



Penerapan Metode Support Vector Machine untuk Analisis Sentimen pada Review Pelanggan Hotel

Siti Azza Amira¹⁾✉, Satria Utama¹⁾, dan Muhammad Hanif Fahmi¹⁾

¹Jurusan Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Raden Rahmat, Indonesia

Info Artikel

Sejarah Artikel:

Diterima: November

2020

Direvisi: Desember 2020

Disetujui: Desember 2020

Keywords:

Support Vector Machine,

Klasifikasi, TF-IDF,

Analisis Sentimen

Abstrak

Perkembangan internet yang semakin pesat membuat banyak orang mengakses internet untuk mendapatkan berbagai macam informasi, salah satunya dalam mencari informasi ulasan hotel. Ulasan tersebut biasanya akan dijadikan sebagai acuan dan sangat berpengaruh bagi calon pengunjung dalam menentukan hotel mana yang terbaik untuk mereka kunjungi ketika bepergian ke luar kota. Maka dari itu dapat dilakukan analisis sentimen untuk mengklasifikasikan ulasan tersebut apakah bersifat positif atau negatif. Proses klasifikasi sentimen menggunakan metode Support Vector Machine dan pembobotan kata menggunakan bantuan metode Term Frequency-Inverse Document Frequency. Data yang digunakan didapatkan dengan melakukan crawling pada situs Tripadvisor dengan menggunakan library scrapy pada python. Berdasarkan hasil pengujian, Tingkat akurasi yang didapatkan sebesar 88%, nilai precision sebesar 98%, nilai recall sebesar 88%, dan nilai F1 Score sebesar 83%.

Abstract

The rapid development of the internet has made many people access the internet to get various kinds of information, one of them is to find information on hotel reviews. These reviews are usually used as a reference and very influential for potential visitors in determining which hotel is best for them to visit when traveling out of town. Therefore a sentiment analysis can be carried out to classify the review whether it is positive or negative. The sentiment classification process uses the Support Vector Machine method and word weighting uses the Term Frequency-Inverse Document Frequency method. Data obtained by crawling the Tripadvisor site using the Python scrapy library. Based on the test results, the accuracy rate obtained is 88%, precision value is 98%, recall value is 88%, and F1 score value is 83%.

© 2020 Universitas Negeri Semarang

✉ Alamat korespondensi:

Jurusan Sistem Informasi, Universitas Islam Raden Rahmat
Jl. Raya Mojosari, No.2, Kepanjen, Malang
E-mail: siti_azza_amira@uniramalang.ac.id

ISSN 2252-6811

E-ISSN 2599-297X

PENDAHULUAN

Pemanfaatan internet sekarang ini sudah mencakup di dalam banyak bidang, seperti sebagai media untuk mendapat berbagai macam informasi, media untuk menjalankan bisnis (berjualan secara *online* maupun berbelanja *online*), media untuk mendapatkan hiburan (musik, film, majalah), hingga untuk mengakses media sosial seperti instagram, twitter dan facebook. Sekarang ini banyak orang yang menggunakan internet untuk mempermudah mereka dalam melakukan sesuatu salah satunya dalam hal untuk memperoleh informasi terkait ulasan hotel. Semakin banyaknya hotel yang dapat dituju saat bepergian ke luar kota dapat membuat calon pengunjung kebingungan untuk menentukan hotel mana yang cocok untuk mereka kunjungi.

Dari sekian banyak situs yang dapat menyajikan ulasan dari sebuah hotel, Tripadvisor adalah salah satu yang paling banyak diakses. Berdasarkan data yang dirilis Skift pada tahun 2013 dengan melihat statistik dari *similiarweb* dalam melihat *traffic* pengunjung selama bulan oktober 2013, Tripadvisor berada di peringkat dua dari 10 situs perjalanan *online* di dunia yang paling banyak diakses dan dimanfaatkan oleh para wisatawan dengan total 48,5 juta pengunjung.

Ulasan tentunya dapat mempengaruhi pengambilan keputusan calon pengunjung selanjutnya dalam menentukan tempat yang ingin mereka kunjungi termasuk salah satunya adalah hotel. Pada penelitian yang dilakukan oleh Rimba Nuzulul Chory, Muhammad Nasrun dan Casi Setianingsih dikatakan perlu adanya analisis klasifikasi terhadap komentar dari pengguna untuk melihat bagaimana kepuasan publik dalam menggunakan layanan (Chory, R. N., dkk, 2018). Penelitian juga dilakukan oleh Khotimah dan Riyanto Sarno pada penelitian dengan judul "Sentimen Detection of Comment Titles in Booking.com Using Probabilistic Latent Semantic Analysis" untuk melakukan klasifikasi terhadap ulasan data pelanggan pada situs booking.com dengan menggunakan metode PLSA (Khotimah, Sarno R., 2018).

