

ANALISIS SENTIMEN REVIEW PRODUK MENGUNAKAN PENDEKATAN BERBASIS KAMUS

Fairuz Ahmad Hirzani¹, Warih Maharani², Moch. Arif Bijaksana³

^{1,2,3}Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom
¹ahmadhirzani@gmail.com, ²wmaharani@gmail.com, ³arifbijaksana@gmail.com

Abstrak

Review produk pada situs *e-commerce* sering kali menjadi tujuan calon pembeli produk dalam menilai kualitas dan tingkat kepuasan konsumen lain terhadap suatu produk. Jumlah review yang sangat banyak tentunya akan menyulitkan pembaca dalam penarikan informasi. Peringkasan dan klasifikasi opini terhadap fitur suatu produk merupakan hal yang dibutuhkan guna mempercepat pembacaan dan penarikan kesimpulan terhadap kualitas suatu produk. Penelitian di bidang ini, khususnya *feature based opinion summarization* telah banyak dikembangkan dengan menggunakan berbagai macam teknik dan pendekatan. Analisis sentimen berbasis fitur membutuhkan proses ekstraksi fitur terlebih dahulu sebelum melakukan identifikasi opini. Proses ini dilakukan dengan melakukan identifikasi terhadap kata benda yang sering dibicarakan menggunakan penerapan *association mining*. *Association mining* yang diterapkan untuk ekstraksi fitur terbukti dapat menghasilkan fitur hasil ekstraksi dengan sebelumnya dilakukan proses pemilihan kata. Fitur yang didapat selanjutnya ditinjau untuk setiap kalimat menggunakan bantuan kamus opini untuk mengidentifikasi orientasi opini fitur yang dibicarakan oleh konsumen. Pendekatan seperti ini dalam analisis sentimen berbasis fitur biasa disebut pendekatan berbasis kamus.

Kata kunci : analisis sentimen, review produk, *feature based opinion summarization*

Abstract

Product review on e-commerce usually becomes a guideline for customer to decide the product's quality and to get the information about another customer satisfaction for the product. However, a huge number of review makes it hard to read. Opinion classification and summarization offer a solution for customer to read the review faster and easier. The research on this field, particularly on feature based opinion summarization has a huge development along with vary of techniques and approach. Feature based sentiment analysis requires a feature extraction process to extract product's feature before identify feature's orientation on the sentence. This process can be done by identifying a noun that talked a lot by customer in the review using association mining. Association mining that applied in the system has proved to gain product's feature as long as term filtering is done before. Then, opinion orientation for feature in each sentence is obtained by utilize opinion dictionary. This kind of approach in sentiment analysis is called dictionary based approach.

Keywords : sentiment analysis, product review, *feature based opinion summarization*

1. Pendahuluan

Review konsumen terhadap suatu produk termasuk faktor yang berpengaruh terhadap perilaku pembelian konsumen. Fakta ini dibuktikan dalam penelitian *Dimensional Research* yang menyatakan bahwa 88% konsumen terpengaruh *online review service* dalam pengambilan keputusan [1]. Jumlah review yang banyak tentunya akan menyulitkan pemahaman dalam membaca review. Ringkasan opini terhadap review produk diharapkan dapat mempercepat pembacaan dan meningkatkan ketepatan penarikan kesimpulan dari review. Solusi yang cocok dengan permasalahan ini adalah pembangunan sistem yang dapat mengidentifikasi opini yang diutarakan untuk fitur. Fitur dalam review produk umumnya merupakan kata benda yang sering dibicarakan dalam review, sehingga proses ekstraksi fitur produk dapat dilakukan menggunakan *frequent pattern generation* pada *association mining* [2]. Keberadaan daftar kata ataupun kamus kata positif dan negatif telah cukup lengkap seiring perkembangan penelitian analisis sentimen. Oleh karena itu, penelitian tugas akhir ini memanfaatkan lexicon dan kamus yang disebut *Dictionary-based Techniques* [3] untuk melakukan klasifikasi opini terhadap fitur produk.

