



## Sistem Pemeringkat Otomatis Berbasis Kata Sifat

Faisal Rahutomo<sup>1</sup>, Diana Mayangsari Ramadhani<sup>2</sup>, Ingrid Yanuar Risca Pratiwi<sup>3</sup><sup>1</sup>Jurusan Teknologi Informasi, Politeknik Negeri Malang<sup>2,3</sup> Magister Terapan Teknik Elektro, Jurusan Teknik Elektro, Politeknik Negeri Malang<sup>1</sup>faisal@polinema.ac.id, <sup>2</sup>diana\_mayangsari@polinema.ac.id, <sup>3</sup>ingrid\_yanuar@polinema.ac.id

### Abstract

This paper exposes a novel method has been developed during these 2 years. The method is named as “adjective based automatic rating system”. This method is developed to utilize the abundant availability of text on the internet for quality and performance rating purpose. The text is processed in such a way and leave only the adjectives. Semantic analysis is done by two knowledge: adjectives of performance definition and Indonesian adjectives database with its synonym-antonym relation. This research proposes several formula steps, therefore the method output is a rating score that can be tunned its scale. The experiment results have been gathered for several objects: tourism, courier service, and organization performance. With detail information in tourism object experiment, this paper cites the other experiment results as well. This paper also provides availability information of the method as Python library. The results show a high correlation score, always more than 0.9. The results also show acceptable error scores, never more than 45%.

Keywords: rating system, automatic, semantic analysis, text processing, internet.

### Abstrak

Makalah ini menjabarkan sebuah usulan metoda baru yang telah dikembangkan selama 2 tahun terakhir. Metode ini disebut sebagai “sistem pemeringkat otomatis berbasis kata sifat”. Metoda yang dikembangkan bertujuan memanfaatkan ketersediaan teks yang berlimpah di internet untuk mendapatkan nilai kualitas atau unjuk kerja. Teks di internet diolah sedemikian hingga hanya menyisakan kata-kata sifatnya. Analisa semantik dilakukan berdasarkan definisi unjuk kerja yang ditentukan dan database kata sifat bahasa Indonesia beserta relasi sinonim-antonimnya. Penelitian ini mengusulkan beberapa proses dan rumusan sedemikian hingga analisa kata dapat memberikan nilai peringkat dengan skala maksimal yang dapat diatur. Pengujian telah dilakukan pada kasus obyek wisata, jasa pengiriman nasional, dan kinerja organisasi. Dengan pembahasan pengujian lebih detail pada kasus obyek wisata, makalah ini juga mencuplik hasil penelitian lainnya. Makalah ini juga memberi penunjuk informasi bahwa metode yang diusulkan di dalam penelitian ini telah tersedia implementasinya secara terbuka di dalam bentuk *library* Python. Hasil pengujian pada semua kasus menunjukkan nilai korelasi yang tinggi, selalu lebih dari 0,9. Pengujian juga menunjukkan tingkat galat yang cukup baik, tidak pernah melebihi angka 45%.

Kata kunci: sistem pemeringkat, otomatis, analisa semantik, pemrosesan teks, internet.

© 2019 Jurnal RESTI

### 1. Pendahuluan

Perkembangan teknologi informasi dan komunikasi yang demikian cepat akhir-akhir ini mempengaruhi perilaku manusia di dalam kehidupannya sehari-hari. Beberapa hal mendorong perubahan perilaku tersebut. Pertama berkembangnya *platform* aplikasi web dari statik menjadi dinamik hingga kini berkembang menjadi sosial media. Perkembangan tersebut membuat pengguna web mudah menyalurkan isi pikirannya, bahkan hingga kegiatannya sehari-hari. Kedua berupa perkembangan teknologi komunikasi perangkat

genggam yang semakin kompleks hingga kini memiliki istilah telepon cerdas (*smartphone*). Perangkat ini membuat akses web menjadi lebih mudah melalui telepon genggam, tidak perlu harus melalui perangkat komputer. Perkembangan ini membuat arus pendapat berupa teks dari masyarakat ke web menjadi jauh lebih besar. Tercatat berdasar hasil registrasi Kementerian Komunikasi dan Informatika, pada 30 April 2019 terdapat lebih dari 250 juta nomor prabayar [1]. Jumlah pelanggan ini bisa meningkat lebih banyak bila ikut memperhitungkan pelanggan pasca bayar. Di sisi

lainnya data menunjukkan jumlah penduduk Indonesia pada tahun 2019 berkisar pada angka sekitar 267 juta jiwa [2]. Bila dibandingkan dengan data pengguna perangkat genggam, bisa diperkirakan hampir seluruh warga negara memilikinya. Akses masyarakat kepada perangkat *smartphone* ini semakin terbuka dengan beredarnya merek-merek dengan harga terjangkau berharga ratusan ribu dengan perangkat lunak android yang gratis dengan banyak aplikasi yang juga gratis. Picuan-picuan faktor teknologi tersebut, baik dari sisi perangkat keras ataupun lunaknya membuat arus informasi berupa teks dari masyarakat menjadi meningkat. Informasi itu bisa berupa opini, kepuasan, komplain, ulasan, dan lain sebagainya.

