

Multi-Aspect Sentiment Analysis Komentar Wisata TripAdvisor dengan Rule-Based Classifier **(Studi Kasus : Bandung Raya)**

Martania Triana Dewi¹, Anisa Herdiani, S.T.,MT², Dana Sulistyio Kusumo, S.T.,MT.,PHD³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹martaniadewi@students.telkomuniversity.ac.id, ²anisaherdiani@telkomuniversity.ac.id,

³danakusumo@telkomuniversity.ac.id

Abstrak

Meningkatnya popularitas situs *web* yang menyajikan informasi tentang tempat wisata dan menyediakan kolom komentar yang dapat diisi oleh pengguna, membuat setiap pembaca komentar dapat menjadikan komentar tersebut sebagai acuan untuk mengunjungi tempat wisata, seperti kolom komentar pada situs *web TripAdvisor*. Tetapi, komentar tempat wisata pada situs *web TripAdvisor* hanya memberi penilaian secara keseluruhan pada setiap komentar dan tidak memberikan penilaian sesuai dengan aspek yang dibicarakan, sehingga menyulitkan pembaca komentar menganalisis aspek yang unggul pada komentar tersebut. Terkait masalah tersebut, penelitian ini bertujuan untuk mengelompokkan kata dari komentar tersebut sesuai dengan aspek yang terkait menjadi sebuah informasi yang lebih rinci dengan menggunakan *multi aspect sentiment analysis*. Metode yang digunakan adalah *rule-based classifier*, *rule-based classifier* dapat mengklasifikasikan sentimen dengan menggunakan aturan yang dapat dibentuk sesuai kondisi dan dibutuhkan dalam penelitian, dari penggunaan aturan tersebut diketahui nilai *coverage* tertinggi 99.00% dan nilai *accuracy* terkecil adalah 97.78 % yang didapat dari setiap aturan pada proses klasifikasi sentimen dengan metode *rule-based classifier*. Pada hasil penelitian, nilai *coverage* pada sentimen positif cukup tinggi, karena banyaknya kata dengan *label* positif, dan untuk nilai *accuracy* cukup baik yang menunjukkan ketepatan klasifikasi dengan *rule-based classifier* pada sistem.

Kata kunci : Bandung, *Multi Aspect Sentiment Analysis*, *Rule-Based Classifier*, *TripAdvisor*, Wisata

Abstract

The increasing popularity of websites that provide information about tourism place and provide comment column, allowing website's visitor to make the comments as a reference to visit the tourism place, such as the comments column on TripAdvisor website. However, the tourist comments on the site the TripAdvisor web only gives an overall rating on each comment and does not provide a rating according to the aspects discussed, it makes difficult for readers to analyze the superior aspects of the comment. Related to the problem, this study aims to group words from the comments in accordance with the related aspects into a more detailed information by using multi aspect sentiment analysis. The method used is the rule-based classifier, rule-based classifier can classify sentiment using rules that can be formed according to conditions and required in the research, from the rule of rules-based classifier known the highest coverage is 98.50% and the smallest accuracy is 99.27%, derived from each rule on the classification process of sentiment by rule-based classifier. In the result, the value of coverage on positive sentiment is quite high, because the number of words with positive label, and accuracy value is good enough that shows the accuracy of classification with rule-based classifier on the system.

Keywords: Bandung, *Multi Aspect Sentiment Analysis*, *Rule-Based Classifier*, *TripAdvisor*, *Tourism*

1. Pendahuluan

1.1 Latar Belakang

Pariwisata merupakan salah satu faktor yang berpengaruh dalam perkembangan perekonomian suatu daerah yang memiliki tujuan meningkatkan devisa negara, meningkatkan pendapatan daerah, memperluas kesempatan kerja, dan tidak mengeksploitasi kekayaan alam. Kementerian Pariwisata memutuskan ada 10 *cluster* pengembangan pariwisata dunia, salah satunya adalah Kota Bandung [1]. Kota Bandung juga memiliki banyak tempat wisata yang menarik untuk dikunjungi wisatawan, baik wisata alam maupun buatan.

