

## PELABELAN KLASTER FITUR SECARA OTOMATIS PADA PERBANDINGAN REVIEW PRODUK

Fahrur Rozi<sup>1</sup>, Satrio Hadi Wijoyo<sup>2</sup>, Septiyan Andika Isanta<sup>3</sup>, Yufis Azhar<sup>4</sup>, Diana Purwitasari<sup>5</sup>

<sup>1,2,3,5</sup>Program Studi Teknik Informatika, Institut Teknologi Sepuluh Nopember Surabaya

<sup>4</sup>Jurusan Teknik Informatika, Universitas Muhammadiyah Malang

E-mail: <sup>1</sup>fahrur.rozi13@mhs.if.its.ac.id, <sup>2</sup>satrio.hadi.wijoyo13@mhs.if.its.ac.id,

<sup>3</sup>septiyan.andika13@mhs.if.its.ac.id, <sup>4</sup>yufi@umm.ac.id, <sup>5</sup>diana@if.its.ac.id

(Naskah masuk: 11 Juni 2014, diterima untuk diterbitkan: 22 Juli 2014)

### Abstrak

Penggunaan *review* produk sebagai suatu sumber untuk mendapatkan informasi dapat dimanfaatkan untuk mengoptimalkan pemasaran suatu produk. Situs belanja *online* merupakan salah satu sumber yang dapat digunakan untuk pengambilan *review* produk. Analisa terhadap produk dapat dilakukan dengan membandingkan antara dua buah produk berbeda berdasarkan fitur produk tersebut. Fitur dari suatu produk didapatkan melalui ekstraksi fitur dengan metode *double propagation*. Fitur yang terdapat dalam sebuah *review* sangat banyak serta terdapat beberapa kata yang memiliki arti yang sama yang mewakili suatu fitur tertentu, sehingga diperlukan suatu pengelompokan terhadap fitur tersebut. Pengelompokan suatu fitur produk dapat dilakukan secara otomatis tanpa memperhatikan kamus kata, yaitu dengan menggunakan teknik *clustering*. *Hierarchical clustering* merupakan salah satu metode yang dapat digunakan untuk pengelompokan terhadap fitur produk. Pengujian dengan metode *hierarchical clustering* untuk pengelompokan fitur menunjukkan bahwa *metode average linkage* memiliki nilai *recall* dan *f-measure* yang paling tinggi. Sementara untuk pengujian pelabelan menunjukkan bahwa *semantic similarity* antar fitur lebih berpengaruh dari pada kemunculan fitur di dokumen.

**Kata kunci:** *clustering, fitur produk, pelabelan*

### Abstract

*Product review can be used as a source for acquire information and to optimize the marketing of product. Online shopping sites are one of source that can be used to get product reviews. Analysis of the product can be done by comparing two different products based on product's features. Features of a product can be obtained through extraction of features with double propagation method. In the product review there are many feature that can be found, and there are some words that have the same meaning which represents a particular feature, so we need a grouping on the feature. Hierarchical clustering is one method that can be used for grouping the features of the product. Based on testing, hierarchical clustering method for grouping feature indicate that the average linkage method has the highest recall and f-measure. As for testing in labeling indicates that the semantic similarity between features is more influential than the appearance of features in the document.*

**Keywords:** *clustering, features of the product, labeling*

## 1. PENDAHULUAN

Pesatnya perkembangan *World Wide Web* dan teknologi informasi telah mempengaruhi perkembangan dalam bidang usaha seperti usaha rumahan, industri, dan perusahaan dengan melalui *review* produk (KUMAR, 2011). *Review* produk dapat dimanfaatkan untuk mendapatkan informasi dan mengoptimalkan pemasaran produk oleh produsen maupun konsumen. Situs belanja *online* menjadi salah satu sumber untuk menganalisis data *review* dari suatu produk.

Data *review* produk dapat dianalisa dengan membandingkan dua produk yang berbeda, sehingga dapat diketahui produk mana yang lebih bagus. Pengekstrakan fitur produk perlu dilakukan sebelum membandingkan dua produk. Permasalahan dalam mendapatkan fitur dari suatu produk cukup

kompleks. Hal ini dikarenakan tidak semua kata benda pada kalimat opini adalah fitur dari produk.

