

Pemanfaatan Filter dalam Object-based Opinion Mining pada Review Produk Pariwisata

Bonda Sisephaputra¹⁾ and Aris Tjahyanto²⁾

*Master's Program in Information System, Institut Teknologi Sepuluh Nopember
Jl. Raya ITS, Keputih, Sukolilo, Surabaya, 60111, Indonesia
e-mail: ¹⁾bonda13@mhs.if.its.ac.id, ²⁾arist@its-sby.edu*

Abstract

The quality of a tourism product can be seen from many objects because each product is a unique individual. Therefore to get a clearer picture about a product, each object has to be evaluated. The information about each object can be extracted by using object-based opinion mining. Based on the previous research, the application of Natural Language Processing (NLP) rules on object-based opinion mining to determine the orientation of the semantic objects has showed good result. However, the performance of the extraction objects task quite bad in the tourism domain. In this study, researchers apply a filter on the objects' extraction process of the hotel and restaurant review data. The utilization of data filter in object-based opinion mining has succeeded in obtaining better objects' extraction result, because the utilization of filter will eliminate each object which is unrelated with the object of the hotel or restaurant domain.

The application of filter in the objects' extraction process can improve the precision of frequent object approach from 45.7% to 64.49% on of the hotel review and from 44.82% to 64.61% on the restaurant review. In the objects' extraction process which is using frequent and infrequent approach, the precision is increased from 22.33% to 63.02% on the hotel review and from 21.6% to 65.4% on the restaurant review. In the performance measurement process, with the usage of all objects that have been successfully extracted, the use of filter got more dominant result compared to no filter on object sentiment classification process. The utilization of filter on frequent object approach got 56.85% accuracy, 60.91% precision, and 79.93% recall on the hotel review, and got 58.85% accuracy, 63.26% precision, and 84.14% recall on the restaurant review. While in the process of sentences sentiment classification, the usage of extracted object as a classification attribute to determine the orientation of the sentence got better results than only using the sentence review directly as an classification attribute, although the results object of object-based opinion mining is worse than the usage of the object of the corpus data as an attribute.

Abstrak

Dalam menentukan kualitas dari suatu produk pariwisata dapat dilihat dari banyak aspek karena masing-masing memiliki keunikan masing-masing, sehingga untuk mendapatkan gambaran yang lebih jelas maka harus dievaluasi masing-masing aspek yang ada. Penggalan informasi yang cocok adalah dengan object-based opinion mining. Penerapan aturan NLP untuk menentukan orientasi semantik dari objek yang berhasil diekstrak terbukti efektif. Akan tetapi, kinerja pada tugas ekstraksi objek cukup buruk dalam domain pariwisata.

Pada penelitian ini peneliti menggunakan filter dalam proses ekstraksi objek yang terdapat pada data review hotel dan restoran. Penggunaan data filter pada object-based opinion mining berhasil memperoleh hasil ekstraksi objek yang lebih baik, karena dapat memfilter objek-objek yang bukan merupakan objek yang berpengaruh terhadap kualitas suatu hotel atau restoran.

Penerapan filter dalam proses ekstraksi objek jelas dapat meningkatkan presisi yaitu pada pendekatan frequent object yang semula 45.7% menjadi 64.49% pada review hotel dan dari 44.82% menjadi 64.61% pada review restoran. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan frequent object dan infrequent object dimana terjadi peningkatan hasil presisi yang semula 22.33% menjadi 63.02% pada review hotel dan dari 21.6% menjadi 65.4% pada review restoran. Pada proses pengukuran performa klasifikasi sentimen objek menggunakan semua objek yang telah berhasil diekstrak, penggunaan filter mendapatkan hasil yang dominan secara keseluruhan dibanding dengan tanpa filter. Pada pendekatan frequent object mendapatkan 56.85% akurasi, 60.91% presisi, dan 79.93% recall pada review hotel, serta mendapatkan hasil 58.85% akurasi, 63.26% presisi, dan 84.14% recall pada review restoran. Pada proses klasifikasi sentimen kalimat penggunaan objek terekstraksi sebagai atribut klasifikasi

untuk menentukan orientasi kalimat mendapatkan hasil yang lebih baik dibanding dengan hanya menggunakan kalimat review secara langsung sebagai atribut klasifikasi, walaupun penggunaan objek dari object-based opinion mining masih kurang dibandingkan dengan penggunaan aspek corpus sebagai atribut.