Dengan berkembangnya media *online*, tidak jarang wisatawan membagikan pengalaman wisata mereka pada situs *web* yang menyediakan kolom komentar pada suatu tempat wisata, seperti pada situs *web TripAdvisor*. Situs *web* wisata seperti *TripAdvisor* cukup berperan penting bagi wisatawan sebagai bahan pertimbangan untuk mengunjungi suatu tempat, karena *TripAdvisor* menjangkau 390 juta pengunjung, serta menampilkan 435 juta komentar tentang 6,8 juta destinasi wisata [2], serta untuk mengetahui kelebihan ataupun kekurangan dari tempat wisata sesuai dengan penulisan pengalaman wisatawan yang berupa komentar pada situs *web TripAdvisor*. Sebagai contoh, salah satu komentar tempat wisata Floating Market yang ditulis oleh *widjokongko* dan

memberikan *rating* 5 (*rating* maksimal 5) : “Unik...kuliner ditepi danau dilengkapi dengan aneka wahana permainan bagi anak2. Kuliner yang tersedia beraneka ragam pilihan, cukup enak. Transaksi pembelian makanan harus menggunakan koin. Terdapat kios oleh2 dengan aneka kerajinan, makanan khas, dll.”, pada komentar tersebut menyampaikan beberapa penilaian pada suatu tempat wisata seperti daya tarik, sarana/prasarana, dan harga. Namun pengguna hanya memberikan *rating* secara keseluruhan tanpa mengetahui aspek apa yang membuat tempat wisata tersebut diberikan *rating* 5 oleh pemberi komentar pada situs *web TripAdvisor*, sehingga menyulitkan pembaca dalam mengambil keputusan untuk mengunjungi suatu tempat wisata atau tidak, karena jumlah komentar pada setiap wisata cukup banyak.

Berdasarkan hasil pengamatan komentar pada situs *web* wisata *TripAdvisor* dibutuhkan *multi aspect sentiment analysis* untuk membantu pembaca komentar *TripAdvisor* mengetahui aspek manakah yang unggul pada suatu tempat wisata, *multi-aspect sentiment analysis* yang bertujuan untuk menganalisis sentimen suatu komentar dengan mempertimbangkan beberapa aspek. *Sentiment analysis* berfokus pada komentar yang mengekspresikan sentimen positif dan sentimen negatif [3].

Tahap yang dilakukan pada penelitian ini adalah penentuan kata kunci pada setiap kata yang memiliki kesamaan, menginisialisasi setiap kategori yang akan dijadikan aspek terkait, dan *labeling* (positif dan negatif) [4]. Untuk proses klasifikasi sentimen, penelitian menggunakan *rule-based classifier* karena aturan dapat ditentukan sesuai kebutuhan penelitian [5], dengan mengklasifikasikan kata yang sudah diberi *label* dan atribut aspek dan menggolongkannya sesuai dengan keputusan yang sudah ditentukan sebelumnya dengan aturan tersebut [6].

1.2 Perumusan Masalah

- Bagaimana melakukan klasifikasi *sentiment analysis* dari komentar tempat wisata di kota Bandung dengan *Rule-Based Classifier*?
- Bagaimana *coverage* dan *accuracy* dari aturan pada *Rule-Based Classifier* terhadap komentar tempat wisata di kota Bandung ?

1.3 Batasan Masalah

- Penelitian ini hanya mencakup komentar tempat wisata di kota Bandung pada situs *web TripAdvisor*.
- Penelitian ini menggunakan 6 tempat wisata di kota Bandung.
- Penelitian ini hanya mencakup komentar tempat wisata pada situs *web TripAdvisor* dengan bahasa Indonesia.
- Komentar yang digunakan pada penelitian ini sebanyak 200 komentar setiap tempat wisata.

1.4 Tujuan

Tujuan penelitian ini adalah melakukan *multi-aspect sentiment analysis* terhadap komentar tempat wisata yang terdapat pada situs *web TripAdvisor* dengan memperinci suatu komentar terhadap beberapa aspek berbeda, menggunakan aturan pada *rule-based classifier*. Dari aturan tersebut kita dapat mengetahui serta menganalisis hasil klasifikasi *sentiment* serta nilai *coverage* dan *accuracy* dari setiap aturan yang telah dibentuk oleh *rule-based classifier*.

1.5 Organisasi Tulisan

Bagian 1 yaitu pendahuluan, membahas latar belakang, perumusan masalah, batasan penelitian, tujuan, serta organisasi tulisan, selanjutnya adalah bagian 2 studi terkait membahas landasan teori pada penelitian ini. Pada bagian 3 yaitu sistem yang dibangun merupakan perancangan sistem yang dibuat untuk implementasi dari penelitian yang akan dilakukan. Bagian 4 yaitu evaluasi menjabarkan hasil dari pengujian berikut analisis hasil pengujian. Bagian 5 yaitu kesimpulan yang menjelaskan kesimpulan dan saran pada penelitian ini.

