OPINION EXTRACTION DENGAN MENGGUNAKAN PENDEKATAN PATTERN RULE

OPINION EXTRACTION USING PATTERN RULE APPROACH

[1]Sendika Panji Anom [2] Warih Maharani,ST, MT. [3]Moch.Arif Bijaksana, IR.,MTECH.
[1 2 3] Fakultas Informatika, Universitas Telkom, Bandung
[1]sendikapanji@gmail.com, [2]wmaharani@gmail.com, [3]arifbijaksana@gmail.com

Abstrak

***Opinion mining* merupakan topik yang sedang marak dilakukan saat ini. *Opinion mining* merupakan cabang penelitian dari text mining yang berfokus dalam analisis opini dari suatu dokumen teks. Dunia bisnis saat ini menggunakan *opinion mining* untuk menganalisis secara otomatis opini pelanggan tentang produk dan pelayanannya dan melakukan klasifikasi opini positif dan negatif. Salah satu cara untuk menerapkan *opinion mining* adalah *Opinion extraction*. Cara ini digunakan untuk mendeteksi dimana terdapat sebuah kata kata opini dalam suatu dokumen atau kalimat. Tahap ini mempunyai beberapa pendekatan dalam penerapannya, pada tugas akhir ini menggunakan *pattern rule* sebagai metode pendekatan. Pedekatan ini digunakan untuk ekstraksi fitur dan opini dalam kalimat dengan mencocokan susunan kalimat dengan pola pengetahuan. Pendekatan ini dapat mengekstraksi kandidat fitur dan opini sekaligus dari sebuah dokumen atau kalimat kemudian dilakukan klasifikasi opini positif negatif berdasarkan *semantic orientation* dengan algoritma PMI-IR**

Kata Kunci: *Opinion Mining, Opinion Extraction, Pattern Rule*

Abstract

Opinion mining is an emerging topic made at this time. Opinion mining is a branch of research that focuses on text mining in opinion analysis of a text document. The business world is currently using opinion mining to automatically analyze customer opinions about products and services and conduct classification of positive and negative opinions. One way to implement opinion mining is the extraction Opinion. This method is used to detect where there is an opinion words or phrases in a document. This phase has several approaches in its application, in this final rule as a method using a pattern approach. This approach is used for feature extraction and matching opinions in a sentence with the sentence structure with a pattern of knowledge. This approach can extract features and opinion candidates at once from a document or sentence then carried classification negative positive opinion is based on semantic orientation with PMI-IR algorithm.

1 Pendahuluan

1.1 Latar Belakang

Kebutuhan akan internet sudah menjadi hal penting untuk mendukung kegiatan manusia saat ini. Internet menjadi sumber hiburan , pendidikan komunikasi , dan transaksi jual beli (*e-commerce*). Pengguna internet saat ini tidak hanya melakukan *browsing* tetapi juga memberikan *feedback* berupa komentar yang berguna terhadap penilaian suatu produk [1].

Tanggapan atau opini dari user tersebut dapat menjadi tolak ukur penilaian suatu produk. Sebagai contoh pada *e-commerce*, jumlah *review* dari pelanggan terhadap suatu produk bisa bertambah dengan cepat. Hal ini membuat pelanggan mengalami kesulitan dalam membaca dengan tujuan untuk menentukan produk yang akan mereka pilih. Dalam hal ini diperlukan cara untuk mengekstraksi bagian fitur beserta opini dari suatu *review* secara otomatis untuk selanjutnya di klasifikasikan menjadi opini positif atau negatif sehingga dapat membantu pelanggan membaca dan menentukan pilihan.

Masalah tersebut dapat teratasi dengan ekstraksi opini. Ekstraksi opini dilakukan untuk mengekstraksi bagian fitur beserta opini yang ada dalam suatu kalimat. Pendekatan yang dapat digunakan dalam *Ekstraksi opini* antara lain *corpus approach , association rule mining* dan *pattern rule*. Namun pendekatan *Corpus approach , dan association rule mining* di desain hanya untuk mendeteksi frekuensi kemunculan suatu kata dalam kalimat , sehingga kata yang lebih banyak muncul yang menjadi kandidat fitur. dengan cara ini memungkinkan banyaknya opini terhadap sebuah fitur yang tidak relevan dikarenakan hanya fitur yang di ekstrak..