Kata Kunci: *object-based opinion mining, opinion mining, analisis sentimen, pariwisata, hotel, restoran*

1. PENDAHULUAN

Informasi tentang pariwisata menjadi salah satu yang diakses oleh masyarakat. Dengan tersedianya informasi ini maka masyarakat tidak lagi kesulitan dalam mencari informasi tempat tujuan mereka menghabiskan masa liburan. Banyak website yang menyediakan informasi ulasan tentang produk pariwisata dan menyediakan fitur untuk para pengunjung website dapat memberikan umpan balik yang berupa ulasan atau opini mereka terhadap objek tersebut. Masalah yang terjadi adalah dengan banyaknya feedback yang ada akan sulit untuk seseorang membaca semua ulasan dan mendapatkan gambaran secara keseluruhan.

Hal yang nantinya akan menjadi tantangan dalam penggalian informasi terhadap data review adalah banyak kalimat yang menyebutkan opini dari obyek yang sedang dikaji atau juga dari setiap fitur dan komponen yang ada didalamnya. Di sisi lain, banyak juga kalimat yang tidak mengandung opini, dan juga menyebutkan benda-benda yang tidak sesuai dengan atribut atau komponen dari obyek yang diulas [1].

Dalam menentukan kualitas dari suatu produk pariwisata dapat dilihat dari banyak aspek karena masing-masing tempat memiliki keunikan masing-masing, sehingga untuk mendapatkan gambaran yang lebih jelas maka harus dievaluasi masing-masing aspek yang ada. Dengan demikian maka yang paling cocok untuk penggalian informasi ini adalah dengan aspect-based opinion mining. Dalam metode ini analisis dilakukan secara mendetail sampai dengan analisis pendapat untuk masing-masing aspek yang ada di dalam suatu opini, sehingga nantinya akan didapat gambaran yang lebih jelas untuk masing-masing aspek suatu produk pariwisata.

Marrese-Taylor [1] melakukan penelitian dimana mereka menerapkan analisis aspect-based opinion mining dalam bidang pariwisata khususnya pengolahan data opini terhadap layanan hotel dan restoran. Hasil penelitian tersebut menghasilkan sebuah prototype pengolahan data opini dan memiliki hasil output analisis secara visual yang dapat dengan mudah dipahami banyak pihak.

Penerapan aturan NLP untuk menentukan orientasi semantik dari aspek yang berhasil diekstrak terbukti sangat efektif, mencapai presisi dan recall rata-rata 90%. Sebaliknya, Hasil dari penelitian tersebut menunjukkan bahwa kinerja pada tugas ekstraksi aspek cukup buruk dalam domain pariwisata. Hal tersebut disebabkan oleh

persentase yang tinggi dari ekspresi diekstraksi yang tidak sesuai dengan ekspresi aspek nyata untuk kedua domain tersebut. Di sisi lain, klasifikasi sentimen menunjukkan hasil yang cukup baik, tetapi dalam kasus ini sebagian besar kesimpulan yang mungkin sulit untuk dibuktikan karena tugas ini hanya dievaluasi dari ekspresi aspek yang diekstrak. Karena ekspresi ini merupakan yang paling sederhana, menentukan orientasi sentimen pada aspek tersebut mungkin lebih mudah. Akibatnya, presisi dan recall bisa menurun ketika semua ekspresi aspek dianggap. Penyebab buruknya hasil ekstraksi aspek adalah pemberi opini menyebutkan benda-benda yang tidak sesuai dengan atribut atau komponen dari produk, hal tersebut juga berakibat diperolehnya presisi yang rendah dalam ekstraksi aspek eksplisit pada kedua kasus.

Dikarenakan hasil yang diperoleh oleh Marrese-Taylor [1] dalam hal ekstraksi fitur masih kurang, maka pada penelitian kali ini peneliti akan mencoba untuk menggunakan filter untuk mendapatkan hasil yang lebih baik. Fokus dari penelitian ini adalah untuk menerapkan filter dengan menggunakan data filter yang disusun berdasarkan faktor dan komponen yang ada dan berhubungan dengan domain hotel dan restoran, sehingga dapat memperbaiki hasil ekstraksi aspek dengan mengeliminasi objek yang dianggap tidak mewakili faktor-faktor yang berhubungan dengan kualitas suatu produk pariwisata.