Pada fitur ulasan di berbagai situs biasanya tidak dapat dilakukan pemilihan untuk menampilkan mana ulasan yang bersifat positif (ulasan dengan opini baik) atau yang bersifat negatif (ulasan dengan opini buruk), sehingga dapat dilakukan analisis sentimen untuk dapat membantu melakukan klasifikasi dari ulasan dari pengunjung tempat tersebut sebelumnya. Ulasan tentunya ditulis berdasarkan apa yang mereka rasakan saat menulis ulasan tersebut sehingga

ulasan tersebut dapat dikatakan jujur tanpa adanya rekayasa. Analisis sentimen adalah sebuah ilmu yang mempelajari bagaimana untuk melakukan analisa terhadap opini, sentimen, evaluasi, penilaian, sikap dan emosi dari sebuah entitas dimana hal tersebut dapat berupa produk, pelayanan, individu, isu-isu, peristiwa, organisasi, dan topik (Sawakoshi, Okada M, Hashimoto K., 2015). Analisis sentimen dapat diterapkan pada beberapa jenis tingkatan yang berbeda entah itu teks yang berupa dokumen ataupun kalimat. Ulasan dari wisatawan dipengaruhi oleh emosi (sentimen) sehingga dapat dikelompokkan atau diklasifikasi agar dapat ditentukan kepolarisasiannya, yaitu positif atau negatif (Indriati, Ridok A., 2016)

Terdapat banyak metode yang dapat digunakan untuk melakukan klasifikasi teks. Salah satu metode yang dapat digunakan adalah *Support Vector Machine (SVM)*. Metode *Support Vector Machine (SVM)* merupakan sebuah metode untuk melakukan klasifikasi, dimana metode ini banyak digunakan untuk membantu di dalam penyelesaian permasalahan dalam banyak bidang, baik penyelesaian dalam masalah *gene expression analysis*, dalam permasalahan bidang finansial, bahkan dalam bidang kedokteran dan cuaca. Pada bidang analisis sentimen sendiri implementasi metode *Support Vector Machine (SVM)* sudah banyak digunakan, dikarenakan metode ini dapat memberikan hasil yang lebih baik jika dibandingkan dengan metode klasifikasi sejenis seperti *Artificial Neural Network (ANN)* terutama di dalam menemukan solusi dikarenakan SVM dapat menemukan solusi yang global optimal (Santosa B., 2010).

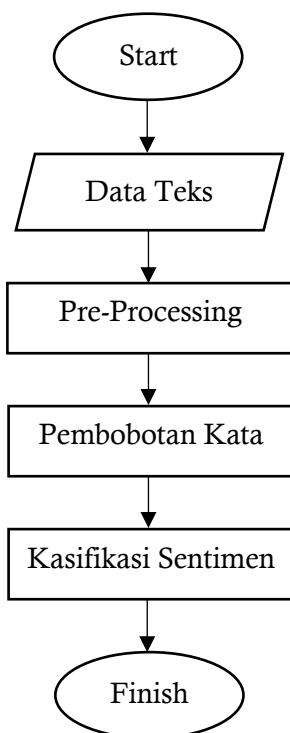
Berdasarkan apa yang sudah dijelaskan sebelumnya maka pada penelitian ini dilakukan analisis sentimen pada ulasan hotel dengan menerapkan metode *Support Vector Machine (SVM)* untuk melakukan klasifikasi data. Dalam melakukan pembobotan kata atau *feature extraction* digunakan bantuan dari metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Sedangkan untuk data yang digunakan dalam proses penelitian ini adalah data ulasan dari Tripadvisor yang diperoleh dengan melakukan proses *crawling* data menggunakan *library scrapy* pada python.