2. Studi Terkait

2.1 Multi-Aspect Sentiment Analysis

Multi-aspect sentiment analysis digunakan sebagai penentuan sentimen pada suatu komentar dengan mempertimbangkan beberapa aspek terkait sehingga proses *sentiment analysis* akan terperinci karena dikelompokkan menjadi aspek yang berbeda, untuk mendapatkan aspek tersebut, membutuhkan *multi-aspect sentence labeling* yaitu mengidentifikasi dan menentukan kata yang memiliki kategori yang sama untuk kemudian ditentukan aspek yang relevan [4].

Tahap yang dilakukan pada proses ini adalah :

- Clustering

Langkah *clustering* bertujuan untuk menentukan kata pada *dataset* yang memiliki kategori yang sama. Seperti pada kata “mahal” dan “murah”, keduanya dapat dikategorikan menjadi kategori harga. *Dataset* yang belum memiliki *label* kemudian akan dikelompokkan pada satu *cluster* [7].

- **Identifying Initial Document for Each Class**

Setelah mendapatkan *cluster*, pada tahap ini adalah mengidentifikasi inisial dari setiap *cluster* atau disebut juga aspek [7].

- **Labeling Document**

Pada tahap ini, kata yang sudah memiliki *cluster* diberi penilaian dan memutuskan *label* (positif atau negatif) pada setiap *cluster* agar dapat dilakukan *multi aspect sentiment analysis* dan selanjutnya akan melalui proses klasifikasi dengan metode *Rule-Based Classification* [7].

2.2 **Rule-Based Classifier**

Rule-Based Classifier merupakan metode untuk mengklasifikasikan data dengan menggunakan aturan yang dapat dibentuk sesuai dengan kondisi yang dibutuhkan dalam penelitian. Sehingga dapat dikatakan *rule-based classifier* sebagai *unsupervised classifier* karena akan melakukan klasifikasi berdasarkan informasi yang dimiliki [8]. *Rule-Based* akan mengklasifikasikan teks yang sudah diberi *label* (positif atau negatif) sesuai dengan pengelompokan yang sudah ditentukan sebelumnya oleh aturan tersebut [5]. Pada *rule-based classifier* menerapkan aturan *if-then* untuk klasifikasi (1)

$$if_{(precondition)} then_{(conclusion)} \tag{1}$$

2.3 **Performansi**

Untuk mengetahui performansi dari aturan *rule-based classifier* untuk *sentiment analysis* dapat dilihat dari nilai *coverage* atau cakupan pada rumus (2) dan *accuracy* atau akurasi pada rumus (3) [9].

$$Coverage(R) = \frac{n_{covers}}{|D|} \tag{2}$$

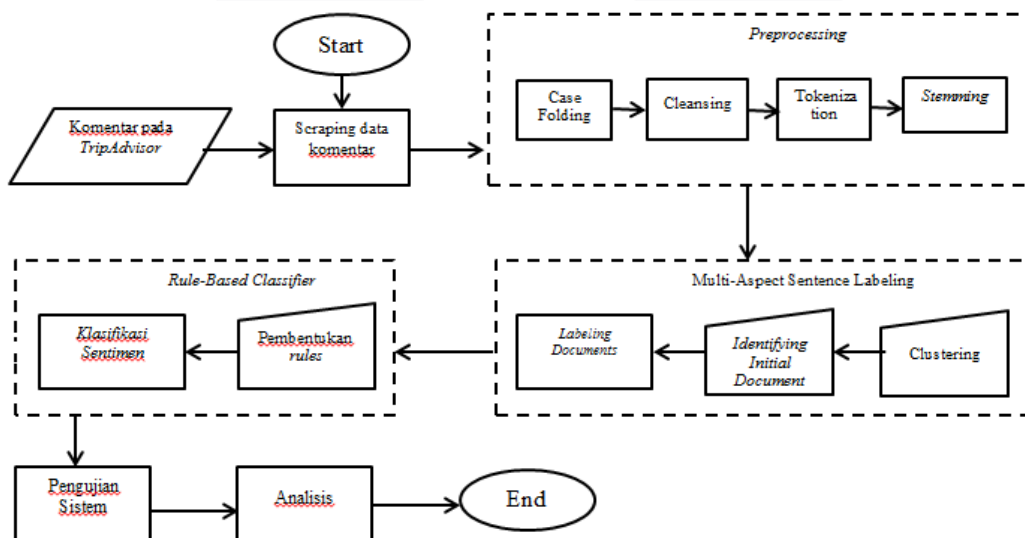
$$Accuracy(R) = \frac{n_{correct}}{n_{covers}} \tag{3}$$

Keterangan Rumus (2) dan (3) :

- n_{covers} = Jumlah data dengan kondisi yang sesuai dengan *precondition* pada aturan *if-then*
- $n_{correct}$ = Jumlah data yang sesuai dengan *precondition* dan *conclusion* pada aturan *if-then*
- $|D|$ = Jumlah seluruh *dataset* yang diujikan.

