

ANALISIS SENTIMEN DESTINASI WISATA DI JAWABARAT PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Riwanti Situmorang^{*1}, Uus Muhammad Husni Tamyis², Lise Sri Andar Muni³
^{1,2,3}Tenik Informatika, S1, Sekolah Tinggi Teknologi Wastukencana Purwakarta.
e-mail : ^{*1}riwantisitumorang84@wastukencana.ac.id, ²mrtamyiz@gmail.com,
³lise@wastukencana.ac.id

Perkembangan teknologi informasi dan komunikasi telah memberikan dampak signifikan pada berbagai aspek kehidupan manusia, termasuk dalam industri pariwisata. Salah satu dampak penting dari perkembangan teknologi adalah kemunculan media sosial, seperti Twitter, yang telah menjadi platform populer bagi pengguna untuk berbagi pengalaman dan pandangan mereka tentang berbagai hal, termasuk destinasi wisata. Namun, dalam jumlah data yang besar dan kompleks, tantangan utama adalah mengklasifikasikan sentimen wisatawan terhadap destinasi wisata di Jawa Barat berdasarkan ulasan dan komentar di Twitter. Oleh karena itu, metode Naive Bayes Classifier dipilih sebagai solusi untuk menganalisis sentimen dengan keahliannya. Penelitian ini bertujuan untuk menganalisis sentimen destinasi wisata di Jawa Barat di Twitter menggunakan metode Naive Bayes Classifier. Dari data 5000 tweet, mayoritas pandangan masyarakat terhadap destinasi wisata tersebut adalah positif (42,24%), diikuti oleh sentimen netral (31,74%), dan negatif (26,02%). Algoritma Naive Bayes memiliki akurasi 0,332, precision 0,658, dan recall 0,332.

Kata Kunci: Wisata Jawa Barat, Review, Analisis Sentimen, Naive Bayes.

I. PENDAHULUAN

Teknologi yang sangat berkembang di dalam informasi dan komunikasi dapat memberikan dampak signifikan ke dalam aspek kehidupan manusia industri pariwisata. Salah satu dampak penting dari perkembangan teknologi adalah kemunculan media sosial, seperti Twitter, yang telah menjadi platform populer bagi pengguna untuk berbagi pengalaman dan pandangan mereka tentang berbagai hal, termasuk destinasi wisata. Dengan semakin meluasnya penggunaan media sosial seperti Twitter, pengguna dapat dengan mudah berbagi pengalaman, ulasan, dan komentar terkait destinasi wisata di Jawa Barat. Informasi yang terkandung dalam data tersebut memiliki potensi untuk memberikan wawasan yang berharga bagi pemerintah dan

pelaku pariwisata dalam mengembangkan destinasi wisata yang lebih baik. Namun, dalam jumlah data yang sangat besar dan kompleks ini, terdapat tantangan dalam menggali informasi yang relevan dan bermanfaat. Dalam konteks ini, analisis sentimen menggunakan metode Naive Bayes Classifier menjadi salah satu solusi yang menarik, karena keahliannya dalam mengklasifikasikan teks berdasarkan sentiment tujuan sentimen terkandung golongan. Penelitian ini bertujuan untuk melakukan analisis sentimen. Melakukan analisis sentimen destinasi wisata di Jawa Barat pada Twitter menggunakan algoritma Naive Bayes Classifier. Hasil analisis sentimen ini diharapkan dapat memberikan wawasan yang berharga kepada pihak berkepentingan pariwisata Jawa Barat, termasuk pemerintah daerah, pengelola destinasi wisata, dan pelaku usaha di sektor pariwisata, untuk menginformasikan keputusan strategis, peningkatan kualitas, dan pengembangan destinasi wisata di Jawa Barat.

