

ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DARI KOMENTAR PENGUNJUNG DENGAN MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) (Studi Kasus: Taman Air Mancur Sri Baduga Purwakarta)

Moch Saoki Syahlan^{*1}, Dede Irmayanti², Syariful Alam³

^{1,2,3} Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana
Jl Cikopak No.53, Mulyamekar, Kec Babakan Cikao, Purwakarta - Jawa Barat - Indonesia
e-mail: ¹mochsaoki65@wastukencana.ac.id, ²dedeirmayanti@wastukencana.ac.id,
³syarifulalam@wastukencana.ac.id

Pemikiran atau pendapat yang diungkapkan dalam ulasan pengunjung dapat dieksplorasi dengan menggunakan teknik pengolahan data yang dikenal dengan analisis sentimen. Dalam rangka menginformasikan kepada dinas pariwisata Purwakarta tentang potensi peningkatan destinasi wisata di Purwakarta, termasuk air mancur Sri Baduga, penelitian ini akan melakukan analisis sentimen tempat wisata dari komentar pengunjung dengan menggunakan metode Support Vector Machine.

Support Vector Machine adalah teknik yang digunakan untuk mengumpulkan ulasan pengguna di Google Maps, yang kemudian dibagi menjadi tiga kategori sentimen—Positif, Negatif, dan Netral—menggunakan Alat Kolaborasi Google Colaboratory.

dapat disimpulkan bahwa hasil dari tanggapan masyarakat mengenai wisata taman air mancur sri baduga purwakarta ini dengan hasil komentar positif 872, negatif 156, Dan netral 127 yang telah melewati tahap cleaning dan text preprocessing. Adapun persentase 81% pada akurasi (accuration), 94% pada nilai precision, dan tingkat keberhasilan (recall) 99%. Nilai ini menunjukkan bahwa klasifikasi algoritma support vector machine dinilai cukup baik dalam pemrosesan dokumen google maps.

Kata Kunci: Destinasi Wisata, Analisis Sentimen, SVM, Google Maps, Colaboratory

I. PENDAHULUAN

Kemajuan teknologi informasi semakin cepat, terutama karena keunggulan internet dalam penyampaian informasi dengan cepat. Pengguna (user) smartphone, tablet, laptop, dan desktop dapat dengan mudah mengoperasikan dan mengakses internet secara terus-menerus. Teknologi informasi modern memiliki berbagai keuntungan dan manfaat bagi orang lain. Pertumbuhan pariwisata berkorelasi erat dengan pertumbuhan teknologi informasi. Komentar tentang perjalanan yang dilakukan oleh wisatawan dipersilakan. Melalui evaluasi aplikasi populer dan situs jejaring sosial untuk melihat dan memposting tentang perjalanan, seperti Google Maps.[1]

Analisis sentimen adalah salah satu teknik pengolahan data yang dapat digunakan untuk menggali sentimen atau opini yang terkandung dalam ulasan dari pengunjung. Salah satu metode dalam analisis sentimen adalah Support Vector Machine (SVM) yang memiliki performa yang baik. Dalam penelitian Analisis sentiment review wisatawan pada objek wisata UBUD menggunakan algoritma support vector machine penelitian ini data yang digunakan pada tools Rapidminer dengan menghasilkan sentimen positif sebanyak 551 data dan sentimen negatif sebanyak 118 data berdasarkan data uji yang digunakan sebanyak 669 data. Berdasarkan hasil tersebut, objek wisata di Ubud lebih cenderung menghasilkan data sentiment positif yang menandakan bahwa wisatawan mendapatkan pengalaman yang positif pada saat mengunjungi objek wisata yang ada di UBUD.[2]

Untuk melakukan analisis sentimen terhadap visitor speech di Kabupaten Purwakarta, penelitian ini akan memberikan solusinya. Dinas Pariwisata dan Kebudayaan Kabupaten Purwakarta dapat melacak perkembangan pariwisata dengan lebih baik dan membantu pengambilan keputusan dengan menggunakan analisis sentimen ini untuk menilai kinerja algoritma Support Vector Machine dalam mengklasifikasikan berdasarkan komentar dari pengunjung objek wisata Taman Air Mancur Sri Baduga. tepat dalam memajukan pariwisata saat ini.

