

ANALISIS SENTIMEN MULTI-ASPEK BERBASIS KONVERSI IKON EMOSI DENGAN ALGORITME NAÏVE BAYES UNTUK ULASAN WISATA KULINER PADA WEB TRIPADVISOR

Sitti Aliyah Azzahra¹, Arief Wibowo^{*2}

^{1,2}Program Studi Magister Ilmu Komputer Fakultas Teknologi Informasi Universitas Budi Luhur

Email: ¹sittialiyahazzahra@gmail.com, ²arief.wibowo@budiluhur.ac.id

*Penulis Korespondensi

(Naskah masuk: 26 Maret 2019, diterima untuk diterbitkan: 22 April 2020)

Abstrak

Wisatawan seringkali mencari informasi tentang obyek wisata pada situs web seperti TripAdvisor. Situs web TripAdvisor memiliki fitur bagi pengguna terdaftar untuk memberi ulasan tentang objek wisata dalam kategori kuliner dari berbagai negara. Ulasan tersebut bisa digunakan wisatawan sebagai pertimbangan sebelum mendatangi objek wisata kuliner yang ingin dituju. Komentar atau ulasan yang ada di situs TripAdvisor dapat dianalisis untuk mengetahui nilai sentimen dari suatu obyek wisata yang diulas. Hasil analisis itu dapat bermanfaat bagi pengelola tempat wisata, pengusaha kuliner maupun bagi wisatawan lain. Ada tantangan yang ditemukan saat analisis sentimen dilakukan pada kalimat ulasan yang mengandung ikon emosi atau *emoticon*, karena ulasan dapat mengandung arti sentimen yang berbeda antara kalimat dengan ekspresi emosi yang ada. Penelitian ini berisi analisis ulasan tentang kuliner kota Bandung pada situs TripAdvisor yang mengklasifikasi sentimen menjadi tiga kelas. Penelitian ini menggunakan teknik klasifikasi data mining dengan *algoritme Naïve Bayes* dikombinasi dengan metode pelabelan multi aspek yang disertai konversi ikon emosi pada teks ulasan. Selain itu, analisis dilakukan pada bobot ulasan berdasarkan jumlah kontribusi pemberi ulasan di web TripAdvisor. Hasil pengujian menunjukkan bahwa penggunaan seluruh kombinasi metode tersebut dalam proses klasifikasi sentimen mampu menghasilkan nilai akurasi sebesar 98,67%.

Kata kunci: Analisis Sentimen, Naïve Bayes Classifier, Multi-aspect sentence labeling, Konversi Ikon Emosi, Ulasan Wisata, Tripadvisor

MULTI-ASPECT SENTIMENT ANALYSIS BASED ON EMOTICON CONVERSION WITH NAÏVE BAYES ALGORITHM FOR CULINARY TOURISM REVIEW ON TRIPADVISOR WEB

Abstract

Tourists often look for information about attractions on websites such as TripAdvisor. The TripAdvisor website has a feature for registered users to provide reviews about attractions in the culinary category from various countries. These reviews can be used by tourists as a consideration before visiting culinary attractions to be addressed. Comments or reviews on the TripAdvisor site can be analyzed to determine the sentiment value of a tourist attraction being reviewed. The results of the analysis can be useful for managers of tourist attractions, culinary entrepreneurs and for other tourists. There are challenges that are found when sentiment analysis is carried out on review sentences that contain emotion icons or emoticons, because reviews may contain different sentiment meanings between sentences and existing emotional expressions. This study contains a review of the culinary analysis of the city of Bandung on the TripAdvisor site which classifies sentiments into three classes. This study uses data mining classification techniques with the Naïve Bayes algorithm combined with a multi-aspect labeling method accompanied by the conversion of emotional icons in the review text. In addition, the analysis is carried out on the weight of the review based on the number of contributing reviewers on the TripAdvisor web. The test results show that the use of all combinations of these methods in the sentiment classification process is able to produce an accuracy value of 98.67%.