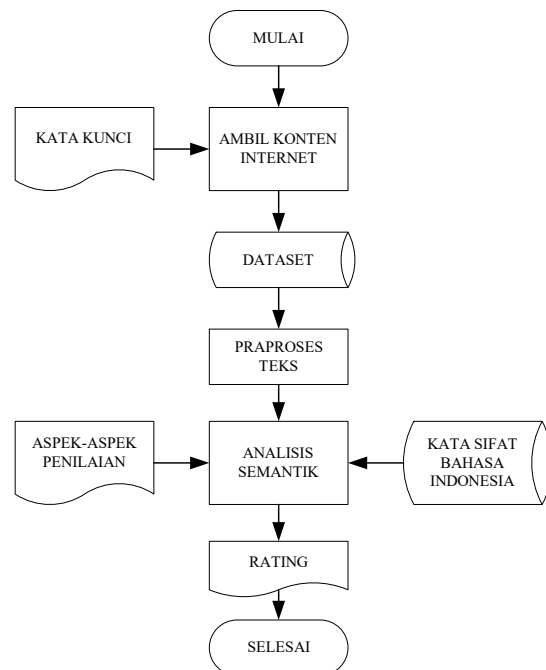
Derasnya informasi yang mengalir tersebut ikut mendukung perilaku dasar masyarakat di dalam perannya sebagai konsumen. Umum terjadi sebelum membeli suatu barang atau mengunjungi suatu tempat, masyarakat meminta pendapat dari orang lain yang sudah memiliki pengalaman. Berlimpahnya informasi di internet tadi termasuk berupa ulasan umpan balik terhadap penyedia barang atau jasa yang ada. Beberapa situs web menyediakan layanan berbagi penilaian ini hingga memberikan nilai pemeringkat (*rating*). Google dan Tripadvisor dapat diambil sebagai contoh ini untuk kasus lokasi penyedia jasa dan barang tertentu. Beberapa lembaga pemeringkat profesional juga memberikan penilaian semacam Skytrax untuk maskapai penerbangan. Portal jual beli semacam Tokopedia dan Bukalapak juga menyediakan nilai pemeringkat ini berupa umpan balik pelanggan yang telah menyelesaikan transaksi, baik kepada toko ataupun barangnya. Dengan demikian masyarakat menjadi lebih mudah untuk mendapat masukan tentang kualitas barang ataupun layanan yang tersedia. Di sisi lainnya penyedia jasa dan barang menjadi harus lebih berhati-hati.

Sayangnya, fasilitas yang telah tersedia tersebut memiliki beberapa kelemahan. Kelemahan pertama dikarenakan koleksi umpan balik penilaian ini yang dilakukan secara manual. Seringkali tidak ada umpan balik atas suatu layanan baik jasa maupun barang. Di sisi lainnya, umpan balik maupun penilaian berupa teks terhadap layanan barang atau jasa tersebut ternyata tersedia di platform yang tidak terstruktur semacam halaman web opini, halaman blog, atau cuitan di Twitter. Sebagai contoh, sebuah tempat wisata tidak ditemukan pemeringkatnya di situs Tripadvisor, tetapi tersedia informasinya baik di WordPress, Blogspot, atau konten blog lainnya [3]. Kelemahan kedua menyangkut fakta bahwa tidak semua sistem pemeringkat layanan jasa dan barang tersedia di internet. Sebagai contoh tidak adanya sistem pemeringkat layanan kurir ekspedisi barang nasional, atau sistem pemeringkat layanan perbankan nasional. Padahal sebagaimana kelemahan pertama tadi, terdapat banyak konten di internet terkait layanan barang ataupun jasa tersebut.

Untuk mengisi ruang kosong ini, makalah ini memaparkan sebuah metoda baru yang diberi nama “sistem pemeringkat otomatis berdasarkan kata sifat”. Idenya diinspirasi pemanfaatan teks di internet untuk dikenali polanya [4]. Makalah ini memaparkan tentang langkah-langkah memanfaatkan kekayaan teks di internet hasil dari arus deras percakapan masyarakat di internet tentang suatu layanan barang dan jasa hingga menghasilkan nilai pemeringkat. Usulan ini memberi penjabaran rinci secara komputasi dari peneliti lain sebelumnya [5]. Kemudian makalah ini menjabarkan cara bekerjanya metoda ini, resume pengujian pada beberapa kasus [6][7][8], hingga implementasinya menjadi sebuah *library* pemrograman Python yang bisa diakses secara terbuka [9][10][11]. Secara khusus makalah ini menjabarkan pengujian metode ini pada implementasi kasus tempat wisata, melanjutkan pembahasan pada makalah sebelumnya yang belum tuntas [6]. Berturut-turut mengikuti alur jurnal ini, bagian selanjutnya membahas metode penelitian, kemudian hasil dan pembahasan, dan ditutup dengan bagian kesimpulan.

## 2. Metode Penelitian

Gambar 1 menunjukkan diagram alir desain sistem yang diusulkan. Bagian awal dari sistem ini adalah pengumpulan konten di internet. Pengumpulan dapat dilakukan secara otomatis dengan menggunakan perangkat *crawling* dengan memanfaatkan *web service* Google *custom search* atau dilakukan secara manual [12]. Pengumpulan konten teks internet ini dilakukan berdasarkan pilihan kata kunci pada kasus yang ditentukan. Data teks ini kemudian diolah pada tahap



Gambar 1. Alur Sistem

Tabel 1. Daftar Tempat Wisata dan Ketersediaanya di Tripadvisor

No	Tempat Wisata	Tersedia
1	Pantai watu godek/gedeg/gedek	Tidak
2	Desa Kenep	Tidak
3	Pantai Puger Kucur	
4	Wisata tubing ledok amprong	Tidak
5	Sumber air jenon	Tidak
6	Waduk karangkates	Tidak
7	Camp Pendakian Latar Ombo	Tidak
8	Coban Manten	Tidak
9	Coban Glotak	Tidak
10	Coban Grojogan Sewu Pujon	Tidak
11	Kebun Teh Wonosari	Tidak
12	Pantai tamban	Tidak
13	Waduk Sidodadi	Tidak
14	Andaman	Tidak
15	Pemandian Cagar	Tersedia
16	Jawa Timur Park	Tersedia
17	Selecta	Tersedia
18	Coban Rondo	Tersedia
19	Coban Talun	Tersedia
20	Taman nasional bromo tengger	Tersedia
21	Museum Brawijaya	Tersedia
22	Taman Rekreasi Kota	Tersedia
23	Wendit	Tersedia
24	Taman Slamet	Tersedia
25	Kampung Warna Warni	Tersedia
26	Malioboro	Tersedia
27	Pantai Pandawa	Tersedia
28	Candi Borobudur	Tersedia
29	Tangkuban Perahu	Tersedia
30	Kawah Putih	Tersedia

praproses. Langkah penting dalam tahap praproses ini adalah tahapan *adjective pass* yang menyisakan hanya kata sifat dari teks yang diolah lebih lanjut. Hasil tahapan praproses kemudian diumpankan kepada bagian analisa semantik kata. Bagian tersebut bekerja dengan masukan data kata sifat bahasa Indonesia beserta relasinya dan data kata aspek penilaian yang tepat pada kasus yang dipilih. Analisa ini dijelaskan lebih lanjut di Bagian 2.3 sehingga dapat memberikan nilai

Tabel 3. Parameter Penilaian Tempat Wisata pada Kuesioner

Nilai	Keterangan
L	Luar Biasa
SG	Sangat
R	Rata-rata
B	Buruk
SB	Sangat

pemeringkat. Nilai pemeringkat dapat diatur skalanya sesuai kebutuhan.