Analisis yang dilakukan pada penelitian tugas akhir ini ditujukan pada empat tahapan utama dalam pembangunan sistem, yakni seleksi kata, ekstraksi fitur, identifikasi kata dan kalimat opini, dan identifikasi opini fitur dengan menggunakan lima dataset produk berbahasa Inggris. Analisis yang dilakukan pada tahapan ekstraksi fitur yaitu penerapan nilai *minimum support* yang berbeda untuk melihat pengaruhnya terhadap fitur yang terekstrak. Fitur yang diekstrak pada proses ini merupakan fitur yang tertera secara eksplisit pada review. Proses lainnya seperti *prunning* dan *infrequent feature extraction* digunakan untuk meningkatkan ketepatan ekstraksi. Beberapa skema yang berbeda diuji dalam proses seleksi kata dan identifikasi kata opini guna melihat pengaruh dan keunggulan dari tiap skema. Analisis identifikasi orientasi opini fitur dilakukan dalam penelitian tugas akhir ini dengan meninjau jarak fitur opini yang berbeda serta penanganan kata negasi yang dapat mempengaruhi orientasi.

2. Landasan Teori

2.1 Analisis Sentimen

Analisis Sentimen, atau dapat juga disebut *Opinion Mining* adalah ranah ilmu yang menganalisis opini seseorang, sentimen, perilaku, penilaian, dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, ataupun terhadap kejadian. Terdapat beberapa nama lain yang juga memiliki *task* yang sedikit berbeda seperti *opinion mining*, *opinion extraction*, *sentiment mining*, *subjectivity analysis*, *emotion analysis*, *review mining*. Secara umum, sentiment analysis dibagi menjadi tiga bagian utama, yaitu pada level dokumen, level kalimat, dan level aspek dan entitas [4]. Level dokumen dan level entitas analisis tidak menemukan secara baik mana yang seseorang suka dan tidak. Sentiment Analisis pada level aspek menunjukkan performa yang lebih baik. Level aspek ini pada awalnya disebut *feature level (feature-based opinion mining and summarization)* [5]. Dalam kalimat “*although the service is not that great, I still love this restaurant*” dapat ditarik orientasinya bahwa kalimat tersebut mengungkapkan opini positif. Padahal, jika dilihat kembali kalimat tersebut memiliki orientasi positif untuk *restaurant*, dan negatif untuk *services*. Dari contoh tersebut, terlihat bahwa tujuan pada level aspek ini adalah untuk menentukan sentiment pada entitas dan pada aspek aspeknya yang berbeda..

2.2 Lemmatization

Penggunaan kata dalam bahasa Inggris dapat memiliki bentuk berbeda sesuai dengan penggunaan dan aturan bahasa. Satu inti kata yang sama dapat memiliki berbagai bentuk seperti *organize*, *organizes*, dan *organizing*, atau *camera* dan *cameras*. Bentuk kata yang berbeda seperti ini dapat menyulitkan untuk digunakan dalam pengolahan komputer. Stemming dan Lemmatization merupakan cara yang digunakan untuk mengubah kata menjadi bentuk dasarnya. Jika stemming melakukan proses ini dengan memotong bagian akhir kata, lemmatization melakukannya dengan cara yang lebih layak dengan tidak menghilangkan makna kata. Gambar berikut mengilustrasikan hasil *lemmatization* pada kalimat.

Sebelum dilakukan Lemmatization:

at first i thought it is only a isolated incident , but i was proven wrong when the player would not read my alias dvds as well

Setelah dilakukan Lemmatization:

at first i think it be only a isolate incident , but i be prove wrong when the player would not read my alias dvd as well

Gambar 1 Contoh Lemmatization

2.3 Natural Language Parser

Natural language parser merupakan program yang dapat melabelkan struktur gramatikal dari sebuah kalimat, seperti identifikasi frase, atau mengidentifikasi kata mana dalam kalimat yang merupakan sebuah subjek atau objek. Stanford Statistical Parser merupakan salah satu library yang dibuat untuk melakukan tugas ini salah satunya dalam mengekstrak *noun phrase* dan *dependencies* dalam kalimat. Identifikasi frase direpresentasikan dalam bentuk tree seperti yang tertera pada Gambar yang merupakan hasil parser dari kalimat “*Bell, based in Los Angeles, makes and distributes electronic, computer and building products*”.

```
(ROOT
(S
(NP
(NP (NNP Bell))
(, ,)
(VP (VBZ based)
(PP (IN in)
(NP (NNP Los) (NNP Angeles))))
(, ,))
(VP (VBZ makes)
(CC and)
(VBZ distributes)
(NP
(UCP (JJ electronic) (, ,) (NN computer)
(CC and)
(NN building))
(NNS products))))))

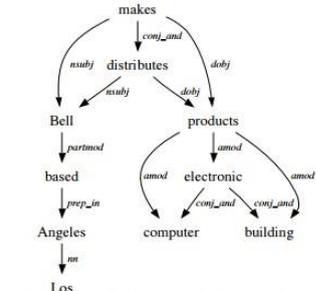
```