Salah satu cara untuk mendapatkan fitur produk dari kalimat opini adalah menggunakan metode *Double Propagation* (DP). Metode *Double Propagation* pertama kali diusulkan oleh Qiu (2011). Metode ini merupakan metode *semi unsupervised* karena masih membutuhkan kamus kata yang berisi sedikit kata sifat. Metode ini dapat melengkapi kamus kata secara otomatis saat proses ekstraksi fitur produk berlangsung.

Setiap orang selalu menggunakan kata-kata yang berbeda dalam mendeskripsikan sebuah fitur dari produk yang direview tersebut. Misalnya, "photo" dan "picture" adalah fitur yang sama atau satu kesatuan dalam kamera (ZHAI, 2011). Penggunaan sebuah kamus kata dapat memperoleh sinonim antar fitur. Akan tetapi, dalam penggunaan

sinonim terdapat beberapa permasalahan, misalnya “*appearance*” dan “*design*”. Kedua kata tersebut bukan sinonim tapi dua kata tersebut adalah satu kesatuan dalam fitur *design*. Masalah tersebut dapat diselesaikan dengan *semantic lexical similarity* berdasarkan *wordnet*.

Pengelompokan terhadap fitur produk sangatlah penting dalam analisa fitur pada *review* produk karena fitur yang sama bisa memiliki lebih dari satu kata. Pengelompokan fitur dapat dilakukan secara manual, sehingga hasil pengelompokan bersifat subjektif yang berbeda antar pakar (AZHAR, 2013). Oleh karena itu diperlukan pengelompokan secara otomatis dan tidak berdasarkan kamus kata yang besar. Salah satu cara mengelompokkan secara otomatis adalah dengan menggunakan teknik *clustering*.

*Hierarchical clustering* merupakan salah satu metode yang dapat digunakan untuk pengelompokan terhadap fitur produk. *Hierarchical clustering* diperlukan untuk menghitung nilai ukuran kesamaan atau kedekatan fitur produk. *Hierarchical clustering* sangat cocok untuk data dokumen teks yang cukup besar dan dapat mempermudah pelabelan kluster fitur (TERRATIPUK, 2006). Label dengan kata-kata yang paling sering muncul dilakukan pengelompokan terlebih dahulu. Algoritma *hierarchical clustering* dapat mendeteksi penggunaan kata yang berbeda di seluruh kategori. Algoritma tersebut lebih unggul dan efektif dalam membanding dokumen yang abstrak (POPESCUL, 2001).

Setelah melakukan proses pengelompokan fitur, permasalahan yang lain adalah bagaimana menentukan label yang dapat merepresentasikan anggota dalam kluster tersebut. Penentuan sebuah label biasanya dapat menggunakan teknik klasifikasi, sedangkan pengelompokan fitur produk menggunakan *semantic lexical similarity* dan *sharing word* untuk menghitung kedekatan antar kalimat (ZHAI, 2011). Akan tetapi pelabelan menggunakan teknik klasifikasi tersebut memerlukan sebuah dataset sebagai *training data*.

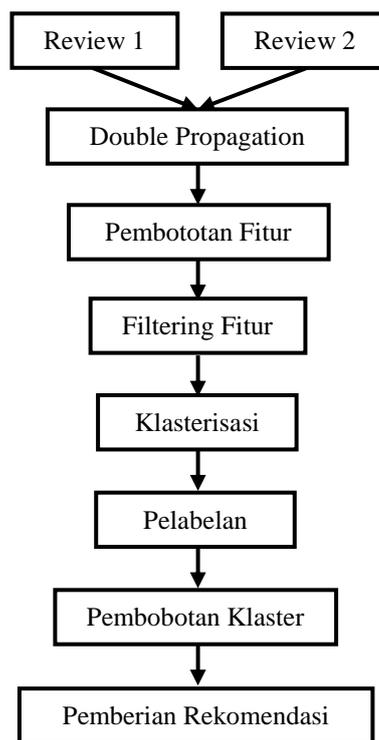
Terdapat beberapa peneliti yang melakukan pelabelan secara otomatis antara lain dengan menggunakan probabilitas kemunculan (LAU, 2010), menggunakan *idf* (*inverse document frequency*) untuk *term weighting based ranking* dan *statistical significance based ranking* untuk perhitungannya (MAO, 2012), selain itu ada juga yang menggunakan nilai *tf* (*term frequency*) dan *idf* dalam perhitungannya (TERRATIPUK, 2006). Pelabelan menggunakan *tf* dan *idf* tidak dapat digunakan dalam pelabelan fitur dikarenakan label harus dapat merepresentasikan semua fitur dalam anggota kluster. Karena tidak dapat menggunakan *tf* dan *idf* maka pelabelan menggunakan nilai *semantic similarity* antar fitur.