METODE PENELITIAN

Analisis sentimen disebut juga *opinion mining* yang berguna di dalam pengelolaan bahasa alami, *text mining*, dan komputasi linguistik dan bertujuan untuk menentukan opini pada suatu topik tertentu, dimana perilaku tersebut dapat mengindikasikan penilaian serta

alasan serta kondisi kecenderungan (Basari, 2013). Analisis sentimen kebanyakan digunakan untuk melakukan analisis atau untuk dapat melakukan penilaian terhadap opini masyarakat baik itu opini yang merujuk kesukaan atau ketidaksukaan terhadap suatu barang maupun jasa. Sentimen tersebut merupakan informasi yang bersifat subjektif dan memiliki nilai polaritas yang positif dan negatif dimana nilai polaritas ini dapat digunakan sebagai parameter untuk dapat menentukan sebuah keputusan (Indriati, Ridok A., 2016).

Metode yang dilakukan dalam penelitian ini yaitu pengumpulan data teks, *pre-processing*, feature extraction (pembobotan kata), dan klasifikasi sentimen.



Gambar 1. Metodologi penelitian

A. Pengumpulan Data

Pada tahap ini akan dilakukan pengumpulan data untuk digunakan dalam melakukan penelitian. Data yang digunakan adalah data yang terkait dengan ulasan pada situs Tripadvisor. Kemudian proses pengambilan data dilakukan dengan menggunakan *library* scrapy pada python dan ulasan mentah yang sudah didapatkan akan disimpan dengan bentuk file CSV. Data mentah yang sudah didapatkan akan dipisahkan lagi berdasarkan tanda titik (.) jadi ketika dalam suatu ulasan terdapat dua kalimat maka data pada satu dokumen tersebut akan di *split* dan dihitung sebagai dua ulasan.

B. *Pre-processing* Data

Setelah melakukan *crawling* data berupa ulasan hotel maka selanjutnya data tersebut akan dilanjutkan ke tahap *pre-processing*. Pada proses *pre-processing* data terdapat beberapa tahapan yang dilakukan yaitu, *case folding*, *tokenizing*, *filtering*, dan *stemming*.

1. *Case Folding*

Pada tahapan ini data ulasan yang sudah didapatkan akan diubah teks yang hurufnya tidak konsisten tadi menjadi huruf kecil semua dan akan dilakukan penghilangan karakter yang tidak penting (selain huruf) yaitu tanda baca dan angka sehingga hasil akhirnya dokumen/teks tersebut sudah kecil semua hurufnya dan tanda bacanya sudah hilang.

2. *Tokenizing*

Pada tahap ini ulasan yang sudah hurufnya sudah kecil semua dan sudah hilang tanda bacanya tersebut akan di-*tokenizing* untuk dipecah katanya menjadi beberapa bagian sehingga akan terpisah menjadi kata-kata.

3. *Filtering*

Pada tahap *Filtering* ini hasil dari tahap sebelumnya akan dieliminasi kata-kata yang dianggap tidak penting.

4. *Stemming*

Setelah di filter maka kata-kata tersebut akan dipangkas jika masih terdapat awalan, akhiran atau sisipan agar menjadi suatu kata dasar. Dari kata dasar ini lah nantinya yang akan dilakukan pembobotan katanya sebelum masuk ke tahap klasifikasi.

C. Pembobotan Kata

Pada tahapan ini akan dilakukan pembobotan terhadap kata dasar yang sudah didapatkan dari hasil proses *pre-processing*. Pembobotan kata atau *feature extraction* dilakukan dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Dimana pada implementasinya, proses pembobotan kata dilakukan dengan memanfaatkan *module* *scikit-learn* yang terdapat pada python.

D. Klasifikasi Sentimen

Setelah dihasilkan topik dari setiap ulasan selanjutnya akan ditentukan *sentiment* dari setiap ulasan tersebut untuk diketahui opininya yaitu bersifat positif atau negatif. Kemudian metode *Support Vector Machine* digunakan dalam penentuan *sentiment* apakah positif atau negatif.

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan adalah data ulasan hotel yang terdapat pada situs TripAdvisor. Data yang didapatkan akan dibagi menjadi data *training* dan data *testing*. Data diambil dengan melakukan *crawling* menggunakan library scrapy pada python. Data mentah yang didapatkan dari hasil *crawling* akan diolah terlebih dahulu dimana data ulasan yang terdapat lebih dari satu kalimat pada setiap ulasannya akan di-*split* dan akan diberikan id pada masing-masing dokumen. Proses untuk *split* data dilakukan jika pada satu ulasan tersebut terdapat tanda baca (.) dikarenakan adanya kemungkinan adanya perbedaan sentimen dari komentar pada setiap kalimat.