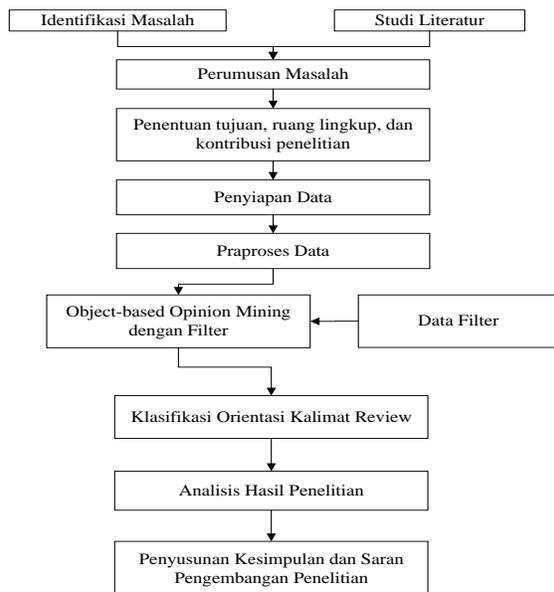
Dalam mempelajari penelitian yang dilakukan Marrese-Taylor [1] dimana menjadi referensi utama dalam penelitian ini, penulis mendapatkan kesulitan dalam memahami dan menerapkan konsep aspek yang dikemukakan oleh Marrese-Taylor dalam tahap ekstraksi aspek. Untuk membuatnya lebih simple dan tidak menyalahi konsep yang ada, akhirnya dalam penelitian ini penulis menyebut komponen atau atribut yang merupakan bagian dari domain hotel dan restoran yang menjadi target opini akan disebut sebagai objek. Objek diekstrak dari kata benda dan frase kata benda yang ada dalam kalimat review yang dianggap mempengaruhi kualitas suatu layanan produk pariwisata. Sehingga metode opinion mining dalam penelitian ini akan disebut sebagai analisis object-based opinion mining dengan acuan berdasarkan analisis aspect-based opinion mining yang dilakukan oleh [1].

Selain itu, pada penelitian ini penulis mencoba untuk menguji apakah hasil dari object-based opinion mining dapat digunakan untuk menentukan orientasi sentiment

secara umum. Selain membutuhkan informasi objek-objek yang dikomentari, tentu saja pembaca review produk pariwisata perlu untuk memahami kesimpulan dari setiap review yang ada. Oleh sebab itu, peneliti mencoba untuk mengklasifikasikan review data ke dalam beberapa kelas: positif, netral, atau negatif. Sehingga dapat ditarik kesimpulan apakah obyek wisata yang layak direkomendasikan atau tidak. Pada penelitian ini akan dilakukan pemrosesan data hasil pengolahan object-based opinion mining dengan melakukan proses klasifikasi menggunakan machine learning. Peneliti akan membandingkan hasil prediksi dengan menggunakan tiga teknik klasifikasi yang berbeda, yaitu Naïve Bayes, support-vector machines (SVM), dan Random Forest (RF).

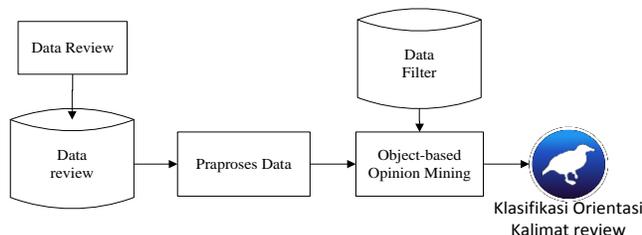
2. METODOLOGI PENELITIAN

Tahap-tahap yang akan dilakukan pada penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Metodologi Penelitian

Sedangkan gambaran umum mengenai sistem yang dibangun dan digunakan pada penelitian ini dapat dilihat pada gambar 2 berikut ini. Setiap bagiannya akan dijelaskan lebih jauh pada bagian berikutnya.



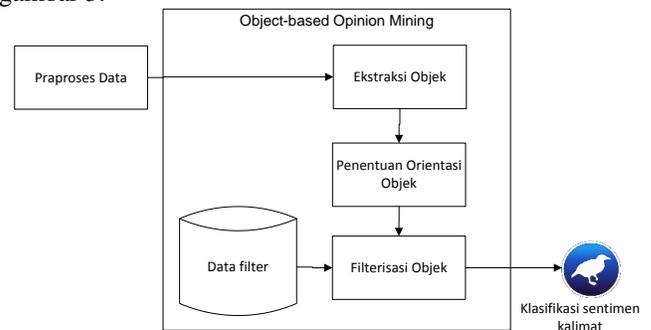
Gambar 2. Gambaran Umum Sistem

Untuk mendapatkan hasil yang secara langsung dapat dibandingkan, maka pada penelitian ini digunakan data *review* dan data uji yang sama dengan yang digunakan pada penelitian Marrese-Taylor [1]. Data uji merupakan data *review* yang telah dilakukan analisis secara manual sehingga dapat digunakan untuk mengukur tingkat keberhasilan dari penelitian yang dilakukan.

Dalam penelitian ini menggunakan data filter restoran dan hotel. Data ini digunakan untuk memfilter objek-objek diluar dari domain yang diteliti. Data filter yang digunakan dalam penelitian ini berasal dari beberapa sumber dan disusun ulang oleh peneliti sesuai dengan kebutuhan penelitian. Data filter yang digunakan terdiri dari 396 komponen yang berhubungan dengan hotel dan 260 komponen yang berhubungan dengan restoran.