Berdasarkan permasalahan tersebut, dalam penelitian ini diusulkan suatu metode untuk

pelabelan kluster secara otomatis pada perbandingan fitur *review* produk. Dalam melakukan pelabelan kluster untuk fitur, diperlukan bobot dari masing-masing fitur produk. Perhitungan nilai bobot dari masing-masing fitur perlu dipertimbangkan nilai *similarity* antar fitur produk dan frekuensi kemunculan fitur di dokumen *review*. Dengan mempertimbangkan kedekatan antar fitur dan frekuensi kemunculan fitur diharapkan dapat merepresentasikan hasil semua fitur anggota kluster.

## 2. METODE YANG DIUSULKAN

Perancangan sistem yang dibangun dalam penelitian ini adalah sistem pemberi rekomendasi produk setelah dilakukan klusterisasi dan pelabelan secara otomatis pada fitur produk dengan memanfaatkan perbandingan *review* produk. Sistem ini memiliki input yang berupa semua dokumen *review* yang dimiliki oleh suatu produk dan output yang dihasilkan merupakan pemberian rekomendasi yang dimiliki oleh setiap produk. Pada sistem ini terdapat delapan tahapan proses yang akan dilakukan. Secara umum, masing-masing tahapan dapat dilihat pada Gambar 1.



Gambar 1. Alur Kerja Metode yang Diusulkan

### 2.1. Preprocessing

Pada tahap ini, *dataset* yang berupa *review* produk diperoleh dari situs Amazon ([www.amazon.com](http://www.amazon.com)). Situs ini dipilih karena merupakan salah satu situs jual beli *online* terbesar

dan memiliki *customer* yang tersebar hampir seluruh dunia. Dokumen *review* ini akan diolah menggunakan aplikasi *POS tagger* dan *dependency parser* untuk mendapatkan jenis kata dan relasi ketergantungan tiap katanya. Keduanya menggunakan *library* yang tersedia di Stanford.edu untuk mengolah data teks. *Library* ini dapat diunduh secara gratis di <http://nlp.stanford.edu/software/>. Hasil *review* dokumen ini akan digunakan dalam tahap ekstraksi fitur.

## 2.2. Ekstraksi Fitur

Tahap ekstraksi fitur digunakan untuk mengakstrasi ciri-ciri dari suatu *review* dokumen. Ekstraksi fitur produk yang terdapat pada dataset menggunakan metode *double propagation* (QUI, 2011). Metode ini tidak hanya akan mengekstrak fitur produk saja tetapi juga dapat mengekstrak kata sifat yang dapat dianggap sebagai kata opini. Jika pada iterasi pertama terdapat beberapa fitur produk yang tidak dapat diekstrak, diharapkan dengan adanya penambahan kata opini baru, fitur tersebut dapat diekstrak pada iterasi berikutnya.

## 2.3. Pembobotan Fitur

Tahap ini merupakan perhitungan pembobotan fitur dari masing-masing fitur produk dokumen sebelum dilakukan proses klusterisasi. Bobot dari masing-masing fitur didapatkan melalui persamaan (1).

$$W_{fi} = \sum_{i=1}^n p_i, \quad (1)$$

dimana  $W_{fi}$  adalah bobot dari fitur  $f$  ke- $i$  yang didapatkan melalui penjumlahan  $n$  skor polaritas kata opini  $p_i$  yang mengomentari fitur  $f$  ke- $i$ . Skor polaritas suatu kata opini  $p_i$  akan bernilai 1 jika kata tersebut adalah kata opini positif, dan bernilai -1 jika kata tersebut adalah kata opini negatif.