Gambar 4 Contoh Hasil Parser

```
nsubj(makes-8, Bell-1)
acl(Bell-1, based-3)
case(Angeles-6, in-4)
compound(Angeles-6, Los-5)
nmod(based-3, Angeles-6)
root(ROOT-0, makes-8)
cc(makes-8, and-9)
conj(makes-8, distributes-10)
amod(products-16, electronic-11)
conj(electronic-11, computer-13)
cc(electronic-11, and-14)
conj(electronic-11, building-15)
dobj(makes-8, products-16)

```

Gambar 2 Contoh Hasil Dependency Parser



Gambar 3 Ilustrasi Dependency

2.4 Association Mining

Association mining dilakukan untuk menemukan hubungan yang menarik dari sekumpulan data yang banyak. Data masukan yang dibutuhkan untuk association mining adalah sekumpulan transaksi yang terdiri berupa itemset. Kasus yang umum dilakukan menggunakan association mining salah satunya adalah menemukan hubungan antar barang yang dibeli oleh pelanggan melalui serangkaian transaksi pembelian. Proses ini dilakukan dengan melalui dua tahapan utama [6] yaitu:

1. Frequent Itemset Generation
Tahapan ini bertugas untuk menemukan itemset yang memenuhi *minimum support threshold*. Itemset yang lolos threshold ini disebut frequent itemset.
2. Rule Generation
Tahapan ini bertugas untuk mengekstrak rule yang memiliki nilai *confidence* diatas nilai tertentu. Rule yang dihasilkan merupakan keluaran terahir dari association mining yang nantinya dianggap sebagai hubungan antar item.

2.5 Evaluasi

Evaluasi performansi yang dilakukan dalam penelitian Tugas Akhir ini adalah precision, recall dan fscore. Fscore sendiri merupakan harmonic mean dari precision dan recall. Evaluasi dilakukan untuk mengetahui tingkat keberhasilan ekstraksi fitur dan klasifikasi opini. Pada penelitian Tugas Akhir ini, evaluasi dilakukan pada level dokumen dan kalimat. Perhitungan hasil ekstraksi fitur dilakukan menggunakan evaluasi berbasis dokumen, sedangkan evaluasi orientasi fitur opini dilakukan menggunakan evaluasi berbasis kalimat. Berikut penjelasan untuk setiap pendekatan evaluasi,

1. Evaluasi Berbasis Dokumen
Hasil ekstraksi atau prediksi dikelompokkan untuk satu dokumen sebelum dilakukan perhitungan evaluasi. Dalam kasus ekstraksi fitur, fitur yang terekstrak dikumpulkan dan dibandingkan dengan daftar fitur yang seharusnya terekstrak.

Tabel 1 Contoh Perhitungan Evaluasi Berbasis Dokumen

Fitur Terekstrak		Fitur Expert Judgement
1. Battery	6. Picture	1. Battery
2. Camera	7. Memory Card	2. Battery Life
3. Battery Life	8. Thing	3. Screen
4. Life	9. Card	4. Picture
5. Screen	10. Amazon	5. Photo
		6. Memory Card

Fitur yang terdaftar dalam tabel diatas merupakan daftar fitur dalam seluruh kalimat pada dataset review. Dalam perbandingan antara fitur terekstrak dengan fitur *expert judgement*, ditemukan 5 fitur benar dari total 10 fitur yang terekstrak. Jumlah ekstraksi fitur melebihi jumlah fitur seharusnya yang berjumlah 6. Sehingga dapat ditentukan nilai precision sebesar 0.5 dan recall 0.83.

2. Evaluasi Berbasis Kalimat
Evaluasi dilakukan pada setiap kalimat pada dataset yang selanjutnya dilakukan perhitungan rata-rata. Nilai rata-rata tersebut yang akan dijadikan nilai evaluasi keseluruhan. Berikut contoh perhitungan evaluasi berbasis kalimat,

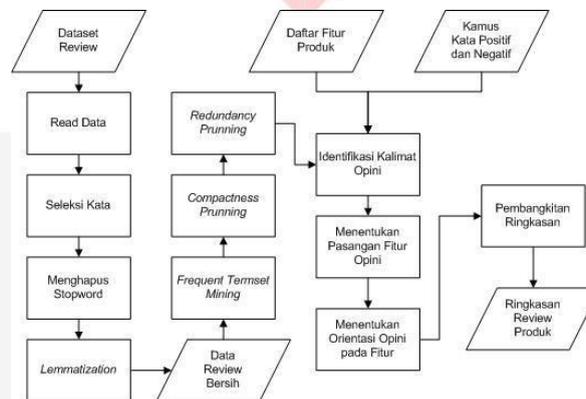
Tabel 2 Contoh Perhitungan Evaluasi Berbasis Kalimat