Pertama kali yang dilakukan pada penelitian adalah menyimpan data *review* yang digunakan untuk penelitian perlu melalui tahap praproses sebelum data tersebut diolah lebih lanjut. Hal ini perlu dilakukan agar hasil klasifikasi memiliki tingkat akurasi yang lebih baik. Untuk semua tahapan praproses data ini peneliti akan menggunakan *NLTK libraries*, dimana untuk menggunakan *libraries* ini peneliti menggunakan Python sebagai bahasa pemrograman. Tahap praproses data pada penelitian ini terdiri dari 5 tahap, yaitu tokenizing, POS Tagging, Chunking, Stopping, dan Stemming.

Bagian utama dari penelitian ini adalah tahap *object-based opinion mining*, dimana data *review* yang sudah melalui praproses data akan diolah. Tujuan utama dari proses ini adalah untuk menemukan objek apa saja yang dikomentari di dalam *review* dan menyimpulkan apakah komentar tersebut positif atau negatif. Dalam *object-based opinion mining* dibagi menjadi dua proses analisis, yaitu proses ekstraksi objek dan proses penentuan orientasi objek. Sehingga setelah proses analisis selesai dapat dilakukan penyusunan rangkuman hasil analisis yang diharapkan menghasilkan informasi dapat dibaca dengan mudah oleh semua pihak. Adapun gambaran dari proses *object-based opinion mining* dapat dilihat pada gambar 3.



Gambar 3. Tahapan Object-based Opinion Mining

Tahap ekstraksi ekspresi objek bertujuan untuk menemukan dan mengekstrak topik penting dalam teks

review yang kemudian akan digunakan untuk membagi-bagi opini kedalam sub-sub berdasarkan objek yang ditemukan. Pertama, pendapat yang diberikan oleh pengguna bisa atau tidak bisa secara eksplisit merujuk pada objek dan sebuah pendapat bisa menggunakan seluruh rangkaian ekspresi yang berbeda [2]. Menimbang bahwa objek adalah atribut atau komponen dari suatu entitas, nama objek berupa kata benda atau frase kata benda.

Tahap penentuan orientasi objek berfungsi untuk menentukan orientasi sentimen pada setiap objek. Proses pada tahap ini bergantung pada kamus kata sentimen yang berisi daftar kata-kata positif dan negatif yang digunakan untuk mencocokkan istilah dalam teks. Proses penentuan orientasi sentimen pada objek juga perlu penjelasan lebih dalam. Liu [3] membuat asumsi bahwa setiap kalimat mengungkapkan pendapat dari pemegang pendapat, yaitu hanya informasi yang terkandung dalam kalimat di mana objek muncul akan dipertimbangkan untuk menentukan apa yang pengguna ceritakan tentang objek ini. Kita akan menganggap bahwa pengguna memberikan pendapat atas objek jika dalam kalimat di mana objek muncul (secara implisit atau eksplisit) perasaan positif atau negatif. Perasaan ini sebagian besar dinyatakan atau ditunjukkan oleh kata-kata khusus, kelompok kata, referensi komparatif atau bahkan tanda baca. Namun demikian, lebih umum, deskripsi, penilaian atau rekomendasi akan di sini dianggap sebagai pendapat.

Pada tahap selanjutnya kita akan menganalisa orientasi dari semua objek yang berhasil diekstrak dengan acuan hasil dari aturan orientasi kata. Setelah orientasi sentimen masing-masing objek diketahui maka akan dilakukan proses filterisasi objek menggunakan data filter, baru setelah itu hasilnya akan disimpan dalam tabel *opinion_mining_result*. Proses filterisasi dilakukan saat terakhir karena peneliti menganggap saat inilah yang paling efisien dalam melakukan filterisasi dimana hasil *object-based opinion mining* sudah terlihat dan hasil filterisasi hanya berupa keterangan saja apakah suatu objek termasuk dalam filter atau tidak.

Setelah seluruh proses *object-based opinion mining* selesai, maka peneliti akan mencoba melakukan penelitian lanjutan untuk menganalisis apakah hasil objek dan orientasinya yang berhasil diekstrak dapat digunakan untuk melakukan klasifikasi untuk menentukan orientasi sentiment *review* secara global tiap kalimat. Untuk dapat melakukan analisis ini maka peneliti akan mencoba melakukan 7 skenario uji coba yang berbeda untuk masing-masing data *review* hotel dan restoran. Peneliti akan menggunakan WEKA sebagai tool untuk melakukan proses klasifikasi. Dari proses klasifikasi ini nantinya akan menghasilkan kesimpulan, apakah hasil *object-based opinion mining* yang telah dihasilkan sebelumnya dapat digunakan untuk menentukan orientasi sentimen data *review* secara global.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Ekstraksi Objek