Hasil dari perhitungan nilai *coverage* (2) merupakan nilai presentase perbandingan antara jumlah data yang sesuai dengan *condition* suatu aturan dan jumlah data yang terdapat pada *dataset*, sedangkan hasil dari nilai *accuracy* (3) merupakan nilai yang merepresentasikan ketepatan sistem bekerja dalam proses klasifikasi dengan nilai sebenarnya [9].

3. **Sistem yang Dibangun**



Gambar 1. Flowchart Sistem yang Dibangun

Pada *flowchart* sistem yang dibangun (Gambar 1) menjelaskan alur sistem. Dengan langkah awal pengumpulan data penelitian dengan *scraping* komentar pada *TripAdvisor*, lalu melakukan *preprocessing* dari data komentar tempat wisata, selanjutnya melakukan tahap *multi-aspect sentence labeling*. Saat data komentar

sudah memiliki aspek dan *label*, data dapat dilakukan klasifikasi sentimen dengan *rule-based classifier*. Setelah proses klasifikasi selesai, maka sistem dapat diujikan dan dianalisis.

3.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan *scraping* dari situs *web* TripAdvisor. *Scraping* merupakan teknik yang digunakan untuk mengambil isi sebuah halaman web secara spesifik seperti komentar. Tempat wisata yang diambil pada penelitian ini berdasarkan tujuan wisata yang ditawarkan oleh *travel agent* yang menyediakan paket wisata di kota Bandung. *Travel agent* yang dipilih mengacu pada *travel agent* yang terdaftar pada *Association of the Indonesian Tours and Travel Agencies (ASITA)*, ASITA adalah suatu perkumpulan yang mewadahi pengusaha atau pelaku usaha di bidang jasa perjalanan wisata di Indonesia yang berlandaskan UUD 1945 dan UU kepariwisataan Indonesia [10]. Sehingga didapatkan 6 tempat wisata di kota Bandung (Tabel 1).

Tabel 1. Daftar Tempat Wisata

No	Tempat Wisata	Jumlah Komentar
1	Floating Market Lembang	200
2	Sari Ater Hot Spring	200
3	Situ Patenggang	200
4	Tangkuban Perahu	200
5	Trans Studio Bandung	200
6	White Crater (Kawah Putih)	200

3.2 Preprocessing

Pada tahap ini dilakukan *preprocessing* data komentar yang sudah didapatkan pada tahap data *scraping*. Proses *preprocessing* dilakukan dengan proses sebagai berikut *case folding*, *cleansing*, *tokenization*, *stemming* [11].

- **Case Folding** : menyeragamkan huruf menjadi *lowercase* (Tabel 2).

Tabel 2. Preprocessing (Case Folding)

Sebelum	Sesudah
Unik...kuliner ditepi danau dilengkapi dengan aneka wahana permainan bagi anak2. Kuliner yang tersedia beraneka ragam pilihan, cukup enak. Transaksi pembelian makanan harus menggunakan koin. Terdapat kios oleh2 dengan aneka kerajinan, makanan khas, dll.	unik...kuliner ditepi danau dilengkapi dengan aneka wahana permainan bagi anak2. kuliner yang tersedia beraneka ragam pilihan, cukup enak. transaksi pembelian makanan harus menggunakan koin. terdapat kios oleh2 dengan aneka kerajinan, makanan khas, dll.

- **Cleansing** : menghapus tanda baca atau simbol (Tabel 3).

Tabel 3. Preprocessing (Cleansing)

Sebelum	Sesudah
unik...kuliner ditepi danau dilengkapi dengan aneka wahana permainan bagi anak2. kuliner yang tersedia beraneka ragam pilihan, cukup enak. transaksi pembelian makanan harus menggunakan koin. terdapat kios oleh2 dengan aneka kerajinan, makanan khas, dll.	unik kuliner ditepi danau dilengkapi dengan aneka wahana permainan bagi anak2 kuliner yang tersedia beraneka ragam pilihan cukup enak transaksi pembelian makanan harus menggunakan koin terdapat kios oleh2 dengan aneka kerajinan makanan khas dll

- **Tokenization** : memisahkan kalimat menjadi satu bagian (Tabel 4).

