EKSTRAKSI FITUR DAN OPINI MENGGUNAKAN PENDEKATAN PATTERN KNOWLEDGE DAN OPINION LEXICON

FEATURE AND OPINION EXTRACTION USING PATTERN KNOWLEDGE AND OPINION LEXICON APPROACHES

[1]I Nyoman Cahyadi Wiratama [2] Warih Maharani, S.T., M.T. [3] Ir. Moch. Arif Bijaksana, M.Tech.
[1 2 3] Fakultas Informatika, Universitas Telkom, Bandung
[1]cahyadi.wiratama@gmail.com, [2]wmaharani@gmail.com, [3]arifbijaksana@gmail.com

Abstrak

Ekstraksi fitur dan opini merupakan suatu cara yang dapat digunakan untuk mengetahui kata fitur dan opini dari suatu *review*. Suatu *tanggapan* dapat mengandung opini positif atau negatif. Dengan mengetahui fitur dan opini dari suatu *review*, dapat membantu seseorang dalam mengambil suatu keputusan. Namun tidak semua kata pada suatu *review* merupakan fitur ataupun opini, dan juga banyaknya *review* semakin menyulitkan seseorang untuk mengetahui fitur dan opini dari *review* tersebut. Maka dari itu diperlukan pengekstraksian fitur dan opini yang akan memudahkan dalam menemukan kata fitur dan opini dari suatu kalimat *review*. Beberapa pendekatan yang dapat digunakan untuk mengekstraksi fitur dan opini, yaitu dengan pendekatan *association rule mining*, *unsupervised pattern mining*, *mutual reinforcement approach*, *opinion lexicon*, dan *pattern knowledge* [1]. Pada tugas akhir ini penulis menggunakan pendekatan *pattern knowledge* dan *opinion lexicon* untuk melakukan prediksi kalimat opini, melakukan ekstraksi fitur dan opini, dan juga menentukan polaritas atau orientasi dari suatu *review* apakah bernilai positif atau negatif yang kemudian akan dikelompokkan berdasarkan fiturnya untuk memudahkan dalam pencarian *review*. Dengan menggunakan pendekatan ini akan didapatkan fitur produk dan polaritas dari suatu kalimat *review* produk.

Kata Kunci: Opini, Fitur, *Pattern knowledge*, *Opinion lexicon*

Abstract

Feature and opinions extraction is a way that can be used to find out the features and opinions of a review. A review can contain positive or negative opinion. However, not all words in a review is a feature or opinion, and also the number of the reviews make it more difficult to know the features and opinion of the review. Thus the required extraction of features and opinion that will facilitate in finding the features and opinions of a sentence review. Feature and opinions extraction is necessary for ease in finding the features and opinion words of a sentence review. Several approaches can be used to extract features and opinion, such as association rule mining approach, unsupervised pattern mining, mutual reinforcement approach, opinion lexicon, and the pattern of knowledge* [1]*. In this final project the author uses pattern knowledge and opinion lexicon approach to make opinions sentence predictions, extracting features and opinions, and also determines the polarity or orientation of a review whether that is positive or negative and group it based on its features. By using this approach we will get the product features and the polarity of a sentence review of the product.

Keywords: Opinion, Features, Pattern knowldege, Opinion lexicon

1 Pendahuluan

1.1 Latar Belakang

Setiap orang memiliki penilaian yang berbeda terhadap baik atau buruknya suatu hal. *Review* produk merupakan penilaian yang diberikan seseorang terhadap fitur suatu produk. Setiap *review* yang diberikan terhadap suatu produk akan memiliki fitur produk yang tidak selalu sama, sehingga akan memiliki informasi yang berbeda untuk setiap fiturnya.

Opini dapat memberikan informasi yang akan membantu seseorang dalam melakukan suatu pengambilan keputusan. Melalui opini kita dapat mengetahui baik atau buruknya penilaian seseorang terhadap suatu produk. Namun tidak semua kata pada *review* produk merupakan opini, sehingga diperlukan perhatian khusus untuk mengetahui bagian mana dari *review* tersebut yang merupakan suatu opini. Banyaknya *review* yang ada juga dapat menyulitkan seseorang untuk mengetahui informasi dari suatu *review* terhadap suatu produk. Maka diperlukan suatu cara untuk mengekstraksi fitur, opini serta menentukan polaritasnya dari suatu tanggapan.