Dari hasil uji coba ini dapat diketahui bahwa pendekatan ekstraksi objek yang telah dilakukan tidak ada yang mendapatkan hasil dominan secara keseluruhan, akan tetapi penerapan filter dalam proses ekstraksi objek jelas dapat meningkatkan presisi dengan cukup meyakinkan yaitu pada pendekatan *frequent object* yang semula 45.7% menjadi 64.49% pada data *review* hotel dan dari 44.82% menjadi 64.61% pada data *review* restoran, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 44.35% menjadi 43.52% pada data *review* hotel dan dari 45.4% menjadi 45.11% pada data *review* restoran. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan *frequent object* dan *infrequent object* dimana pada kedua data *review* terjadi peningkatan hasil presisi yang cukup meyakinkan yaitu yang semula 22.33% menjadi 63.02% pada data *review* hotel dan dari 21.6% menjadi 65.4% pada data *review* restoran, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 59.97% menjadi 55.95% pada data *review* hotel dan dari 61.06% menjadi 59.2% pada data *review* restoran. Waktu yang dibutuhkan untuk keempat pendekatan tersebut tidak lebih dari 4 detik.

Tabel 1. Hasil Uji Ekstraksi Objek Hotel

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	36.80%	45.70%	44.35%	45.01%
Frequent and Infrequent Object	31.82%	22.33%	59.97%	32.54%
Frequent Object with filter	38.74%	64.49%	43.52%	51.97%
Frequent and Infrequent Object with Filter	46.15%	63.02%	55.95%	59.27%
Filter Object	46.76%	62.36%	57.39%	59.77%

Tabel 2. Hasil Uji Ekstraksi Objek Restoran

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	35.86%	44.82%	45.40%	45.11%
Frequent and Infrequent Object	31.37%	21.60%	61.06%	31.91%
Frequent Object with filter	39.91%	64.61%	45.11%	53.13%
Frequent and Infrequent Object with Filter	48.93%	65.40%	59.20%	62.15%
Filter Object	49.16%	63.76%	60.92%	62.31%

Hasil yang cukup menarik didapatkan dari hasil ekstraksi objek dengan menggunakan pendekatan data filter yang secara langsung digunakan untuk mengekstrak objek yang ada pada data *review*, dimana pendekatan ini mendapatkan hasil akurasi dan f-measure tertinggi baik pada kedua data *review* yaitu mendapatkan 46.76% akurasi dan 59.77% f-measure pada data *review* hotel dan

mendapatkan 49.16% akurasi dan 62.31% f-measure pada data *review* restoran. Akan tetapi hasil tersebut dapat diperoleh dengan menghabiskan waktu yang jauh lebih lama dibandingkan dengan pendekatan yang lain yaitu 3051 detik pada data hotel dan 1559 detik pada data restoran, padahal pendekatan lain hanya memakan waktu tidak lebih dari 4 detik untuk melakukan ekstraksi objek. Hal ini terjadi dikarenakan pada pendekatan data filter secara langsung setiap kalimat yang sudah dipecah ke dalam token-token akan dilakukan scanning dan dicocokkan pada semua data filter yang ada, dimana tentu saja hal itu akan membutuhkan waktu yang cukup lama.

3.2. Hasil Klasifikasi Sentimen Objek

Pengujian dilakukan dengan menggunakan seluruh objek yang berhasil diekstrak dalam *object-based opinion mining*. Berbeda dengan hasil pada tahap sebelumnya yang hanya menggunakan objek yang sesuai dengan corpus saja, disini penggunaan filter dalam *object-based opinion mining* benar-benar terlihat penannya dengan mendapatkan hasil yang dominan secara keseluruhan dibanding dengan tanpa filter. Hasil tertinggi didapatkan pada pendekatan *frequent object* dengan filter, yaitu 56.85% akurasi, 60.91% presisi, 79.93% recall, 69.14% f-measure pada data *review* hotel, serta mendapatkan hasil 58.85% akurasi, 63.26% presisi, 84.14% recall, 72.22% f-measure pada data *review* restoran. Sedangkan untuk pendekatan data filter secara langsung, hasil yang diperoleh tidak lebih baik dari pendekatan lain yang menggunakan filter dan ditambah lagi waktu yang dibutuhkan juga jauh lebih lama yaitu 2700 detik pada data *review* hotel dan 1291 detik pada data *review* restoran. Padahal pendekatan lain hanya memakan waktu tidak lebih dari 202 detik.

