Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 7, Juli 2025, Hal. 1993-2003
DOI: https://doi.org/10.52436/1.jpti.892
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Analisis Sentimen Media Sosial Twitter terhadap Destinasi Wisata Krui Menggunakan Algoritma Support Vector Machine dan SMOTE

Kelvin indra Richardo[1], Amarudin*[2]

[1,2]Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
Email: [1]kelvin_indra_richardo@teknokrat.ac.id, [2]amarudin@teknokrat.ac.id

Abstrak

Pariwisata di Krui, Pesisir Barat, Lampung, semakin berkembang, namun pemahaman terhadap opini wisatawan masih terbatas. Analisis sentimen media sosial seperti Twitter dapat memberikan wawasan berharga, tetapi ketidakseimbangan data antara sentimen positif dan negatif menjadi tantangan dalam klasifikasi. Model klasifikasi cenderung lebih akurat dalam mengenali kelas mayoritas dibandingkan kelas minoritas, sehingga hasil analisis dapat menjadi bias. Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap destinasi wisata di Krui menggunakan *Support Vector Machine* (SVM). Untuk mengatasi ketidakseimbangan data, diterapkan *Synthetic Minority Over-sampling Technique* (SMOTE). Tahapan penelitian mencakup *preprocessing* teks, ekstraksi fitur TF-IDF, pembagian data latih dan uji, serta pelatihan model SVM sebelum dan sesudah SMOTE. Hasil sebelum SMOTE, model memiliki akurasi 90%, tetapi recall untuk sentimen negatif hanya 43%, terdapat kesulitan dalam mengenali kelas minoritas. Setelah penerapan SMOTE, *recall* negatif meningkat menjadi 52% dan *F1-score* negatif menjadi 66%, dengan akurasi keseluruhan mencapai 91%. Penelitian ini membuktikan bahwa SMOTE dapat meningkatkan keseimbangan klasifikasi sentimen. Hasil ini dapat dimanfaatkan oleh pemangku kepentingan pariwisata di Krui untuk memahami opini wisatawan, meningkatkan strategi promosi, dan meningkatkan kualitas layanan wisata.

Kata kunci: *Analisis Sentimen, Krui, Pariwisata, Support Vector Machine, Synthetic Minority Over-sampling Technique, Twitter.*

Sentiment Analysis of Twitter Users Toward Tourist Destinations in Krui Using the SVM Algorithm

Abstract

Tourism in Krui, West Coast, Lampung, is growing, but understanding of tourist opinions is still limited. Sentiment analysis of social media such as Twitter can provide valuable insights, but the imbalance of data between positive and negative sentiments is a challenge in classification. Classification models tend to be more accurate in recognizing the majority class than the minority class, so the analysis results can be biased. This study aims to analyze Twitter user sentiment towards tourist destinations in Krui using Support Vector Machine (SVM). To overcome the data imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was applied. The research stages include text preprocessing, TF-IDF feature extraction, training and test data division, and SVM model training before and after SMOTE. The results show that before SMOTE, the model has an accuracy of 90%, but the recall for negative sentiment is only 43%, indicating difficulty in recognizing the minority class. After applying SMOTE, the negative recall increased to 52% and the negative F1-score to 66%, with an overall accuracy reaching 91%. This study proves that SMOTE can improve the balance of sentiment classification. These results can be utilized by tourism stakeholders in Krui to understand tourist opinions, improve promotional strategies, and improve the quality of tourism services.

Keywords: *Krui, Sentiment Analysis, Support Vector Machine, Synthetic Minority Over-sampling Technique, Tourism, Twitter.*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