## 2.4. Filtering Fitur

Tahap *filtering* fitur bertujuan untuk menghilangkan fitur yang tidak digunakan dan bisa menjadi pencilan dalam sebuah kluster. Biasanya fitur yang demikian memiliki komentar yang sedikit. *Filtering* fitur dilakukan berdasarkan jumlah kemunculan fitur tersebut dalam semua *review* produk.

## 2.5. Klusterisasi Fitur

Setelah dilakukan *filtering* fitur produk terhadap semua dokumen, langkah selanjutnya adalah proses pengklusteran dokumen menggunakan *hierarchical clustering*. *Hierarchical clustering* dimulai dengan menjadikan tiap obyek menjadi sebuah kluster dan secara iterasi menggabungkan tiap kluster yang mirip. Terdapat beberapa kriteria *linkage* umum yang digunakan, seperti *single linkage clustering*, *average linkage clustering*, dan

*complete linkage clustering* (TAN, 2006). Pengukuran jarak *similarity* dalam *hierarchical clustering* ini menggunakan model dari *Wu and Palmer's* (SHENOY, 2012) seperti persamaan (2). Perhitungan  $sim(i, j)$  adalah nilai *semantic similarity* fitur ke- $i$  dengan fitur ke- $j$  dengan mempertimbangkan *information content* (IC) dari *least common subsume* (LSC) untuk dua konsep.

$$sim(i, j) = \frac{2 * IC(LCS)}{IC(concept_1) + IC(concept_2)} \quad (2)$$

dalam metode modifikasi ini nilai *similarity* antara 0 sampai 1, sedangkan dalam penelitian ini menggunakan *dissimilarity* maka *dissimilarity* dihitung  $1 - sim(i, j)$ .

## 2.6. Pelabelan Kluster

Pelabelan kluster fitur dihitung berdasarkan total *semantic similarity* antar fitur dan frekuensi kemunculan fitur dari dokumen. Semakin banyak fitur tersebut muncul pada suatu dokumen maka fitur tersebut dianggap penting (AZHAR, 2013).

Perhitungan total *semantic similarity* antar fitur didapatkan melalui persamaan (3).

$$Tsim_i = \sum_{j=0}^n sim(i, j), \quad (3)$$

dimana  $Tsim_i$  adalah total nilai *semantic similarity* fitur ke- $i$  dan  $sim(i, j)$  adalah nilai *semantic simlirity* fitur ke- $i$  dengan fitur ke- $j$ .

Perhitungan frekuensi kemunculan fitur dalam fitur dokumen *review* didapatkan melalui persamaan (4).

$$Dfr_i = \frac{D_i}{N}, \quad (4)$$

dimana  $Dfr_i$  adalah frekuensi dokumen *review* kemunculan fitur ke- $i$ ,  $D_i$  adalah total dokumen yang memiliki fitur ke- $i$ , dan  $N$  adalah total semua dokumen *review*.

Dengan kedua rumus dapat diperoleh perhitungan bobot dari masing-masing fitur dengan persamaan (5).

$$W_i = \alpha * Tsim_i + (1 - \alpha) * Dfr_i, \quad (5)$$

dimana  $W_i$  adalah bobot dari fitur ke- $i$ ,  $Tsim_i$  adalah total nilai *semantic simlirity* fitur ke- $i$ ,  $Dfr_i$  adalah frekuensi dokumen *review* kemunculan fitur ke- $i$ , dan  $\alpha$  adalah konstanta pertimbangan nilai kemunculan fitur di dokumen. Setelah didapatkan bobot masing-masing fitur anggota kluster, lalu dipilih bobot tertinggi untuk dijadikan label.

## 2.7. Pembobotan Kluster

Tahap ini merupakan perhitungan pembobotan kluster, bobot tersebut di dapat dari masing-masing bobot fitur produk dokumen yang ada di kluster tersebut. Bobot dari masing-masing kluster ini didapatkan melalui persamaan (6),

$$W_{ci} = \sum_{i=1}^n W_{fi} \tag{6}$$

dimana  $W_{ci}$  adalah bobot dari kluster ke- $i$  yang didapatkan melalui penjumlahan  $n$  skor fitur anggota kluster tersebut.