### 2.1. Pengumpulan Dataset

Makalah ini memaparkan proses pembentukan empat dataset yang diperlukan untuk berjalannya sistem: dataset penilaian sebuah obyek, dataset kata sifat aspek penilaian, dataset daftar kata sifat dengan relasinya, dan dataset konten blog. Dataset penilaian ini nantinya digunakan pada bagian akhir penelitian, yaitu membandingkan penilaian manual dan penilaian program. Dataset kata sifat aspek penilaian dan daftar kata sifat digunakan di dalam proses perhitungan pemeringkat. Sedangkan dataset konten blog adalah teks yang diolah untuk diperoleh nilai pemeringkatnya. Keempatnya disusun sebagai bahan aspek penilaian suatu obyek pada metoda sistem pemeringkat berdasarkan kata sifat.

Kemudian makalah ini menunjukkan sebuah contoh dalam penelitian ini secara lebih mendetail, yaitu proses pengumpulan dataset penilaian obyek wisata.

#### 2.1.1. Dataset Penilaian

Terdapat dua langkah dalam pengumpulan dataset penilaian objek wisata. Yang pertama adalah mengumpulkan 30 nama tempat wisata secara acak di Indonesia. Penelitian ini fokus pada lokasi wisata di Jawa dan Bali. Daftar objek wisata tersebut dibagi menjadi 2 bagian. Bagian pertama adalah objek wisata yang tidak memiliki penilaian di Tripadvisor dan sebagian sisanya adalah tempat wisata yang memiliki penilaian di Tripadvisor. Tabel 1 menunjukkan daftar ini berdasarkan waktu penelitian dilakukan. Perubahan

Tabel 2. Potongan Bagian Formulir Kuesioner

Destinasi Wisata	Akses jalan	Kuliner	Fasilitas	Biaya Masuk	Usia layak kunjungan	Penilaian Keseluruhan
Pantai Watu godekLumajang	L / SG / R / B / SB	L / SG / R / B / SB	L / SG / R / B / SB	a. gratis b. membayar ± Rp ...	A / B / C / D / E	L / SG / R / B / SB
Pantai Kanep, Banyuwangi	L / SG / R / B / SB	L / SG / R / B / SB	L / SG / R / B / SB	a. gratis b. membayar ± Rp ...	A / B / C / D / E	L / SG / R / B / SB
Pantai Puger Kucur, Puger-Jember	L / SG / R / B / SB	L / SG / R / B / SB	L / SG / R / B / SB	a. gratis b. membayar ± Rp ...	A / B / C / D / E	L / SG / R / B / SB
..	..	..	..	..	..	..
Bendungan Karangates, Malang	L / SG / R / B / SB	L / SG / R / B / SB	L / SG / R / B / SB	a. gratis b. membayar ± Rp ...	A / B / C / D / E	L / SG / R / B / SB

pada saat ini dapat terjadi karena sifat data Tripadvisor yang dinamis.

Langkah kedua adalah mengumpulkan penilaian pemeringkat objek wisata secara langsung di Tripadvisor bila nilai pemeringkatnya ada pada laman tersebut. Bila tidak ada, dilakukan penilaian secara manual kepada beberapa responden dengan menggunakan kuesioner yang ditampilkan pada Tabel 2. Kuesioner tersebut diisi oleh 11 responden yang telah memiliki pengalaman berwisata di berbagai lokasi. Responden mengisi dan memberikan nilai dengan melingkari parameter-parameter yang sesuai dengan pendapat responden pada masing-masing kolom yang disediakan dengan parameter pilihan pada Tabel 3.

#### 2.1.2. Dataset Daftar Kata Sifat Aspek Penilaian

Pengumpulan dataset ini bertujuan untuk menentukan kategori dan sub kategori kata sifat yang tepat dan berelasi dengan peringkat suatu tempat wisata. Hasil dari dataset ini digunakan dalam metoda sistem pemeringkat berdasarkan kata sifat untuk menentukan nilai pemeringkat suatu tempat wisata berdasarkan konten blog. Dataset didapat dengan membagikan kuesioner kepada 11 responden. Formulir yang dibagikan diisi oleh responden. Responden berjumlah ganjil, agar tidak terjadi penilaian yang ambigu dan dapat dilakukan pengambilan keputusan akhir dengan *voting*. Responden mengisi dengan menyetujui atau tidak menyetujui kata sifat positif yang sesuai dengan penilaian suatu tempat wisata berdasarkan kategorinya masing-masing, serta dapat memberikan saran kata sifat. Hasil dari umpan balik responden ini disarikan sebagai kata sifat kategori dan sub kategori aspek penilaian yang ditunjukkan pada Tabel 4 kolom 1 dan 2.

#### 2.1.3. Dataset Relasi Kata Sifat Bahasa Indonesia

Dataset relasi kata sifat bahasa Indonesia ini masih belum tersedia secara terbuka. Sedangkan hal ini sangat dibutuhkan di dalam analisa semantik kata untuk memperluas analisa kata sifat yang ada. Sehingga diperlukan daftar kata sifat berikut relasinya. Merujuk kepada dataset serupa, yaitu WordNet, yang tersedia di dalam bahasa-bahasa lain di kalangan penelitian internasional, perlu adanya dataset kata sifat dengan relasi sinonim dan antonim [13]. Sehingga penelitian ini berusaha membangun daftar kata sifat bahasa Indonesia dan relasinya yang ramah ilmu komputer. Penyusunan daftar ini merupakan bagian penelitian sebelumnya yang mengembangkan alat penelitian sistem temu kembali bahasa Indonesia [14]. Penelitian ini kemudian menyusun *synonym set* (*synset*) kata sifat dan mendata adanya antonim antara satu *synset* dengan *synset* lainnya. Langkah-langkah yang dilakukan sebagai berikut. Menyiapkan database Tesaurus Indonesia dari penelitian sebelumnya [15]. Membuat *query* dengan mengambil kolom id *synset* dan *synset* dengan kategori kata sifat ("a"). Terdapat 4.045 *synset* dengan kategori kata sifat ("a") yang ditemukan. Mengunduh hasil ke