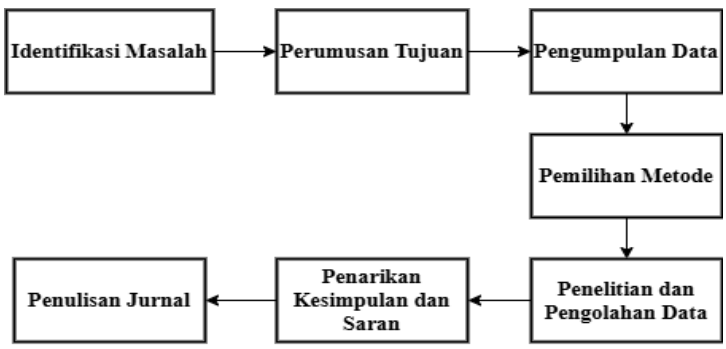
1. PENDAHULUAN

Pariwisata merupakan sektor vital dalam perekonomian Indonesia, berkontribusi signifikan terhadap pendapatan nasional dan penciptaan lapangan kerja [1]. Perkembangan teknologi informasi dan komunikasi telah mengubah cara wisatawan mencari informasi dan berbagi pengalaman mereka. Media sosial, khususnya Twitter, menjadi platform utama bagi pengguna untuk menyampaikan opini dan ulasan terkait destinasi wisata. Analisis sentimen terhadap data yang diperoleh dari Twitter dapat memberikan wawasan mendalam mengenai persepsi publik terhadap destinasi wisata tertentu. Pendekatan analisis sentimen telah diterapkan dalam berbagai penelitian untuk memahami opini masyarakat terhadap objek wisata maupun topik lain. Misalnya, Rahutomo *et al.* (2016) mengimplementasikan analisis sentimen pada ulasan film menggunakan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan sentimen positif dan negatif [2]. Lestari (2018) juga menggunakan algoritma SVM dalam menganalisis sentimen di Twitter terkait topik cadar dan hijab, menunjukkan efektivitas metode ini dalam mengkategorikan sentimen [3]. SVM dikenal memiliki kinerja yang baik dalam klasifikasi teks, termasuk analisis sentimen [4]. Tarımer *et al.* (2019) membandingkan beberapa algoritma machine learning pada data komentar film di IMDB dan Twitter, dan menemukan bahwa SVM menghasilkan akurasi tertinggi dibandingkan algoritma lain [5]. Penelitian lain oleh Chairunnisa dan Putra (2021) menganalisis sentimen masyarakat terhadap tempat wisata di Indonesia dengan akurasi mencapai 82,35% menggunakan SVM [6]. Singgalen *et al.* (2021) menekankan pentingnya analisis sentimen dan pemodelan topik untuk optimalisasi strategi pemasaran destinasi wisata [7]. Sementara itu, Asiva *et al.* (2015) menganalisis sentimen wisatawan terhadap Pantai Tanjung Bira di Sulawesi Selatan yang hasilnya memberikan wawasan penting terkait pelayanan wisata [8]. Penelitian oleh Fatimah *et al.* (2021) juga menunjukkan bahwa data dari media sosial dapat mencerminkan preferensi dan kepuasan wisatawan terhadap sebuah destinasi [9]. Krui, yang terletak di Pesisir Barat Lampung, merupakan destinasi wisata yang menawarkan keindahan alam luar biasa, seperti pantai eksotis dan hutan lindung. Meskipun memiliki potensi wisata yang besar, pemahaman mengenai persepsi wisatawan terhadap Krui masih terbatas. Oleh karena itu, penelitian ini penting untuk mengetahui opini masyarakat secara langsung dari media sosial. Dalam pengolahan data sentimen, ketidakseimbangan jumlah data sentimen positif dan negatif seringkali terjadi. Hal ini dapat mengganggu performa model klasifikasi. Salah satu metode untuk mengatasi masalah ini adalah **Synthetic Minority Over-sampling Technique (SMOTE)**. Menurut Prasetya *et al.* (2022), SMOTE secara signifikan meningkatkan akurasi dan sensitivitas model klasifikasi dalam domain data yang tidak seimbang [10]. Oleh karena itu, integrasi SMOTE dalam proses klasifikasi diharapkan dapat memperbaiki hasil klasifikasi, khususnya dalam mendeteksi sentimen minoritas. Dengan demikian, penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap destinasi wisata di Krui menggunakan algoritma SVM. Secara khusus, tujuan dari penelitian ini adalah mengevaluasi performa model SVM dalam mengklasifikasikan sentimen positif dan negatif, menganalisis dampak penerapan SMOTE terhadap performa model dalam mengatasi ketidakseimbangan data, dan memberikan wawasan kepada pemangku kepentingan pariwisata di Krui mengenai opini wisatawan yang dapat dimanfaatkan untuk perbaikan strategi pemasaran dan pengelolaan destinasi. Penelitian ini diharapkan dapat memberikan kontribusi nyata bagi pengembangan pariwisata Krui melalui pendekatan berbasis data, sehingga meningkatkan daya tarik dan kepuasan wisatawan.

2. METODE PENELITIAN

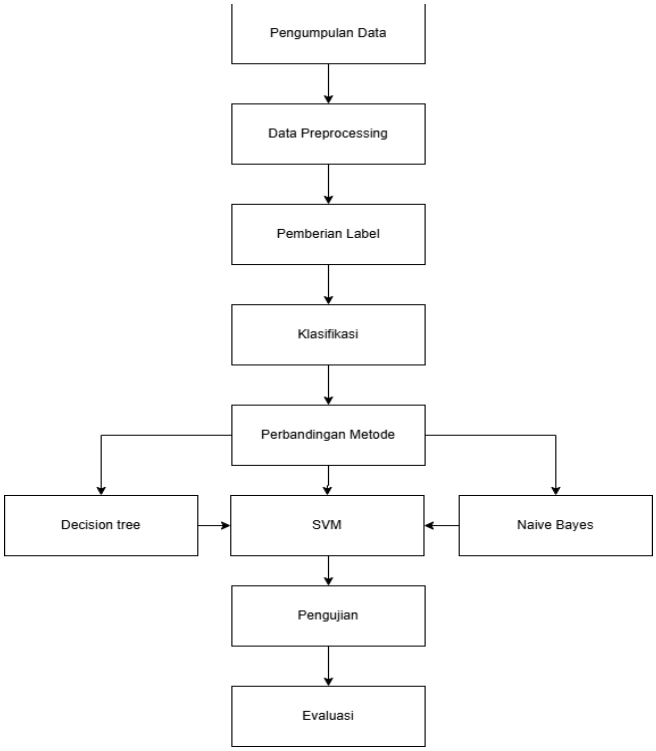
2.1. Tahapan Penelitian

Tahapan Penelitian merupakan alur penelitian yang dilakukan secara berurutan, berikut adalah gambaran tahapan penelitian dan Metode pada Gambar 1 Tahapan Penelitian dan Gambar 2 Metode Penelitian yang diusulkan.



Gambar 1. Tahapan penelitian

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 2. Metode Penelitian yang Diusulkan

2.2. *Support Vector Machine (SVM) Classifier*

Salah satu tugas dalam data mining adalah klasifikasi data, yaitu mengelompokkan data ke dalam satu atau beberapa kelas yang telah didefinisikan sebelumnya. *Support Vector Machine* (SVM) *Classifier* adalah salah satu metode pembelajaran mesin yang sering digunakan untuk tugas klasifikasi. SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas dalam suatu ruang fitur dengan margin maksimal.
Dalam kasus data yang dapat dipisahkan secara linear, SVM mencari *hyperplane* dengan rumus sebagaimana yang disajikan pada formula (1), dan dengan syarat formula (2).