II. METODE PENELITIAN

Adapun perancangan sistem, diperlukan metode yang menjadi acuan dan pedoman untuk diimplementasikan pada rancangan yang dibuat. Berikut teori yang akan digunakan antara lain Analisis Sentimen, Algoritma Support Vector Machine, TF-IDF, Confusion Matrix.

A. Analisis Sentimen

Analisis sentimen adalah proses penggalian, pengolahan, dan pemahaman data berupa teks tidak terstruktur secara otomatis untuk mendapatkan kembali informasi sentimen yang terkandung dalam sebuah opini atau kalimat opini. Analisis sentimen dapat diterapkan pada opini di segala bidang seperti ekonomi, politik, sosial dan hukum. Media sosial Twitter membuka jendela bagi para peneliti untuk mempelajari emosi, suasana hati, dan opini publik melalui analisis sentimen.[3]

B. Support Vector Machine

Support Vector Machine (SVM) adalah sistem pembelajaran berbasis optimisasi yang menggunakan ruang fiktif berupa fungsi linier dalam fitur berdimensi tinggi. Jika dibandingkan dengan teknik lainnya, metode Support Vector Machine (SVM) merupakan teknologi baru. Karena fungsi kernel akan menentukan ruang fitur dimana fungsi classifier akan dicari, pemilihan fungsi kernel yang tepat dan sesuai sangatlah penting.[4].

Kelas dataset diwakili oleh variabel y_i , sedangkan jumlah datanya diwakili oleh variabel x_i . Dataset dibagi menjadi dua kelas menggunakan algoritma SVM. Kelas yang tersisa memiliki nilai -1, sedangkan kelas pertama yang dibagi dengan hyperplane memiliki nilai 1.

$$X_i.W + b \geq 1 \text{ untuk } Y_i = 1$$

$$X_i.W + b \leq -1 \text{ untuk } Y_i = -1$$

Keterangan :

X_i = data ke -i

W = nilai bobot *support vector* yang tegak lurus dengan hyperplane

b = nilai bias

Y_i = kelas data ke- i

Bobot Garis vektor yang tegak lurus dengan garis *hyperplane* dan titik pusat koordinat dikenal sebagai vektor bobot (w). Posisi garis yang berhubungan dengan titik koordinat dinyatakan dengan bias (b). Pada persamaan (3) dihitung nilai b , dan pada persamaan (4) ditentukan nilai w .

$$b = -12(w.x_+ + w.x_-)$$

$$w = \sum a_i y_i x_i = 1$$

Keterangan :

b = nilai bias

$w.x_+$ = nilai bobot untuk kelas data positif

$w.x_-$ = nilai bobot untuk kelas data negative

w = bobot *vector*

a_i = nilai bobot data ke-i

y_i = kelas data ke-i

x_i = data ke-i

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

TF-IDF adalah metode untuk menekankan hubungan antara kata (istilah) dan dokumen. Frekuensi sebuah kata dalam dokumen tertentu dan kebalikan dari frekuensi kertas yang mengandung kata tersebut digabungkan dalam metode ini untuk menghitung bobot. Frekuensi sebuah kata muncul dalam dokumen memberi tahu kita seberapa penting kata itu bagi dokumen itu. Jumlah dokumen yang menggunakan kata mengungkapkan seberapa sering kata itu digunakan. Akibatnya, jika sebuah kata sering muncul dalam sebuah dokumen sementara jarang muncul di kumpulan makalah secara keseluruhan, hubungan antara istilah dan dokumen akan menjadi sangat signifikan.[5].

$$W_{i,j} = \frac{n_{i,j}}{\sum_{j=1}^p n_{j,i}} \log_2 \frac{D}{d_j}$$

dengan masing-masing keterangan yaitu:

$W_{i,j}$: pembobotan tf-idf untuk term ke-j pada dokumen ke-i
$n_{i,j}$: jumlah kemunculan term ke-j pada dokumen ke-i
P	: banyaknya term yang terbentuk
$\sum_{j=1}^p n_{j,i}$: jumlah kemunculan seluruh term pada dokumen ke-i
D	: jumlah keseluruhan dokumen
d_j	: Banyak nya dokumen yang mengandung term ke-j.