Terdapat beberapa kekurangan pada beberapa metode yang biasa digunakan untuk melakukan ekstraksi opini, seperti tidak mampu menentukan apakah suatu opini negatif atau positif, tidak dapat menentukan fiturnya, tanggapan yang akan ditulis harus sesuai dengan format tanggapan (positif/negatif) dan kurang efektif pada banyak kasus.

Pattern knowledge dan *opinion lexicon* merupakan salah satu pendekatan yang dapat digunakan untuk mengekstraksi opini dan fitur dari suatu tanggapan [1]. *Pattern knowledge* mampu untuk menentukan kandidat fitur dan opini. Namun tidak semua kata hasil dari *pattern rule* merupakan suatu opini, sehingga dibutuhkan

opinion lexicon untuk dapat menentukan kata opini yang tepat. Kedua metode ini mampu untuk menentukan fitur, mengatasi penulisan tanggapan secara *free format*, dan dapat menentukan polaritas suatu opini termasuk opini positif atau negatif. Penentuan fitur dan polaritas opini menjadi opini positif atau negatif akan memudahkan seseorang untuk menemukan informasi yang mereka cari.

2 Dasar Teori

2.1 Opinion Mining

Opinion mining atau yang juga sering disebut *sentiment analys*, merupakan bidang studi yang menganalisis opini, sentimen, evaluasi, sikap, dan emosi seseorang terhadap suatu objek seperti produk, jasa, organisasi, individu, masalah, peristiwa, topik, dan atribut mereka. *Opinion mining* berfokus pada opini yang mengekspresikan perasaan positif atau negatif [2].

2.1 Dataset

Dataset yang digunakan yaitu *customer review product* yang berjumlah 5 data teks dengan jenis produk yang berbeda. Semua *review* diambil dari *amazon.com* yang telah digunakan oleh Minqing Hu dan Bing Liu dalam penelitiannya. Setiap *dataset* yang digunakan memiliki karakteristik yang berbeda-beda. Dimana setiap *dataset* akan memiliki kalimat *review* yang kata fiturnya tidak terdapat pada kalimat *review* (implisit) dan jumlah kalimat yang berbeda-beda. Berikut menunjukkan karakteristik setiap *customer review product* pada *dataset* :

Tabel 2.1 *Dataset customer review product*

Dataset	Jumlah kalimat	Fitur Implisit	
		[p]	[u]
Apex AD2600 Progressive-scan DVD player	739 kalimat	45	38
Canon G3	597 kalimat	10	20
Creative Labs Nomad Jukebox Zen Xtra 40GB	1716 kalimat	52	60
Nikon coolpix 4300	346 kalimat	4	14
Nokia 6610	546 kalimat	5	24

2.2 *POS Tagging*

POS Tagging atau yang juga dikenal dengan *Part-Of-Speech Tagging* merupakan proses mengklasifikasikan kata ke dalam *part-of-speech* nya yang kemudian akan diberi label sesuai dengan kelompok katanya [3]. Label yang digunakan untuk menandai setiap kata di sebut juga dengan *tagset*. Terdapat beberapa metode yang bisa digunakan untuk melakukan proses *POS Tagging*, seperti *Standford Tagger* dan *Brill Tagger*. Pada tugas akhir ini penulis menggunakan *Stanford Tagger*.

2.3 Stopwords

Stopwords merupakan daftar kata-kata yang tidak mengandung suatu informasi. *Stopwords* pertama kali diperkenalkan oleh Hans Peter Luhn yang merupakan seorang ilmuwan komputer dan ahli informasi pada tahun 1958. Karena kata-kata pada *stopwords* sering muncul atau digunakan, maka ketika kata ini tidak digunakan dalam melakukan indexing maka dapat mengurangi 30-50% *space and time* yang diperlukan pada saat itu [4]. *Stopwords* yang digunakan yaitu *SEO Stopwords*[1] bahasa Inggris dengan melakukan sedikit perubahan pada *list*-nya seperti simbol (“@”, “#”, “&”, dan lainnya).