$$f(x) = w^T x + b \qquad (1)$$

$$yi(wTxi + b) \geq 1, \forall i \qquad (2)$$

Keterangan:
w adalah vektor bobot,
x adalah vektor fitur,
b adalah bias,
y_i adalah label kelas (+1 atau −1).
SVM bertujuan untuk memaksimalkan margin antara dua kelas, yang diformulasikan sebagai berikut:

$$\min_{w,b} \quad \frac{1}{2}|w|^2 \qquad (3)$$

dengan batasan:

$$y_i(w^T x_i + b) \geq 1\ (4) \qquad (4)$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Jika data tidak dapat dipisahkan secara linear, SVM menggunakan kernel trick, yang memetakan data ke dimensi yang lebih tinggi agar dapat dipisahkan dengan lebih baik. Beberapa jenis kernel yang sering digunakan adalah:

a. *Linear Kernel*: $K(x_i, x_j) = x_i^T x_j$
b. *Polynomial Kernel*: $K(x_i, x_j) = (x_i^T x_j + c)^d$
c. *Radial Basis Function (RBF) Kernel*: $K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2)$

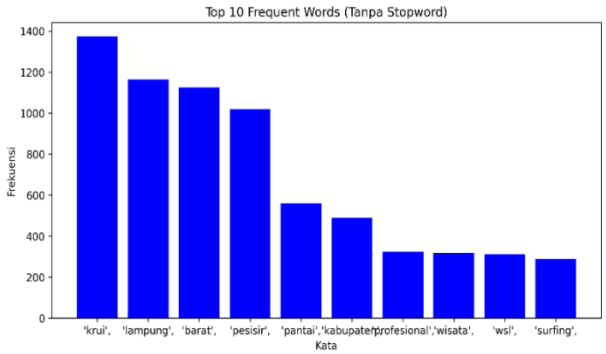
Metode SVM ini memiliki beberapa keunggulan dibandingkan algoritma lain, seperti kemampuannya untuk menangani data berdimensi tinggi dan memiliki risiko *overfitting* yang lebih rendah. Namun, SVM juga memiliki tantangan, terutama dalam pemilihan parameter seperti C (*regularization parameter*) dan jenis kernel yang sesuai dengan dataset yang digunakan. Dalam penelitian ini, SVM digunakan untuk mengklasifikasikan sentimen pengguna Twitter terhadap destinasi wisata di Krui. Data yang digunakan telah melalui tahapan *preprocessing*, yang terdiri dari *cleansing, tokenizing, case folding, stopword removal, dan stemming*. Setelah itu, fitur teks dikonversi menggunakan TF-IDF *Vectorizer*, lalu dilakukan pembagian data menjadi data latih (*training set*) dan data uji (*testing set*). Untuk menangani ketidakseimbangan data, penelitian ini menerapkan *Synthetic Minority Over-sampling Technique* (SMOTE) sebelum melatih model SVM. Model SVM kemudian dievaluasi menggunakan *confusion matrix* dan *classification report* untuk mengukur tingkat akurasi, presisi, *recall*, dan *F1-score.*

2.3 Pengumpulan Data

Penelitian ini melakukan pengumpulan data dengan cara *Crawling Data* di twitter dengan *keywords* "Krui, Pantai Krui, Pantai Mandiri. Pantai Tanjung Setia, *surfing* Krui, Pesisir Barat, wisata Krui, Pulau Pisang, Labuhan Jukung, Ekowisata Krui, Pantai Melasti, Pantai Batu Tihang" melalui *keywords* ini, *Crawling Data* dilakukan dengan pemanfaatan Twitter API menggunakan bahasa pemtograman python. Data yang di kumpulkan berupa *comment* dan *reply*, yang setelahnya disimpan dalam bentuk CSV untuk di proses lebih lanjut.

2.4. Dataset

Penelitian ini menggunakan dataset yang terdiri dari 3.717 data yang berasal dari pengguna Twitter yang membahas destinasi wisata di Krui, Pesisir Barat, Lampung. Sebelum dilakukan analisis, data komentar terlebih dahulu melalui tahap preprocessing, yang mencakup pembersihan teks, seperti penghapusan tanda baca, normalisasi huruf menjadi huruf kecil, serta eliminasi kata-kata yang tidak memiliki makna signifikan. Setelah proses ini selesai, setiap komentar diklasifikasikan ke dalam sentimen positif atau negatif. Dalam penelitian ini, *Model Support Vector Machine* (SVM) digunakan untuk melakukan klasifikasi sentimen. Berdasarkan analisis data, ditemukan 10 kata dengan frekuensi kemunculan tertinggi, yaitu "krui" (1.373 kali), "lampung" (1.164 kali), "barat" (1.125 kali), "pesisir" (1.020 kali), "pantai" (560 kali), "kabupaten" (489 kali), "profesional" (323 kali), "wisata"(318 kali), "wsl"(312 kali),dan "surfing" (288 kali). Untuk menggambarkan distribusi kemunculan kata-kata tersebut, dibuat grafik Top 10 Frekuensi Kata, yang memberikan gambaran visual mengenai kata-kata yang paling sering muncul dalam *dataset*. Grafik ini membantu dalam memahami fokus utama percakapan yang terjadi terkait destinasi wisata di Krui. Contoh distribusi frekuensi kata-kata tersebut dapat dilihat pada Gambar 2 di bawah ini.



Gambar 3. Frekuensi kemunculan kata tertinggi

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.5 *Data Preprocessing*

Data yang di ambil dari twitter ini pada awalnya tidak terstruktur dan tidak konsisten jadi perlu di bersihkan terlebih dahulu dengan melakukan *text preprocessing* sebelum di lakukan analisis sentiment pada data. Setelah dilakukannya pembersihan pada data, proses pelabelan dilakukan pada data guna memberikan label sentiment pada setiap data teks. *Preprocessing* ini dilakukan menggunakan *library NLTK* dan *spaCy* didalam Bahasa pemrograman Python dengan beberapa tahapan sebagai berikut: *Cleansing, Case Folding, Tokenizing, Stopword Removal, Stemming*, dan *Data Labeling*. Penjelasan tahapan dari preprocessing data adalah sebagai berikut:

a. *Cleansing*: Dataset yang masih belum diproses dibersihkan terlebih dahulu dari symbol, tanda baca, *link url*, *emoticon*, dan angka. Berikut adalah contoh hasil *cleansing* pada data dapat dilihat pada Tabel 1.