Tabel 4. Preprocessing (Tokenization)

Sebelum	Sesudah
unik kuliner ditepi danau dilengkapi dengan aneka wahana permainan bagi anak2 kuliner yang tersedia beraneka ragam pilihan cukup enak transaksi pembelian makanan harus menggunakan koin terdapat kios oleh2 dengan aneka kerajinan makanan khas dll	unik bagi enak kios kuliner anak2 transaksi oleh2 ditepi kuliner pembelian dengan danau yang makanan aneka dilengkapi tersedia harus kerajinan dengan beraneka menggunakan makanan aneka ragam koin khas wahana pilihan terdapat dll permainan cukup

- **Stemming** : mencari akar kata, menghilangkan imbuhan pada kata (contoh : lokasinya menjadi lokasi) (Tabel 5)

Tabel 5. Preprocessing (Stemming)

Sebelum				Sesudah			
unik	bagi	enak	kios	unik	bagi	enak	kios
kuliner	anak2	transaksi	oleh2	kuliner	anak2	transaksi	oleh2
ditepi	kuliner	pembelian	dengan	tepi	kuliner	beli	dengan
danau	yang	makanan	aneka	danau	yang	makan	dengan
dilengkapi	tersedia	harus	kerajinan	lengkap	sedia	harus	aneka
dengan	beraneka	menggunakan	makanan	dengan	aneka	guna	rajin
aneka	ragam	koin	khas	aneka	ragam	koin	makan
wahana	pilihan	terdapat	dll	wahana	pilih	dapat	khas
permainan	cukup			main	cukup		dll

3.3 Multi-Aspect Sentence Labeling

Setelah data melalui *preprocessing* maka didapat ribuan komentar pada tempat wisata yang sudah siap untuk tahap penelitian selanjutnya yaitu *multi-aspect sentence labeling*. Komentar tersebut kemudian akan melalui beberapa tahap yaitu *clustering*, *labeling document*, dan *identifying initial document for each class*. Setelah tahap tersebut maka akan menghasilkan kamus kata pada setiap aspek (Lampiran 9).

- **Clustering**

Menentukan kata pada komentar tempat wisata di kota Bandung yang memiliki *cluster* yang sama untuk dijadikan *keyword* pada setiap *cluster*, sehingga didapat 11 *cluster* yaitu :

- Cluster 1 : buah, wortel, stroberi
- Cluster 2 : binatang, kelinci, kuda
- Cluster 3 : danau, gunung, kawah
- Cluster 4 : tunggang, foto, main
- Cluster 5 : bersih, kotor, asri, aman
- Cluster 6 : liku, tanjak, curam
- Cluster 7 : dekat, jauh, macet
- Cluster 8 : hotel, camp, homestay, cottage
- Cluster 9 : bandara, parkir, masjid, bank
- Cluster 10 : bayar, mahal, murah, rugi, diskon
- Cluster 11 : profesional, kualitas, kelola, pemandu

- **Identifying Initial Document for Each Class**

Mengidentifikasi aspek yang terkait pada *cluster*, aspek yang diambil berdasarkan pedoman dari Analisis Daerah Operasi Objek dan Daya Tarik Wisata Alam (ADO-ODTW) Dirjen PHKA tahun 2003 [12] dan [13](Lampiran 7).

- **Labeling Document**

Setelah didapatkan nama dari setiap *cluster* yang dijadikan aspek, dapat ditentukan *label* (positif dan negatif) pada setiap kata [14] (Lampiran 8). Karena pada penelitian ini menggunakan *label* positif dan negatif, maka untuk kata yang bersifat netral akan diberi *label* positif.

3.4 Rule-Based Classifier

Pada penelitian ini penetapan aturan *rule-based* menggunakan aturan *if-then* untuk klasifikasi (1). Pada klausa *if* pada aturan *rule-based* merupakan *precondition* atau kondisi yang harus dipenuhi, dapat terdiri satu atau lebih atribut *test*. Untuk klausa *then* (*conclusion/rule consequent*) pada aturan *rule-based* merupakan keputusan dari kondisi yang sudah dibuat pada klausa *if*. Pada *then* (*conclusion/rule consequent*) terdapat prediksi kelas (penelitian ini akan memprediksikan sentimen pada setiap aspek yang terdapat pada setiap komentar) dengan aturan pada rumus(4) dan (5) [8], sehingga didapatkan klasifikasi sentimen yang dapat dilihat contohnya pada tabel 8. Untuk komentar yang tidak terklasifikasi dengan aturan baik R1 atau R2 maka tidak terhitung pada hasil sentimen.

$$R1 : (\text{positif} \geq 2) \wedge (\text{negatif} < \text{positif}) \rightarrow \text{sentimen_positif} \quad (4)$$

$$R2 : (\text{negatif} \geq 2) \wedge (\text{positif} < \text{negatif}) \rightarrow \text{sentimen_negatif} \quad (5)$$

Keterangan Rumus (4) dan (5):