dalam bentuk file *spreadsheet*. Menghapus *synset* yang kosong. Bila terdapat id *synset* namun tidak memiliki kumpulan *lemma*, maka data tersebut dihapus. Ada kemungkinan terdapat kesalahan input data. Kemudian Mengeliminasi *synset* yang kembar. Bila terdapat 2 atau lebih id *synset* yang berbeda namun isi lemma-nya sama, maka id dan *synset* akan dihapus dan disisakan 1 *synset* saja. Ini dimungkinkan karena satu *synset* dapat didefinisikan berulang di kamus karena anggota lemma yang berawalan abjad yang berbeda-beda. Selanjutnya menggabungkan *synset* yang masing-masing anggota sinonimnya memiliki makna yang sama atau satu pengertian yang sama, sehingga dapat dijadikan 1 *synset*. Terdapat kemungkinan beberapa *synset* dengan makna yang sama, tetapi anggotanya berbeda-beda. Penggabungan dilakukan dengan menggunakan logika "or". Terakhir, mendata antonim antar *synset* dengan bantuan Kamus Tesaurus Bahasa Indonesia. Cara mencari antonim suatu *synset* dengan mencari simbol "ant" *synset* yang kita maksud di kamus. Akhirnya dicatat relasi antonim antar satu *synset* dengan *synset* lainnya.

#### 2.1.4. Dataset Konten Blog Ulasan Tempat Wisata

Pengguna aplikasi memasukan kata kunci pencarian, kata kunci berupa bahasa Indonesia tempat wisata di Indonesia. Dengan memanfaatkan *web service* Google API, sistem mencari data halaman web blog *traveler* berdasar kata kunci tempat wisata yang sudah diberikan sebelumnya. Setelah data tautan laman web dari Google muncul, kemudian link dan kontennya diambil untuk disimpan sebagai dataset.

#### 2.2. Praproses Teks

Pada tahap ini dilakukan pembersihan data dengan penghapusan teks yang tidak diperlukan seperti tag *HTML* dan hanya mengambil kontennya. Kemudian proses pembersihan juga menghapus angka, tanda baca, dan mengubah semua huruf besar ke huruf kecil. Langkah selanjutnya adalah *tokenizing*, mengubah suatu kalimat atau paragraf menjadi bagian-bagian kata. Langkah selanjutnya yang paling penting adalah menyeleksi kata-kata sifat yang muncul di teks sehingga mendapatkan kumpulan kata sifat dengan memanfaatkan *adjective pass list*. *Adjective pass list* di dalam penelitian ini disusun berdasarkan Kamus Tesaurus Bahasa Indonesia [16]. Praproses di dalam sistem ini tidak melibatkan langkah *stemming* karena proses ini bisa mengakibatkan dua kata berbeda makna menjadi satu. Langkah *stop word* diganti dengan *adjective pass* dengan tujuan hanya memanfaatkan kata sifat saja. Pemanfaatan kata sifat saja di dalam metode ini mengikuti langkah peneliti lain sebelumnya [3].

#### 2.3. Analisa Semantik Pemeringkat Berdasar Kata Sifat

Algoritma baru yang diterapkan oleh peneliti ini memproses kata-kata konten dari blog yang telah dilakukan praproses. Pada metoda ini proses dilakukan

Tabel 4. Contoh Proses Perhitungan Metode Pemeringkat Berbasis Kata Sifat

Kategori	Sub kategori	Jumlah		SC (1)	TH (2)	C (3)	A (4)	CR (5)
		$\sum$ Sinonim	$\sum$ Antonim					
akses jalan	lancar	1	0	1	1	2	1	5
	mudah	1	0	1	1			
fasilitas	indah	1	0	1	1	4	0,44	2,222
	bersih	0	0	0	0			
	alami	0	0	0	0			
	terpesona	1	0	1	1			
	unik	0	0	0	0			
	jernih	1	0	1	1			
	bagus	1	0	1	1			
	luas	2	2	0	0			
	megah	1	1	0	0			
	enak	1	0	1	1	4	0,67	3,333
kuliner	nikmat	1	0	1	1			
	sedap	1	0	1	1			
	asin	0	0	0	0			
	lezat	0	0	0	0			
	khas	1	0	1	1			
biaya masuk	murah	2	0	2	1	3	1	5
	ekonomis	2	0	2	1			
	gratis	1	0	1	1			
usia pengunjung	muda	4	1	3	1	2	0,5	2,5
	dewasa	1	0	1	1			
	tua	0	4	-4	-1			
	remaja	1	0	1	1			
emosional	senang	3	0	3	1	6	0,85	4,286
	bahagia	3	0	3	1			
	gembira	3	0	3	1			
	seru	0	0	0	0			
	asyik	3	0	3	1			
	menarik	1	0	1	1			
	takjub	0	0	0	1			
Pemeringkat Akhir (RA) (6)								3,724

dengan hanya menggunakan kata sifat. Kata-kata sifat yang ditemukan dibandingkan dengan daftar aspek kata sifat yang mewakili nilai-nilai pembentuk pemeringkat. Analisa dan perbandingan ini juga memanfaatkan relasi sinonim dan antonim kata sifat dengan bantuan data relasi kata sifat. Akhirnya sistem dapat memberikan nilai pemeringkat suatu tempat wisata berdasarkan teks ulasan yang tersedia bebas di internet.

Tahapannya dapat diwakilkan dengan Persamaan 1 hingga 6. Contoh proses metoda sistem pemeringkat berbasis kata sifat ini dapat dilihat pada Tabel 4. Kategori dan sub kategori merupakan daftar kata sifat yang diusulkan dari sistem penilaian suatu tempat wisata. Daftar kata sifat hasil praproses dari konten blog web dihitung dengan mencocokkan dengan daftar kata sifat aspek penilaian. Antonim dan sinonim dari daftar kata sifat aspek penilaian juga dihitung. Setiap kata sinonim yang tersedia pada konten blog dihitung 1. Berkebalikan dengan antonim, setiap kata antonim yang ditemukan dihitung -1. Jumlah sinonim dan antonim yang tersedia dalam teks kemudian disimpan. Setelah