Tabel 1. *Cleansing*

Text	***Cleansing***
Tk kira itu di pulau pisang banyak pohon pisangnya eh ternyata pohon kelapa temenku juga bertanya tanya kemarin pas liburan 😜 😜	Tk kira itu di pulau pisang banyak pohon pisangnya eh ternyata pohon kelapa temenku juga bertanya tanya kemarin pas liburan
Alhamdullilah nostalgia saya dulu 98 - 2012 kerja survey harimau & satwa mangsanya di taman nadional bukit barisan selatan dgn menggunakan camera trap...jalur aspal yg kalian lewati itu sebenarnya membelah kawasan taman nasional...sampe thn 2000an jalan malam bisa berhenti deh krn dilewati kawanan gajah 👍 👍 👍	alhamdullilah nostalgia saya dulu kerja survey harimau amp satwa mangsanya di taman nadional bukit barisan selatan dengan menggunakan camera trap jalur asli tetapi palsu yang kalian lewati itu sebenarnya membelah kawasan taman nasional sampai tahun an jalan malam bisa berhenti deh karena dilewati kawanan gajah

b. *Casefolding*: Proses ini guna merubah huruf besar yang berada pada dataset menjadi huruf kecil. Berikut contoh dari proses *Casefolding* pada Tabel 2.

Tabel 2. *Casefolding*

Text	***Casefolding***
Krui yang terletak di Kabupaten Pesisir Barat Lampung memiliki segudang potensi khususnya di ranah wisata bahari. Krui punya ombak kiri terbaik dunia	krui yang terletak di kabupaten pesisir barat lampung memiliki segudang potensi khususnya di ranah wisata bahari krui punya ombak kiri terbaik dunia
Jadi kangen mau ke Krui walaupun saya bukan orang sana. Saya senang sama lingkungan disana	jadi kangen mau ke krui walaupun saya bukan orang sana saya senang sama lingkungan disana

c. *Tokenizing*: Proses dimana teks dipecah menjadi kata perkata yang dipisahkan dengan tanda “,”. Berikut adalah contoh dari proses *tokenizing* pada Tabel 3.

Tabel 3. *Tokenizng*

Casefolding	***Tokenizing***
dan belakangan ini arus kunjungan wisata ke krui pun semakin meningkat dengan semakin gencarnya promosi baik di dalam negeri maupun di luar negeri baik melalui media cetak maupun media elektronik kruiprosurfwsl	['dan', 'belakangan', 'ini', 'arus', 'kunjungan', 'wisata', 'ke', 'krui', 'pun', 'semakin', 'meningkat', 'dengan', 'semakin', 'gencarnya', 'promosi', 'baik', 'di', 'dalam', 'negeri', 'maupun', 'di', 'luar', 'negeri', 'baik', 'melalui', 'media', 'cetak', 'maupun', 'media', 'elektronik', 'kruiprosurfwsl']

p-ISSN: 2775-4227
e-ISSN: 2775-4219

angkat wisata bahari persaingan di krui profesional dimulai lampung persaingan menjadi surfer terbaik dunia di krui profesional resmi dimulai wonderfulindonesia pesonaindonesia	['angkat', 'wisata', 'bahari', 'persaingan', 'di', 'krui', 'profesional', 'dimulai', 'lampung', 'persaingan', 'menjadi', 'surfer', 'terbaik', 'dunia', 'di', 'krui', 'profesional', 'resmi', 'dimulai', 'wonderfulindonesia', 'pesonaindonesia']

d. *Stopwords Removal*: Proses penghapusan beberapa kata hubung yang tidak perlu didalam kalimat. Contoh dari *Stopword Removal* dapat dilihat pada Tabel 4.

Tabel 4. *Stopwords Removal*

Casefolding	***Stopwords Removal***
angkat wisata bahari persaingan di krui profesional dimulai genpinews	angkat wisata bahari persaingan krui profesional dimulai genpinews
serasa milik sendiri wisata pantai krui pesisir barat	serasa milik wisata pantai krui, pesisir barat

e. Stemming: Proses penghapusan imbuhan pada kata yang memiliki imbuhan. Contohnya dapat dilihat pada Tabel 5 di bawah.

Tabel 5. *Stemming*

Casefolding	***Stemming***
setelah seluruh peserta parade budaya sasuduk sampai di kawasan wisata labuhan jukung krui mereka bersiap siap membuat selimpok gelamay dan nyulam tapis pesonafestivaltelukstabas	['seluruh', 'serta', 'parade', 'budaya', 'sasuduk', 'kawasan', 'wisata', 'labuh', 'jukung', 'kru', 'siap', 'siap', 'buat', 'selimpok', 'gelamay', 'nyulam', 'tapis', 'pesonafestivaltelukstabas']
pantai mandiri adalah tempat wisata yang terletak di kecamatan krui selatan kabupaten pesisir barat lampung	['pantai', 'mandiri', 'tempat', 'wisata', 'letak', 'camat', 'kru', 'selatan', 'kabupaten', 'pesisir', 'barat', 'lampung']