Pada tugas akhir ini dengan menggunakan pendekatan *pattern rule* dilakukan pengecekan dua kata berturut turut berdasarkan pola pengetahuan dan mendeteksi sekaligus kandidat fitur dan opini dalam satu kalimat dengan kata sifat atau kata keterangan pada bagian pertama dan kata kedua memberikan konteks yang selanjutnya menjadi fitur. Pengecekan dengan mengunakan 2 kata berturut turut terbukti menangani *isolated*

adjectives seperti contoh kata *adjective* "*unpredictable*" mempunyai makna negatif pada bidang *automotive* pada frasa "*unpredictable steering*" sedangkan mempunyai makna positive pada bidang film seperti "*unpredictable plot*"[1] .Pola pengetahuan yang ada pada *pattern rule* seperti ini dapat menentukan fitur sekaligus orientasi positif atau negatif pada suatu kalimat..

2 Dasar Teori

2.1 Data Mining

Data mining adalah proses ekstraksi informasi dalam pola data yang besar sehingga dapat berguna. Hal ini juga disebut proses penemuan ilmu , pertambangan ilmu dari data , ataupun ekstraksi pengetahuan dan analisis data/pola. Penemuan data selanjutnya dapat digunakan untuk membuat keputusan tertentu untuk pengembangan penelitian. [2]

2.2 Opinion Mining

Opinion mining dapat di definisikan sebagai bagian dari komputasi linguistik yang berfokus pada penggalian opini masyarakat dari web. Banyak nya informasi di dalam web menarik sebagaian besar masyarakat untuk menganalisis dan mencari informasi di dalam web [3].

2.3 POS Tagging

Part of speech (POS) tagging adalah teknik untuk mengidentifikasi kelas morfosintaksis yang dilakukan dengan penandaan tata bahasa atau kategori kata. *POS tagging* dilakukan dalam konteks linguistik komputasi yang betujuan untuk mengindetifikasi setiap kata yang ada di setiap kalimat dan di uraikan untuk menghasilkan *part of speech* [5].

2.4 PreProcess

Preprocess adalah tahapan dimana data mentah akan diolah menjadi data yang berkualitas. Tahapan ini sangat dibutuhkan untuk mendapatkan data yang berkualitas dan meningkatkan efisiensi proses penggalian informasi, salah satunya dengan cara mengurangi ukuran data yang bisa di dapatkan dengan memperbaiki data yang salah serta menggabungkan dan menghilangkan fitur yang berlebihan dalam sebuah kalimat [7].

2.4.1 Stopword

Stopword adalah metode dasar dalam penyaringan kata dalam dokumen untuk mendapatkan data yang berkualitas. *Stopword* dapat berupa kata tanpa arti dalam bahasa khusus. Kata kata dalam *stopword* dianggap teralu sering berada diantara dokumen dan faktanya kata yang 80% muncul dalam dokumen itu tidak berguna dalam membantu pemahaman suatu dokumen dan tentunya harus dihilangkan. Secara normal *stopword* merupakan elemen penting dalam dokumen yang harus di saring. Artikel, preposisi, konjungsi merupakan kandidat alami stopword. Berikut merupakan beberapa contoh stopword dalam bahasa inggris [8]:

2.4.2 Stemming

Stemming digunakan sebagai proses untuk mengurangi jumlah kata yang bertujuan membantu pemahaman sebuah dokumen . Hal ini berdasarkan observasi bahwa kata kata dalam dokumen sering mempunyai variasi morfologi. Contoh dari variasi morfologi adalah penggunaan kata *computing, computer, computation, computes, computational, computable, dan computability* mempunyai makna sama dalam dokumen. Kata kata tersebut tentunya mempunyai akar linguistik yang sama. Menempatkan kata kata tersebut secara bersama sebagai kejadian satu kata seperti *"comput"* akan memberikan indikasi kuat untuk memahami isi dari dokumen dan mengurangi jumlah representasi kata kata ke dalam jumlah yang relatif mudah dikelola dibandingkan dengan meletakan secara terpisah [9].