II. METODE

Dalam sebuah perancangan sistem hal yang paling diperlukan sebuah metode yang akan menjadi tujuan dan pedoman yang ingin di implementasikan dalam sebuah perancangan. Hal selanjutnya yang di perlukan antara lain, Analisis Sentimen, Algoritma Naive Bayes TF-IDF Confusion matrix dan pengumpulan data pada pembuatan rancang bantuan analisis sentiment.

A. Analisis Sentimen

Analisis sentimen adalah menganalisis teks digital untuk menentukan sebuah nada emosional pesan tersebut positif, negatif atau netral. Perusahaan memiliki data teks dalam volume besar seperti, transaksi obrolan dengan pelanggan, komentar media sosial dan ulasan analisis sentimen adalah teknik yang digunakan mengidentifikasi sentimen analisis juga termasuk proses yang berguna menentukan opini dimana sentimen dapat di gunakan dalam memahami ulasan atau komentar yang digunakan oleh masyarakat pengguna internet.

B. Naive Bayes Classifier

Naive Bayes adalah salah satu pengklarifikasian probabilitas sederhana yang dapat menghitung sekumpulan probabilitas yang telah dapat menjumlahkan frekuensi dan kombinasi nilai yang terdapat di data set algoritma naive bayes juga salah satu metode mining termasuk 10 klarifikasi data mining yang populer di antara algoritma lainnya. Metode naive bayes dapat dinilai sangat berpotensi baik dalam mengklarifikasi data di bandingkan pengklarifikasian dalam hal akurasi eviensi komputer.

Naive bayes adalah metode yang sangat efektif dalam penyelesaian masalah penggolongan metode ini dapat bekerja sesuai dengan probabilitas untuk mentukan probabilitas yang akan walaupun dokumen yang digunakan jutaan jumlahnya.

C. Terms Frequency-Inverse Document Frequency (TF-IDF)

Metode TF-IDF adalah sebuah cara yang dapat memberikan bobot yang berhubungan dengan (trem) terhadap data frekuensi data dapat mengandung kata untuk menunjukkan untuk seberapa luas atau beredarnya kata tersebut sehingga bobot dapat dihubungkan kesebuah kata dan sebuah data akan sangat tinggi jika frekuensi kata tersebut semakin meninggi dalam data. Frekuensi didalam semua dokumen terdapat kandungan kata yang sangat rendah pada kumpulan dokumen metode ini dapat menggabungkan atau menyatukan dua konsep dalam perhitungan bobot. Rumus untuk TF- IDF.

$$W_{ij} = \frac{n_{ij}}{\sum_j = 1 n_{j,i}} \log 2 \frac{D}{d_j}$$

Keterangan :

- W_{ij} : Pembobotan tf-idf untuk trem ke-j pada dokumen-i
- N_{i,j} : Jumlah kemunculan untuk trem-j pada dokumen-i
- P : Banyak trem yang terbentuk
- $\sum_j = 1 n_{j,i}$: Jumlah kemunculan seluruh tram pada dokumen ke-j
- D : Jumlah keseluruhan dokumen
- D_j : Banyak dokumen yang mengandung trem ke-j

D. ConfusionMatrix

Dalam perancangan cifusion metrix yang telah digunakan sebagai metode evaluasi dari akurasi yang di rancang. Confision matrix adalah salah satu metode yang berguna dalam melakukan perhitungan yang sangat tinggi dalam akurasi data mining. Confusion matrinx telah di informasikan tentang pengklarifikasian yang dapat di prediksi dengan sangat besar oleh sebuah klarifikasi dari 4 istila representasi hasil yang telah di proses klarifikasi confusion matrix sebagai berikut, true positif (TP) , true negatif (TN) ,false positif(FL), false negatif(FN) terlihat seperti gambar di bawah.

