ensembles in classification model*, vol. 56. Springer Singapore, 2019.
- [30] S. Chen, G. I. Webb, L. Liu, dan X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Syst.*, vol. 192, no. xxxx, hal. 105361, 2020, doi: 10.1016/j.knosys.2019.105361.
- [31] J. Lin, H. Chen, S. Li, Y. Liu, X. Li, dan B. Yu, "Accurate prediction of potential druggable proteins based on genetic

- algorithm and Bagging-SVM ensemble classifier,” *Artif. Intell. Med.*, vol. 98, no. March, hal. 35–47, 2019, doi: 10.1016/j.artmed.2019.07.005.
- [32] I. K. Nti, O. Nyarko-Boateng, dan J. Aning, “Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation,” *Int. J. Inf. Technol. Comput. Sci.*, vol. 13, no. 6, hal. 61–71, 2021, doi: 10.5815/ijitcs.2021.06.05.
- [33] K. O. Fapohunda *et al.*, “Evaluation Of Infectious Diseases Outbreak Among Prisoners Using Machine Learning,” *J. Multidiscip. Eng. Sci. Technol.*, vol. 10, no. 2, hal. 2458–9403, 2023, [Daring]. Tersedia pada: www.jmest.org.
- [34] A. Peryanto, A. Yudhana, dan R. Umar, “Klasifikasi Citra Menggunakan Convolutional Neural Network dan K Fold Cross Validation,” *J. Appl. Informatics Comput.*, vol. 4, no. 1, hal. 45–51, 2020, doi: 10.30871/jaic.v4i1.2017.