2.4 Lemmatization

Lemmatization merupakan salah satu teknik lain dari normalisasi untuk mengubah setiap *form* kata yang terinfleksi dalam suatu dokumen ke bentuk dasarnya, yang disebut *lemma* [5]. Yaitu dengan menghilangkan akhiran infleksi dan mengembalikannya ke bentuk dasar atau kamus [6]. *Lemmatization* yang digunakan yaitu *stanford lemmatization*[2].

2.5 N-gram

N-gram merupakan mekanisme pemotongan N-karakter dari suatu *string* yang panjangnya lebih dari N. Contoh dari penggunaan *N-gram* pada kata “TEXT” yaitu sebagai berikut :

1. *bi-gram* : TE, EX, XT
2. *tri-gram* : TEX,EXT
3. *quad-gram* : TEXT

Penggunaan *N-gram* telah berhasil menyelesaikan beberapa permasalahan seperti menangani *noisy* ASCII input, *text retrieval*, dan dalam berbagai aplikasi yang menggunakan *natural language processing* di dalamnya [7]. *Lemmatization* yang digunakan yaitu *stanford lemmatization.*

[1] http://www.link-assistant.com/seo-stop-words.html

[2] http://nlp.stanford.edu/software/corenlp.shtml

2.5 Pattern Knowledge

Pattern knowledge merupakan salah satu metode pendekatan yang digunakan untuk mengekstrak fitur dan opini yang ada pada kata benda atau frase kata benda dengan menggunakan pola pengetahuan berdasarkan *linguistic rule* [1]. Pada tugas akhir ini penulis menggunakan *pattern rule*, *noun phrase parser*, dan *type dependency* sebagai *pattern knowledge* yang digunakan untuk menentukan fitur dan opini.

2.5.1 Pattern Rule

Pattern rule digunakan untuk mengekstrak fitur yang merupakan *noun* atau *noun phrase* [1]. Berikut merupakan *rule* yang digunakan pada *pattern rule* [8]:

Tabel 2.2 Rule pada *pattern rule*

Pattern	The first word	The second word	The third word
Pattern 1	JJ	NN/NNS	-
Pattern 2	JJ	NN/NNS	NN/NNS
Pattern 3	RB/RBR/RBS	JJ	-
Pattern 4	RB/RBR/RBS	JJ/RB/RBR/RBS	NN/NNS
Pattern 5	RB/RBR/RBS	VBN/VBD	-
Pattern 6	RB/RBR/RBS	RB/RBR/RBS	JJ
Pattern 7	VBN/VBD	NN/NNS	-
Pattern 8	VBN/VBD	RB/RBR/RBS	-
Pattern 9	NN/NNS	JJ	-
Pattern 10	NN	-	-
Pattern 11	VB	-	-
Pattern 12	VB	RP	-
Pattern 13	DT	NN	-
Pattern 14	NN	NN	-
Pattern 15	JJ	VB	NN
Pattern 16	NN	VB	NN
Pattern 17	NN	IN	NN
Pattern 18	NN	NN	NN

Pengambilan 2-3 kata pada data *input* dilakukan dengan menggunakan *n-gram*, yaitu *bi-gram* dan *tri-gram*. Kemudian akan dicek kecocokan pola kata-nya dengan *pattern rule* untuk mendapatkan kata atau frase fitur. *Noun* merupakan kata yang paling menunjukkan bahwa kata tersebut merupakan fitur. Dan *adjective* atau *adverb* merupakan kata yang paling menunjukkan bahwa kata tersebut merukapan opini.

2.6.2 Noun Phrase Parser

Stanford parser merupakan suatu parser bahasa alami yang bekerja pada struktur gramatikal dari suatu kalimat, misalnya, kelompok kata (seperti "frasa") dan kata-kata yang merupakan subjek atau objek dari suatu kata kerja. Probabilistik parser menggunakan pengetahuan tentang bahasa yang diperoleh dari kalimat yang telah melalui proses *hand-parser* untuk menghasilkan analisis yang paling memungkinkan dari kalimat-kalimat baru [9]. Stanford parser dapat digunakan untuk mendapatkan noun phrase dari suatu kalimat. Proses *noun phrase parser*, akan digunakan untuk mendapatkan kata atau frasa fitur dengan menggunakan *phrase structure trees*.