Tabel 1. Data Ulasan Hasil Crawling pada Tripadvisor

Doc	Data Ulasan Hotel
1.	Staff hotel ramah dan sigap membantu saat pertama kali tiba di hotel, Sarapan variatif dan rasanya enak, Harganya relatif lebih murah.
2.	Tapi akan lebih bagus lagi seandainya disekitar hotel ada minimarket sehingga tidak menyulitkan jika ingin belanja.
3.	Hotel bisnis yang sangat oke disini staffnya sangat ramah membantu dari mulai masuk sudah disambut makanan di restoran juga enak enak kamar nya nyaman
4.	Pelayanan menyenangkan, Ketika masuk disambut dengan hangat dan langsung diberikan minuman.
5.	Mungkin perlu di tambah semacam minimarket di area restoran dan kolam renang supaya fasilitas lebih seru dan tidak membuat bosan.
6.	Lokasinya strategis di tengah kota.

B. Pre-processing Data

Pre-processing data dilakukan dengan melakukan beberapa tahap yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Tahap *pre-processing* dilakukan dengan menggunakan *Natural Language Toolkit* (NLTK) pada python. Hasil dari tahapan adalah dokumen yang kalimatnya sudah terpecah setiap katanya dan menjadi kata dasar.

Tahapan pertama yang dilakukan adalah proses *case folding*, yaitu mengubah teks yang awalnya mempunyai huruf tidak konsisten atau masih tercampur antara huruf besar dan kecil menjadi huruf kecil semua serta dilakukan penghilangan karakter selain huruf, seperti angka dan tanda baca. Hasil akhir dari proses ini adalah

ulasan yang hurufnya sudah menjadi kecil semua serta angka dan tanda baca nya sudah hilang.

Tabel 2. Data Ulasan Setelah *Case folding*

Doc	Ulasan Setelah Proses <i>Case folding</i>
1.	staff hotel ramah dan sigap membantu saat pertama kali tiba di hotel sarapan variatif dan rasanya enak harganya relatif lebih murah
2.	tapi akan lebih bagus lagi seandainya disekitar hotel ada minimarket sehingga tidak menyulitkan jika ingin belanja
3.	hotel bisnis yang sangat oke disini staffnya sangat ramah membantu dari mulai masuk sudah disambut makanan di restoran juga enak enak kamar nya nyaman
4.	pelayanan menyenangkan Ketika masuk disambut dengan hangat dan langsung diberikan minuman
5.	mungkin perlu di tambah semacam minimarket di area restoran dan kolam renang supaya fasilitas lebih seru dan tidak membuat bosan
6.	lokasinya strategis di tengah kota

Selanjutnya dilakukan proses *tokenizing* yaitu pemecahan kalimat menjadi beberapa bagian sehingga akan terpisah menjadi kata-kata.

Tabel 3. Data Ulasan Setelah *Tokenizing*

Doc	Ulasan Setelah Proses <i>Tokenizing</i>
1.	['staff', 'hotel', 'ramah', 'dan', 'sigap', 'membantu', 'saat', 'pertama', 'kali', 'tiba', 'di', 'hotel', 'sarapan', 'variatif', 'dan', 'rasanya', 'enak', 'harganya', 'relatif', 'lebih', 'murah']
2.	['tapi', 'akan', 'lebih', 'bagus', 'lagi', 'seandainya', 'disekitar', 'hotel', 'ada', 'minimarket', 'sehingga', 'tidak', 'menyulitkan', 'jika', 'ingin', 'belanja']
3.	['hotel', 'bisnis', 'yang', 'sangat', 'oke', 'disini', 'staffnya', 'sangat', 'ramah', 'membantu', 'dari', 'mulai', 'masuk', 'sudah', 'disambut', 'makanan', 'di', 'restoran', 'juga', 'enak', 'enak', 'kamar', 'nya', 'nyaman']
4.	['pelayanan', 'menyenangkan', 'Ketika', 'masuk', 'disambut', 'dengan', 'hangat', 'dan', 'langsung', 'diberikan', 'minuman']
5.	['mungkin', 'perlu', 'di', 'tambah', 'semacam', 'minimarket', 'di', 'area', 'restoran', 'dan', 'kolam', 'renang', 'supaya', 'fasilitas', 'lebih', 'seru', 'dan', 'tidak', 'membuat', 'bosan']
6.	['lokasinya', 'strategis', 'di', 'tengah', 'kota']

Tahapan selanjutnya adalah proses *Filtering* yaitu melakukan eliminasi kata-kata yang dianggap tidak penting atau tidak memiliki makna seperti kata penghubung.