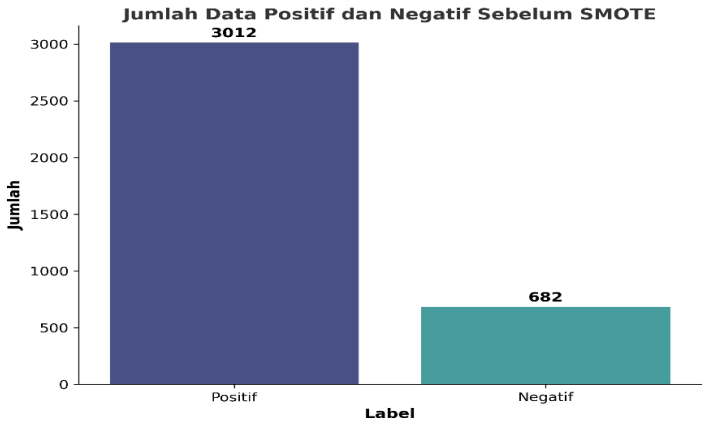
f. *Labeling*: Pada proses ini, pelabelan akan di lakukan pada data dilabeli dengan positif, dan negatif. Tolok ukur penentuan label pada dataset ditentukan secara manual dengan memberi sentimen mengenai teks yang ada di data. Tabel 6 berikut adalah contoh dari *Labeling.*

Tabel 6. *Labeling*

Text	***Sentiment***
warga pesisir barat lampung mendapatkan keberkahan adanya ajang kejuaraan selancar internasional atau word surf league krui profesional di pantai tanjung setia kruiprosurfing	*Positive*
setelah beberapa jam perjalanan akhirnya sampai juga taman nasional bukit barisan jalur ini salah satu jalur ekstrem yang kita lewatin mulai dari tanjakan turunan satwa liar pohon tumbang rawan longsor truck dan jalanan rusaknya	*Negative*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Frekuensi pada Gambar 3 mengidentifikasikan adanya ketidakseimbangan jumlah data antara sentimen positif dan negatif dalam *dataset* yang digunakan. Terdapat 3.012 data sentimen positif dan 682 data sentimen negatif, yang menunjukkan dominasi sentimen positif dalam komentar. Ketidakseimbangan ini dapat berdampak pada performa model *machine learning*, karena model cenderung lebih fokus pada kelas mayoritas dan mengabaikan kelas minoritas. Untuk mengatasi hal tersebut, penelitian ini menerapkan metode SMOTE (*Synthetic Minority Over-sampling Technique*). Metode ini bekerja dengan menambahkan data sintetis pada kelas minoritas, dalam hal ini sentimen negatif, berdasarkan pola yang ditemukan dalam data asli. Dengan cara ini, distribusi data menjadi lebih seimbang, sehingga model dapat belajar dengan lebih baik dan meningkatkan akurasi klasifikasi sentimen. Jumlah klasifikasi sentimen sebelum penerapan SMOTE dapat dilihat pada Gambar 3 di bawah ini.



Gambar 4. Jumlah Data Positif dan Negatif Sebelum SMOTE

2.6. *Data Imbalance Optimization*

Dalam proses klasifikasi, seringkali ditemui masalah ketidakseimbangan data di mana jumlah sampel dalam satu kelas jauh lebih banyak dibandingkan kelas lainnya [4]. Ketidakseimbangan ini dapat menyebabkan model pembelajaran mesin menjadi bias terhadap kelas mayoritas, sehingga performa klasifikasi terhadap kelas minoritas menurun. Misalnya, dalam analisis sentimen, jumlah opini positif, negatif, atau netral mungkin tidak seimbang, yang dapat mempengaruhi akurasi model dalam mengidentifikasi sentimen tertentu.
Untuk mengatasi masalah ini, berbagai teknik telah dikembangkan, salah satunya adalah *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE adalah metode *oversampling* yang bertujuan untuk menyeimbangkan distribusi kelas dengan menghasilkan sampel sintetis untuk kelas minoritas. Alih-alih hanya menduplikasi sampel yang ada, SMOTE menciptakan sampel baru berdasarkan sampel minoritas yang ada. Proses kerja SMOTE melibatkan beberapa langkah. Pertama, untuk setiap sampel dalam kelas minoritas, algoritma akan mengidentifikasi beberapa tetangga terdekatnya dalam ruang fitur. Kemudian, sampel sintetis dibuat dengan memilih salah satu tetangga terdekat dan menghasilkan titik baru yang berada di antara sampel asli dan tetangganya. Secara matematis, jika x_i adalah sampel minoritas dan x_{tetangga} adalah salah satu tetangga terdekatnya, maka sampel sintetis x_{baru} dapat dihitung sebagai:

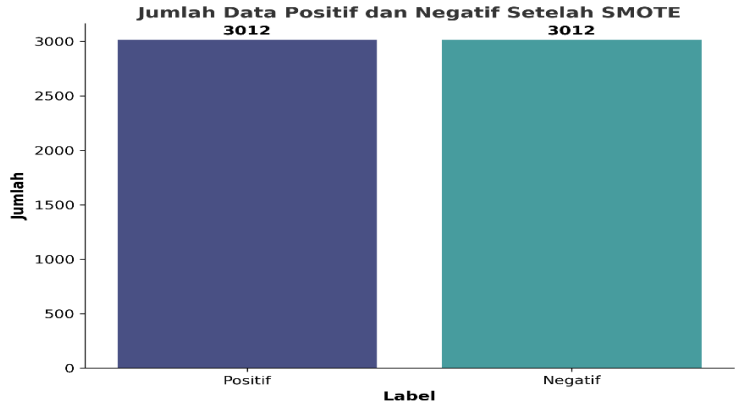
$$x_{\text{baru}} = x_i + \lambda \times \left(x_{\text{tetangga}} - x_i\right) \qquad (5)$$

di mana λ adalah bilangan acak antara 0 dan 1. Dengan cara ini, SMOTE memperluas ruang keputusan untuk kelas minoritas, memungkinkan model untuk belajar lebih baik dan mengurangi bias terhadap kelas mayoritas.