Keywords: Sentiment Analysis, Naïve Bayes Classifier, Multi-aspect sentence labeling, Emoticon Converting, Travel Reviews, Tripadvisor

1. PENDAHULUAN

Sebagian besar wisatawan dapat mencari informasi tentang destinasi objek wisata yang ingin mereka kunjungi dari salah satu situs web seperti TripAdvisor. Situs web TripAdvisor tidak hanya menampilkan hasil review atau ulasan tentang obyek wisata, namun juga aspek kuliner, akomodasi (hotel dan transportasi) menuju dan dari suatu obyek wisata. Situs ini juga menyediakan kolom komentar dari pengguna yang mengulas tentang objek wisata maupun kuliner yang telah dikunjungi dan dapat dibaca oleh wisatawan lain. Data komentar yang ada di situs web tersebut dapat dimanfaatkan bagi pengelola obyek wisata dan kuliner, berdasarkan hasil analisis sentimen yang kelas sentimen positif, negatif, atau netral. Hasil analisis opini itu selanjutnya dapat digunakan sebagai pertimbangan wisatawan dalam penentuan keputusan memilih objek wisata maupun kuliner yang akan dikunjungi.

Terdapat beberapa penelitian terdahulu yang melakukan analisis sentimen terhadap data ulasan online, antara lain Analisis Sentimen Terhadap Review Restoran Fish Streat Pada Aplikasi Zomato Menggunakan *Stemming* Nazief Adriani dan *Naive Bayes Classifier* (Yusnitasari, 2017), dengan nilai akurasi mencapai 87,50%. Keterbatasan penelitian ini adalah pada ketersediaan korpus bahasa Indonesia yang masih terbatas, sehingga pengklasifikasian sentimen menjadi relatif kurang maksimal.

Penelitian lain mengkombinasi Algoritma Genetika dan *Information Gain* untuk analisis Sentimen Review Hotel Menggunakan Algoritma *Naive Bayes* (Abdillah, 2018) dengan akurasi mendekati 90%. Penelitian ini menganalisis sentimen dengan menggunakan metode *Naive Bayes Classifier* dan seleksi fitur *Query Expansion Ranking* untuk data komentar wisata Malang di Tripadvisor dan menghasilkan nilai akurasi sebesar 86.6 % (Fanissa, Fauzi, dan Adinugroho, 2018), Namun penelitian ini dirasa belum cukup untuk menangani kompleksitas kalimat yang memiliki kandungan *emoticon* atau ikon emosi. Penelitian lain yang telah menggunakan analisis ikon emosi mampu mencapai akurasi sebesar 86,67% dengan *algoritme Naive Bayes*. Perolehan akurasi yang relatif rendah, diduga karena jumlah maupun pemilihan ikon emosi yang masih terbatas (Ruhana, 2019).

Berkaitan dengan penelitian yang telah dilakukan tersebut, maka dilakukan penelitian lanjutan untuk meningkatkan nilai akurasi pada penelitian sebelumnya. Penelitian ini akan menganalisis ulasan dari komentar wisatawan tentang objek wisata kuliner kota Bandung dari web TripAdvisor melalui analisis sentimen multi aspek yang akan diklasifikasikan menjadi tiga label yaitu positif, negatif, dan netral dengan menggunakan metode *Naive Bayes Classifier*.

Perbedaan penelitian ini adalah perluasan fitur konversi *emoticon* dengan menambahkan jumlah

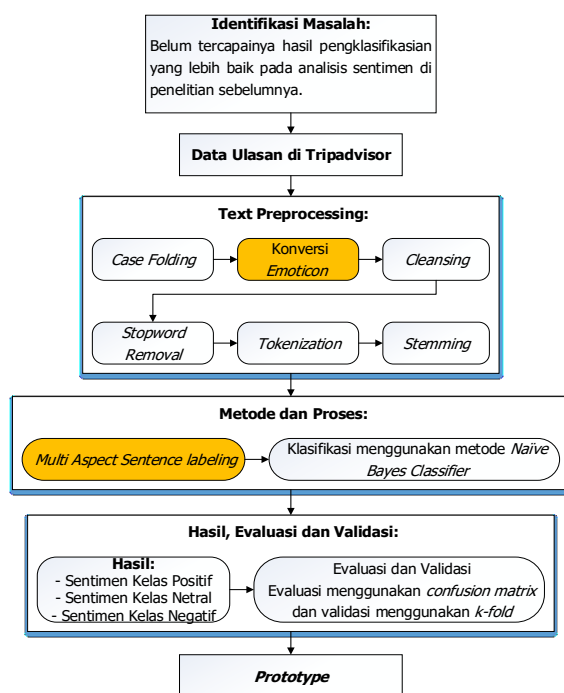
ikon emosi, perluasan fitur pembobotan dari ulasan balasan dari suatu ulasan (*reply comment*) dan pelabelan sentimen multi aspek di dalam pemrosesan teks atau corpus bertujuan untuk meningkatkan nilai akurasi model prediksi sentimen dari penelitian terdahulu. Penambahan fitur ini menjadi penting karena sentimen yang disampaikan oleh wisatawan (*traveler*) umumnya sangat singkat dan padat, menyesuaikan basis platform perangkat keras yang digunakan untuk menulis review tentang makanan. Selain itu, penggunaan konversi jenis *emoticon* yang lebih variatif diduga turut memberikan nilai sentimen, sehingga perlu dianalisis lebih lanjut.