memperoleh jumlah sinonim dan antonim yang cocok dengan masing-masing sub kategori untuk tiap kategori, maka jumlah sinonim yang cocok dikurangi dengan jumlah antonim sesuai seperti yang dijelaskan dalam Persamaan 1, dan dikenal sebagai variabel sub kategori (SC). Hasil dari proses tersebut kemudian dihitung dengan proses *thresholding* (TH) dalam Persamaan 2. Jika nilainya lebih dari 1 maka nilainya dipangkas menjadi 1. Jika nilainya kurang dari -1 maka nilainya dipangkas menjadi -1. Jika nilainya 0 maka nilai ditetapkan sebagai 0. Setelah proses *thresholding*, nilai setiap kategori dapat diperoleh dengan penjumlahan semua nilai sub kategori, (C) sesuai dengan Persamaan 3. Dengan  $n$  menunjukkan banyaknya sub kategori dari sebuah kategori. Setelah mendapatkan nilai perkategori, proses selanjutnya adalah menghitung nilai rata-rata dari kategori (A). Nilai kategori dibagi dengan jumlah sub kategorinya seperti yang dijelaskan dalam Persamaan 4. Langkah ini memberikan nilai pemeringkat ternormalisasi untuk masing-masing kategori, dengan nilai kontinu antara 0 hingga 1. Nilai pemeringkat ini dapat diatur berdasarkan pemeringkat maksimum sistem

( $R_{max}$ ) yang diinginkan. Peringkat masing-masing kategori ( $CR$ ) kemudian dikalikan dengan  $R_{max}$  seperti yang dijelaskan dalam Persamaan 5. Dengan perhitungan ini, jika nilai rata-rata dari kategori adalah 1 maka peringkat didefinisikan sebagai  $R_{max}$ . Jika nilai rata-rata dari kategori adalah 0 maka peringkatnya juga didefinisikan sebagai 0. Setelah mendapatkan peringkat kategori, langkah terakhir adalah menghitung nilai peringkat akhir ( $RA$ ) seperti yang dijelaskan dalam Persamaan 6. Jumlah nilai peringkat kategori untuk semua kategori dibagi dengan banyaknya kategori.

$$SC = \sum Synonym - \sum Antonym \quad (1)$$

$$TH = \begin{cases} 1 & \text{if } SC \geq 1 \\ -1 & \text{if } SC \leq -1 \\ 0 & \text{if } SC = 0 \end{cases} \quad (2)$$

$$C = \sum_{k=1}^n TH_k \quad (3)$$

$$A = \frac{C}{n} \quad (4)$$

$$CR = A \times R_{max} \quad (5)$$

$$RA = \frac{\sum_{j=1}^j CR_j}{j} \quad (6)$$

#### 2.4. Metrik Evaluasi Performa

Pengujian sistem dilakukan untuk memastikan bahwa sistem yang dirancang dijalankan seperti yang diharapkan. Pengujian sistem terdiri dari dua tahap pengujian: persentase galat dan korelasi koefisien Pearson ( $PCC$ ). Persamaan 7 mendeskripsikan perhitungan galat dengan memeriksa perbedaan nilai pemeringkat manual dari sistem yang diperoleh dari sistem. Hasilnya dibagi dengan nilai peringkat manual dan dikalikan dengan 100%. Sedangkan perhitungan  $PCC$  dijabarkan pada Persamaan 8. Pengujian ini dilakukan dengan memproses hasil tautan Google berturut-turut sebanyak 10, 20, dan 30 konten.

$$Galat = \frac{|x-y|}{x} \times 100\% \quad (7)$$

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (8)$$

### 3. Hasil dan Pembahasan

Tabel 5 dan 6 menunjukkan hasil perhitungan prosentasi galat dan korelasi  $PCC$  antara pemeringkat sistem dengan nilai pemeringkat kuesioner menggunakan Persamaan 7 dan 8. Data nilai pemeringkat manual disimpan ke dalam variabel  $x$ , sedangkan nilai pemeringkat program disimpan ke dalam variabel  $y$ . Rata-rata hasil dari pengujian sistem menggunakan  $PCC$  sangat bagus, dikarenakan nilai  $PCC$  mendekati nilai 1 yang bermakna sistem memiliki nilai korelasi yang tinggi dengan penilaian manual. Nilai  $PCC$  terendah berada dinilai 0,968 pada pengujian dengan 10 konten. Sedangkan nilai tertinggi adalah 0,980 pada pengujian dengan 30 konten. Nilai galat yang didapat juga relatif rendah dengan maksimal nilai galat adalah sebanyak

31,8% pada sistem dengan jumlah konten 30. Nilai galat terendah sebesar 20,4% pada pengujian dengan 10 konten. Yang menarik di sini adalah nilai  $PCC$  yang cenderung semakin baik dengan lebih banyak konten teks yang terlibat berkebalikan dengan nilai galat yang cenderung semakin besar.

Nilai galat tidak nampak berpola antara tempat wisata yang telah ada pemeringkat-nya di Tripadvisor maupun yang belum ada. Tampak nilai galat sangat rendah bahkan 0% pada kata kunci “Bendungan Sidodadi” dengan 20 konten. Nilai terbesar pada kata kunci “Ledok amprong river tubing” dengan 10 dan 20 konten. Suatu ketika pada 10 konten kata kunci “Gunung Tangkuban Perahu” memiliki nilai galat 0,9 menjadi naik 17,3 pada 20 konten.

Tabel 7 dan 8 menunjukkan hasil perhitungan prosentasi galat dan korelasi  $PCC$  antara pemeringkat sistem dengan nilai pemeringkat Tripadvisor menggunakan Persamaan 7 dan 8. Data nilai pemeringkat Tripadvisor disimpan ke dalam variabel  $x$ , sedangkan nilai pemeringkat program disimpan ke dalam variabel  $y$ . Rata-rata hasil dari pengujian sistem menggunakan  $PCC$  mendekati nilai 1 yang bermakna sistem memiliki nilai korelasi yang tinggi dengan penilaian manual. Nilai  $PCC$  terendah berada dinilai 0,96 pada pengujian dengan 10 konten. Sedangkan nilai tertinggi adalah 0,979 pada pengujian dengan 30 konten. Nilai galat yang didapat juga relatif rendah dengan maksimal nilai galat adalah sebanyak 20,3% pada sistem dengan jumlah konten 20. Nilai galat terendah sebesar 18,9% pada pengujian dengan 10 konten. Sebagaimana pengujian sebelumnya di Tabel 5 dan 6, hasil di Tabel 7 dan 8 menunjukkan nilai  $PCC$  yang cenderung semakin baik dengan lebih banyak konten teks yang terlibat. Akan tetapi, berbeda dengan percobaan sebelumnya yang nilai galat yang cenderung semakin besar dengan bertambahnya konten, pada pembandingan data Tripadvisor ini tidak nampak adanya pola.