Penerapan SMOTE telah terbukti efektif dalam berbagai studi. Misalnya, penelitian yang dipublikasikan di *Scientific Reports* menunjukkan bahwa algoritma SMOTE dapat meningkatkan efektivitas klasifikasi pada data yang tidak seimbang dengan menghasilkan sampel minoritas baru secara acak untuk meningkatkan representasi kelas minoritas [11]. Selain itu, artikel di *Machine Learning Mastery* menjelaskan bahwa SMOTE adalah teknik augmentasi data untuk kelas minoritas yang menghasilkan sampel sintetis dengan menginterpolasi antara sampel minoritas yang ada [12] Dalam penelitian ini, SMOTE diterapkan sebelum pelatihan model *Support Vector Machine* (SVM) untuk memastikan bahwa model tidak bias terhadap kelas mayoritas. Dengan menyeimbangkan distribusi kelas, diharapkan model dapat mengklasifikasikan data dengan lebih akurat dan adil, khususnya dalam analisis sentimen terhadap opini pengguna Twitter terkait destinasi wisata di Krui. Gambar 4 menunjukkan adanya ketidakseimbangan jumlah data antara sentimen positif dan negatif. Dari total data yang digunakan, terdapat 3.012

p-ISSN: 2775-4227
e-ISSN: 2775-4219

data berlabel positif dan 682 data berlabel negatif. Ketidakseimbangan ini dapat memengaruhi kinerja model klasifikasi, karena model cenderung lebih akurat dalam mengenali kelas mayoritas dibandingkan kelas minoritas. Untuk mengatasi masalah ini, penelitian ini menerapkan metode SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE digunakan untuk menyeimbangkan distribusi data dengan cara menambahkan data sintetis pada kelas minoritas (dalam hal ini, sentimen negatif). Data sintetis yang dihasilkan berdasarkan karakteristik data asli, sehingga model dapat mengenali pola lebih baik dan meningkatkan akurasi prediksi. Setelah menerapkan SMOTE, jumlah data untuk kedua kelas menjadi seimbang, masing-masing sebanyak 3.012 data. Dengan demikian, model *Support Vector Machine* (SVM) dapat belajar dari distribusi data yang lebih seimbang dan menghasilkan prediksi yang lebih optimal. Contoh setelah SMOTE tertera pada gambar 4 berikut.



Gambar 5. Jumlah Data Positif dan Negatif Setelah SMOTE

2.4 Evaluasi

Evaluasi dilakukan untuk mengukur performa model dalam mengklasifikasikan sentimen positif dan negatif pada data yang telah diproses. Dalam penelitian ini, evaluasi dilakukan menggunakan *confusion matrix* serta metrik akurasi, presisi, *recall*, dan *F1-score*. *Model Support Vector Machine* (SVM) dibandingkan sebelum dan setelah penerapan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk melihat efektivitas teknik *oversampling* dalam menangani ketidakseimbangan data. Sebelum menerapkan SMOTE, model lebih dominan dalam mengenali sentimen positif, tetapi kurang akurat dalam mengenali sentimen negatif. Berdasarkan *confusion matrix* sebelum SMOTE, model mengklasifikasikan 608 data positif dengan benar dan 56 data negatif dengan benar, tetapi salah mengklasifikasikan 73 data negatif sebagai positif dan 2 data positif sebagai negatif. Hal ini menunjukkan bahwa model cenderung bias terhadap kelas positif, karena jumlah data positif jauh lebih banyak dibandingkan dengan data negatif. Setelah menerapkan SMOTE, jumlah data positif dan negatif menjadi lebih seimbang. Evaluasi model setelah SMOTE menunjukkan peningkatan dalam mengenali sentimen negatif, meskipun terdapat sedikit peningkatan kesalahan dalam mengenali sentimen positif. *Confusion matrix* setelah SMOTE menunjukkan bahwa model mampu mengklasifikasikan 604 data positif dengan benar dan 67 data negatif dengan benar, tetapi masih salah mengklasifikasikan 62 data negatif sebagai positif dan 6 data positif sebagai negatif. Hasil evaluasi menunjukkan bahwa penerapan SMOTE membantu model untuk lebih seimbang dalam mengenali sentimen positif dan negatif. Meskipun terdapat sedikit peningkatan kesalahan dalam mengklasifikasikan sentimen positif, model menjadi lebih akurat dalam mengidentifikasi sentimen negatif. Dengan demikian, penggunaan SMOTE dalam penelitian ini berhasil meningkatkan performa klasifikasi sentimen pengguna Twitter terkait destinasi wisata di Krui.

3. HASIL DAN PEMBAHASAN

3.1. *Word Cloud Visualization*

Wordcloud digunakan sebagai alat visualisasi untuk menampilkan kata-kata yang paling sering muncul dalam suatu *dataset*. Semakin besar ukuran sebuah kata dalam *wordcloud*, semakin tinggi frekuensi kemunculannya dalam data di gambar ini frekuensi wordcloud didominasi oleh kata “pesisir barat”. Dalam penelitian ini, *wordcloud* dihasilkan dari dataset yang digunakan dan divisualisasikan dalam Gambar 5.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Gambar 5. Visualisasi *WordCloud*