No	Fitur Terekstrak	Fitur Expert Judgement	Kalimat	Prec	Rec
1	dvd player	dvd player	for the price , it's a very nice dvd player	1.0	1.0
2	-	-	what more could you ask for	1.0	1.0
3	dvd disc	dvd disc, read	the one i purchased for a frend as a gift will not read any dvd disc	1.0	0.5
4	-	apex	my dad also has an apex tv and it was another great value for the money	0.0	0.0
5	look, feature, load	look, feature	it looks great - and is loaded with features	0.7	1.0
				0.74	0.7

3. Pembahasan

3.1 Gambaran Umum Sistem

Sistem yang dibangun dalam tugas akhir ini adalah sistem yang dapat menentukan orientasi opini terhadap suatu fitur berdasarkan data komentar yang diberikan sehingga pada akhirnya dapat menghasilkan ringkasan untuk mempermudah pembacaan review produk. Tahapan penentuan opini ini dibangun oleh empat tahapan utama yaitu, (1) Ekstraksi fitur berdasarkan data review, (2) Identifikasi orientasi kata dan kalimat opini, (3) Penentuan orientasi opini terhadap fitur, dan (4) Pembangkitan ringkasan. Gambaran umum sistem secara lengkap dapat dilihat pada diagram dibawah,



Gambar 5 Gambaran Umum Sistem

3.2 Seleksi Kata

Penyeleksian kata atau *term filtering* dilakukan untuk menyeleksi kata yang dibutuhkan pada ekstraksi fitur. Kemungkinan besar kata fitur merupakan kata benda atau frase kata benda, sehingga kata yang bukan merupakan kata benda atau frase kata benda tidak akan dilibatkan pada proses selanjutnya. Berikut merupakan beberapa cara dalam menyeleksi kata yang berkaitan dengan fitur produk:

1. Seleksi Kata Benda (NN)

Kalimat review dilakukan POS Tagging terlebih dahulu untuk mengetahui Part of Speech tag untuk setiap kata seperti kata benda, kata sifat, dll. Selanjutnya, kata yang bukan merupakan kata benda dan stopword dihapus dari kalimat. Kalimat yang telah dilakukan proses POS Tagging seperti "this_DT is_VBZ my_PRP\$ first_JJ digital_JJ camera_NN and_CC what_WP a_DT toy_NN it_PRP is_VBZ" akan menghasilkan kalimat "camera_NN toy_NN".

2. Seleksi Frase Kata Benda (NP)

Frase kata benda tidak hanya dibentuk oleh kata benda saja, melainkan dapat dibentuk oleh gabungan kata sifat dan kata benda, atau gabungan jenis kata lain. Proses ekstraksi frase kata benda dilakukan pada lingkungan pemrograman Java dengan bantuan library Stanford Parser. Stanford parser dapat memotong kalimat menjadi bagian bagian frasa dalam bentuk tree. Sebagai contoh kalimat "i recently purchased the canon powershot g3 and am extremely satisfied with the product" menghasilkan frase (DT the) (JJ canon) (NN powershot) (NN g3) dan (DT the) (NN purchase).

3. Seleksi Frase Kata Benda dan Subjek Objek (NPSO)

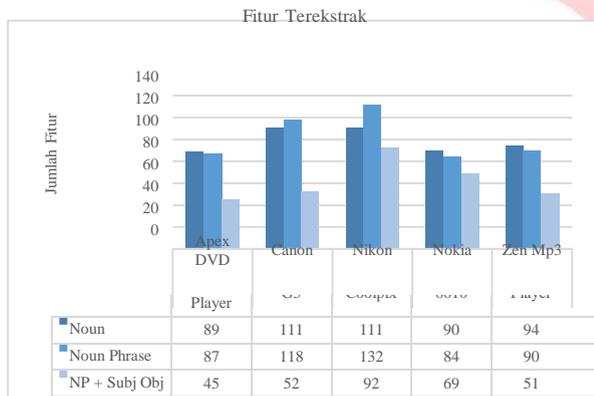
Ide dasarnya adalah kata yang muncul dalam bentuk subjek atau objek kemungkinan besar adalah sebuah fitur dalam suatu kalimat review. Frasa yang bukan merupakan subjek atau objek akan dihapus, sehingga menyisakan frasa yang sekaligus merupakan subjek atau objek dalam kalimat. Dalam Stanford Dependency Parser, subjek dilabelkan dengan *subj* sedangkan objek dilabelkan dengan *dobj*.

```
nsubj(purchased-3, i-1), nsubjpass(satisfied-11, i-1), advmod(purchased-3,
recently-2), root(ROOT-0, purchased-3), det(g3-7, the-4), amod(g3-7, canon-5),
nn(g3-7, powershot-6), dobj(purchased-3, g3-7), auxpass(satisfied-11, am-9),
advmod(satisfied-11, extremely-10), conj_and(purchased-3, satisfied-11),
det(purchase-14, the-13), prep_with(satisfied-11, purchase-14)
```