R1 : Jika pada suatu aspek pada suatu komentar terdapat lebih dari sama dengan 2 kata positif dan jumlah kata negatif kurang dari kata positif, maka klasifikasi sentimen menjadi *sentimen positif*

R2 : Jika pada suatu aspek pada suatu komentar terdapat lebih dari sama dengan 2 kata negatif dan jumlah kata positif kurang dari kata negatif, maka klasifikasi sentimen menjadi *sentimen negatif*

Tabel 6. Contoh Hasil Klasifikasi Multi-Aspect Sentimen Analysis pada Sistem

No	Daya Tarik				Aksesibilitas				Akomodasi				Sarana/Prasarana				Harga				Pelayanan			
	P	N	R1	R2	P	N	R1	R2	P	N	R1	R2	P	N	R1	R2	P	N	R1	R2	P	N	R1	R2
1	2	0	R1	-	1	0	-	-	0	0	-	-	0	0	-	-	0	0	-	-	1	2	-	R2
2	2	0	R1	-	2	0	R1	-	0	0	-	-	2	0	R1	-	3	0	R1	-	0	0	-	-
3	5	0	R1	-	0	0	-	-	2	0	R1	-	0	0	-	-	0	0	-	-	0	0	-	-
4	1	0	-	-	2	1	R1	-	0	0	-	-	3	0	R1	-	0	0	-	-	0	0	-	-
5	2	2	R1	-	5	1	R1	-	0	0	-	-	6	0	R1	-	2	1	R1	-	0	0	-	-

Keterangan Tabel 6 :

P : label Positif R1 : Sentimen Positif
N : label Negatif R2 : Sentimen Negatif

4. Evaluasi

Pada penelitian ini, hal yang dievaluasi adalah hasil klasifikasi sentimen dengan *rule-based classifier* sesuai dengan aturan yang sudah dibentuk pada R1 dan R2, serta pengujian dari sistem dengan perhitungan *coverage* pada setiap wisata dan *accuracy* pada setiap aspek.

4.1 Hasil Pengujian

Pengujian pada penelitian ini adalah menghitung dan menganalisis hasil klasifikasi sentimen pada sistem (tabel 7) atau pada pengujian ini disebut dengan *ncover*s, hasil klasifikasi sebenarnya atau pada penelitian ini disebut dengan *ncorrect*, dan jumlah komentar yang masing-masing terdapat 200 komentar setiap tempat wisata atau pada penelitian ini disebut nilai |D| (Lampiran 11). Sehingga dengan nilai *ncover*s, *ncorrect*, dan |D| didapatkan nilai *coverage* R1 (tabel 8) dan R2 (tabel 9) yang dihitung dengan menggunakan rumus (2), nilai *accuracy* per aspek pada setiap tempat wisata untuk R1 (Lampiran 12) dan R2 (Lampiran 13) yang dihitung dengan rumus (3), dan rata-rata nilai *accuracy* pada R1 dan R2 setiap aspek tempat wisata (gambar 2).

Tabel 7. Hasil Klasifikasi Sentimen Floating Market

	Sentimen Positif	Sentimen Negatif
Daya tarik	100 % - 198	0 % - 0
Aksesibilitas	95.83 % - 23	4.17 % - 1
Akomodasi	0 % - 0	0 % - 0
Sarana/Prasarana	100 % - 70	0 % - 0
Harga	100 % - 73	0 % - 0
Pelayanan	0 % - 0	0 % - 0

Keterangan Tabel 7:

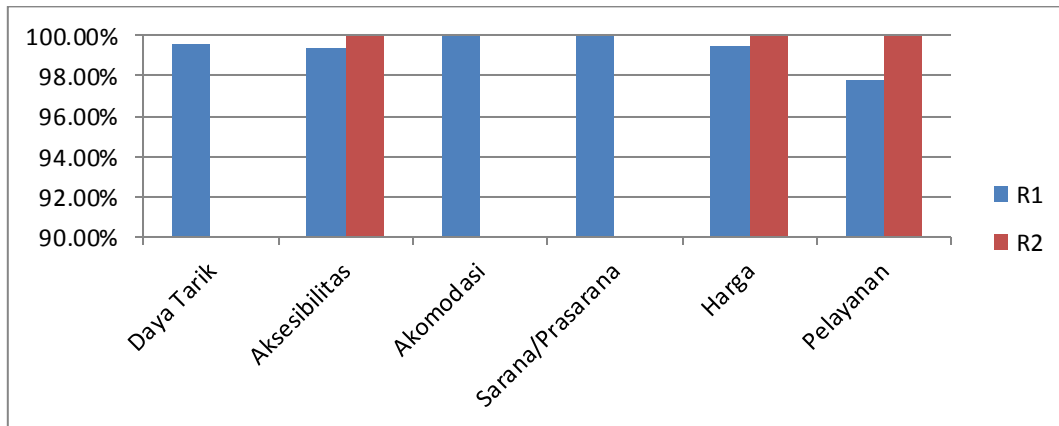
100% - 198 : dimana presentase sentimen positif aspek daya tarik adalah 100% dan jumlah komentar dengan sentimen positif pada aspek daya tarik terdapat pada 198 komentar.