**2.8. Pemberian Rekomendasi**

Tahap pemberian rekomendasi tiap produk dengan membandingkan bobot kluster dari kedua produk. Bobot kluster yang didapatkan oleh setiap produk harus dibandingkan untuk mengetahui fitur dari produk mana yang lebih unggul. Sehingga tiap produk akan memiliki  $n$  buah fitur unggul. Setelah itu dihitung skor masing-masing produk tersebut dengan menggunakan persamaan (7).

$$S_p = \sum_{i=1}^n W_{ci} \tag{7}$$

dimana  $S_p$  adalah skor dari produk  $p$ ,  $W_{ci}$  adalah bobot dari kluster ke- $i$ , dan  $n$  adalah jumlah kluster yang dimiliki oleh produk  $p$ .

**3. HASIL UJI COBA DAN PEMBAHASAN**

Uji coba dilakukan dengan menggunakan dataset dari situs amazon.com. Sebanyak 6 buah produk dari 3 kategori yaitu *smartphone*, *tablet*, dan *camera*. Masing-masing produk memiliki rata-rata 100 komentar. Komentar untuk tiap produk tersebut nantinya disebut sebagai dokumen review yang berisi teks opini.

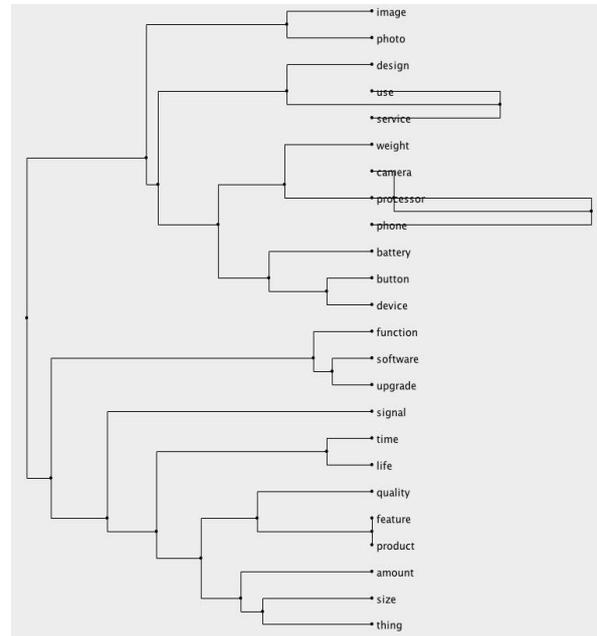
Uji coba pertama dilakukan menggunakan *precision*, *recall*, dan *f-measure* untuk melihat seberapa baik performa kriteria *lingkage* dari *hierarchical clustering* dalam mengelompokan fitur produk. Nilai *precision* yang dimaksud disini adalah proporsi dari suatu set yang diperoleh yang relevan, nilai *recall* adalah proporsi dari semua dokumen yang relevan di koleksi termasuk dokumen yang diperoleh, sedangkan nilai *f-measure* adalah *harmonic mean* dari *precision* dan *recall* (TAN, 2006).

Tabel 1. Uji *Precision*, *Recall*, dan *f-measure* dalam *Clustering* Fitur Produk