D. Confusion Matrix

Salah satu teknik untuk mengevaluasi kinerja metode klasifikasi adalah matriks konfusi. Intinya, matriks kebingungan terdiri dari data yang membandingkan hasil klasifikasi sistem dengan hasil klasifikasi yang seharusnya. Terdapat 4 (empat) istilah yang dapat digunakan untuk menunjukkan hasil dari proses kategorisasi saat mengukur kinerja dengan menggunakan matriks konfusi. True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) adalah empat kata. Banyaknya data negatif yang diketahui secara akurat disebut dengan True Negative (TN), sedangkan data False Positive (FP) adalah data negatif yang salah diidentifikasi sebagai positif[6]. Berikut adalah tabel dari sebuah perhitungan *confusion matrix*:

Tabel 1. Perhitungan *Confusion Matrix*

Nilai Prediksi	Nilai Aktual	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Keterangan:

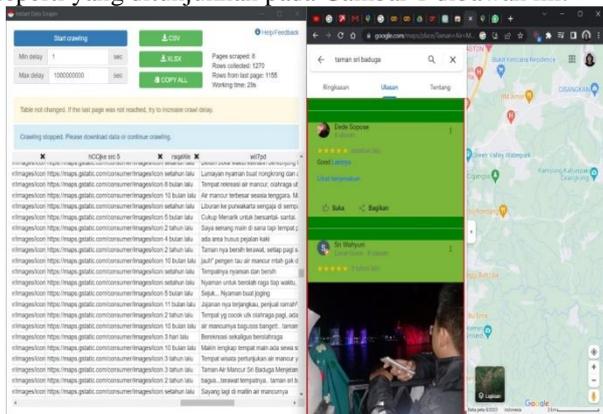
- 1. True Positive (TP)**
Hasil prediksi untuk menyatakan positif dan hasil sesungguhnya positif.
- 2. True Negative (TN)**
Hasil prediksi untuk menyatakan negatif dan hasil sesungguhnya negatif.
- 3. False Positive (FP)**
Hasil prediksi untuk menyatakan positif dan hasil sesungguhnya negatif.
- 4. False Negative (FN)**
Hasil prediksi untuk menyatakan negatif dan hasil sesungguhnya positif

III. HASIL PEMBAHASAN

Pada penelitian analisis sentimen taman air mancur sri baduga purwakarta di aplikasi google maps ini menggunakan Bahasa pemograman Python dengan menggunakan tools Google Colaboratory. Untuk pelabelan sentimen ini dilakukan secara manual dengan membaca seluruh komentar dan mengkategorikan nya 3 sentimen yaitu positif, negatif dan netral

A. Hasil *Scraping* Data

Pada penelitian ini dilakukan proses pengumpulan data, dimana data yang dikumpulkan merupakan data review dari pengguna aplikasi *Google Maps* untuk mengetahui komentar dari tempat wisata Taman Air Mancur Sri Baduga Purwakarta. Proses pengambilan data dilakukan dengan teknik *scraping* melalui *Instant Data Scraper* versi 1.0.8 yang dimiliki oleh *Google Chrome*. seperti yang ditunjukkan pada Gambar 1 dibawah ini:



Gambar 1. Hasil *Scraping* Data

B. *Preprocessing* Text

Preprocessing Tekt dipecah menjadi potongan-potongan kecil menggunakan teknik termasuk transformasi, tokenisasi, dan penyaringan. Langkah-langkah analisis dapat diaktifkan atau dinonaktifkan dan diterapkan secara berurutan. Teks yang diproses sebelumnya akan melaporkan sejumlah hal yang berbeda,

termasuk jumlah dokumen (tentang jumlah input dokumen), jumlah total token (menghitung semua token dalam korpus (kumpulan dokumen), dan token yang dilaporkan secara eksklusif pada token unik dalam korpus daripada token duplikat.[7].

Data yang dimasukkan pada tahap ini masih mentah dan belum mengalami pemrosesan apa pun, proses tersebut akan menghasilkan data bersih yang akan mempermudah proses kategorisasi. Prosedur *preprocessing* mencakup langkah-langkah seperti *case folding*, *tokenizing*, dan *filtering*.

Tabel 2. Proses *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
Air mancurnya keren, rame banget yang nonton	air mancurnya keren rame banget yang nonton
Biar Bintang yg berbicara.. BTW ada Mie Ayam & Bakso enak juga, lokasi dekat museum diorama (gd.kembar).	biar bintang yg berbicara btw ada mie ayam bakso enak juga lokasi dekat museum diorama gdkembar ...