2. METODE PENELITIAN

Penelitian yang dilakukan menggunakan metode penelitian eksperimen. Metode ini merupakan metode sistematis yang digunakan untuk membangun hubungan yang mengandung sebab akibat. Metode penelitian eksperimen ini menggunakan pendekatan kuantitatif dimana dalam metode Eksperimen, peneliti harus melakukan tiga persyaratan yaitu kegiatan mengontrol, kegiatan memanipulasi, dan observasi (Nazir, 2017).

2.1 Kerangka Konsep Penelitian

Tahap penelitian dimulai dari identifikasi masalah penelitian yang dilanjutkan dengan usulan alternatif pemecahan masalah dilanjutkan pemrosesan teks *corpus* mulai dari pra-pemrosesan, pemodelan dengan metode hingga didapatkan luaran yang diharapkan yaitu nilai sentimen ulasan wisata kuliner dengan kelas negatif, netral atau positif.



Gambar 1. Kerangka Konsep Penelitian

2.2 Sampling/Metode Pemilihan Sampel

Metode *sampling* yang digunakan dalam penelitian ini adalah metode *Non Random Sampling* atau Sampel Tidak Acak. *Non Random Sampling* adalah metode dimana setiap elemen/anggota populasi tidak mempunyai peluang yang sama untuk dijadikan sampel dikarenakan pengambilan sampel mempunyai kriteria tertentu (Triyono, 2018).

Jumlah sampel yang akan digunakan adalah data ulasan kuliner kota Bandung. Data penelitian ini berupa 900 corpus tentang kuliner kota Bandung yang diunduh dari TripAdvisor selama tahun 2018, dengan temuan bahwa 63,7% data mengandung ikon emosi.

2.3 Pengumpulan Data Tripadvisor

Pengumpulan data ulasan wisata kuliner Bandung pada web Tripadvisor menggunakan teknik *web scrapping* yaitu teknik di mana program komputer mengekstraksi data dari output yang dapat dibaca manusia yang berasal dari program lain. Teknik ini digunakan untuk memperoleh seluruh data yang dibutuhkan dalam penelitian. *Web scraping* merupakan proses pengumpulan data melalui interaksi dengan API.

2.4 Instrumentasi Penelitian

Instrumentasi yang digunakan dalam penelitian ini adalah berupa data ulasan terbaru yang diposting oleh pengunjung di situs *www.tripadvisor.com* yang dilakukan *scrapping* data menggunakan *tools* Data Mining dan Software yang digunakan sebagai alat bantu pada pengolahan data dalam penentuan metode untuk menghasilkan nilai akurasi yang akurat adalah *Rapidminer* versi 5.3. Setelah dilakukan pengolahan data, akan dilanjutkan dengan tahap evaluasi dan validasi. Pada tahap evaluasi digunakan instrumen *confusion matrix*, sedangkan untuk teknik validasinya menggunakan metode *k-fold*.

2.5 Teknik Analisis dan Pengujian Data

2.5.1. Teknik Analisis

Teknik Analisis yang digunakan adalah pengumpulan data awal. Data yang digunakan adalah data ulasan dari Tripadvisor. Data ulasan tersebut diambil menggunakan teknik *crawling* melalui API, dengan hasil data ulasan yang berhasil diperoleh berjumlah tiga ratus komentar. Teknik analisis dalam penelitian ini akan menggunakan metode *Naïve Bayes Classifier* dan *Multi-Aspect Sentence Labeling* dan konversi *emoticon*.

2.5.2. Pengolahan Data Awal

Data yang telah dikumpulkan diolah melalui beberapa tahapan seperti *Text Preprocessing* yang terdiri dari *Case Folding*, *Convert Emoticon*, *Cleansing*, *Stopword Removal*, *Tokenizing* dan *Stemming*, *multi-aspect sentence labeling*, pembobotan *Term Frequency* (Raw-TF) serta pemodelan klasifikasi opini menggunakan *Naïve Bayes Classifier* (NBC).