3.3. Pengujian

Pengujian dalam penelitian ini dilakukan untuk mengevaluasi performa tiga model klasifikasi, yaitu *Support Vector Machine* (SVM), *Naive Bayes*, dan *Decision Tree*, baik sebelum maupun setelah penerapan SMOTE. Evaluasi model dilakukan menggunakan metrik utama, yaitu akurasi, presisi, *recall*, dan *F1-score*, guna memahami sejauh mana model mampu mengklasifikasikan sentimen positif dan negatif secara akurat. Berdasarkan hasil pengujian, model SVM menunjukkan performa terbaik dibandingkan model lainnya, dengan akurasi mencapai 90% sebelum SMOTE dan meningkat menjadi 91% setelah SMOTE. Pada pengujian sebelum SMOTE, SVM memiliki presisi 93%, *recall* 72%, dan *F1-score* 77%, yang menunjukkan bahwa meskipun model memiliki presisi tinggi, kemampuan dalam mengenali kelas negatif (minoritas) masih terbatas. Setelah SMOTE diterapkan, SVM mengalami peningkatan performa, dengan presisi dan recall mencapai 91% dan 75%, serta F1-score meningkat menjadi 81%, mengindikasikan keseimbangan yang lebih baik dalam klasifikasi sentimen. Model Naive Bayes, meskipun lebih sederhana, menunjukkan akurasi 71% sebelum SMOTE dan 74% setelah SMOTE. Penerapan SMOTE membantu meningkatkan presisi dan recall model ini, namun F1-score yang hanya mencapai 81% menunjukkan bahwa model ini kurang optimal dalam mengenali pola data yang kompleks, terutama dalam kondisi ketidakseimbangan data. Model Decision Tree menunjukkan performa yang lebih baik dibandingkan Naive Bayes, dengan akurasi 85% sebelum SMOTE dan meningkat menjadi 86% setelah SMOTE. F1-score Decision Tree juga mengalami peningkatan dari 77% menjadi 82%, menunjukkan bahwa model ini mampu meningkatkan keseimbangan klasifikasi setelah data diseimbangkan menggunakan SMOTE. Namun, model ini masih sedikit tertinggal dibandingkan SVM, terutama dalam akurasi dan presisi secara keseluruhan. Hasil pengujian ini menunjukkan bahwa SVM merupakan model yang paling cocok digunakan untuk data sentimen wisata di Krui. Keunggulan SVM dalam menangani data berdimensi tinggi dan kemampuannya mencari *hyperplane* optimal untuk memisahkan kelas positif dan negatif menjadi alasan utama mengapa model ini mampu memberikan hasil yang lebih konsisten. Selain itu, penerapan SMOTE berhasil meningkatkan keseimbangan model, terutama dalam mengenali kelas minoritas, sehingga SVM mampu menghasilkan akurasi dan F1-score tertinggi dibandingkan model lain. Secara keseluruhan, SVM memberikan kombinasi terbaik antara presisi, recall, dan F1-score, sehingga model ini dapat diandalkan dalam analisis sentimen wisata di Krui, memberikan wawasan yang lebih akurat dan representatif bagi pemangku kepentingan pariwisata dalam mengambil keputusan berbasis data.

Tabel 7. Hasil Perbandingan Metode Sebelum SMOTE

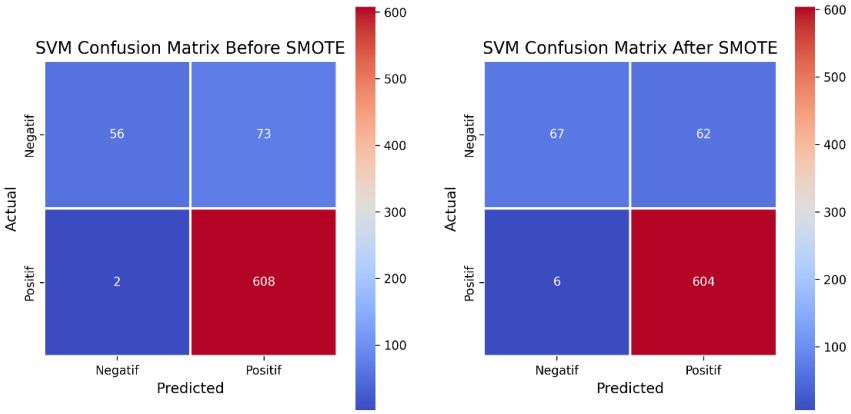
Performance Metrics	***Decision Tree***	***Naïve Bayes***	***SVM***
Accuracy	85%	71%	**90%**
Precision	88%	80%	**93%**
Recall	73%	**79%**	72%
F1-Score	**77%**	75%	**77%**

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Tabel 8. Hasil Perbandingan Metode Setelah SMOTE

Performance Metrics	***Decision Tree***	***Naïve Bayes***	***SVM***
Accuracy	86%	74%	**91%**
Precision	89%	81%	**91%**
Recall	76%	**82%**	75%
F1-Score	**82%**	81%	81%

Confusion matrix merupakan sebuah tabel yang berfungsi untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label asli dari data. Dalam penelitian ini, digunakan *confusion matrix* berukuran 2x2, di mana model mengelompokkan data ke dalam dua kategori, yaitu positif dan negatif. Untuk melihat perbandingan hasil *Confusion Matrix* dari algoritma SVM sebelum dan sesudah penerapan SMOTE, dapat dilihat pada Gambar 6 di bawah ini.