Tabel 4. Data Ulasan Setelah Filtering

Doc	Ulasan Setelah Proses Filtering
1.	['hotel', 'ramah', 'sigap', 'membantu', 'pertama', 'hotel', 'sarapan', 'variatif', 'rasanya', 'enak', 'harganya', 'relatif', 'murah']
2.	['lebih', 'bagus', 'hotel', 'minimarket', 'menyulitkan', 'belanja']
3.	['hotel', 'hotel', 'bisnis', 'ramah', 'membantu', 'makanan', 'restoran', 'enak', 'enak', 'kamar', 'nyaman']
4.	['pelayanan', 'menyenangkan', 'hangat', 'minuman']
5.	['tambah', 'semacam', 'minimarket', 'area', 'restoran', 'kolam', 'renang', 'fasilitas', 'seru', 'membuat', 'bosan']
6.	['lokasinya', 'strategis', 'tengah', 'kota']

Tahap yang terakhir adalah proses *stemming*, yaitu kata-kata tersebut selanjutnya akan di pangkas jika awalan, akhiran atau sisipan masih terdapat di dalam kata sehingga dapat menjadi suatu kata dasar. Kata dasar ini yang nantinya akan digunakan pada proses pembobotan.

Tabel 5. Data Ulasan Setelah Stemming

Doc	Ulasan Setelah Proses Stemming
1.	['hotel', 'ramah', 'sigap', 'bantu', 'pertama', 'hotel', 'sarapan', 'variatif', 'rasa', 'enak', 'harga', 'relatif', 'murah']
2.	['lebih', 'bagus', 'hotel', 'minimarket', 'sulit', 'belanja']
3.	['hotel', 'bisnis', 'ramah', 'bantu', 'makanan', 'restoran', 'enak', 'enak', 'kamar', 'nyaman']
4.	['layan', 'senang', 'hangat', 'minuman']
5.	['tambah', 'macam', 'minimarket', 'area', 'restoran', 'kolam', 'renang', 'fasilitas', 'seru', 'buat', 'bosan']
6.	['lokasi', 'strategis', 'tengah', 'kota']

C. Pembobotan Kata

Setelah dilakukan *pre-processing* maka *term list* yang didapat dari proses tersebut akan dihitung bobot angka sesuai dengan jumlah kemunculannya di setiap dokumen. Proses pembobotan kata dilakukan dengan menggunakan TF-IDF.

Tahap pertama dalam metode TF-IDF adalah dengan menghitung nilai *term frequency* (TF) dari setiap kata. TF adalah jumlah kemunculan kata di dokumen yang tersebut. Contoh dari perhitungan nilai TF yaitu jika pada sebuah kalimat 'hotel' muncul sebanyak 2 kali pada dokumen maka nilai TF dari 'hotel' adalah 2.

Tabel 6. Nilai TF

Term	TF					
	D1	D2	D3	D4	D5	D6
hotel	2	1	2	0	0	0
ramah	1	0	1	0	0	0
sigap	1	0	0	0	0	0
bantu	1	0	1	0	0	0
pertama	1	0	0	0	0	0
sarapan	1	0	0	0	0	0
variatif	1	0	0	0	0	0
rasa	1	0	0	0	0	0
enak	1	0	2	0	0	0
harga	1	0	0	0	0	0
relatif	1	0	0	0	0	0
murah	1	0	0	0	0	0
lebih	0	1	0	0	0	0
bagus	0	1	0	0	0	1
minima	0	1	0	0	1	0
rket	0	1	0	0	1	0
...
Kota	0	0	0	0	0	1

Setelah menghitung nilai TF maka akan dilakukan perhitungan nilai IDF. Nilai *Inverse Document Frequency* (IDF) didapatkan dengan menghitung menggunakan persamaan:

$$idf_t = \log_{10} \frac{N}{df_t} \quad (1)$$

Berikut adalah contoh perhitungan nilai IDF pada kata 'hotel'. Hasil perhitungan selengkapnya terdapat pada Tabel 7.

$$idf_t = \log_{10} \frac{6}{3} = 0,30103$$

Tabel 7. Nilai IDF

Term	DF	IDF
hotel	3	0,30103
ramah	2	0,47712
sigap	1	0,77815
bantu	1	0,47712
pertama	1	0,77815
sarapan	1	0,77815
variatif	1	0,77815
rasa	1	0,77815
enak	2	0,47712
harga	1	0,77815
relatif	1	0,77815
murah	1	0,77815
lebih	1	0,77815
bagus	2	0,47712
minimarket	2	0,47712
...
Kota	1	0,77815

Selanjutnya dilakukan perhitungan nilai TF-IDF. Nilai TF-IDF didapatkan dengan mengkalikan hasil dari nilai *document frequency (TF)* dengan *inverse document frequency (IDF)*.