2.6.3 Type Dependency

Stanford typed dependencies dirancang untuk memberikan gambaran sederhana dari hubungan gramatikal dalam suatu kalimat yang dapat dengan mudah dipahami dan efektif untuk digunakan oleh seseorang tanpa keahlian linguistik yang ingin mengetahui keterhubungan tekstual [10]. Untuk mengidentifikasi kandidat fitur dan opini, perlu dilakukan pengecekan hasil dari *type dependency* dengan *rule* dari *dependency relation templates*. Berikut merupakan *rule* dari *dependency relation templates* yang digunakan pada proses *type dependency* [11]:

Tabel 2.3 *Dependency relation templates*

Dependency relation template	Feature word	Opinion word
NN - amod – JJ	NN	JJ
NN - nsubj – JJ	NN	JJ
NN - nsubj - VB - dobj – NN	The first NN	The last NN
VB - advmod – RB	VB	RB

2.7 Corpus

Corpus merupakan suatu koleksi sistematis dari bahasa alami secara lisan ataupun tulisan di dalam suatu konteks, yang disimpan pada komputer untuk analisis kualitatif dan kuantitatif. Suatu corpus dihasilkan sesuai dengan prinsip desain eksplisit untuk berbagai tujuan [12]. *Corpus* fitur menggunakan kata fitur yang ada pada *dataset*. Kata fitur pada *dataset* tersebut didapatkan dari hasil pada penelitian sebelumnya. *Corpus* diguanakan untuk melakukan pengecekan terhadap hasil *pattern knowledge* sehingga didapatkan kandidat fitur produk.

2.8 WordNet

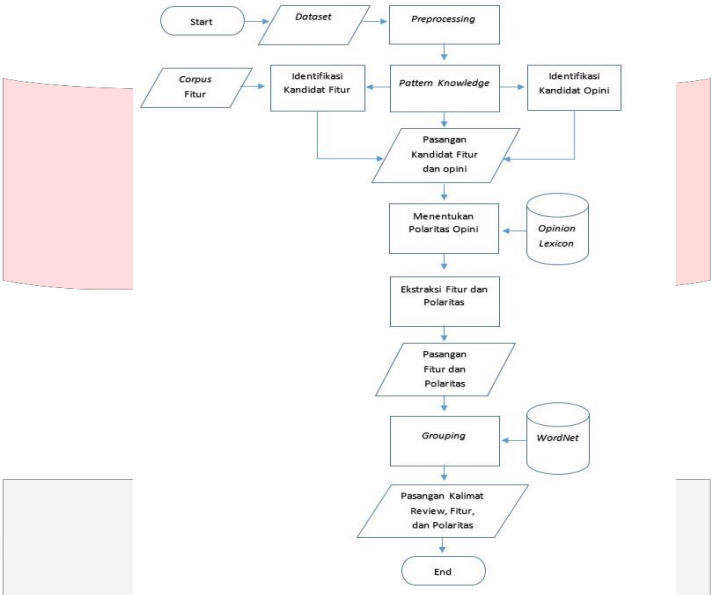
WordNet merupakan kamus bahasa Inggris yang berorientasi semantik. Terdapat sebanyak 155.287 kata dan 117.659 *set* sinonim yang disebut synset di dalam *WordNet* [3]. Dengan *WordNet* kita dapat mengetahui sinonim dari suatu kata. Pencarian kata sinonim digunakan pada proses *grouping*, yaitu mengelompokkan hasil ekstraksi sesuai masing-masing fiturnya.

2.9 Opinion Lexicon

Opinion lexicon adalah kumpulan kata dan frase yang berupa kata sifat, kata keterangan, kata kerja, dan kata benda yang merupakan kata opini [13]. *SentiWordNet* 3.0.0 merupakan *opinion lexicon* yang digunakan dalam penelitian ini. Opinion lexicon digunakan untuk mendapatkan nilai polaritasnya dan untuk memastikan pahwa kata opini merupakan kata *sentiment.*

3 Perancangan dan Implementasi

Sistem yang akan dibuat meliputi pengekstraksian fitur dan opini, menentukan polaritas, dan *grouping*. Berikut merupakan skema umum sistem :