Sub pekerjaan dari tahap *Text Preprocessing* yang dilakukan adalah sebagai berikut:

1. *Case Folding*
Tahap mengkonversi seluruh teks ke dalam bentuk huruf kecil / *lower-case*.
2. *Emoticon Converting*
Tahap mengkonversi *emoticon* ke dalam bentuk teks.
3. *Cleansing*
Tahap menghapus tanda baca dan angka.
4. *Stopword Removal*
Menghilangkan kata-kata yang tidak penting.
5. *Tokenizing*
Tahap pemotongan string input berdasarkan tiap kata yang menyusunnya.
6. *Stemming*
Tahap menghilangkan semua kata imbuhan baik sufiks dan prefiks. Berfungsi untuk mengurangi variasi pada kata yang memiliki kata dasar sama.

Sebelum dilakukan pembelajaran data teks menggunakan *Naïve Bayes Classifier* (NBC), dilakukan proses pembobotan menggunakan *Term Frequency Murni* (Raw-TF). *Term Frequency* adalah frekuensi dari kemunculan suatu term atau kata dalam sebuah dokumen. Semakin besar jumlah kemunculan suatu term (TF tinggi) dalam dokumen, maka semakin besar juga bobotnya. *Raw-TF* digunakan agar data dapat dianalisis dengan menggunakan *Naïve Bayes Classifier* (NBC).

2.5.3. Teknik Pengujian

Dalam penelitian ini dilakukan eksperimen dan proses pengujian metode. Eksperimen yang dilakukan peneliti, menggunakan *framework RapidMiner* 5.3 untuk mengolah data yang akan digunakan dalam penentuan metode yang menghasilkan nilai akurasi yang akurat. Setelah dilakukan pengujian metode dengan *rapidminer*, maka akan dilanjutkan dengan tahap evaluasi. Dimana untuk evaluasi menggunakan *Confusion Matrix*, sedangkan untuk teknik validasinya menggunakan *K-Fold Cross Validation*.

2.5.4. Evaluasi dan Validasi Data Akhir

Evaluasi menggunakan *confused matrix* dengan kurva ROC untuk melakukan pengujian tingkat akurasi metode diukur, teknik validasi menggunakan *k-fold cross validation*. Evaluasi menggunakan *confused matrix* dengan beberapa rumus yang dapat digunakan untuk menilai dan mengevaluasi model klasifikasi diantaranya (Mentari, Fauzi, dan Muflikhah, 2018).

3. TEXT MINING

Text Mining merupakan proses analisa dari teks menjadi informasi yang digunakan untuk tujuan tertentu, sebagaimana didefinisikan oleh Bramer

(Fanissa, Fauzi, dan Adinugroho, 2018). Syarat umum teks mining adalah informasi yang diambil dapat menjadi data yang berguna, dimana data tersebut harus bersifat jelas dan eksplisit. *Text mining* adalah penemuan dari pengetahuan yang menarik pada dokumen teks. Hal ini merupakan tantangan untuk menemukan pengetahuan yang akurat pada teks dokumen yang diinginkan, dimana penemuan pengetahuan dapat menjadi efektif digunakan dan dapat memperbaharui pola penemuan pada proses penerapan ke *text mining* (Hidayat, 2017).

3.1 Analisis Sentimen Multi Aspek

Analisis sentimen multi aspek digunakan untuk menentukan jenis-jenis aspek yang terdapat pada teks komentar, sehingga proses analisis sentimen akan lebih terperinci dan mendalam, karena teks telah dikelompokkan menjadi beberapa aspek yang berbeda - beda. Untuk mendapatkan beberapa aspek-aspek yang berbeda tersebut, maka dibutuhkan *multi-aspect sentence labeling* yang bertujuan untuk melakukan pengidentifikasian dan menentukan kata yang memiliki kategori yang sama dan nantinya dapat ditentukan aspek yang relevannya (Martania, Anisa, dan Dana, 2018).

3.2 Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier adalah sebuah metode pengklasifikasian yang dapat diterapkan pada data teks, dengan menggunakan probabilitas sederhana yang berakar pada *Teorema Bayes* dan memiliki asumsi ketidaktergantungan (*independent*) yang tinggi dari masing – masing kondisi atau kejadian (Risca, Asmara, dan Rahutomo, 2017).