2.4.3 Lemmatiazation

Lemmatization adalah proses untuk menemukan bentuk normalisasi dari sebuah kata. Lemmatization merupakan langkah *preprocessing* yang penting bagi banyak *text mining*. Hal ini juga digunakan didalam pengolahan bahasa dan bidang lainnya yang berhubungan dengan bidang linguistic. *Lemmatization* hamper mirip dengan *stemming* tetapi tidak butuh untuh menghasilkan suatu aturan untuk mengganti kata, tetapi *Lemma* menggunakan penggantian elemen kata pada akhiran kata [10]

2.4 Pattern Rule

Pattern Rule adalah salah satu pendekatan untuk mendapatkan ektraksi opini dan fitur. Pendekatan ini berfokus pada pengecekan dua kata berturut turut pada frasa yang mengandung kata sifat, kata keterangan, dan kata benda yang sudah mengalami proses pengolahan *POS Tag* dan di cocokan dengan komposisi *rule tag* dalam tabel untuk menentukan fitur dan opini, *rule* yang digunakan adalah sebagai berikut [11] :

Tabel 2.1 *List Pattern Rule [10]*

Pattern	First Word	Second Word	Third Word
Pattern 1	JJ	NN / NNS	-
Pattern 2	RB/RBR/RBS	JJ	not NN nor NNS
Pattern 3	JJ	JJ	not NN nor NNS
Pattern 4	NN/NNS	JJ	not NN nor NNS
Pattern 5	RB/RBR/RBS	VB/VBD/VBN/VBG	-
Pattern 6	RR/RBR/RBS	RR/RBR/RBS	JJ
Pattern 7	VBN/VBD	NN/NNS	-
Pattern 8	VBN/VBD	RB/RBR/RBS	-
Pattern 9	RR/RBR/RBS	VBN/VBD	-

2.5 PMI-IR Algorithm

Algoritma *PMI-IR (Pointwise Mutual Information and Information Retrieval)* adalah algoritma untuk menentukan orientasi positif dan negatif dari frasa dalam kalimat. Proses penentuan dilakukan dengan cara menghitung kedekatan kata dalam setiap frasa dalam kalimat dengan referensi kata yang ditentukan sebelumnya. Kata yang ditentukan sebagai referensi positif adalah *"excellent"* dan kata dengan referensi negatif adalah *"poor"* sesuai dengan *five star rating system* .Kedekatan akan dihitung dengan jumlah hits yang dihasilkan dari kedekatan kata yang telah diuji .Hasil dari perhitungan kedekatan kata selanjutnya akan menentukan *semantic orientation* dalam suatu kata atau frasa. [10] . Rumus yang digunakan untuk menentukan *Semantic Orientation* pada algoritma PMI- IR adalah :

$$\log_2\left(\frac{h\text{[illegible]}N\text{[illegible]}\text{Excellent}")h\text{[illegible]}("\text{[illegible]})}{h\text{[illegible]}N\text{[illegible]}\text{Poor}")h\text{[illegible]}("\text{[illegible]})}\right)$$

Gambar 2-1 Perhitungan PMI-IR

2.6 Precission and Recall

Melakukan pengecekan atas dokumen yang relevan atau tidak adalah pekerjaan yang sulit. Jumlah dokumen yang tidak relevan akan lebih banyak dibandingkan dengan dokumen yang relevan sehingga akan menurunkan tingkat akurasi kecocokan dokumen yang relevan. Terdapat 4 kategori dalam suatu proses pencarian dokumen, Kategori tersebut adalah:

- *True Positive* : Dokumen yang relevan dan diidentifikasi dengan benar sebagai dokumen yang relevan
- *True Negative* : Dokumen yang tidak relevan dan diidentifikasi dengan benar sebagai dokumen yang tidak relevan
- *False Positive* : Dokumen yang tidak relevan dan diidentifikasi salah sebagai dokumen relevan (*Error Type 1*)
- *False Negative* : Dokumen yang relevan dan diidentifikasi salah sebagai dokumen yang tidak relevan (*Error Type 2*)