Gambar 3.1 Skema umum sistem

Pertama pada *Dataset* akan dilakukan *preprocessing* terlebih dahulu sebelum digunakan pada proses ekstraksi fitur dan opini. *Dataset* yang telah melalui proses *preprocessing* akan diproses menggunakan *pattern knowledge* untuk mengidentifikasi kandidat fitur dan opini. Proses identifikasi kandidat fitur dilakukan menggunakan *pattern knowledge* dan *corpus* fitur. Kata fitur dan opini yang berhasil teridentifikasi akan menjadi pasangan kandidat fitur dan opini. Pasangan kandidat fitur dan opini ini kemudian akan dilakukan pengecekan nilai sentiment untuk dapat menentukan polaritasnya. Kemudian akan dilakukan ekstraksi terhadap fitur dan polaritasnya yang selanjutnya akan dihasilkan pasangan fitur dan polaritasnya. Proses terakhir yaitu melakukan proses grouping berdasarkan fiturnya terhadap hasil ekstraksi tersebut.

4 Pengujian dan Analisis Hasil Implementasi

Pengujian dilakukan dengan melakukan kombinasi antara *pattern rule*, *noun phrase parser*, dan *type dependency* pada pendekatan *pattern knowledge* serta menggunakan *opinion lexicon* dalam melakukan pengekstraksian fitur dan polaritas pada kalimat *review* sehingga dapat diketahui berapa besar tingkat keberhasilannya.

4.1 Analisis Ekstraksi Fitur dan Opini

Berikut merupakan hasil dari pengujian ekstraksi fitur dan opini yang dilakukan terhadap 5 dataset customer review product dengan menggunakan *pattern rule*, *noun phrase parser*, *type dependency* dan kemudian kombinasinya :

Tabel 4.1 *F1-score* ekstraksi fitur dan opini

	F1 Fitur dan Opini						
	Rule	NP parser	Type dependecies	Rule & NP parser	Rule &Type dependencies	NP parser & Type dependencies	Rule, NP parser & Type dependencies
Apex AD2600 Progressive-scan DVD player	52.2	53.28	50.69	52.2	51.3	51.44	51.42
Canon G3	57.39	55.94	56.89	55.31	55.23	54.82	54.15
Creative Labs Nomad Jukebox Zen Xtra 40GB	55.98	56.24	55.64	55.49	55.11	55.32	55
Nikon coolpix 4300	52.58	53	54.26	52.02	51.61	52.93	51.92
Nokia 6610	56.04	54.46	52.93	56.72	55.72	54.79	56.81

Berdasarkan hasil perhitungan *F1-score* pada ekstraksi fitur dan opini yang dapat dilihat pada gambar 4.1, dapat dilihat bahwa setiap *dataset* memberikan hasil yang tidak konstan dalam pengekstraksian fitur dan opini, ini dipengaruhi oleh karakteristik yang berbeda yang dimiliki oleh setiap *dataset*. Dan dapat diketahui ekstraksi fitur dan opini menggunakan *pattern rule* menghasilkan performansi yang paling baik yaitu dengan nilai *F1-score* paling tinggi sebesar 57,39% pada *dataset* canon G3. Ini dikarenakan *pattern rule* mampu mengekstrak kata fitur maupun frasa fitur karena pola-pola yang digunakan mencakup kata maupun frasa fitur, dan mampu menemukan pasangan kata opininya. Sedangkan ekstraksi fitur dan opini dengan menggunakan *type dependency* menghasilkan performansi paling rendah *F1-score* paling rendah yaitu sebesar 50,69% pada *dataset* apex AD2600. Ini dikarenakan *type dependency* hanya melakukan pengekstaksian fitur dan opini yang memiliki keterhubungan gramatikal saja, dan *type dependency* tidak mampu untuk mengekstrak frasa fitur.