Naïve Bayes Classifier memiliki kelebihan, antara lain mampu menganalisis jumlah data pelatihan yang sedikit untuk penentuan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Abdilah, Mardiyani, dan Safudin, 2018). Bentuk umum *teorema bayes* adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan:

X = Data dengan kelas yang belum diketahui
H = Hipotesis data X merupakan suatu kelas spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (*posterior probability*)

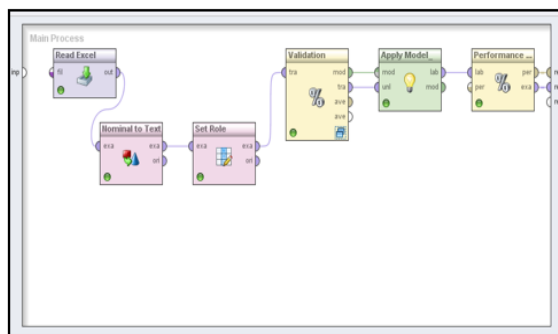
$P(H)$ = Probabilitas hipotesis H (*prior probability*)

4. HASIL DAN PEMBAHASAN

4.1 Pemodelan Klasifikasi

Proses pemodelan dalam penelitian ini dilakukan dengan menggunakan *tools RapidMiner* 5.3. *Tools* ini dipilih karena fleksibilitasnya dalam membangun model prediksi data mining

berdasarkan atribut-atribut yang dianalisis. Berikut ini adalah skema pemodelan dalam *RapidMiner*.



Gambar 2. Pemodelan dengan *RapidMiner*

Pada Gambar 2 terlihat bahwa data pemodelan dibaca berdasarkan format .xls yang menjadi basis pembacaan *corpus* ulasan di TripAdvisor. Selanjutnya dilakukan transformasi *corpus* dari bentuk nominal ke teks dan memilih metode yang digunakan yaitu pengklasifikasi Naïve Bayes dengan penambahan fitur *Multi - Aspect Sentence Labeling* maupun konversi *emoticon*. Hasil pelabelan sentimen di validasi menggunakan metode *k-fold cross validation* dengan nilai $k=10$. Dari validasi ini didapatkan indikator pengujian berupa *accuracy*, *precision* dan *recall*.

4.2 Hasil Pengolahan Data

4.3.1. Pra-pemrosesan Teks

Berikut tabel hasil pra-pemrosesan teks yang dilakukan:

Tabel 1. Pra-pemrosesan Teks

Teks Asli	Disini banyak orang2 yang piknik ke Bandung tumpah ruah....Bayangin aja, 1 bis pada masuk belanja kue2 dan snack.....untuuunggg....kami lebih cepat dikit dari mereka...Jadi bisa dapat kue lapis philipine dan kue picnic roll...karena setelah aku bayar, ada beberapa orang bilang,yachhh picnic roll nya habis.....☹️
Case Folding	disini banyak orang2 yang piknik ke bandung tumpah ruah....bayangin aja, 1 bis pada masuk belanja kue2 dan snack.....untuuunggg....kami lebih cepat dikit dari mereka...jadi bisa dapat kue lapis philipine dan kue picnic roll...karena setelah aku bayar, ada beberapa orang bilang,yachhh picnic roll nya habis.....☹️
Cleansing	disini banyak orang yang piknik ke bandung tumpah ruah bayangin aja bis pada masuk belanja kue dan snack untuuunggg kami lebih cepat dikit dari mereka jadi bisa dapat kue lapis philipine dan kue picnic rol karena setelah aku bayar ada beberapa orang bilang yachhh picnic roll nya habis emotesenang
Stop-word Removal	piknik bandung tumpah ruah bayangin aja bis masuk belanja kue snack untuuunggg lebih cepat dikit jadi bisa dapat kue lapis philipine

	kue picnic roll bayar beberapa orang bilang yachhh picnic rollnya habis emotsenang
Tokenizing	piknik bandung tumpah ruah bayangin aja bis masuk belanja kue snack untuuunggg lebih cepat dikit jadi bisa dapat kue lapis philipine kue picnic roll bayar beberapa orang bilang yachhh picnic rollnya habis emotsenang
Stemming	Sini banyak orang piknik bandung tumpah ruah bayan aja bis masuk belanja kue untung lebih cepat dikit jadi bisa dapat kue lapis philipine kue picnic roll bayar berapa orang bilang yah roll habis senang