Tabel 3. Hasil Uji Klasifikasi Sentimen Objek Hotel

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	40.23%	44.94%	65.22%	53.21%
Frequent and Infrequent Object	19.57%	21.59%	40.43%	28.15%
Frequent Object with filter	56.85%	60.91%	79.93%	69.14%
Frequent and Infrequent Object with Filter	55.25%	59.41%	76.27%	66.79%
Filter Object	53.29%	58.46%	73.79%	65.24%

Tabel 4. Hasil Uji Klasifikasi Sentimen Objek Restoran

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	40.85%	44.71%	73.80%	55.68%
Frequent and Infrequent Object	19.16%	21.06%	50.68%	29.76%
Frequent Object with filter	58.85%	63.26%	84.14%	72.22%
Frequent and Infrequent Object with Filter	57.94%	61.51%	84.24%	71.10%
Filter Object	55.49%	59.32%	82.75%	69.10%

3.3. Hasil Deteksi Kalimat Opini

Dari hasil uji coba ini dapat diketahui bahwa penerapan filter dalam proses deteksi kalimat opini hanya dapat meningkatkan sedikit saja presisi yaitu pada pendekatan *frequent object* yang semula 88.15% menjadi 91.35% pada data *review* hotel dan dari 90.41% menjadi 93.13% pada data *review* restoran, dengan mengorbankan sedikit penurunan juga pada hasil recall yaitu yang semula 75.7% menjadi 71.1% pada data *review* hotel dan dari 71.54% menjadi 66.12% pada data *review* restoran. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan *frequent object* dan *infrequent object* dimana pada kedua data *review* terjadi peningkatan sedikit saja presisi yaitu yang semula 85.76% menjadi 89.9% pada data *review* hotel dan dari 89.71% menjadi 93.2% pada data *review* restoran, dengan mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 83.09% menjadi 76.03% pada data *review* hotel dan dari 82.66% menjadi 74.25% pada data *review* restoran. Waktu yang dibutuhkan untuk keempat pendekatan tersebut tidak lebih dari 22 detik.

Tabel 5. Hasil Uji Deteksi Kalimat Opini Hotel

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	73.38%	88.15%	75.70%	81.45%
Frequent and Infrequent Object	76.30%	85.76%	83.09%	84.40%
Frequent Object with filter	72.50%	91.35%	71.10%	79.96%
Frequent and Infrequent Object with Filter	74.90%	89.90%	76.03%	82.39%
Filter Object	75.16%	89.94%	76.35%	82.59%

Tabel 6. Hasil Uji Deteksi Kalimat Opini Restoran

Jenis Ekstraksi Objek	Accuracy	Precision	Recall	F-measure
Frequent Object	71.70%	90.41%	71.54%	79.88%
Frequent and Infrequent Object	78.94%	89.71%	82.66%	86.04%
Frequent Object with Filter	69.57%	93.13%	66.12%	77.33%
Frequent and Infrequent Object with Filter	75.53%	93.20%	74.25%	82.65%
Filter Object	76.60%	93.31%	75.61%	83.53%

Untuk hasil proses deteksi kalimat opini dengan menggunakan pendekatan data filter yang secara langsung mendapatkan hasil cukup bagus yaitu 75.38% akurasi, 89.94% presisi, 76.35% recall, dan 82.59% f-measure pada data *review* hotel. Sedangkan untuk data *review* restoran mendapatkan hasil 76.6% akurasi, 93.31% presisi, 75.61% recall, dan 83.53% f-measure. Hasil tersebut mendapatkan hasil f-measure lebih baik jika dibandingkan dengan pendekatan lain yang menggunakan data filter. Seperti tahap analisis sebelumnya pendekatan ini memakan waktu lebih lama dibandingkan pendekatan lain yaitu 35 detik pada hotel dan 21 detik pada restoran.

3.4. Hasil Klasifikasi Sentimen Kalimat Review

Dari hasil uji coba ini dapat diketahui bahwa penerapan kalimat review secara langsung ke dalam proses klasifikasi sentimen menghasilkan hasil yang rendah yaitu 56.40% akurasi, 31.8% presisi, 56.4% recall, dan 40.7% f-measure pada data review hotel dan 64.26% akurasi, 41.3% presisi, 64.3% recall, 50.3% f-measure pada data review restoran. Hasil yang didapatkan semua sama persis pada ketiga teknik klasifikasi yang digunakan, hal ini terjadi karena semua kalimat hanya diklasifikasikan ke dalam kelas positif saja.