4.2 Analisis Penentuan Fitur dan Polaritas

Penentuan polaritas dilakukan dengan menghitung nilai sentimen dari kata opini hasil dari proses ekstraksi fitur dan opini menggunakan *pattern rule, noun phrase parser, type dependency* dan kombinasinya pada *opinion lexicon* yaitu *SentiWordNet*. Berikut merupakan hasil dari pengujian dalam menentukan fitur dan polaritas. :

Tabel 4.2 *F1-score* fitur dan polaritas

	F1 Fitur Polaritas						
	Rule	NP parser	Type dependecies	Rule & NP parser	Rule &Type dependencies	NP parser & Type dependencies	Rule, NP parser & Type dependencies
Apex AD2600 Progressive-scan DVD player	49.35	50.23	49.74	48.32	48.58	48.01	47.62
Canon G3	54.55	52.52	56.11	51.24	52.61	51.29	50.14
Creative Labs Nomad Jukebox Zen Xtra 40GB	53.98	52.96	54.76	51.74	52.86	51.78	51.05
Nikon coolpix 4300	50.54	50.42	52.81	48.78	49.57	49.84	48.69
Nokia 6610	52.63	51.42	50.8	52.16	51.88	50.73	51.81

Berdasarkan hasil perhitungan *F1-score* pada penentuan fitur dan polaritas yang dapat dilihat pada gambar 4.2, dapat dilihat bahwa perbedaan karakteristik *dataset* juga mempengaruhi hasil dari penentuan fitur dan polaritas, karena penentuan fitur dan polaritas dipengaruhi oleh fitur dan opini yang telah terekstrak sebelumnya, sehingga mengakibatkan hasil penentuan fitur dan polaritas juga tidak konstan untuk setiap *dataset*-nya. Dan dapat diketahui performansi paling baik dengan nilai *F1-score* paling tinggi dalam menentukan fitur dan polaritas yaitu sebesar 56.11% diperoleh dari hasil *type dependency* pada *dataset* canon G3. Dapat dilihat juga bahwa *F1-score* pada *type dependency* memiliki nilai *F1-score* paling tinggi pada 3 dari 5 *dataset*. *Type dependency* mampu menghasilkan pasangan kata opini yang tepat karena *type dependency* melihat keterhubungan gramatikal setiap katanya, dan juga digunakan *dependency relation template* untuk mendapatkan pasangan kata fitur dan opini yang tepat, sehingga mampu menghasilkan polaritas yang tepat pula

Sedangkan performansi paling rendah dengan nilai *F1-score* paling rendah dalam menentukan fitur dan polaritas yaitu sebesar 47,62% yang diperoleh dari hasil kombinasi *pattern rule*, *noun phrase parser* dan *type dependency* yaitu pada *dataset* apex AD2600. Hasil *F1-score* dalam menentukan fitur dan polaritas yang diperoleh dari hasil kombinasi *pattern rule*, *noun phrase parser* dan *type dependency* juga menghasilkan nilai *F1-score* paling rendah pada 4 dari 5 *dataset*. Ini dikarenakan pengekstraksian pasangan fitur dan polaritas dengan menggunakan kombinasi dari ketiga cara tersebut akan menghasilkan jumlah pasangan yang banyak, namun akan memiliki akurasi yang kurang baik karena setiap cara tersebut memiliki kelebihan dan kekurangan tersendiri tergantung dari karakteristik *dataset* yang digunakan.

4.3 Analisis Prediksi Kalimat Opini

Prediksi kalimat opini dilakukan dengan membandingkan kalimat yang memiliki fitur dan opini hasil ekstraksi dari sistem dengan kalimat yang memiliki fitur dan opini pada *dataset*. Berikut merupakan hasil dari pengujian prediksi kalimat opini :

Tabel 4.3 *F1-score* prediksi kalimat opini

	F1 Kalimat Opini Prediksi						
	Rule	NP parser	Type dependecies	Rule & NP parser	Rule &Type dependencies	NP parser & Type dependencies	Rule, NP parser & Type dependencies
Apex AD2600 Progressive-scan DVD player	48.18	53.7	28.57	58.49	51.57	55.9	59.6
Canon G3	56.07	61.66	28.83	65.22	55.7	63.16	64.53
Creative Labs Nomad Jukebox Zen Xtra 40GB	51.24	57.4	34.63	62.27	54.23	60.58	62.98
Nikon coolpix 4300	55.41	58.44	28.57	63.31	55.15	59.49	63.53
Nokia 6610	62.92	64.31	40.22	69.84	64.49	68.05	71.35