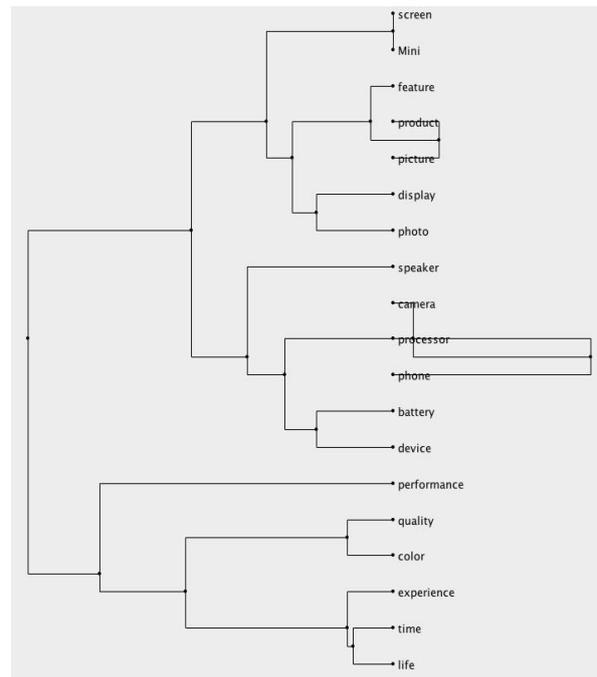
	Precision	Recall	f-measure
<b>Single linkage</b>	81%	37%	53%
<b>Average linkage</b>	63%	77%	70%
<b>Complete linkage</b>	57%	63%	60%

Tabel 1 menunjukkan bahwametode yang paling bagus untuk *clustering* fitur produk adalah *Avarage linkage clustering* karena nilai *recall* = 77% dan *f-measure* = 70% dari metode *Avarage linkage clustering* tertinggi dari pada metode lainnya. Jadi algoritma tersebut digunakan dalam uji coba selanjutnya. Sedangkan untuk hasil dendrogram dari kluster fitur produk menggunakan metode

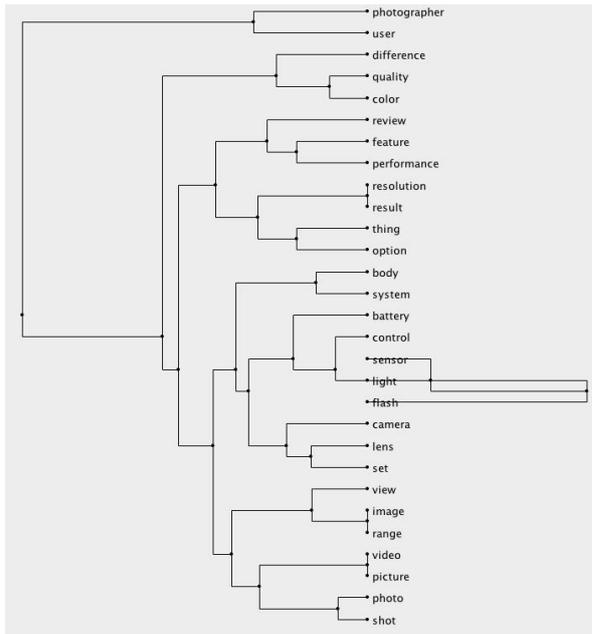
*average linkage clustering* dapat dilihat pada Gambar 2, 3, dan 4.



Gambar 2. Dendrogram dari Klasterisasi Fitur Produk untuk Domain *Smartphone*



Gambar 3. Dendrogram dari Klasterisasi Fitur Produk untuk Domain *Tablet*



Gambar 4. Dendrogram dari Klasterisasi Fitur Produk untuk Domain *Camera*

Uji coba kedua dilakukan untuk melihat seberapa baik hasil dari pelabelan kelompok fitur produk dan menentukan nilai  $\alpha$ . Untuk evaluasi kebenaran pelabelan fitur yang dihasilkan, penelitian ini menggunakan definisi *exact match* pelabelan yang benarseperti Gambar 5 (TERRATIPUK, 2006).

Definisi dari label yang benar, dihitung kinerjanya menggunakan *Match at top N* (*Match@N*) menunjukkan hasil *top N* mengandung label yang benar ini adalah indikator biner dan

monoton meningkat sebagai peningkatan  $N$  (TERRATIPUK, 2006).