Hasil yang cukup bagus didapatkan pada uji coba dengan melakukan klasifikasi menggunakan objek yang ada pada data corpus, dimana hasil yang diperoleh cukup bagus pada semua klasifikasi. Pada data review hotel ketiga classifier mendapatkan hasil sama persis yaitu 86.44% akurasi, 89.1% presisi, 86.4% recall, 83.9% f-measure. Hasil berbeda ditunjukkan pada klasifikasi data review restoran dimana SVM dan Random Forest mendapatkan akurasi lebih baik dibanding dengan Naïve Bayes yaitu 89.15% dibandingkan dengan 88.51%. Kesimpulan yang dapat ditarik dari hasil ini adalah bahwa penggunaan objek yang terekstrak dari kalimat review dapat digunakan untuk proses klasifikasi dalam menentukan orientasi sentimen kalimat review. Berhubung data aspek pada corpus adalah data uji yang digunakan dalam menentukan performa object-based opinion mining, maka dalam proses klasifikasi ini kita juga perlu menguji hasil dari masing-masing pendekatan yang telah dilakukan pada object-based opinion mining untuk mengetahui apakah hasil tersebut dapat menghasilkan hasil yang bagus jika dijadikan atribut klasifikasi penentuan orientasi sentimen kalimat.

Pada uji coba klasifikasi dengan menggunakan objek yang ada pada hasil object-based opinion mining sebagai atribut menunjukkan bahwa hasil yang diperoleh tidaklah lebih bagus dibanding dengan hasil klasifikasi yang menggunakan aspek corpus sebagai atribut. Akan tetapi bila dibandingkan dengan hasil klasifikasi yang hanya menggunakan kalimat review saja sebagai atribut, maka bisa dibilang hasil klasifikasi dari objek hasil object-based opinion mining mendapatkan hasil yang lebih bagus.

Secara keseluruhan tidak ada teknik klasifikasi yang mendapatkan hasil yang dominan lebih baik dibanding dengan yang lain pada semua hasil pendekatan object-based opinion mining. Akan tetapi dapat disimpulkan bahwa klasifikasi yang lebih banyak mendapatkan hasil yang lebih baik adalah teknik klasifikasi Random Forest, dimana pada hasil klasifikasi dengan atribut data frequent object dan infrequent object dengan filter, Random Forest mendominasi seluruh hasil dibanding dengan Naïve Bayes dan SVM. Pada data review hotel Random Forest mendapatkan hasil 61.34% akurasi, 62.2% presisi, 61.3% recall, 57.1% f-measure dan pada data review restoran mendapatkan hasil 64.68% akurasi, 65.5% presisi, 64.7% recall, 61.7% f-measure.

Dari uji coba secara keseluruhan dari proses klasifikasi sentimen kalimat dapat ditarik kesimpulan bahwa penggunaan objek terekstraksi sebagai atribut klasifikasi untuk menentukan orientasi kalimat mendapatkan hasil yang lebih baik dibanding dengan hanya menggunakan kalimat review secara langsung sebagai atribut klasifikasi. Walaupun hasil objek dari object-based opinion mining masih kurang dibandingkan dengan penggunaan aspek corpus sebagai atribut, akan tetapi hasil pemanfaatan object-based opinion mining dalam klasifikasi sentimen kalimat menghasilkan lebih baik dibanding menggunakan kalimat review. Selain itu yang paling penting adalah hasil dari proses klasifikasi menggunakan objek dari object-based opinion mining dengan filter mendapatkan hasil yang lebih baik dibanding dari tanpa menggunakan data filter sama sekali.

4. KESIMPULAN DAN SARAN

Setelah penelitian mengenai pemanfaatan filter pada object-based opinion mining yang telah selesai dilaksanakan, maka berikut ini merupakan kesimpulan dari penelitian ini.

- a. Penerapan filter dalam proses ekstraksi objek jelas dapat meningkatkan presisi dengan cukup meyakinkan yaitu pada pendekatan frequent object yang semula 45.7% menjadi 64.49% pada data review hotel dan dari 44.82% menjadi 64.61% pada data review restoran, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 44.35% menjadi 43.52% pada data review hotel dan dari 45.4% menjadi 45.11% pada data review restoran. Hal serupa juga terjadi pada pendekatan ekstraksi objek dengan menggunakan frequent object dan infrequent object dimana pada kedua data review terjadi peningkatan hasil presisi yang cukup meyakinkan yaitu yang semula 22.33% menjadi 63.02% pada data review hotel dan dari 21.6% menjadi 65.4% pada data review restoran, dengan hanya mengorbankan sedikit penurunan pada hasil recall yaitu yang semula 59.97% menjadi 55.95% pada data review hotel dan dari 61.06% menjadi 59.2% pada data review restoran.
- b. Penggunaan filter dalam object-based opinion mining mendapatkan hasil yang dominan secara keseluruhan dibanding dengan tanpa filter pada proses klasifikasi sentimen objek. Hal ini terbukti saat dilakukan pengukuran performa menggunakan semua objek yang telah berhasil diekstrak pada semua pendekatan yang telah dilakukan. Hasil tertinggi didapatkan pada pendekatan frequent object dengan filter, yaitu 56.85% akurasi, 60.91% presisi, 79.93% recall, 69.14% f-measure pada data review hotel, serta mendapatkan hasil 58.85% akurasi, 63.26% presisi, 84.14% recall, 72.22% f-measure pada data review restoran.