Selanjutnya, makalah ini akan mengutip hasil-hasil penelitian sebelumnya yang ikut melakukan evaluasi terhadap metoda yang diusulkan. Kasus pertama yang dibahas adalah “Sistem Pemeringkat Otomatis Jasa Pengiriman Barang Nasional” [8]. Sedangkan kasus selanjutnya adalah “Menilai Kinerja Suatu Organisasi” [7]. Kedua kasus tersebut dipilih dengan dasar tidak adanya sistem pemeringkat yang tersedia sebagaimana tempat wisata. Sedangkan ketersediaan teks yang menilai keduanya berlimpah di internet.

#### 3.1. Hasil Pengujian Pada Kasus Lainnya

Penelitian yang telah dilakukan oleh Rahutomo dkk [8] bertujuan untuk mengembangkan sistem pemeringkat secara otomatis pelayanan jasa pengiriman nasional berdasar konten teks. Beberapa lembaga jasa pengiriman nasional dipilih sebagai kata kuncinya. Pengujian nilai

Tabel 5. Data Pemeringkat untuk 30 Tempat Wisata dari 11 Responden

No	Nama	Kuesioner		10 konten		x.y	Galat	20 konten		x.y	Galat	30 konten		x.y	Galat
		x	x <sup>2</sup>	y	y <sup>2</sup>			y	y <sup>2</sup>			y	y <sup>2</sup>		
1	Pantai Watu Godek/gedeg	2,3	5,4	3,5	12,3	8,2	50,0	4,2	17,6	9,8	80,0	4,2	17,6	9,8	80,0
2	Pantai Kenep	3,0	9,0	3,6	13,0	10,8	20,0	4,4	19,4	13,2	46,7	4,7	22,1	14,1	56,7
3	Pantai Puger Kucur	2,7	7,1	3,4	11,6	9,1	27,5	3,5	12,3	9,3	31,2	4,0	16,0	10,7	50,0
4	Ledok amprong river tubing	2,3	5,4	4,4	19,4	10,3	88,6	4,4	19,4	10,3	88,6	4,2	17,6	9,8	80,0
5	Sumber Air Jenon	3,0	9,0	3,1	9,6	9,3	3,3	4,3	18,5	12,9	43,3	4,4	19,4	13,2	46,7
6	Bendungan Karangates	3,3	10,6	3,3	10,9	10,7	1,5	4,0	16,0	13,0	23,1	4,0	16,0	13,0	23,1
7	Latar Ombo (Gunung Panderman)	3,2	10,2	3,4	11,6	10,9	6,2	4,1	16,8	13,1	28,1	4,2	17,6	13,4	31,3
8	Air Terjun Manten	3,3	11,1	3,9	15,2	13,0	17,0	4,0	16,0	13,3	20,0	4,1	16,8	13,7	23,0
9	Air Terjun Glotak	3,0	9,0	3,9	15,2	11,7	30,0	4,1	16,8	12,3	36,7	4,1	16,8	12,3	36,7
10	Air Terjun Grojogan Sewu Pujon	3,7	13,4	3,6	13,0	13,2	1,8	4,4	19,4	16,1	20,0	4,4	19,4	16,1	20,0
11	Kebun Teh Wonosari	3,6	12,6	2,8	7,8	10,0	21,3	4,4	19,4	15,6	23,7	4,5	20,3	16,0	26,6
12	Pantai Tamban	3,0	9,0	3,2	10,2	9,6	6,7	3,3	10,9	9,9	10,0	4,0	16,0	12,0	33,3
13	Bendungan Sidodadi	3,0	9,0	2,9	8,4	8,7	3,3	3,0	9,0	9,0	0,0	4,3	18,5	12,9	43,3
14	Andaman	3,3	10,6	3,7	13,7	12,0	13,8	4,4	19,4	14,3	35,4	4,4	19,4	14,3	35,4
15	Air Hangat Cangar	3,3	10,9	3,1	9,6	10,2	6,1	4,0	16,0	13,2	21,2	4,3	18,5	14,2	30,3
16	Jawa Timur Park	4,3	18,3	4,0	16,0	17,1	6,4	4,2	17,6	17,9	1,7	4,2	17,6	17,9	1,7
17	Selecta	3,7	13,7	3,9	15,2	14,4	5,4	4,2	17,6	15,5	13,5	4,2	17,6	15,5	13,5
18	Air Terjun Rondo	3,6	13,0	3,9	15,2	14,0	8,3	4,0	16,0	14,4	11,1	4,3	18,5	15,5	19,4
19	Air Terjun Talun	3,5	12,3	3,1	9,6	10,9	11,4	4,1	16,8	14,4	17,1	4,4	19,4	15,4	25,7
20	Gunung Bromo	4,0	16,0	3,6	13,0	14,4	10,0	4,2	17,6	16,8	5,0	3,9	15,2	15,6	2,5
21	Museum Brawijaya	3,1	9,7	3,4	11,6	10,6	9,3	3,3	10,9	10,3	6,1	3,6	13,0	11,2	15,7
22	Taman Rekreasi Kota	3,0	9,0	4,1	16,8	12,3	36,7	4,7	22,1	14,1	56,7	4,4	19,4	13,2	46,7
23	Wendit	3,1	9,9	4,0	16,0	12,6	27,3	3,9	15,2	12,3	24,1	4,4	19,4	13,8	40,0
24	Taman Slamet	3,3	11,1	4,1	16,8	13,7	23,0	4,3	18,5	14,3	29,0	4,3	18,5	14,3	29,0
25	Kampung Warna Warni	3,3	10,7	4,4	19,4	14,4	34,4	4,5	20,3	14,7	37,5	4,5	20,3	14,7	37,5
26	Malioboro	4,3	18,1	2,7	7,3	11,5	36,5	2,7	7,3	11,5	36,5	2,6	6,8	11,1	38,8
27	Pantai Pandawa	3,8	14,1	4,1	16,8	15,4	9,3	4,5	20,3	16,9	20,0	4,5	20,3	16,9	20,0
28	Candi Borobudur	4,3	18,4	1,9	3,6	8,1	55,7	3,3	10,9	14,1	23,0	3,4	11,6	14,6	20,7
29	GunungTangkuban Perahu	3,7	13,4	3,7	13,7	13,6	0,9	4,3	18,5	15,8	17,3	4,3	18,5	15,8	17,3
30	Putih crater	4,0	16,0	2,4	5,8	9,6	40,0	2,3	5,3	9,2	42,5	4,4	19,4	17,6	10,0
		100,8	345,9	105,1	378,1	350,1	20,4		481,6	397,6	28,3		527,1	418,6	31,8