Pada Tabel 2 yaitu menunjukkan hasil *Case Folding* dimana data tersebut sebelum di *Case Folding* terdapat huruf kapital, Lalu setelah melewati tahapan *Case Folding* semua huruf kapital diubah menjadi huruf kecil.

Tabel 3. Proses *Tokenizing*

Adem	tempatnya	Bisa	foto
Dan	Belajar	Budaya	sunda

Pada Tabel 3 setelah melakukan tahap *Case Folding* data akan dilakukan tahap selanjutnya yaitu melakukan tokenisasi untuk memecah teks atau memisahkan per-kata atau bisa disebut pemenggalan kalimat menjadi kata.

Tabel 4. Proses *Stop Words*

Sebelum <i>Stop words</i>	Sesudah <i>Stop words</i>
Bagus... Cuman sayang masih di tutup	['bagus', 'cuman', 'sayang', 'tutup']

Pada Tabel 4 Memfilter/menghentikan kata-kata, yang mencakup menghapus kata-kata berhenti dari bacaan (misalnya, menghapus dan, atau, ini...), dilakukan setelah prosedur tahapan tokenizing. Untuk mengujinya, urutkan bahasa yang akan disaring (filter). Untuk pembelajaran bahasa, server NLTK menawarkan kata berhenti yang dapat diunduh

C. *TF-IDF*

Sesudah melakukan proses term weighting setelah *preprocessing* data (transformasi, tokenization, dan filtering), setelah itu akan diberi bobot atau nilai. Bobot setiap frase dihitung untuk setiap potongan data dengan maksud untuk menentukan ketersediaan dan kemiripannya; semakin sering suatu istilah muncul dalam data, semakin tinggi bobot atau nilainya. Pendekatan *TF-IDF* diterapkan saat pembobotan kata. Dengan membagi bobot setiap kata di seluruh dokumen dengan nilai *IDF*-nya, seseorang dapat menentukan nilai *TF-IDF*.

Term Frequency

Pada tahap ini semakin besar jumlah kemunculan istilah dalam dokumen, semakin besar bobotnya. Kata yang tidak muncul akan bernilai 0, dan jika ada maka bernilai 1. Berikut hasil term frequency dapat dilihat pada tabel dibawah ini :

No	Kata	Positive	Negative	Neutral
1	air	1	0	0
2	mancur	1	0	0
3	taman	1	0	0
4	ramai	1	0	0
5	bersih	1	0	0
		5	0	0

D. Klasifikasi Support Vector Machnie

Seteah Mengikuti prosedur TF-IDF, model SVM diuji dengan Python, dan kumpulan data pelatihan dan pengujian kemudian dibagi 80:20. Seperti yang terlihat pada gambar di bawah ini, setelah mendapatkan data pelatihan untuk pengujian model tersebut.

```
#Training Model
from datetime import datetime
start_time = datetime.now()
from joblib import dump
#algoritme fitting

#text_algorithm = MultinomialNB()
text_algorithm = svm.SVC(kernel="rbf", C=1.0)

model = text_algorithm.fit(x_train, y_train)

# save the model to disk
dump(model, filename="model_sentiment_naive.joblib")

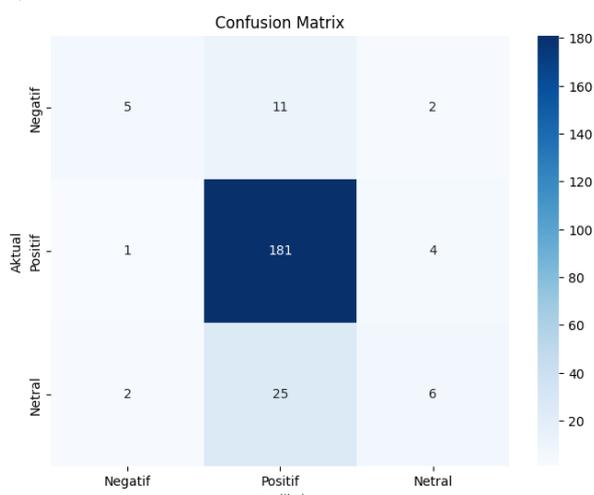
end_time = datetime.now()
result_time = end_time-start_time
print("Duration:",result_time)

#y_train
Duration: 0:00:00.639705
```