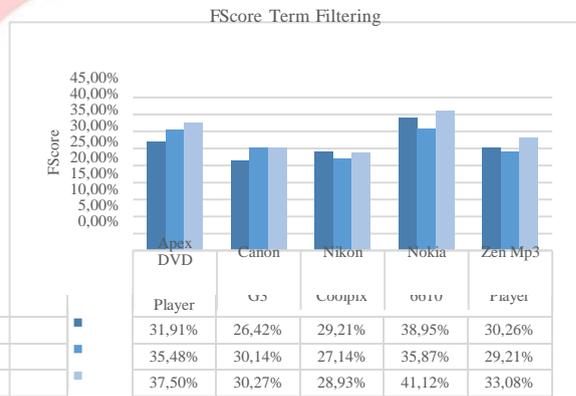
Gambar 6 Keluaran Stanford Dependency Parser

Dari hasil Dependency Parser diatas dapat diambil *nsubj(purchased-3, i-1)* dan *dobj(purchased-3, g3-7)*. Subjek pada kalimat review tersebut adalah "I" dan objek pada kalimat tersebut adalah "g3". Sebelumnya frasa kata benda dari kalimat tersebut telah didapatkan, yaitu "the canon powershot g3" dan "the purchased". Satu dari dua frasa kata benda merupakan objek yaitu "g3" sehingga diambil satu frasa yaitu frasa "the canon powershot g3".

Penerapan tiga skema seleksi kata diukur nilai akurasi dan hasil fitur yang dihasilkan. Berdasarkan pengujian untuk lima dataset dengan *minimum support* 1%, NPSO menghasilkan jumlah fitur paling kecil dibandingkan dengan skema lain, seperti yang terlihat pada Gambar 8. Hal ini membutuhkan penentuan nilai *minimum support* yang lebih besar untuk NPSO. Jika fitur yang terekstrak setara, NPSO menghasilkan nilai Fscore lebih tinggi dibandingkan dua skema lainnya. Perbandingan nilai Fscore tertera pada Gambar 7.



Gambar 8 Perbandingan Jumlah Fitur Terekstrak



Gambar 7 Perbandingan nilai Fscore Seleksi Kata

3.3 Ekstraksi Fitur Produk

Sistem ini melakukan penentuan orientasi opini terhadap fitur suatu produk, sehingga ekstraksi fitur merupakan tahapan yang harus dilakukan sebelum dapat menentukan orientasi opini. Kata benda ataupun frasa kata benda yang sering dibicarakan oleh konsumen dalam sebuah review produk menjadi kandidat fitur yang potensial [2]. Ekstraksi fitur tersebut dapat dilakukan menggunakan *association mining*, dengan tujuan menemukan *frequent itemset* yang dapat mewakili fitur sebuah produk. Kata-kata yang muncul bersamaan dengan urutan tertentu dalam bahasa manusia biasanya memberikan sebuah arti tertentu atau biasa disebut frasa. *Association mining* sendiri menghasilkan himpunan kata yang sering muncul sendiri atau bersamaan dalam sebuah kalimat, tidak peduli kedekatan antar kata tersebut. Sehingga sangat mungkin menghasilkan gabungan beberapa kata yang tidak memiliki arti. *Compactness pruning* dilakukan untuk mengeliminasi gabungan kata tersebut. Eliminasi fitur dapat dilakukan dengan menghitung jarak antar kedua kata yang muncul, apabila jaraknya terlalu jauh dan melewati *threshold* tertentu maka fitur tersebut akan dieliminasi. Gambar 10 menjelaskan algoritma dalam melakukan *compactness pruning* dan *redundancy pruning* pada Gambar 9.