Tabel 6. Hasil Perhitungan *PCC* dan Galat

	10 konten	20 konten	30 konten
PCC	0,968	0,974	0,980
Galat (%)	20,4	28,3	31,8

pemeringkat manual juga diukur dengan program diukur dengan metrik galat dan *PCC*. Pengujian juga dilakukan sebanyak tiga kali dengan jumlah konten yang berbeda: 10, 20, dan 30 konten. Nilai *PCC* juga cenderung baik mendekati 1, dengan hasil berturut-turut untuk 10, 20 dan 30 konten adalah 0,95931, 0,98896, dan 0,99076. Akan tetapi nilai galat cukup besar. Berturut-turut nilai galat untuk 10, 20, dan 30 konten sebesar 41,7%, 40,3%, dan 35,6%. Pada kasus ini cukup menarik, data menunjukkan nilai *PCC* dan galat semakin baik dengan semakin banyaknya konten yang terlibat.

Pada pemeringkat otomatis untuk penilaian terhadap kinerja organisasi berdasarkan umpan balik persepsi

masyarakat yang tertuang di internet juga memberikan hasil yang sangat positif [7]. Berturut-turut nilai *PCC* untuk pengujian dengan 10, 30, dan 50 konten adalah 0,9859, 0,9984, dan 0,9821. Hasil terbaik diperoleh pada kasus 30 konten. Sedangkan hasil terburuk pada kasus 50 konten. Meskipun demikian, seluruh nilai korelasi lebih dari 0,98 menunjukkan adanya korelasi yang kuat antara penilaian manual dengan penilaian sistem. Hasil ini menunjukkan hasil yang positif dari penerapan sistem pemeringkat kinerja organisasi berdasarkan kata sifat. Penelitian tersebut juga menghitung nilai galat dalam penilain nilai pemeringkat secara manual dan nilai pemeringkat yang dihasilkan oleh program. Berturut-



Tabel 7. Data Pemeringkat untuk 15 Tempat Wisata dari Tripadvisor

No	Nama	Kuesioner		10 konten		x.y	Galat	20 konten		x.y	Galat	30 konten		x.y	Galat
		x	x <sup>2</sup>	y	y <sup>2</sup>			y	y <sup>2</sup>			y	y <sup>2</sup>		
1	Air Hangat Cangar	3,0	9,0	3,1	9,6	9,3	3,3	4,0	16,0	12,0	33,3	4,3	18,5	12,9	43,3
2	Jawa Timur Park	4,0	16,0	4,0	16,0	16,0	0,0	4,2	17,6	16,8	5,0	4,2	17,6	16,8	5,0
3	Selecta	4,0	16,0	3,9	15,2	15,6	2,5	4,2	17,6	16,8	5,0	4,2	17,6	16,8	5,0
4	Air Terjun Rondo	4,0	16,0	3,9	15,2	15,6	2,5	4,0	16,0	16,0	0,0	4,3	18,5	17,2	7,5
5	Air Terjun Talun	4,5	20,3	3,1	9,6	14,0	31,1	4,1	16,8	18,5	8,9	4,4	19,4	19,8	2,2
6	Gunung Bromo	4,5	20,3	3,6	13,0	16,2	20,0	4,2	17,6	18,9	6,7	3,9	15,2	17,6	13,3
7	Museum Brawijaya	3,0	9,0	3,4	11,6	10,2	13,3	3,3	10,9	9,9	10,0	3,6	13,0	10,8	20,0
8	Taman Rekreasi Kota	2,5	6,3	4,1	16,8	10,3	64,0	4,7	22,1	11,8	88,0	4,4	19,4	11,0	76,0
9	Wendit	3,5	12,3	4,0	16,0	14,0	14,3	3,9	15,2	13,7	11,4	4,4	19,4	15,4	25,7
10	Taman Slamet	4,0	16,0	4,1	16,8	16,4	2,5	4,3	18,5	17,2	7,5	4,3	18,5	17,2	7,5
11	Kampung Warna Warni	4,0	16,0	4,4	19,4	17,6	10,0	4,5	20,3	18,0	12,5	4,5	20,3	18,0	12,5
12	Malioboro	4,0	16,0	2,7	7,3	10,8	32,5	2,7	7,3	10,8	32,5	2,6	6,8	10,4	35,0
13	Pantai Pandawa	4,0	16,0	4,1	16,8	16,4	2,5	4,5	20,3	18,0	12,5	4,5	20,3	18,0	12,5
14	Candi Borobudur	4,5	20,3	1,9	3,6	8,6	57,8	3,3	10,9	14,9	26,7	3,4	11,6	15,3	24,4
15	GunungTangkuban Perahu	3,5	12,3	3,7	13,7	13,0	5,7	4,3	18,5	15,1	22,9	4,3	18,5	15,1	22,9
16	Putih crater	4	16	2,4	5,76	9,6	40	2,3	5,29	9,2	42,5	4,4	19,36	17,6	10
		237,5		206,3		213,4	18,9	250,9		237,4	20,3	273,7		249,8	20,2

Tabel 8. Hasil Perhitungan PCC dan Galat

	10 konten	20 konten	30 konten
PCC	0,96408	0,972373	0,979822
Galat (%)	18,9	20,3	20,2

turut hasil rata-rata galat untuk 10, 30, dan 50 konten adalah 45,2%, 40,4%, dan 30,73%. Rata-rata nilai galat terbesar ada pada 10 konten, yaitu 45,2%. Tetapi nilai galat terkecil juga ada pada salah satu kata kunci di percobaan 10 konten, sebesar 3,38%.