Gambar 2. Hasil Uji Testing

E. Confusion Matrix

Setelah tahap evaluasi pada titik ini menggunakan pendekatan validasi silang, dan hasil akurasi, presisi, dan daya ingat nantinya akan diuji menggunakan Matriks Kebingungan. Matriks konfusi yang dihasilkan setelah menggunakan alat *Google Collab* untuk mengevaluasi keseluruhan data peta. dapat dilihat pada Gambar dibawah ini:



Gambar 3. Hasil Confusion Matrix

F. Word Cloud

Selanjutnya tahap terakhir adalah visualisasi yang digambarkan dengan *word cloud* pada sentiment keseluruhan dokumen negatif positif dan netral tentang wisata taman air mancur sribaduga purwakarta. Hasil dari Visualisasi ditampilkan pada gambar dibawah ini :

Word Cloud Positive pada data *maps* sentimen keseluruhan:



Gambar 4. Hasil Word Cloud Sentimen Positive

Dari hasil Gambar diatas, merupakan hasil visualisasi dari *word cloud Positive*. Terdapat 22 kata yang sering muncul diantaranya “purwakarta”, “mancur”, “nyaman” dan sebagainya.

Word Cloud Negative pada data *maps* sentimen keseluruhan:



Gambar 5. Hasil Word Cloud Sentimen Negative

Dari hasil Gambar diatas, merupakan hasil visualisasi dari *word cloud Negative*. Terdapat 20 kata yang sering muncul diantaranya “tutup”, “covid”, “pandemi” dan sebagainya.

Word Cloud Neutral pada data *maps* sentimen keseluruhan



Gambar 6. Hasil Word Cloud Sentimen Neutral

Dari hasil Gambar diatas, merupakan hasil visualisasi dari word cloud *Neutral*. Terdapat 18 kata yang sering muncul diantaranya “tutup”, “mancur”, “buka” dan sebagainya.

IV. KESIMPULAN

Dari hasil penelitian diatas mengenai analisis *sentiment* pada wisata taman air mancur sribaduga purwakarta pada *platform google maps* yang berjumlah 1155 data dan sudah melewati tahap *preprocessing text* yang terdiri dari filtering, tokenization, transformation dan klasifikasi menggunakan algoritma *support vector machine* serta evaluasi data dengan *confusion matrix* menggunakan *tools Google Collab* dapat disimpulkan bahwa hasil dari tanggapan masyarakat mengenai wisata taman air mancur sri baduga dengan hasil komentar positif 872, negative 156 Dan netral 127. Adapun persentase 81% pada akurasi (accuracy), 94% pada nilai precision, dan tingkat keberhasilan (recall) 99%. Nilai ini menunjukkan bahwa klasifikasi algoritma support vector machine dinilai cukup baik dalam pemrosesan dokumen google maps, dikarenakan hasil persentase akurasi sebesar 81%. Berdasarkan nilai tersebut membuktikan sentiment pada wisata taman air mancur sribaduga purwakarta pada *platform google maps* tergolong baik.

DAFTAR PUSTAKA

- [1] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.
- [2] I Wayan Budi Suryawan, Nengah Widya Utami, and Ketut Queena Fredlina, “Analisis Sentimen Review Wisatawan Pada Objek Wisata Ubud Menggunakan Algoritma Support Vector Machine,” *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 133–140, 2023, doi: 10.51401/jinteks.v5i1.2242.
- [3] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [4] F. A. Muttaqin and A. M. Bachtiar, “Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak ‘Dodo Kids Browser,’” *J. Ilm. Komput. dan Inform.*, pp. 1–8, 2016.
- [5] M. Nurjannah and I. Fitri Astuti, “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,” *J. Inform. Mulawarman*, vol. 8, no. 3, pp. 110–113, 2013.
- [6] Karsito and Susanti Santi, “Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naive Bayes Di Perumahan Azzura Residence,” *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.
- [7] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, “Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes,” *INSOLOGI J. Sains dan Teknol.*, vol. 1, no. 4, pp. 411–423, 2022, doi: 10.55123/insologi.v1i4.770.