**Algoritma *Exact Match* :**  
 Label *C* adalah label benar  
 Label *P* adalah label induk  
 Label *L* adalah *exact match* dari label *C*  
 Jika ada sinonim *SL* dari *L* sehingga *SL* adalah salah satu dari “*C*”, “*CP*”, dan “*PC*”

Gambar 5. Algoritma *Exact Match*

Tabel 2. Uji *Match@N* Pelabelan Fitur Produk untuk Menentukan Nilai  $\alpha$

	<b>Match@N</b>
$\alpha=0$	0.43
$\alpha=0.25$	0.57
$\alpha=0.5$	0.57
$\alpha=0.75$	0.71
$\alpha=1$	0.71

Tabel 2 menunjukkan bahwa nilai  $\alpha$  yang terbaik dari beberapa uji coba yang dilakukan adalah 0.75, yang berarti nilai *semantic similarity* lebih besar pengaruhnya dari pada nilai kemunculan di dokumen. Hasil *Match@N* dari nilai  $\alpha = 0.75$  dan  $\alpha = 1$  adalah sama. Nilai  $\alpha = 0.75$  dipilih karena dalam penelitian ini masih mempertimbangkan nilai kemunculan fitur di dokumen tersebut. Jika memilih nilai  $\alpha = 1$  maka kemunculan fitur di dokumen tidak dipertimbangkan.

Tabel 3. Perbandingan Skor Pemberian Rekomendasi dengan  $\alpha = 0.75$  untuk Domain *Tablet*

Label	Anggota Kluster	Skor Pemberian Rekomendasi	
		Ipad Mini	Samsung Tab 3
Performance	Performance	35.0	2.0
Speaker	Speaker	33.5	3.0
Phone	Camera	60.5	-2.0
	Processor		
Screen	Phone	142.0	206.0
	Screen		
Product	Mini	98.5	65.5
	Feature		
Color	Product	35.5	50.5
	Picture		
	Quality		
Time	Color	101.0	32.5
	Experience		
	Time		
Device	Life	161.5	80.0
	Battery		
Display	Device	60.0	13.5
	Display		
	Photo		

Tabel 4. Perbandingan Skor Pemberian Rekomendasi dengan  $\alpha = 0.75$  untuk Domain *Smartphone*

Label	Anggota Klaster	Skor Pemberian Rekomendasi	
		iPhone 5	Samsung Galaxy
Signal	Signal	6.5	-8.5
Time	Time	12.0	28.5
Software	Life		
	Function	6.0	43.5
Design	Upgrade		
	Design		
Photo	Use	36.0	26.5
	Service		
Phone	Image	9.0	21.5
	Photo		
Device	Weight		
	Camera	113.5	34.0
Feature	Processor		
	Phone		
Thing	Battery		
	Button	-7.5	52.5
Thing	Device		
	Quality		
Thing	Feature	60.5	73.0
	Product		
Thing	Amount		
	Size	23.5	1.0
	Thing		

Tabel 5. Perbandingan Skor Pemberian Rekomendasi dengan  $\alpha = 0.75$  untuk Domain *Camera*

Label	Anggota Klaster	Skor Pemberian Rekomendasi	
		Canon EOS	Nikon D3200
System	Body	36.0	30.5
Range	System		
	View		
Flash	Image	51.0	94.0
	Range		
Camera	Battery		
	Control		
Quality	Sensor	-4.5	54.5
	Light		
Quality	Flash		
	Camera		
Feature	Lens	86.0	501.0
	Set		
Picture	Difference		
	Quality	-42.0	98.0
Resolution	Color		
	Review		
Photographer	Feature	-49.0	57.0
	Performance		
Photographer	Video		
	Picture		
Photographer	Photo	42.5	128.5
	Shot		
Photographer	Resolution, Result		
	Thing, Option	37.5	123.0
Photographer	Photographer		
	User	27.5	39.0

Tabel 3 menunjukkan bahwa skor pemberian rekomendasi label *phone* anggota kluster *camera*, *processor*, dan *phone* dari *Samsung Tab 3* adalah negatif sebesar -2.0 yang mengartikan bahwa *Samsung Tab 3* memiliki opini yang negatif terhadap fitur *phone* dan memiliki skor label *screen* anggota kluster *screen* dan *mini* tertinggi sebesar 206.0. Sedangkan hampir semua skor pemberian rekomendasi label fitur dari *Ipad Mini* rata-rata bernilai positif dan lebih unggul daripada yang dimiliki *Samsung Tab 3*.