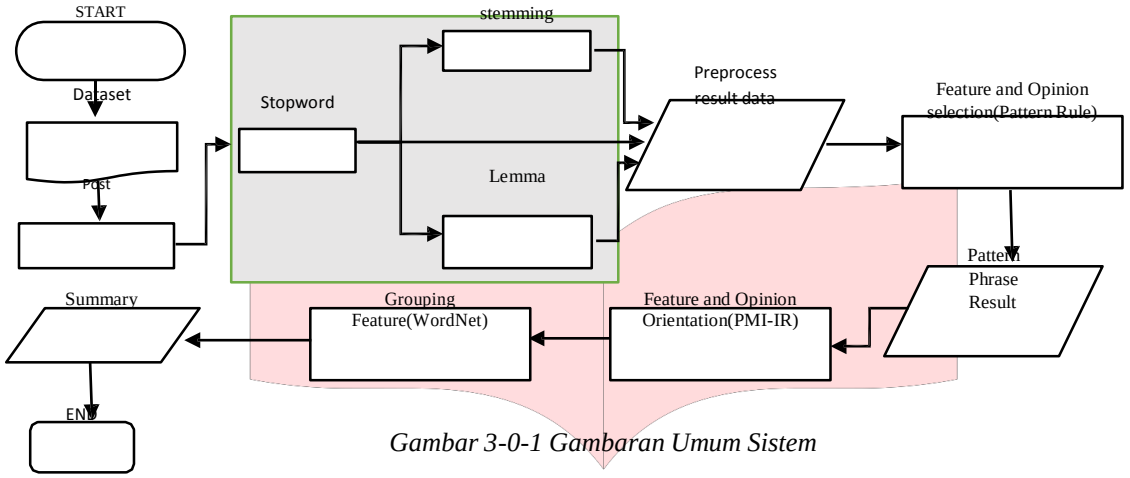
Precision and Recall adalah proses evaluasi untuk menghitung tingkat akurasi pencarian informasi relevan pada suatu dokumen. *Precision* digunakan untuk menghitung banyaknya informasi atau item yang teridentifikasi sebagai item yang relevan, *TP / (TP+FP)* dan *Recall* digunakan untuk menghitung banyaknya informasi atau item yang relevan yang berhasil diidentifikasi *TP/(TP+FN)* [13].

2.7 Grouping

Wordnet adalah sistem referensi leksikal berbentuk kamus bahasa inggris yang berisi set sinonim atau arti kata . *Wordnet* berfokus pada arti kata, bukan bentuk kata. Saat ini *wordnet* berisi sekitar 95.600 bentuk kata yang berbeda dan disusun dalam 70.100 arti kata atau set sinonim. Pencarian sinonim atau persamaan kata bisa kita lakukan dengan akurat dengan *wordnet* [11]. Wordnet dalam tugas akhir ini membantu proses *grouping* untuk kalimat berdasarkan kelas

3 Pembahasan

Rancangan sistem untuk ekstraksi fitur dan opini dengan menggunakan pendekatan *pattern rule* secara umum , hasil akhirnya adalah suatu ringkasan yang dapat mengelompokan fitur berserta sentiment positif dan negatif setiap fitur dapat digambarkan sebagai berikut :



Gambar 3-0-1 Gambaran Umum Sistem

Gambar diatas merupakan rancangan sistem untuk ekstraksi fitur dan opini dengan menggunakan pendektan pattern rule. Dalam penelitian tugas akhir ini , pendekatan ekstraksi fitur dan opini dengan pattern rule di buat menjadi 3 tahapan preprocess untuk mengetahui tahapan mana yang bisa menghasilkan fitur dan opini yang relevan dengan tingkat akurasi yang tinggi. Tahapan di bagi manjadi tahapan preprocess stopword , stopword dan stemming serta tahapan stopword dan lemma. Setelah melalui preprocess sistem akan dilanjutkan dengan ekstraksi kandidat pasangan fitur dan opini dengan menggunakan pendekatan pola pengetahuan pattern rule. Hasil ekstraksi dengan pattern rule , frasa akan masuk ke dalam pengujian orientasi opini dengan menggunakan metode PMI-IR guna mengetahui polaritas fitur dan opini dari frasa tersebut. Tahap selanjutnya adalah tahapan untuk meringkas hasil ekstraksi dengan cara mengelompokan fitur yang mempunyai arti dan makna yang sama menggunakan teknik Grouping dibantu dengan tools wordnet untuk memudahkan review fitur dari suatu produk. Akhir dari tahapan sistem akan di peroleh kesimpulan dengan mengklasifikasikan fitur dan opini beserta kalimat ke dalam tanggapan positif atau tanggapan negatif.