Gambar 6. Perbandingan Confusion Matrix Optimasi SMOTE

Confusion matrix yang ditampilkan pada gambar 6 menunjukkan hasil evaluasi kinerja *model Support Vector Machine* (SVM) sebelum dan setelah penerapan SMOTE. Pada *confusion matrix* sebelum SMOTE, model memiliki akurasi yang tinggi dalam mengklasifikasikan kelas positif, dengan jumlah prediksi benar yang mencapai 608 dan hanya terdapat 2 kesalahan klasifikasi ke kelas negatif. Namun, model mengalami kesulitan dalam mengenali kelas negatif, yang ditunjukkan oleh tingginya *jumlah false positive*, yaitu 73, dibandingkan dengan *true negative* yang hanya 56. Hal ini menunjukkan adanya ketidakseimbangan kelas yang mempengaruhi performa model. Setelah penerapan SMOTE, terjadi peningkatan dalam kemampuan model dalam mengenali kelas negatif. Jumlah *true negative* meningkat dari 56 menjadi 67, sementara jumlah *false positive* menurun dari 73 menjadi 62. Perbaikan ini menunjukkan bahwa model lebih akurat dalam mengklasifikasikan data negatif setelah dilakukan *oversampling* untuk menyeimbangkan distribusi kelas. Meskipun terdapat sedikit peningkatan pada *false negative* dari 2 menjadi 6, secara keseluruhan model tetap mampu mempertahankan performa yang baik dalam mengklasifikasikan kelas positif, dengan jumlah *true positive* yang tetap tinggi, yaitu 604. Dengan demikian, penerapan SMOTE terbukti membantu meningkatkan keseimbangan dalam klasifikasi, sehingga model tidak terlalu bias terhadap kelas mayoritas dan lebih efektif dalam mengenali pola dalam data.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen wisata di Krui menggunakan algoritma *Support Vector Machine* (SVM) dengan optimasi *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan kelas dalam data. Data yang digunakan telah melalui tahap *preprocessing*, termasuk pembersihan teks dan *case folding*, sebelum dibagi menjadi 80% data latih dan 20% data uji. Evaluasi model dilakukan dengan akurasi, presisi, *recall, F1-score*, serta *confusion matrix* untuk memahami distribusi prediksi model terhadap data aktual. Hasil analisis menunjukkan bahwa sebelum penerapan SMOTE**,** *Model Support Vector Machine* (SVM) memiliki akurasi 90%, dengan kelemahan dalam mengenali kelas negatif**,** yang terlihat

p-ISSN: 2775-4227
e-ISSN: 2775-4219

dari rendahnya nilai recall untuk kelas tersebut. Setelah penerapan SMOTE**,** terjadi peningkatan performa model, terutama dalam recall dan F1-score, dengan akurasi meningkat menjadi 91%. *Confusion matrix* menunjukkan bahwa model setelah SMOTE lebih efektif dalam mengklasifikasikan data negatif dibandingkan sebelumnya, sehingga mampu mengatasi ketidakseimbangan kelas dengan lebih baik. Hasil penelitian ini memiliki manfaat dalam mendukung pengembangan sektor pariwisata di Krui, khususnya bagi pemerintah daerah, pelaku usaha, serta pemangku kepentingan di industri pariwisata. Dengan memahami persepsi wisatawan berdasarkan analisis sentimen, pihak terkait dapat mengidentifikasi aspek-aspek positif yang dapat ditingkatkan dalam promosi wisata Krui serta mengetahui keluhan yang perlu diperbaiki untuk meningkatkan kepuasan wisatawan. Selain itu, hasil penelitian ini juga dapat digunakan sebagai dasar dalam perumusan kebijakan berbasis data guna meningkatkan daya tarik wisata Krui di tingkat nasional maupun internasional. Dengan demikian, penelitian ini memberikan wawasan yang berharga bagi pengembangan pariwisata berbasis data, sehingga dapat mendukung pertumbuhan ekonomi daerah serta meningkatkan pengalaman wisatawan yang berkunjung ke Krui.

DAFTAR PUSTAKA

[1] Badan Keahlian Sekretariat Jenderal DPR RI, "Analisis Ringkas Cepat: Urgensi Penguatan Daya Saing Pariwisata Untuk Meningkatkan Perekonomian Nasional," *Www.Bk.Dpr.Go.Id*, pp. 1–44, 2023, [Online]. Available: https://berkas.dpr.go.id/pa3kn/analisis-ringkas-cepat/public-file/analisis-ringkas-cepat-public-41.pdf

[2] Faisal Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "IMPLEMENTASI TWITTER SENTIMENT ANALYSIS UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," *J. Inform. Polinema*, vol. 4, no. 2, p. 93, Feb. 2018, doi: 10.33795/jip.v4i2.152.

[3] Lusiana Lestari, "Analisis Sentimen Twitter Dengan Menggunakan Algoritma Svm," 2018.

[4] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: 10.26418/jp.v6i3.42896.

[5] İ. Tarımer, A. Çoban, and A. E. Kocaman, "Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques," pp. 1–8, 2019, [Online]. Available: http://arxiv.org/abs/1903.11983

[6] A. Chairunnisa, T. Informatika, F. I. Komputer, and U. L. Kuning, "Analisis Sentimen Masyarakat Terhadap Tempat Wisata Di Indonesia Menggunakan Metode," 2022.

[7] Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 459–470, 2021, doi: 10.51519/journalisi.v3i3.171.

[8] Asiva Noor Rachmayani, "ANALISIS SENTIMEN WISATAWAN TERHADAP DESTINASI PANTAI TANJUNG BIRA PADA ULASAN DI GOOGLE MAPS MENGGUNAKAN ALGORITMA NAIVE BAYES," p. 6, 2015.

[9] S. Fatimah, E. Purwanto, and H. Permatasari, "Analisis Sentimen Twitter Tentang Wisata di Kota Solo," *Techno.Com*, vol. 22, no. 4, pp. 854–869, Nov. 2023, doi: 10.33633/tc.v22i4.9154.

[10] J. Prasetya and A. Abdurakhman, "COMPARISON OF SMOTE RANDOM FOREST AND SMOTE K-NEAREST NEIGHBORS CLASSIFICATION ANALYSIS ON IMBALANCED DATA," *MEDIA Stat.*, vol. 15, no. 2, pp. 198–208, Apr. 2023, doi: 10.14710/medstat.15.2.198-208.

[11] S. Wang, Y. Dai, J. Shen, and J. Xuan, "Research on expansion and classification of imbalanced data based on SMOTE algorithm," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-03430-5.

[12] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE for Imbalanced Classification with Python," 2002. [Online]. Available: https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/