Fitur-fitur yang tidak lolos dalam *association mining* karena tingkat kemunculannya dianggap kecil disebut dengan *infrequent feature*. Proses ekstraksi fitur ini dilakukan dengan melihat kata benda atau frasa kata benda yang dekat dengan kata opini dalam sebuah kalimat. Kedekatan kata benda dengan kata opini tersebut dapat diukur dalam parameter *minimum distance*, jika kata benda memenuhi batas jarak yang ditentukan maka kata benda tersebut disimpan menjadi kandidat *infrequent feature*. Tahapan selanjutnya yaitu menghitung kemunculan kandidat fitur, apabila kemunculannya diatas batas yang ditentukan atau *minimum occurrence*, maka kata benda tersebut dianggap sebagai fitur.

```

For each sentences do
  For each featurePhraseInSentece do
    Words <- Tokenize(featurePhraseInSentence)
    Calculate distance between two Words
    If distance>minDistance then
      Add to listNewFeaturePhrase
    EndIf
  EndFor
End For
For each featurePhrase do
  Count feature occurence in listNewFeaturePhrase
  If occurence < minOccurence then
    Remove(featurePhrase)
  EndIf
EndFor
    
```

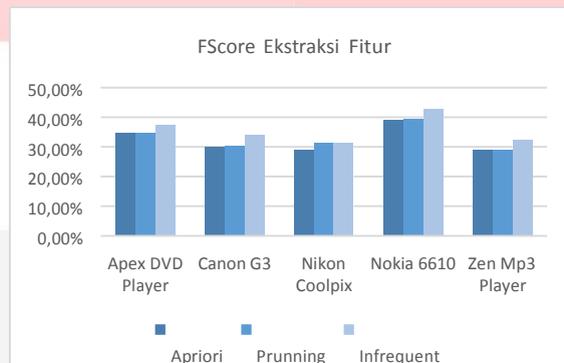
Gambar 10 Algoritma Compactness Pruning

```

for each sentences do
  for each featureSingleWord in Sentence do
    check whether its superset appear in sentence
    if not appeared then
      increment pureSupport of featureSingleWord
    endif
  endfor
endfor
for each pureSupport of featureSingleWord do
  if pureSupport of featureSingleWord < minPureSupport then
    remove(featureSingleWord)
  endif
endfor
    
```

Gambar 9 Algoritma Redundancy Pruning

Hasil pengujian nilai evaluasi tahapan ekstraksi fitur ditunjukkan pada Tabel dan Gambar yang menunjukkan peningkatan nilai Fscore pada setiap tahapannya. Peningkatan ini sangat bergantung pada nilai parameter yang diberikan. Dengan kondisi data dan parameter tertentu, *prunning*, dan ekstraksi *infrequent feature* dapat mengurangi nilai fscore.



Gambar 11 Perbandingan Fscore Ekstraksi Fitur

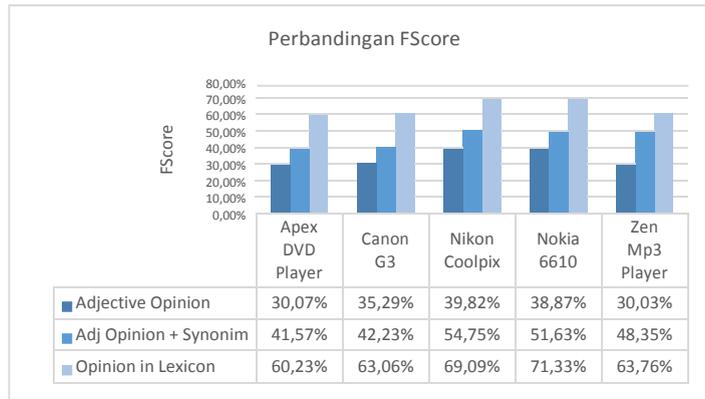
Tabel 3 Perubahan Fscore pada Setiap Tahapan Ekstraksi

Produk	Apriori			Prunning			Infrequent		
	Precision	Recall	Fscore	Precision	Recall	Fscore	Precision	Recall	Fscore
Apex DVD Player	55,56%	25,25%	34,72%	55,56%	25,25%	34,72%	46,27%	31,31%	37,35%
Canon G3	44,23%	22,77%	30,07%	46,00%	22,77%	30,46%	34,74%	33,67%	34,20%
Nikon Coolpix	25,00%	34,33%	28,93%	28,75%	34,33%	31,29%	28,75%	34,33%	31,29%
Nokia 6610	47,83%	33,00%	39,05%	48,53%	33,00%	39,29%	44,57%	41,00%	42,71%
Zen Mp3 Player	64,71%	18,64%	28,95%	64,71%	18,64%	28,95%	38,89%	27,68%	32,34%