Tabel 8. Coverage R1 (Sentimen Positif)

Tempat Wisata	Coverage R1					
	Daya Tarik	Aksesibilitas	Akomodasi	Sarana/Prasarana	Harga	Pelayanan
Floating Market	99.00%	11.50%	0.00%	35.00%	36.50%	0.00%
Sari Ater	96.00%	13.50%	10.50%	10.50%	17.00%	0.50%
Situ Patenggang	98.50%	25.00%	2.00%	17.00%	17.00%	0.50%
Tangkuban Perahu	97.50%	26.00%	0.50%	25.00%	11.50%	0.50%
Trans Studio Bandung	97.00%	8.50%	4.50%	15.00%	21.50%	0.50%
White Crater	95.50%	40.00%	2.50%	38.00%	28.00%	4.50%

Tabel 9. Coverage R2 (Sentimen Negatif)

Tempat Wisata	Coverage R2					
	Daya Tarik	Aksesibilitas	Akomodasi	Sarana/Prasarana	Harga	Pelayanan
Floating Market	0.00%	0.50%	0.00%	0.00%	0.00%	0.00%
Sari Ater	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
Situ Patenggang	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Tangkuban Perahu	0.00%	1.00%	0.00%	0.00%	0.00%	1.00%
Trans Studio Bandung	0.00%	0.00%	0.00%	0.00%	0.50%	0.00%
White Crater	0.00%	2.00%	0.00%	0.00%	1.50%	0.50%



Gambar 2. Accuracy

4.2 Analisis Hasil Pengujian

- Klasifikasi sentimen pada tabel 7 menghasilkan hasil klasifikasi dengan perbedaan yang cukup signifikan. Pada sistem lebih banyak menangkap klasifikasi sentimen positif (R1). Hal tersebut terjadi karena pada penelitian ini hanya menggunakan *label* positif dan negatif, maka kata yang bersifat netral seperti “tiket” dan “bayar” digolongkan sebagai *label* positif, sehingga jumlah komentar sentimen positif (R1) lebih banyak dibandingkan dengan sentimen negatif (R2).
- Suatu komentar (sudah melalui *preprocessing*) yang memiliki lebih dari satu kata yang sama (tabel 10), kata “tiket” akan terhitung menjadi tiga kata bukan satu kata pada proses klasifikasi sentimen.
- Dengan aturan R1 (Sentimen Positif) dan R2 (Sentimen Negatif), sistem akan mengklasifikasikan sentimen jika terdapat dua kata atau lebih dengan *label* positif maupun negatif pada suatu aspek. Sehingga, komentar yang hanya memiliki satu kata pada suatu aspek tidak dilakukan klasifikasi atau tidak dihitung. Seperti pada komentar Floating Market tidak mencakup klasifikasi sentimen baik R1 maupun R2 pada aspek pelayanan. Padahal, terdapat komentar yang menyinggung aspek pelayanan, contohnya pada tabel 11 terdapat satu kata “therapis” yang termasuk pada aspek pelayanan.
- Besar atau kecilnya nilai *coverage* tidak berpengaruh pada kinerja sistem, karena *coverage* bergantung pada data komentar yang dimiliki pada setiap tempat wisata.
- Nilai *accuracy* bernilai rendah karena kurangnya ketepatan sistem dalam proses klasifikasi, dilihat dari jumlah data yang tercakup dengan R1 atau R2 pada sistem berbeda dengan jumlah data yang tercakup dengan R1 atau R2 sebenarnya. Seperti aspek daya tarik nomor 5 (tabel 6). Pada hasil klasifikasi sistem data terklasifikasi sebagai R1, yang sebenarnya data tersebut tidak terklasifikasi R1 maupun R2.
- Nilai *accuracy* yang kosong menunjukkan bahwa tidak adanya data komentar yang sesuai dengan R1 maupun R2 pada aspek tersebut.