Tabel 8. Nilai TF*IDF

Term	TF*IDF					
	D1	D2	D3	D4	D5	D6
hotel	0,602	0,301	0,602	0	0	0
ramah	0,477	0	0,477	0	0	0
sigap	0,778	0	0	0	0	0
bantu	0,477	0	0,477	0	0	0
pertama	0,778	0	0	0	0	0
sarapan	0,778	0	0	0	0	0
variatif	0,778	0	0	0	0	0
rasa	0,778	0	0	0	0	0
enak	0,477	0	0,954	0	0	0
harga	0,778	0	0	0	0	0
relatif	0,778	0	0	0	0	0
murah	0,778	0	0	0	0	0
lebih	0	0,778	0	0	0	0
bagus	0	0,477	0	0	0	0,477
minimarket	0	0,477	0	0	0,477	0
...
Kota	0	0	0	0	0	0,653

Tabel 8 menyajikan data hasil perhitungan nilai TF-IDF. Perhitungan dilakukan menggunakan persamaan:

$$W_{t,d} = Wtf_{t,d} \times idf_t \quad (2)$$

Berikut adalah contoh perhitungan nilai TF-IDF pada kata 'hotel'.

$$W_{t,d} = * 0,30103 = 0,60206$$

D. Klasifikasi Sentimen dan Pengujian

Klasifikasi dihitung menggunakan metode *Support vector machine*. Tahap pertama yang dilakukan adalah menghitung *training*. Tahap pertama dari perhitungan *training* dengan melakukan perhitungan kernel. Fungsi kernel yang digunakan adalah kernel *polynomial*.

$$K(x_i, x_d) = (X_i^T X_d + 1)^d, y > 0 \quad (3)$$

Tabel 9. Nilai Kernel

K(xi,xi)	D1	D2	D3	D4	D5	D6
D1	87,6	1,5	4,8	1	1	1
D2	1,5	20,3	1,5	1	2,03	1,5
D3	4,8	1,55	36,3	1	2,03	1
D4	1	1	1	21,5	1	1
D5	1	2,0	2,03	1	100,9	1
D6	1	1,5	1	1	1	1

Kemudian dilakukan inisialisasi terhadap parameter yang akan digunakan, yaitu nilai $C, \lambda, \gamma, \epsilon$, dan iterasi maksimum.

Table 10. Nilai Insialisasi Parameter

c	1
λ	0,5
γ	0,1
ϵ	0,0001
Iterasi Max	3

Selanjutnya dilakukan inisialisasi nilai $\alpha=0$ dan menghitung matriks hessian. Matriks hessian didapat dari perkalian antar kernel polynomial dan y dimana merupakan vector bernilai 1 dan -1. Perhitungan matriks hessian dilakukan dengan menggunakan persamaan:

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (4)$$

Tabel 11. Matriks Hessian

K(xi,xi)	D1	D2	D3	D4	D5	D6
D1	87,8	-1,80	5,08	1,25	-1,250	1,2
D2	-1,80	20,5	-1,80	-1,25	2,285	-1,7
D3	5,08	-1,80	36,6	1,25	-2,285	1,2
D4	1,25	-1,25	1,25	21,8	-1,250	1,2
D5	-1,25	2,28	-2,28	-1,25	101,2	-1,2
D6	1,25	-1,75	1,25	1,25	-1,250	1,2

Setelah mendapatkan nilai matriks hessian maka akan dilakukan perhitungan iterasi, $\delta\alpha_i$ dan memperbarui nilai α_i dengan menggunakan persamaan:

$$E_i = \sum_{j=1}^n \alpha_i D_{ij} \quad (5)$$

Menghitung nilai $\delta\alpha_i$ dapat dihitung menggunakan persamaan:

$$\delta\alpha_i = \min(\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i) \quad (6)$$

Memperbarui nilai α_i dapat dihitung menggunakan Persamaan 7.