3.4 Identifikasi Kata dan Kalimat Opini

Identifikasi kata opini merupakan tahapan dalam menentukan apakah suatu kata opini dalam kalimat merupakan opini yang mengungkapkan orientasi positif atau negatif. Tahapan ini diperlukan untuk melabelkan positif atau negatif terhadap suatu fitur dalam kalimat. Pelabelan positif dan negatif pada suatu kata dilakukan menggunakan daftar kata positif dan negatif yang tersedia secara bebas untuk digunakan dalam penelitian *sentiment analysis*. Bing Liu pada penelitian sebelumnya telah melakukan pencatatan daftar kata opini positif dan negatif yang terdapat pada review produk. Daftar tersebut berjumlah 6789 kata, yang tersusun oleh 2006 kata positif dan 4783 kata negatif. Daftar kata inilah yang digunakan pada penelitian ini guna mentukan orientasi opini sebuah kata.

Pengujian tahapan ini dilakukan menggunakan tiga skema pemanfaatan kamus yang digunakan. Tiga skema tersebut yakni identifikasi berdasarkan kata sifat yang selanjutnya dicocokkan ke kamus, identifikasi berdasarkan kata sifat yang dilengkapi sinonim dari WordNet, dan skema terakhir yaitu pengecekan secara langsung ke kamus. Hasil pengujian menunjukkan bahwa skema pencocokan langsung ke kamus memiliki nilai Fscore tertinggi. Tingginya hasil evaluasi skema ini karena kamus yang digunakan memiliki kosakata opini yang banyak dan dapat mengidentifikasi kata opini pada kalimat. Penggunaan skema kata sifat memiliki nilai Fscore yang kecil karena tidak semua kata opini merupakan kata sifat. Sedangkan pemanfaatan sinonim kata sifat banyak menghasilkan hasil identifikasi yang tidak tepat. Perbandingan nilai Fscore pada setiap skema dapat dilihat pada Gambar 12



Gambar 12 Perbandingan Fscore Kalimat Opini

3.5 Klasifikasi Orientasi Opini Fitur

Penentuan orientasi terhadap fitur dalam kalimat dilakukan dengan menemukan pasangan fitur dan opini terlebih dahulu. Setelah pasangan ditemukan, orientasi fitur akan mengikuti orientasi kata opini pasangannya yang telah ditentukan pada proses sebelumnya. Sebagai contoh pada kalimat “*picture quality* [+] ## *the picture quality are absolutely amazing*” memiliki fitur *picture quality* dengan kata *amazing* sebagai pasangannya. Pasangan fitur *player* dengan opini *loved* ditemukan pada kalimat “*player* [+] ## *i took it to my father's house to play a tom jones concert dvd and he loved the player so much i gave it to him*”. Kata *amazing* dan *loved* berdasarkan lexicon memiliki orientasi positif, hal ini menyebabkan fitur *picture quality* dan *player* memiliki orientasi opini positif. Pasangan fitur dan opini pada penelitian tugas akhir ini dilakukan mengikuti salah satu metode dalam penelitian Kam Tong Chan. Metode yang digunakan termasuk kedalam metode sederhana dalam penelitian tersebut, yaitu dengan menghitung jarak antara fitur dan opini. Nilai relasi antara fitur dan opini dilambangkan dengan dengan fungsi $rel(f, w)$, dengan f adalah fitur produk dan w adalah kata opini. Nilai $rel(f, w)$ yang melebihi batas *threshold* tertentu akan dijadikan sebagai pasangan fitur dan opini yang valid. Fungsi tersebut dihitung menggunakan formula,

$$rel(f, w) = \frac{1}{|f \cap w|} \quad (1)$$

Jumlah fitur dan opini dalam kalimat tidak selalu genap berpasangan, melainkan dalam jumlah yang berbeda. Opini yang ditujukan untuk fitur yang telah memiliki pasangan akan dipaksa mencari fitur terdekat lainnya. Seluruh rangkaian tahapan ini dilakukan menggunakan algoritma pada Gambar 13.

```

F <- features in sentence
W <- opinion words in sentence
For each w in opinion words W do
Score <- highest rel(f,w)
If score > threshold then
    If the same word is already assign to f then
        Try another f with the next highes rel score
    Else
        Associate w to f
    Endif
Endif
endfor
    
```