4.3.2. Konversi Emoticon

Fitur pertama yang ditambahkan dalam pemrosesan teks adalah konversi *emoticon*. Penelitian ini telah memperluas jenis simbol *emoticon*, menjadi sebanyak dua puluh tiga jenis dan lebih banyak dari penelitian sebelumnya yang hanya sebanyak delapan belas (Ruhana, 2019). Berikut adalah tabel konversi *emoticon* yang digunakan dalam penambahan fitur yang diusulkan:

Tabel 2. Hasil Konversi Emoticon

No.	Emoticon	Teks Hasil Konversi	Kelas
1.	:)	Senang	Positif
2.	:(Sedih	Negatif
3.	:D	Senang	Positif
4.	:s	Sedih	Negatif
5.	:')	Senang	Positif
6.	:'))	Senang	Positif
7.	:))	Senang	Positif
8.	:")	Senang	Positif
9.	:)	Senang	Positif
10.	:((Sedih	Negatif
11.	^^	Senang	Positif
12.	;(Kecewa	Negatif
13.	:'(Kecewa	Negatif
14.	: '((Sedih	Negatif
15.	: '(Sedih	Negatif
16.	:-x	Kesal	Negatif
17.	:x	Kesal	Negatif
18.	:o	Ragu	Negatif
19.	:-/	Ragu	Netral
20.	:/	Ragu	Netral
21.	:-	Ragu	Netral
22.	:-/	Ragu	Netral
23.	:	Ragu	Netral

Perluasan fitur konversi *emoticon* dengan menambah jenis varian ikon emosi ini didasarkan atas kondisi data *corpus* yang dianalisis, bahwa secara dominan mengandung *emoticon*. Dengan perluasan fitur ini diduga akan memperkuat proses klasifikasi yang dilakukan pada tahap selanjutnya.

4.3.3. Multi-Aspect Sentence Labeling

Setelah data melalui tahap text preprocessing, maka didapatkan ribuan kata yang sudah siap untuk diproses ke tahap selanjutnya yaitu *multi-aspect*

sentence labeling. Proses ini akan melalui beberapa tahap yaitu klustering (pengelompokan), dan mengidentifikasi dokumen awal untuk setiap kelas. Hasil dari pelabelan kalimat multi aspek dapat dilihat pada Tabel 3.

Tabel 3. Multi Aspect-Sentence Labeling

Ikon	Suasana	Pelayanan	Harga	Fasilitas	Rasa
kecewa	sempit	lumayan	jangkau	tradisional	enak
sedih	ramai	ramah	mahal	sederhana	asin
senang	super	pas	murah	paket	lembut
kesal	susah	suka	cocok	legendaris	panas
	bingung	favorit	gratis	sewa	gurih
	penasaran	rekomendasi	spesial	pribadi	nikmat
	penuh	istimewa	biaya	pondok	Segar
	duduk	cepat	uang	privat	hangat
	sabar	unik		minimum	sejuk
	tenang	oke		mewah	lezat
	keren	bagus		bersih	pedas
	sulit	layak		kotor	manis
	antri	nyaman		lengkap	enak
	sesak	buruk		miskin	doyan
	heran	sopan		asli	beku
	strategis	jude		andong	muntah
	alam	masalah		restoran	hambar

Dari Tabel 3 terlihat bahwa pada dataset diberikan enam label berdasarkan aspek-aspek yang dianalisis dari *corpus*. Namun demikian, sebenarnya digunakan sembilan label yang meliputi ikon, suasana, pelayanan, harga, fasilitas, rasa, waktu, menu dan aspek lain. Pelabelan multi aspek ini akan memudahkan dalam proses pembobotan kata berdasarkan kategori aspek yang ada.

4.3.4. Pembobotan Term Frequency

Penelitian ini menggunakan metode *Term Frequency* jenis *TF*- Murni (*Raw TF*). Dimana pada *Raw TF*, nilai *TF* yang diberikan berdasarkan jumlah kemunculan suatu term di dokumen. Hasil perhitungan pembobotan *Term Frequency* pada pemrosesan teks dengan penambahan fitur konversi *emoticon* dan *Multi-Aspect Sentence Labeling*, terlihat pada Tabel 4.