Tabel 4 menunjukkan bahwa skor pemberian rekomendasi label *signal* dengan anggota kluster *signal* dari *Samsung Galaxy 4* adalah negatif sebesar -8.5 yang mengartikan bahwa *Samsung Galaxy 4* memiliki opini yang negatif terhadap fitur *signal*. Sedangkan skor label yang lainnya dari *Samsung Galaxy 4* bernilai positif dan lebih unggul skornya di label *time*, label *software*, label *photo*, label *device*, dan label *feature* daripada yang dimiliki *Iphone 5*. Sebaliknya *Iphone 5* juga lebih unggul di dalam label *signal*, label *design*, label *phone*, dan label *thing*. Skornya *Iphone 5* tertinggi di dalam label *phone* sebesar 113.5. Tetapi skornya *Iphone 5* di label *device* anggota kluster *battery*, *button*, dan *device* bernilai -7.5.

Tabel 5 menunjukkan bahwa skor pemberian rekomendasi label *feature* anggota kluster *review*, *feature*, dan *performance* dari *Canon EOS* sebesar -49.0 yang menyatakan bahwa *Canon EOS* memiliki opini buruk terhadap label fitur *feature*. Sedangkan skor tertinggi untuk label *camera* anggota kluster *camera*, *lens*, dan *set* adalah *Nikon D3200* sebesar 501.0 dan hampir semua skor label fiturnya lebih unggul daripada yang dimiliki *Canon EOS*.

Penelitian ini memiliki kelemahan untuk nilai *semantic similarity* belum optimal dari fitur produk karena ada beberapa hasil yang tidak sesuai untuk klusterisasi fitur produk. Misalnya “*camera*” dan “*phone*” jaraknya lebih dekat dari pada *camera* dan *photo* atau *picture* padahal seharusnya *camera* ,*photo* dan *picture* adalah satu fitur yang sama, hal itu dapat menjadikan hasil klusterisasi tidak sesuai

#### 4. KESIMPULAN

Hasil percobaan dari penelitian ini menunjukkan bahwa metode yang diusulkan dapat melakukan pelabelan kluster fitur secara otomatis. Metode yang bagus untuk pengelompokan menggunakan *hierarchical clustering* adalah *avarege linkage*, dan nilai  $\alpha$  untuk metode pelabelan yang diajukan adalah 0.75. Pengelompokan dan pelabelan secara otomatis dapat mempermudah dalam melakukan analisa fitur produk *smartphone*, *tablet*, dan *camera* seperti pada Tabel 3, 4, dan 5. Penelitian selanjutnya adalah memperbaiki nilai dari *semantic simaliry* untuk fitur produk.

#### 5. DAFTAR PUSTAKA

- AZHAR, Y., AGUS Z. A., & DIANA P. 2013. Otomatisasi Perbandingan Produk Berdasarkan Bobot Fitur Pada Teks Opini.
- KUMAR, A. K. M. & SURESHA. 2011. Analyzing Web user' Opinion from Phrases and Emoticons. *IJCA Special Issue on "Computational Science – New Dimensions & Perspectives"*.
- LAU, J. H., DAVID N., SARVNAZ K., & TIMOTHY B. 2010. Best Topic Word Selection for Topic Labeling. *Coling*. Hal. 605-613.
- MAO, X. L., ZHAO, Y. M., ZHENG, J. Z., TAT S. C., HONGFEI, Y., & XIAOMING, L. 2012. Automatic Labeling Hierarchiccal Topics.
- POPESCU, A. & LYLE H. U. 2001. Automatic Labeling of Document Clusters
- QIU, G., LIU, B., BU, J., & CHEN, C. 2011. Opinion Word Expansion and Target Extraction Through Double Propagation. *Computational Linguistics*, 37(1). Hal.9-27.
- SHENOY, M. K., K. C. SHET, & DINESH A. 2012. A New Similarity Measure For Taxonomy Based on Edge Counting. *International Journal of Web & Semantic Technology (IJWesT)*. Vol. 3, No. 4
- TAN, P. N., MICHEAL S., & VIPIN K. 2006. Introduction to Data Mining. Pearson Education : India.
- TERRATIPUK, P. & JAMIE C. 2006. Automatically Labeling Hierarchical Cluster.
- ZHAI, Z., BING L., HUA X., & PEIFA J. 2011. Clustering Product Features for Opinion Mining.