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (7)$$

Keterangan :

α_i = alfa ke- i

E_i = Error rate

D_{ij} = Matriks Hessian

C = Konstanta C

γ = konstanta $gamma$

$\delta\alpha_i$ = delta alfa ke- i

Hasil dari perhitungan nilai E pada iterasi ke-0 terdapat pada Tabel 12. Sedangkan hasil dari perhitungan $\delta\alpha$ pada iterasi ke-0 terdapat pada Tabel 13.

Tabel 12. Nilai E pada Iterasi ke-0

Iterasi	Ei	D1	D2	D3	D4	D5	D6
		0	0	0	0	0	0

Tabel 13. Nilai $\delta\alpha$ pada Iterasi ke-0

$\delta\alpha$	D1	D2	D3	D4	D5	D6
	0	0	0	0	0	0

Setelah didapatkan nilai $\delta\alpha$ dilakukan perhitungan untuk mendapatkan α_i . Hasil terdapat pada Tabel 14.

Tabel 14. Nilai α pada Iterasi ke-0

D1	D2	D3	D4	D5	D6
0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

Kemudian dilakukan perhitungan terus menerus sampai mencapai iterasi maksimum agar mendapatkan nilai α_i yang diperlukan untuk mencari *support vector*. Tabel 15 menunjukkan hasil dari perhitungan α_i pada iterasi ke-3.

Tabel 15. Nilai α pada Iterasi ke-3

D1	D2	D3	D4	D5	D6
0,084	0,091	0,084	0,082	0,113	0,264

Selanjutnya dilakukan perhitungan untuk mencari *support vector* masing-masing dokumen. Berdasarkan nilai α_i terakhir yang didapat, maka diambil nilai α_i terbesar dari masing-masing kelas. Berikut hasil dari pemilihan α_i terbesar dari masing-masing kelas.

Tabel 16. α_i Terbesar dari Masing-Masing Kelas

Alfa	D1	D2	D3	D4	D5	D6
Kelas	1	-1	1	1	-1	1
α	0	0	0	0	0	0
α	0,10	0,10	0,10	0,10	0,10	0,10
α	0	0,037	0	0	0	0,180
α	0,08	0,091	0,08	0,08	0,11	0,264
X+ =						0,26
X- =						0,11

Kemudian dilakukan perhitungan nilai bias b dengan menggunakan Persamaan 8. Pada Tabel 17 ditunjukkan hasil perhitungan nilai bias.

$$b = -\frac{1}{2} \left(\sum_{i=1}^N \alpha_i y_i K(x_i, x^-) \right) + \sum_{i=1}^N \alpha_i y_i K(x_i, x^+) \quad (8)$$

Tabel 17. Nilai Bias

xi	K(xi,x+)	K(xi,x-)	$\alpha_i y_i K(x_i, x^+)$	$\alpha_i y_i K(x_i, x^-)$
D1	1	1,557	0,084	0,131
D2	1,507	20,341	-0,138	-1,870
D3	1	1,557	0,0842	0,1312
D4	1	1	0,0821	0,0821
D5	1	2,035	-0,1139	-0,2319
D6	1	1,507	0,2641	0,3980
To-tal	$\sum_{i=0}^N \alpha_i y_i K(x_i, x^+)$		0,26229	
	$\sum_{i=0}^N \alpha_i y_i K(x_i, x^-)$			-1,35984
bias			-0,548778948	

Setelah dilakukan perhitungan training, maka selanjutnya barulah dilakukan perhitungan *testing*. Perhitungan *testing* dilakukan dengan menggunakan persamaan berikut:

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x) + b \quad (9)$$

Sebagai contoh *Testing* pada 3 dokumen yang diambil dari dokumen training 1, 2 dan 3. Dimana Nilai $f(x)$ didapat dari hasil penjumlahan α_i dikali hasil perhitungan kernel dikali kelas ditambah b. Jika fungsi bernilai positif maka diklasifikasikan pada kelas sentimen positif, sebaliknya jika fungsi bernilai negatif maka diklasifikasikan pada kelas sentimen negatif. berikut hasil klasifikasi *support vector machine* terhadap data 3 data *testing*.