Berdasarkan hasil perhitungan *F1-score* pada penentuan fitur dan polaritas yang dapat dilihat pada gambar 4.3, dapat dilihat hasil yang tidak kostan pada prediksi kalimat opini di setiap *dataset*-nya, ini dipengaruhi oleh karakteristik yang dimiliki oleh setiap *dataset* yang berbeda-beda. Dan dapat diketahui bahwa performansi paling baik untuk prediksi kalimat opini yaitu menggunakan hasil dari kombinasi *pattern rule*, *noun phrase parser*, dan *type dependency* menghasilkan nilai *F1-score* paling tinggi yaitu sebesar 71,35%. Menggunakan kombinasi dari *pattern rule*, *noun phrase parser*, dan *type dependency* mampu menghasilkan jumlah fitur dan

opini yang paling banyak dibandingkan yang lain, karena ketiga cara tersebut masing-masing memiliki kelebihan tersendiri dalam melakukan ekstraksi fitur dan opini tergantung dari karakteristik dataset yang digunakan, sehingga mampu memprediksi kalimat review yang memiliki fitur dan opini dengan baik

Dan dengan nilai *F1-score* paling rendah di hasilkan menggunakan hasil dari *type dependency* yaitu sebesar 28,57%. Performansi paling rendah untuk prediksi kalimat opini yaitu menggunakan hasil dari *type dependency*, yang memiliki nilai *F1-score* paling rendah pada seluruh *dataset*. Ini dikarenakan *type dependency* hanya mampu memprediksi kalimat opini yang memiliki keterhubunguan gramatikal antara fitur dan opininya dan juga keterhubungan gramatikalnya harus memenuhi aturan pada *dependency relation template* sehingga hanya mampu memprediksi sedikit kalimat review dengan benar.

4.4 Analisis Penentuan Grouping

Penentuan grouping dilakukan dengan mencari kata sinonim dari fitur yang telah terekstraksi hasil dari ekstraksi fitur dan opini menggunakan *pattern rule, noun phrase parser, type dependency* dan kombinasinya lalu mengelompokkannya sesuai dengan fitur dan sinonimnya. Berikut merupakan hasil dari penentuan *grouping* :

Fitur : picture	Fitur : image
Kalimat review : [picture,[+]]##they fired away and the picture turned out quite nicely . (as all of my pictures have thusfar) . [quality,[+], image,[+]]##i chose the g3 because of its reputation for very high quality , clean images . [picture,[+]]##it takes great pictures , operates quickly , and feels solid .	Kalimat review : [picture,[+]]##they fired away and the picture turned out quite nicely . (as all of my pictures have thusfar) . [quality,[+], image,[+]]##i chose the g3 because of its reputation for very high quality , clean images . [picture,[+]]##it takes great pictures , operates quickly , and feels solid .

Gambar 4.1 Hasil *grouping*

Dapat dilihat dari gambar 4.1 hasil *grouping* berdasarkan fitur "*picture*" dan "image" akan menghasilkan grouping yang sama untuk kalimat review dengan kata fitur "picture" maupun "image", karena kata "picture" bersinonim dengan "image" sehingga akan memiliki kalimat *review* dengan fitur "*picture*" dan "*image*" yang sama pada *list group*-nya. hasil ekstraksi yang memiliki fitur "*picture*" ataupun sinonim dari kata "*picture*". Namun memungkinkan untuk terdapat kalimat *review* yang berbeda pada *group* fitur yang bersinonim, karena setiap kata akan memiliki jumlah sense yang tidak selalu sama, misalnya "*picture*" memiliki jumlah *senses* yaitu 11, sedangkan "*image*" memiliki jumlah *senses* 9, sehingga pada *list group* dapat memiliki jumlah yang berbeda tergantung dari sinonim fitur masing-masing kalimat *review*.