Tabel 4. Hasil Pembobotan Term Frequency

Dokumen	Ikon	Suasana	Pelayanan	Harga	Fasilitas	Rasa	Waktu	Menu	Lainnya
text1	0	0	1	0	2	2	0	8	10
text2	0	3	1	0	4	2	0	2	10
text3	1	1	0	1	3	3	2	3	9
text4	0	2	4	0	2	1	1	7	9
text5	0	5	1	0	3	1	1	10	9
text6	1	1	3	1	1	2	2	3	8
text7	1	1	2	1	3	1	1	5	12
text8	0	3	4	0	5	2	1	2	7

Pada Tabel 4 terlihat bahwa komponen pembobotan dokumen menganalisis sembilan nilai kriteria, yang terdiri dari ikon, suasana, pelayanan, harga, fasilitas, rasa, waktu, menu dan lainnya. Komponen lainnya ini merupakan teknik analisis yang mengandung kebaruan metode, yaitu menghitung bobot penilaian kriteria berdasarkan jumlah kontribusi dan jumlah *like* yang diterima pengguna web TripAdvisor. Analisis tambahan pada fitur ini diduga akan meningkatkan nilai akurasi karena jumlah kontribusi

pemberi ulasan dapat memperkuat validitas nilai sentimen yang dianalisis.

4.3.5. Naïve Bayes Classification

Tahap inti dari pemodelan adalah penerapan algoritma *Naïve Bayes* dalam proses penentuan label kelas sentimen dari setiap dokumen. Berikut adalah hasil perhitungan *Naïve Bayes Classification*:

Tabel 5. Hasil Perhitungan *Naïve Bayes Classification*

Dokumen	Ikon	Suasana	Pelayanan	Harga	Fasilitas	Rasa	Waktu	Menu	Lainnya	Kelas
text1	0	0	1	0	2	2	0	8	10	positif
text2	0	3	1	0	4	2	0	2	10	positif
text3	1	1	0	1	3	3	2	3	9	positif
text4	0	2	4	0	2	1	1	7	9	negatif
text5	0	5	1	0	3	1	1	10	9	negatif
text6	3	1	1	0	1	0	0	10	10	negatif
text7	1	1	3	1	1	2	2	3	8	netral
text8	1	1	2	1	3	1	1	5	12	netral
text9	0	2	1	3	3	3	3	1	5	?

Pada Tabel 5 terlihat bahwa pengklasifikasi mencoba melakukan prediksi kelas untuk 'text9' yang memiliki bentuk asli sebagai berikut:

Tempat makan ini terletak dekat dengan rumah penduduk lokal. Biaya masuk murah, IDR 15000. Biaya masuk adalah IDR 20000 selama liburan. Dari pintu masuk/gerbang, mereka menyediakan layanan antar jemput gratis. Ada begitu banyak kegiatan untuk dilakukan, terutama untuk anak-anak. Anak-anak bermain, memberi makan kelinci, dll....Tetapi harus membayar untuk itu :D Ini adalah tempat yang bagus untuk makan siang atau makan malam di Purbasari dengan makanan khas sunda yang sangat lezat dan enak , di samping danau.

Berdasarkan hasil pengamatan, pemberi ulasan 'text9' telah memiliki empat kontribusi di web TripAdvisor dengan jumlah *like* sebanyak satu.

Berdasarkan bentuk *corpus* 'text9', dilakukan perhitungan probabilitas menggunakan metode *Naïve Bayes* dari persamaan 1 dengan tahapan sebagai berikut:

- Hitung probabilitas bersyarat (*likelihood*) dokumen 'text9' pada kelas positif, negatif dan netral. Hasil perhitungan probabilitas kelas menggunakan persamaan 1 memiliki hasil sebagai berikut:

$$\text{Kelas Positif} = 19.029206$$

$$\text{Kelas Negatif} = 0$$

$$\text{Kelas Netral} = 449.121081$$

- Probabilitas prior* dari kelas positif, negatif dan netral dihitung dengan proporsi dokumen pada tiap kelas:

$$P(\text{positif}) = 3/9 = 0.33$$

$$P(\text{negatif}) = 3/9 = 0.33$$

$$P(\text{netral}) = 3/9 = 0.33$$

- Hitung Probabilitas *posterior* menggunakan rumus *Naïve Bayes Classifier*:

$$P(\text{positif} | \text{text9}) = 0.697737$$

$$P(\text{negatif} | \text{text9}) = 0$$

$$P(\text{netral} | \text{text9}) = 16.467772$$

Berdasarkan perhitungan probabilitas akhir di atas, maka dapat disimpulkan bahwa dokumen 'text9' termasuk dalam kelas **Netral**, karena nilai $P(\text{Netral} | \text{text9})$ lebih besar dari nilai $P(\text{Positif} | \text{text9})$ dan $P(\text{Negatif} | \text{text9})$.