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values 0 (Negative) 1 (Positive)	1 (Positive)	TP (True Positive)	FP (False Positive)
	0 (Negative)	FN (False Negative)	TN (True Negative)

Gambar.1 Confusion Matrix

Keterangan :

1. True positif (TP)
 Sesungguhnya hasil prediksi menyatakan bagus atau positif.
2. True negatif (TN)
 Sesungguhnya hasil prediksi dinyatakan dengan hasil negatif atau tidak baik.
3. False positif (FP)
 Sesungguhnya hasil dari prediksi dinyatakan positif tetapi hasil yang sesungguhnya menjadi negatif
4. False negatif (FN)
 Sesungguhnya hasil dari prediksi dinyatakan negatif tetapi hasil sesungguhnya di nyatakan positif.

Preses yang sudah dilalui tersebut akan dapat dihitung kemudian nilai accuracy dengan rumus dibawah ini:

$$Accuracy = (TP + TN)/(TP + FP + FN + T)$$

III. PEMBAHASAN

A. Sistem Yang Dirancang

Prerocessing dalam melakukan analisis yang di implementasikan menjadi sebuah aplikasi berbasis website yang memiliki sistem dalam menganalisis sentimen terhadap masyarakat terkait dengan destinasi wisata di jawa barat. Sistem ini dibuat meliputi menarik data prerocessing klarifikasi data dan visualisasi data yang akan menjalankan fungsinya masing-masing. Sistem ini sangat membutuhkan input data dengan cara crawling data menggunakan aplikasi orange data yang di ambil meliputi nama pengguna, tanggal twitt serta komentar di twitt. Setelah data ditemukan sebuah sistem akan melanjutkan proses yang akan dapat di olah data di proses yang akan kemudian menjadi class terdiri dari komentar positif, netral dan negatif. Pengambilan data secara crawling pada twitter dalam pem prosesan pada aplikasi orange. Pengolahan data dengan prerocessing teks mencakup proses transformation, tokenization, filtering pada aplikasi orange

setelah itu data sudah melakukan preprocessing lalu melakukan pembobotan dengan menggunakan metode TF-IDF dan akan di konversi dan melakukan perhitungan menggunakan algoritma naive bayes untuk memberikan kelas setiap komentar.

B. Pengujian Tahap Metode

Pengujian menggunakan confusion matrix dilakukan dalam penghitungan hasil proses metode yang dijalankan oleh algoritma naive bayes. Perhitungan ini dapat di liputi sebagai berikut adalah data klarifikasi data trening dan data testing klarifikasi data dapat dijadikan data trening yang akan di olah menggunakan aplikasi orange.

Pengklarifikasiannya dapat dinyatakan otomatis dengan menggunakan sistem. Setelah klarifikasi data dilakukan selanjutnya dapat ditentukan oleh kelas –kelas yang sesuai dengan parameter dan confusion matrix yaitu sebagai berikut true positif, true negatif, false positif dan false negatif. Data didapatkan dengan cara klarifikasi dan label yang di sesuaikan dengan ketentuan confusion matrix.

$$accuracy = \frac{TP+TN}{TP+FN+TN+FN} \times 100\%$$

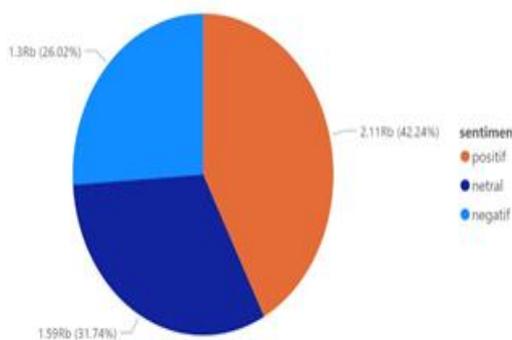
$$error\ rate = \frac{FP+FN}{TP+FP+TN+FN} \times 100\%$$

Sistem hasil dari perhitungan dari akurasi sistem yang dapat di lihat pada tabel 1.