Tabel 18. Nilai Bias

SV	T1	T2	T3
D1	0,0027	0,00273	0,00498
D2	-0,00600	-0,00856	-0,00660
D3	0,00586	0,00835	0,00644
D4	0,00293	0,00293	0,00293
D5	-0,00271	-0,00271	-0,00271
D6	-0,00299	-0,00299	-0,00299
Total	-0,00019	-0,00025	0,00205
Total+ bias	-0,02269	-0,00025	0,00205
Kelas	-1	-1	1

Berdasarkan 223 data ulasan yang digunakan untuk penelitian, setelah dilakukan klasifikasi menggunakan *support vector machine* terdapat 177 ulasan yang diklasifikasikan ke kelas positif dan 46 ulasan diklasifikasikan ke kelas negatif. Selanjutnya dilakukan pengukuran evaluasi untuk mengukur tingkat keberhasilan suatu *classifier*. Pada umumnya untuk mengukur evaluasi digunakan *confusion matrix*.

Tabel 19. *Confusion Matrix*

		Kelas Prediksi		Total
		TRUE	FALSE	
Kelas	TRUE	TP (True Positive)	FN (False Negative)	P
	Sebenarnya	FP (False Positive)	TN (True Negative)	
Kelas	FALSE		N (True Negative)	
	Sebenarnya	(False Positive)		
Total		P'	N'	P+N

Pengujian dilakukan dengan menghitung nilai akurasi, *precision*, *recall* dan F1 score. Nilai *confusion matrix* berdasarkan 223 data ulasan adalah TP = 173, TN = 24, FP = 4, FN = 22.

Perhitungan akurasi menggunakan rumus:

$$Akurasi = \frac{TP+TN}{TN+TP+FP+FN} \quad (10)$$

Berdasarkan perhitungan nilai akurasi yang didapatkan sebesar:

$$Akurasi = \frac{173+24}{24+173+4+22} = 0,88 = 88\%$$

Perhitungan nilai *precision* menggunakan rumus:

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

Berdasarkan perhitungan nilai *precision* yang didapatkan sebesar:

$$Precision = \frac{173}{173+4} = 0,98 = 98\%$$

Perhitungan nilai *recall* menggunakan rumus:

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

Berdasarkan perhitungan nilai *recall* yang didapatkan sebesar:

$$Recall = \frac{173}{173+22} = 0,89 = 89\%$$

Perhitungan nilai *F1 score* menggunakan rumus:

$$F1 Score = 2 \times \frac{Precision \times recall}{Precision + recall} \quad (13)$$

Berdasarkan perhitungan nilai *F1 score* yang didapatkan sebesar:

$$F1 Score = 2 \times \frac{0,98 \times 0,89}{0,98 + 0,89} = 0,93 = 93\%$$

SIMPULAN

Metode *Support vector machine* yang dikombinasikan dengan TF-IDF dapat menyelesaikan permasalahan dalam klasifikasi sentimen. Hal ini dibuktikan dengan mampunya TF-IDF dalam memberikan nilai bobot pada suatu kata dan kemampuan *Support vector machine* dalam memberikan label sentimen pada masing-masing ulasan.

DAFTAR PUSTAKA

- Amira, S. A., Irawan, M. I. (2019). *Opinion Analysis of Traveler Base on Tourism Site Review Utilize Sentiment Analysis Using Support vector machine Method*. IPTEK The Journal for Technology and Science.
- Basari (2013). *Opinion Mining of Movie Review using Hybrid Method of SVM and partical swarm optimazion*.
- Chory, R. N., Nasrun, M., Setianingsih, C., (2018). *Sentiment Analysis on User Satisfaction Level of Mobile Data Services Using Support vector machine (SVM) Algorithm*. The 2018 IEEE IoTaIS.
- Indriati, Ridok A. (2016). *Sentiment Analysis For Review Mobile Application Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN)*. JEEST 2016.
- Khotimah, Sarno R. (2018). *Sentimen Detection of Comment Titles in Booking.com Using Probabilistic Latent Semantic Analysis*. 6th ICoICT 2018.
- Liu & Bing (2012). *Sentiment Analysis and Opinion Mining*.
- Liu & Bing (2010). *Sentiment Analysis and Subjectivity, Handbook of Natural Language Processing*.
- Santosa B. (2010). Tutorial Support vector machine. Teknik Industri ITS.
- Sawakoshi, Okada M, Hashimoto K. (2015). *An Investigation of Effectiveness of Opinion and Fact sentences for Sentiment Analysis of Customer Reviews*. 2015 CCATS.
- Shandra, E. N., Setiawan, B. D., Sari, Y. (2019). Klasifikasi Pola Sidik Bibir untuk Menentukan Jenis Kelamin Manusia dengan Metode "Gray Level Co-Occurrence Matrix dan Support Vector Machine". Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.