- c. Penggunaan data filter secara langsung dalam object-based opinion mining sebenarnya dapat mendapatkan hasil yang cukup bagus, akan tetapi efek negatif yang ditimbulkan adalah proses analisis menghabiskan waktu yang jauh lebih lama dibandingkan dengan pendekatan yang lain contohnya yaitu pada proses ekstraksi objek waktu yang dibutuhkan mencapai 3.051 detik pada data hotel dan 1.559 detik pada data restoran, padahal pendekatan lain hanya memakan waktu tidak lebih dari 4 detik untuk melakukan ekstraksi objek.
- d. Pada proses klasifikasi sentimen kalimat dapat ditarik kesimpulan bahwa penggunaan objek terekstraksi sebagai atribut klasifikasi untuk menentukan orientasi kalimat mendapatkan hasil yang lebih baik dibanding dengan hanya menggunakan kalimat review secara langsung sebagai atribut klasifikasi. Walaupun hasil objek dari object-based opinion mining masih kurang dibandingkan dengan penggunaan aspek corpus sebagai atribut, akan tetapi hasil pemanfaatan object-based opinion mining dalam klasifikasi sentimen kalimat menghasilkan lebih baik dibanding menggunakan kalimat review. Selain itu yang paling penting adalah hasil dari proses klasifikasi menggunakan objek dari object-based opinion mining dengan filter mendapatkan hasil yang lebih baik dibanding dari tanpa menggunakan data filter sama sekali. Secara keseluruhan tidak ada teknik klasifikasi yang mendapatkan hasil yang dominan lebih baik dibanding dengan yang lain, tetapi teknik klasifikasi yang lebih banyak mendapatkan hasil yang lebih baik adalah teknik klasifikasi Random Forest, dimana pada hasil klasifikasi dengan atribut data frequent object dan infrequent object dengan filter, Random Forest mendominasi seluruh hasil dibanding dengan Naïve Bayes dan SVM. Pada data review hotel Random Forest mendapatkan hasil 61.34% akurasi, 62.2% presisi, 61.3% recall, 57.1% f-measure dan pada data review restoran mendapatkan hasil 64.68% akurasi, 65.5% presisi, 64.7% recall, 61.7% f-measure.

Pada penelitian ini memang peran dari data filter dapat meningkatkan hasil object-based opinion mining, akan

tetapi dalam penelitian ini masih banyak yang bisa digali yaitu seperti belum adanya proses untuk menganalisa objek yang secara implisit terdapat dalam data review, karena dalam konteks data review pariwisata cukup banyak opini terhadap objek yang disampaikan secara implisit.

Belum adanya konsep dan landasan ilmiah yang pasti dalam menyusun dan menentukan data filter yang sesuai, sehingga proses penyusunan data filter yang digunakan dalam penelitian ini dilakukan secara manual oleh peneliti, yaitu berdasarkan data ontologi yang didapat dari internet dan juga berdasarkan data aspek yang ada pada data corpus. Oleh karena itu apabila ada peneliti lain mencoba melakukan penelitian serupa maka kemungkinan besar data filter yang dihasilkan akan berbeda. Hal ini termasuk merupakan kelemahan yang ada dalam penelitian ini dan diharapkan pada penelitian selanjutnya dapat diperbaiki dengan adanya landasan ilmiah yang pasti.

Selain itu permasalahan lain adalah belum adanya proses koreksi untuk menangani penulisan kata yang salah. Permasalahan yang menjadi kekurangan pada penelitian ini diharapkan dapat menjadi masukan bagi penelitian mendatang sehingga perbaikan dari proses analisis object-based opinion mining akan menghasilkan performa yang lebih baik. Tentu saja meningkatnya performa object-based opinion mining akan berdampak meningkat pula hasil klasifikasi sentimen kalimat review untuk mendapatkan orientasi review secara global.

5. DAFTAR RUJUKAN

- [1] Marrese-Taylor, E., Velásquez, J. D., Bravo-Marquez, F. (2014). A novel deterministic approach for aspect-based opinion mining in tourism products reviews. *Expert Systems with Applications*.
- [2] Marrese-Taylor, E., Velásquez, J. D., Bravo-Marquez, F., & Matsuo, Y. (2013). *Identifying customer preferences about tourism products using an aspect-based opinion mining approach*. In *Proceedings of the knowledge-based and intelligent information and engineering systems* (pp. 182–191). KES.
- [3] Liu, B. (2007). *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer Verlag. 107