Aktual	Prediksi		
	Negatif	Netral	Positif
Negatif	1231	70	0
Netral	1167	420	0
Positif	1750	352	10

Tabel. 1 Hasil Evaluasi Naive Bayes

Hasil akurasi tersebut sistem menghasilkan proses 5000 sentimen dapat di lihat pada gambar 2.



Gambar 2 Distribusi Jumlah Data Setiap Kelas

IV. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan tentang analisis sentimen destinasi wisata di Jawa Barat pada Twitter menggunakan metode Naive Bayes, dapat disimpulkan bahwa algoritma Naive Bayes merupakan pilihan yang efektif untuk memproses data terkait sentimen destinasi wisata di Jawa Barat. Dari data sample 5000 tweet yang dianalisis, sekitar 42,24% tweet masuk dalam kategori sentimen positif, 31,74% tweet masuk dalam kategori sentimen netral, dan 26,02% tweet masuk dalam kategori sentimen negatif. Hal ini menunjukkan bahwa mayoritas pandangan masyarakat terhadap destinasi wisata di Jawa Barat adalah positif, dengan sebagian kecil yang netral dan negatif.

Selanjutnya, untuk pengujian algoritma Naive Bayes menggunakan metode Orange, diperoleh hasil akurasi sebesar 0,332, precision sebesar 0,658, dan recall sebesar 0,332. Hasil pengujian ini menunjukkan bahwa algoritma Naive Bayes dapat dengan baik mengklasifikasi dokumen tweet dan menentukan sentimen destinasi wisata di Jawa Barat.

Kesimpulan ini menegaskan bahwa analisis sentimen dengan menggunakan algoritma Naive Bayes memberikan wawasan yang berharga tentang pandangan dan respon masyarakat terhadap destinasi wisata di Jawa Barat. Hasil penelitian ini dapat menjadi dasar penting bagi pemerintah daerah, pengelola destinasi wisata, dan pelaku usaha di sektor pariwisata.

DAFTAR PUSTAKA

- [1] Hadna, M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016(Sentika), 57–64. <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf>
- [2] Kurniawan, T. (2017). Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naive Bayes Classifier Dan Support Vector Machine Media Mainstream Menggunakan Naive Machine. IT Journal, 23, 1.
- [3] Nurjannah, M., & Fitri Astuti, I. (2013). PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman. Jurnal Informatika Mulawarman, 8(3), 110–113.
- [4] Priyanto, A., & Ma'arif, M. R. (2018). Implementasi Web Scrapping dan Text Mining untuk Akuisisi dan Kategorisasi Informasi dari Internet (Studi Kasus: Tutorial Hidroponik). Indonesian Journal of Information Systems, 1(1), 25–33. <https://doi.org/10.24002/ijis.v1i1.1664>
- [5] Rosandy, T. (2016). PERBANDINGAN METODE NAIVE BAYES CLASSIFIER DENGAN METODE DECISION TREE (C4.5) UNTUK MENGANALISA KELANCARAN PEMBIAYAAN (Study Kasus : KSPPS / BMT AL-FADHILA. Jurnal Teknologi Informasi Magister Darmajaya, 2(01), 52–62.
- [6] Wiguna, R. A. raffaidy, & Rifai, A. I. (2021). Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining. Journal of Information Systems and Informatics, 3(1), 1–12. <https://doi.org/10.33557/journalisi.v3i1.78>
- [7] Yasid, M. (2019). Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naive Bayes. Jurnal Ilmiah Informatika, 7(02), 82. <https://doi.org/10.33884/jif.v7i02.1329>

- [8] Saputra, Tufik, Ramadani, Oktaviani & Marsusanti (2018), M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016(Sentika), 57–64. <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf>
- [9] Astari, N. M. A. J., Dewa Gede Hendra Divayana, & Gede Indrawan. (2020). Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Dan Informatika (JSI)*, 15(1), 27–29. <https://doi.org/10.30864/jsi.v15i1.332>
- [10] Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. I. M. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes. 8–9.