### 3.2. Ketersediaan Implementasi sebagai *Library* Python

Untuk memudahkan penggunaan metoda ini bagi semua pihak yang membutuhkan, sistem ini juga telah dikembangkan dan tersedia dalam bentuk library Python. Paket dapat dilihat di link berikut: <https://pypi.org/project/PySpoks> [10]. Kemudian library dapat digunakan siapapun dengan menggunakan perintah “pip install PySpoks”. Library juga dapat diakses di GitHub [11]. Sedangkan pengembangan library ini diceritakan lebih detail pada publikasi “Pengembangan *Library* Sistem Pemeringkat otomatis Berbasis Kata Sifat” [9]. Library tersebut juga diuji pada kasus wisata dengan kata kunci lainnya selain [6] dengan hasil sangat baik. Nilai korelasi mendekati 1 dengan tingkat galat hanya berkisar dalam angka belasan [9].

## 4. Kesimpulan

Makalah ini telah memaparkan sebuah metoda baru yang dikembangkan dalam waktu 2 tahun terakhir. Berdasarkan penelitian yang telah dilakukan, nilai pemeringkat suatu obyek dapat di lakukan secara otomatis dari berbagai sumber teks di internet dengan menggunakan metoda sistem pemeringkat berbasis kata sifat. Metoda ini telah diuji dan digunakan pada 3 obyek kasus berbeda dengan hasil yang sangat menjanjikan.

Nilai PCC mendekati 1 dengan penilaian manual, sedangkan nilai galat tidak pernah melebihi angka 45%. Tingkat korelasi yang mendekati 1 serta baiknya prosentase galat yang didapat dengan metode ini membuktikan bahwa metoda ini sangat baik digunakan dalam sistem pemeringkat secara otomatis berdasar sumber teks di internet.

Selanjutnya penelitian ini dapat dikembangkan untuk berbagai manfaat terapan yang ada. Pengembangan lainnya yang mungkin dilakukan adalah modifikasi proses penghitungan baik itu dilakukan dan diuji ulang dengan menggunakan kasus-kasus yang sama atau berbeda. Pelibatan tipe kata selain kata sifat juga menarik untuk dikaji lebih jauh dampaknya terhadap unjuk kerja.

## Ucapan Terimakasih

Makalah ini merupakan hasil penelitian skim inovasi yang dibiayai dengan sumber dana DIPA BLU Politeknik Negeri Malang dengan nomer: DIPA 042.01.2.401004/2018 tanggal 05 Desember 2017 dengan Surat Perjanjian No: 6332/PL2.1/ HK/2018. Para penulis menghaturkan banyak terima kasih kepada Direktur Politeknik Negeri Malang dan UPT P2M atas dukungan pendanaan yang telah dilakukan.

## Daftar Rujukan

- [1] D. Hutabarat, “Inilah rincian jumlah pelanggan prabayar masing-masing operator,” 2018. [Online]. Available: <https://kominfo.go.id/content/detail/13131/inilah-rincian-jumlah->



- pelanggan-prabayar-masing-masing-operator/0/sorotan\_media.
- [2] Katadata.co.id, "Jumlah Penduduk Indonesia 2019 Mencapai 267 Juta Jiwa," 2019. [Online]. Available: <https://data.boks.katadata.co.id/datapublish/2019/01/04/jumlah-penduduk-indonesia-2019-mencapai-267-juta-jiwa>.
- [3] B. Pan, T. MacLaurin, and J. C. Crotts, "Travel Blogs and the Implications for Destination Marketing," *J. Travel Res.*, vol. 46, no. 1, pp. 35–45, 2007.
- [4] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, Jan. 2008.
- [5] M. Jabreel, A. Moreno, and A. Huertas, "Semantic comparison of the emotional values communicated by destinations and tourists on social media," *J. Destin. Mark. Manag.*, vol. 6, no. 3, pp. 170–183, 2017.
- [6] D. M. Ramadhani, C. Rahmad, and F. Rahutomo, "Tourism destination pemeringkat system based on social media analysis (proposal and dataset development in Indonesian language)," in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, 2017, pp. 41–46.
- [7] F. Rahutomo, Z. H. R. Adi, I. F. Rozi, and P. Y. Saputra, "Implementasi Text Mining Pada Website/Blog Di Internet Untuk Menilai Kinerja Suatu Organisasi," *Inovtek Polbeng Seri Inform.*, vol. 3, no. 2, pp. 101–109, 2018.
- [8] F. Rahutomo, I. P. Hidayati, P. Y. Saputra, I. F. Rozi, U. Nurhasan, and D. M. Ramadhani, "Automatic Pemeringkat System for National Delivery Services," in *2018 International Conference on Applied Science and Technology (iCAST)*, 2018, pp. 226–230.
- [9] M. E. R. Putra, F. Rahutomo, and Y. Pramitarini, "Pengembangan Library Sistem Pemeringkat Otomatis Berbasis Kata Sifat," *J. Inform.*, vol. 6, no. 2, 2019.
- [10] M. E. R. Putra, F. Rahutomo, and Y. Pramitarini, "Sistem Pemeringkat Otomatis Berbasis Kata Sifat," 2019. [Online]. Available: <https://pypi.org/project/PySpoks/>.
- [11] M. E. R. Putra, F. Rahutomo, and Y. Pramitarini, "Library Sistem Pemeringkat Otomatis Berbasis Kata Sifat," 2019. [Online]. Available: <https://github.com/ElfanRodh/PySpoks>.
- [12] E. Costa *et al.*, "A framework for building web mining applications in the world of blogs: A case study in product sentiment analysis," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 4813–4834, 2012.
- [13] G. A. Miller, "WordNet: A Lexical Database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995.
- [14] F. Rahutomo and E. Rohadi, "Pengembangan Piranti Penelitian Sistem Temu Kembali Informasi Bahasa Indonesia," in *Seminar Nasional Sistem Informasi Indonesia (SESINDO)*, 2015, pp. 313–319.
- [15] D. Zamzami, F. Rahutomo, and D. Puspitasari, "Aplikasi Wordnet Indonesia Berdasarkan Kamus Thesaurus Bahasa Indonesia menggunakan Algoritma Rule Based Text Parsing," in *Seminar Informatika Aplikatif Polinema*, 2016.
- [16] P. Bahasa, *Kamus Tesaurus Bahasa Indonesia*. Departemen Pendidikan Nasional, 2008.