4.3.6. Pengujian Model dengan 10-Fold Cross Validation

Pada penelitian ini dilakukan pengujian model dengan menggunakan teknik *10-fold cross validation*, dimana proses ini membagi data ke dalam sepuluh bagian secara acak. Proses awal dimulai dengan membentuk model data pada bagian pertama. Model yang terbentuk akan diujikan dengan sembilan bagian data sisanya. Setelah itu dilakukan proses perhitungan nilai akurasi untuk melihat seberapa banyak data yang sudah terklasifikasi dengan benar.

4.3.7. Evaluasi dan Validasi

Pada proses evaluasi dan validasi, dilakukan analisis menggunakan *Confusion Matrix*. Pada tahapan ini ditampilkan hasil akurasi, nilai positif, negatif dan netral hasil pemodelan prediksi menggunakan *Naïve Bayes Classifier*.

Tabel 6. *Confusion Matrix Dengan Konversi Emoticon*

	Pred Positif	Pred Negatif	Pred Netral
True Positif	283	6	1
True Netral	3	289	2
True Negatif	0	0	316

Dari Tabel 6 terlihat bahwa nilai akurasi untuk seluruh fitur yang digunakan mencapai 98,67%. Hal ini merupakan suatu perbaikan nilai akurasi bila dibandingkan dengan pemodelan tanpa menggunakan fitur konversi *emoticon* dan perluasan fitur pembobotan ulasan, dengan nilai akurasi hanya mencapai 88,78%.

5. KESIMPULAN

Dari semua proses rangkaian kegiatan penelitian di atas yaitu analisis sentimen yang dilakukan menggunakan metode *Naïve Bayes Classifier* dengan penambahan fitur konversi *emoticon* dan *Multi-Aspect Sentence Labeling* mendapatkan nilai akurasi sebesar 98.67% yang sebelumnya hanya sebesar 88.78%, sehingga dapat disimpulkan bahwa penggunaan kombinasi metode tersebut telah terbukti memperbaiki hasil penelitian sebelumnya.

Untuk pengembangan penelitian sejenis di masa mendatang, tantangan untuk meningkatkan nilai akurasi dapat dilakukan dengan menambahkan atau memperbanyak kamus kata pada setiap aspeknya.

Selain itu, dapat memperluas simbol *emoticon* untuk mendapatkan lebih banyak indikator sentimen, serta penekanan pada *pre-processing* terutama pada singkatan-singkatan kata, kalimat ambigu maupun teks yang bernada sarkastik.

DAFTAR PUSTAKA

- ABDILAH, A., MARDIYANI, E., & SAFUDIN, M. 2018. Integrasi Algoritma Genetika Dan Information Gaint Untuk Menganalisis Sentimen Review Hotel Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Komputer*, 4(1), 186–193.
- FANISSA, S., FAUZI, M. A., & ADINUGROHO, S. 2018. Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2766–2770.
- HIDAYAT, A. M. 2017. Klasifikasi Analisis Sentimen Menggunakan Algoritma Naive Bayes Pada Layanan Internet First Media. Universitas Budiluhur.
- MARTANIA, T. D., ANISA, H., DANA, S. K., 2018. Multi-Aspect Sentiment Analysis Komentar Wisata TripAdvisor dengan Rule-Based Classifier (Studi Kasus : Bandung Raya). *e-Proceeding of Engineering*, Vol.5, No.1.
- MENTARI, N. D., FAUZI, M. A., & MUFLIKHAH, L. 2018. Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2739–2743.
- NAZIR, M. 2017. *Metode Penelitian*. Bogor: Penerbit Ghalia Indonesia.
- RISCA PRATIWI, I. Y., ASMARA, R. A., & RAHUTOMO, F. 2017. Study of Hoax News Detection Using Naive Bayes Classifier in Indonesia Language. *International Conference on Information & Communication Technology and System (ICTS)*, 73–78.
- RUHYANA, N. 2019. Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes. *Jurnal IKRA-ITH Informatika* Vol 3 No 1 Maret 2019 ISSN 2580-4316
- TRIYONO. 2018. Teknik Sampling Dalam Penelitian.[Online]Tersedia:<https://www.researchgate.net/publication/324029597>

Halaman ini sengaja dikosongkan