5 Kesimpulan dan Saran

5.1 Kesimpulan

Setelah melakukan pengujian dan menganalisis hasil uji pada bab 4, maka dapat diambil kesimpulan sebagai berikut :

1. Kalimat *review* pada *dataset* memiliki karakteristik yang berbeda-beda, dimana setiap dataset memiliki jumlah kalimat dan jumlah kalimat dengan fitur implisit yang berbeda-beda sehingga memberikan hasil performansi yang tidak konstan untuk setiap *dataset*-nya.
2. Pendekatan dengan *pattern knowledge* dan *opinion lexicon* menggunakan *pattern rule* mampu memberikan performansi paling baik dalam melakukan pengekstraksian fitur dan opini yaitu dengan nilai *F1-score* yang dimiliki sebesar 57,39%. *Pattern rule* mampu mengekstrak kata fitur maupun frasa fitur karena pola-pola yang digunakan mencakup kata maupun frasa fitur, dan mampu menemukan pasangan kata opininya.
3. Pendekatan dengan pattern knowledge dan *opinion lexicon* menggunakan *type dependency* mampu memberikan performansi yang paling baik dalam proses penentuan fitur dan polaritas yaitu dengan nilai *F1-score* yang dimiliki sebesar 56.11%. *Type dependency* mampu menghasilkan pasangan kata opini yang tepat karena *type dependency* melihat keterhubungan gramatikal setiap katanya, dan juga digunakan *dependency relation template* untuk mendapatkan pasangan kata fitur dan opini yang tepat, sehingga mampu menghasilkan polaritas yang tepat pula.
4. Pendekatan dengan *pattern knowledge* dan *opinion lexicon* menggunakan kombinasi *pattern rule, noun phrase parser*, dan *type dependency* serta menggunakan *SentiWordNet* sebagai *opinion lexicon* mampu memberikan performansi yang paling baik dalam melakukan prediksi kalimat opini yaitu dengan nilai *F1-score* yang dimiliki sebesar 71,35%. Ini karena ketiga cara tersebut saling menutupi kelemahannya masing-masing, dan setiap caranya memiliki kelebihan tersendiri dalam melakukan ekstraksi fitur dan opini tergantung dari karakteristik dataset yang digunakan, sehingga mampu memprediksi kalimat review yang memiliki fitur dan opini dengan baik.

5. Penentuan *grouping* dapat dilakukan dengan menggunakan *WordNet* sebagai kamus kata bahasa Inggris untuk mendapatkan persamaan kata (sinonim) fitur produk sehingga dapat memudahkan dalam pencarian suatu fitur pada kalimat *review*.

5.2 Saran

Saran yang diberikan antara lain sebagai berikut :

1. Mengidentifikasi fitur produk pada kalimat *review* yang bersifat implisit.
2. Mengidentifikasi negasi kata opini.
3. Mengidentifikasi kata ganti benda untuk mendapatkan kata fitur bersifat implisit.

Daftar Pustaka :

[1] S. S. Htay and K. T. Lynn, "Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews," *The Scientific World Journal*, 2013.

[2] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.

[3] S. Bird, E. Klein and E. Loper, Natural Language Processing with Python, 2009.

[4] A. Blanchard, "Understanding and customizing stopword lists for enhanced patent mapping*," 2007.

[5] T. Korenius, J. Laurikkala, K. Järvelin and M. Juhola, "Stemming and Lemmatization in the Clustering of Finnish".

[6] V. Balakrishnan and E. Lloyd-Yemoh, "Stemming and Lemmatization: A Comparison of Retrieval Performances," 2014.

[7] W. B. Cavnar and J. M. Trenkle, "N-Gram-Based Text Categorization".

[8] S. S. Htay, "Biologically Inspired Opinion Mining: Features Extraction based on Linguistic Patterns for Customer Reviews," *International Journal of Information Systems and Engineering (online)*, 2014.

[9] "The Stanford Natural Language Processing Group," Stanford University Natural Language Processing, [Online]. Available: http://nlp.stanford.edu/software/lex-parser.shtml. [Accessed 04 06 2015].

[10] M.-C. de Marneffe and C. D. Manning, "Stanford typed dependencies manual," 2008.

[11] L. Zhuang, F. Jing and X.-Y. Zhu, "Movie Review Mining and Summarization," 2006.

[12] M. N. KAYAOĞLU, "The Use of Corpus for Close Synonyms," *The Journal of Language and Linguistic Studies*, 2013.

[13] X. Ding, B. Liu and P. S. Yu, "A Holistic Lexicon-Based Approach to Opinion Mining," 2008.

[14] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," 2002.