Gambar 13 Algoritma Pasangan Fitur Opini

Tabel 4 Evaluasi Pengaruh Penanganan Negasi

Produk	Tanpa Negation			Dengan Negation			Δ Fscore
	Precision	Recall	Fscore	Precision	Recall	Fscore	
Apex DVD Player	53,47%	53,65%	53,56%	53,67%	53,76%	53,71%	0,16%
Canon G3	59,35%	59,23%	59,29%	59,35%	59,23%	59,29%	0,00%
Nikon Coolpix	52,75%	52,55%	52,65%	53,03%	52,85%	52,94%	0,29%
Nokia 6610	54,50%	54,03%	54,26%	54,50%	54,03%	54,26%	0,00%
Zen Mp3 Player	48,02%	48,74%	48,38%	48,57%	49,32%	48,94%	0,56%

Peningkatan nilai evaluasi pada identifikasi opini fitur dapat dilihat pada Tabel 4. Nilai evaluasi yang dihasilkan pada tahapan ini tergantung pula pada nilai *threshold* jarak yang diberikan. Pengujian untuk nilai *threshold* dilakukan untuk setiap dataset guna mendapatkan hasil evaluasi terbaik. Hasil evaluasi terbaik untuk setiap dataset dapat dilihat pada Tabel 5.

Tabel 5 Evaluasi Akhir Opini Fitur

Produk	Evaluasi		
	Precision	Recall	Fscore
Apex DVD Player	60,61%	60,61%	60,61%
Canon G3	59,35%	59,23%	59,29%
Nikon Coolpix	53,18%	53,13%	53,15%
Nokia 6610	54,50%	54,03%	54,26%
Zen Mp3 Player	59,83%	59,67%	59,75%

4. Kesimpulan

Berdasarkan hasil pengujian dan analisa yang telah dilakukan sebelumnya, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Seleksi kata NPSO cocok digunakan pada dataset review yang memiliki mayoritas kalimat kompleks dengan tata bahasa yang sesuai. Sedangkan seleksi kata NN dan NP baik digunakan untuk dataset yang terdiri dari kalimat sederhana.
2. Ekstraksi *infrequent feature* dan *prunning* dapat meningkatkan nilai Fscore dengan menggunakan parameter yang sesuai. Penentuan nilai parameter yang salah dapat mengurangi nilai dari perhitungan evaluasi.
3. Ekstraksi berbasis kamus untuk identifikasi kata opini dapat menentukan orientasi dengan baik dibandingkan dengan identifikasi kata sifat. Sedangkan penggunaan sinonim untuk identifikasi dapat menimbulkan kesalahan identifikasi orientasi opini.
4. Penggunaan *negation handling* pada identifikasi kata opini terbukti dapat meningkatkan nilai evaluasi dalam identifikasi opini fitur.
5. Penentuan orientasi opini pada kalimat menghasilkan hasil yang baik apabila jarak antara opini dan fiturnya dekat.

5. Saran

Saran yang diperlukan dari tugas akhir ini untuk pengembangan sistem selanjutnya adalah sebagai berikut:

1. Mampu menangani *misspelling* dan singkatan.
2. Mampu menangani fitur *implicit* yang diutarakan dalam dataset
3. Menggunakan penerapan *coreference resolution* guna meningkatkan akurasi sistem.

6. Daftar Pustaka

- [1] Zendesk, "The Impact of Customer Service on Customer Lifetime Value," Dimensional Research, 2013. [Online]. [Diakses 20 October 2014].
- [2] S. H. Ghorashi, R. Ibrahim, S. Noekhah dan S. Dastjerdi, "A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Review," *IJCSI International Journal of Computer Science Issues*, vol. IX, no. 4, pp. 29-35, 2012.
- [3] V. S. Jagtap dan K. Pawar, "Analysis of Different Approach to Sentence Level Sentiment Classification," *International Journal of Scientific Engineering and Technology*, vol. 2, no. 3, pp. 164-170, 2013.
- [4] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012.
- [5] M. Hu dan B. Liu, "Mining and Summarizing Customer Review," 2004.
- [6] P.-N. Tan, M. Steinbach dan V. Kumar, *Introduction to Data Mining*, Boston: Addison-Wesley Longman Publishing, 2005.
- [7] The Stanford Natural Language Processing Group, "Stanford Log-linear Part-Of-Speech Tagger," [Online]. Available: <http://nlp.stanford.edu/software/tagger.shtml>. [Diakses 23 October 2014].
- [8] G. A. Miller, R. Beckwith dan C. Fellbaum, "Intoduction to WordNet: An On-line Lexical Database," 1993.