Tabel 10. Contoh Komentar dengan Tiga Kata yang sama

rendam, paling, benar, turut, sebelum, malam, kena, mobil, **tiket**, kolam, harus, **tiket**, kaya, sungai, malam, **tiket**, langsung, sriwedari

Tabel 11. Contoh Komentar Floating Market

enak, banget, pijat, di, tengah, rumah, apung, sambal, makan, es, krim, **therapis**, sopan, dan, nggak, cekikikan, nggak, buru2, dan, harga, org2, yg, antri, kalau, makan, enak, semua, dan, jangkau

5. Kesimpulan

Kesimpulan pada penelitian ini adalah :

- Klasifikasi sentimen pada *rule-based classifier* dilakukan dengan tahap *preprocessing* yang meliputi *case folding*, *cleansing*, *tokenization*, dan *stemming*, dan tahap *multi-aspect sentence labeling* yang meliputi *clustering*, *identifying initial document for each class*, dan *labeling*. Sehingga data komentar dapat dilakukan klasifikasi sentimen dengan *rule-based classifier*.
- *Rule-Based Classifier* dapat mengklasifikasikan sentimen positif dan sentimen negatif sesuai *rules* atau aturan yang dibentuk dengan nilai *accuracy* terkecil 97.78%.
- Rata-rata nilai terbesar *coverage* yang dihasilkan klasifikasi *rule-based classifier* pada klasifikasi sentimen positif (R1) yaitu 97.25% dan pada klasifikasi sentimen (R2) yaitu 0.33%.
- Banyaknya jumlah kata dengan *label* positif membuat hasil klasifikasi sentimen positif (R1) menjadi lebih banyak dibanding klasifikasi sentimen negatif (R2).

Saran pada penelitian ini adalah :

- Dapat mempertimbangkan penggunaan *label* netral untuk menyeimbangkan hasil klasifikasi sentimen.
- Untuk mendapatkan hasil klasifikasi sentimen lebih baik, dapat dilakukan penambahan atau penyesuaian *rules* yang sesuai dengan data penelitian.
- Lebih memperbanyak kamus kata pada setiap aspek yang terkait, sehingga klasifikasi sentimen lebih baik.
- Untuk mendapatkan nilai *coverage* yang beragam atau lebih rinci, dapat menambahkan jumlah data komentar dari setiap wisata.

Daftar Pustaka

- [1] A. H. Putra, "Bandung Masuk 10 Kluster Pengembangan Pariwisata Dunia," *Tempo*, 07 September 2016.
- [2] "About Us: TripAdvisor," TripAdvisor, [Online]. Available: <https://tripadvisor.mediaroom.com/us-about-us>. [Accessed 24 March 2017].
- [3] S. A. AL-Ani, S. A. El-Ghany and H. M.El-Bakry, "A Collective Intelligence Based Multi Aspect Sentiment," *International Journal of Advanced Research in*, vol. IV, no. 1, p. 1, 2016.
- [4] B. Lu, M. Ott, C. Cardie and B. Tsou, "Multi-aspect Sentiment Analysis with Topic Models," 2011.
- [5] P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, 2013.
- [6] B. Qin, Y. Xia, S. Prabhakar and Y. Tu, "A Rule-Based Classification Algorithm for Uncertain Data," *IEEE International Conference on Data Engineering*, 2009.
- [7] B. Liu, X. Li, W. S. Lee and P. Yu, "Text Classification by Labeling Words," *Association for the Advancement of Artificial Intelligence*, 2004.
- [8] P. Chikersal, S. Poria and E. Cambria, "Sentiment Analysis of Tweets by Combining a Rule-based Classifier," *SeNTU*, 2015.
- [9] J. Han, M. Kamber and J. Pei, "Rule-Based Classification," in *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2012, pp. 355-357.
- [10] "Tentang ASITA - Assosication of the Indonesian Tours and Travel Agencies," [Online]. Available: <https://asitaindonesia.org/tentang/>. [Accessed 28 September 2017].
- [11] S. Vijayarani, J. Ilamathi and Nithya, "Preprocessing Techniques for Text Mining - An Overview," *International Journal of Computer Science & Communication Networks*, vol. V, no. 1, pp. 7-16, 2015.
- [12] I. A. Ginting, P. Patana and R., "Assessment and Development of Object Potency and Pull Factor of Ecotourism at Sibolangit," *Peronema Forestry Science Journal*, vol. II, 2013.
- [13] P. Naidoo, P. Ramseook-Munhurrun and P. Seegoolam, "AN ASSESSMENT OF VISITOR SATISFACTION WITH NATURE-BASED TOURISM ATTRACTIONS," *INTERNATIONAL JOURNAL OF MANAGEMENT AND MARKETING RESEARCH*, vol. 4, p. 6, 2011.
- [14] J. Bata, Suyoto and Pranowo, "LEKSIKON UNTUK DETEKSI EMOSI," *Seminar Nasional Informatika 2015*, 2015.