4.Pengujian dan Analisis Hasil Implementasi

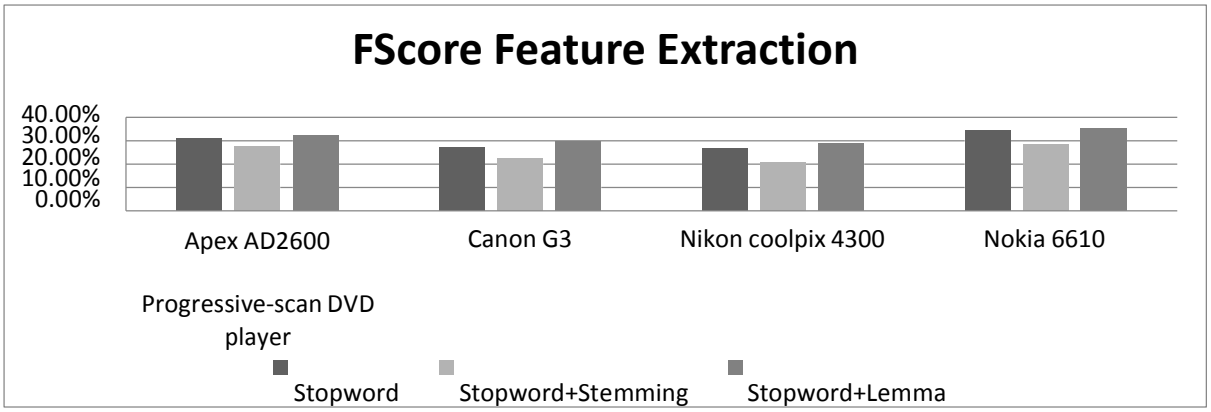
Pengujian ini dilakukan untuk melihat hasil dari keseluruhan proses yang berjalan pada sistem yang dibangun. Hasil keluaran yang akan di uji adalah ekstraksi fitur dan opini dengan menggunakan pendekatan *pattern rule* yang selanjutnya di evaluasi *precission* dan *recall* dari ekstraksi fitur dan opini dengan beberapa *preprocess*. Tahapan preprocess yang dilakukan adalah tahapan dengan hanya menggunakan preprocess stopword , tahapan dengan stopword dan stemming , dan tahapan dengan menggunakan stopword dan lemmatization . Hal ini dilakukan dengan tujuan menilai hasil *preprocess* yang sudah di bedakan guna mencari hasil ekstraksi yang terbaik bagi ekstraksi fitur dan opini.

4.1 Analisis Hasil Pengujian

Pada bagian ini diperlihatkan hasil pengujian yang telah dilakukan berikut dengan analisis terhadap hasil yang telah didapatkan. Hasil evaluasi pengujian ini menggunakan akurasi berdasarkan nilai Fscore yang merupakan rata rata dari precission recall fitur dan polaritas opini dari evaluasi terhadap hasil ekstraksi fitur dan opini

4.1.1 Analisis Ekstraksi Fitur

Evaluasi ekstraksi fitur di lakukan dengan menghitung precission recall hasil ekstraksi dengan hasil dataset asli. Evaluasi ini sekaligus berguna untuk mendapatkan fitur yang relevan dari proses ekstraksi. Berikut adalah hasil perhitungan dari ekstraksi fitur :



Gambar 4-1 Fscore feature Extraction

Dari gambar diatas dapat dilihat bahwa rata rata precission dan recall hasil ekstraksi fitur dengan menggunakan tahapan preprocess *stopword* dengan *lemma* lebih besar di banding dua tahapan preprocess yang lain serta hasil ekstraksi fitur dengan tahapan *preprocess stopword* dan *stemming* paling kecil akurasinya di banding yang lain. Hal ini menunjukan normalisasi kata pada *Lemma* lebih baik di banding dengan *stemming*. Hasil *Precission* juga di pengaruhi perhitungan dari *precission* yaitu *TP / (TP+FP)* . *TP* adalah jumlah kandidat fitur yang teridentifikasi benar dan relevan antara hasil ekstraksi dengan label dataset asli, sementara *TP+FP* adalah jumlah kata yang berhasil terekstrak oleh pola pengetahuan *pattern rule*. semakin banyak hasil yang terekstrak *TP+FP* sementara kandidat fitur yang benar sedikit *TP* maka hasil *precission recall* akan semakin kecil , begitupun sebaliknya.

kemudian untuk *recall* ekstraksi fitur dengan menggunakan tahapan *preprocess stopword* dengan lemma terbukti lebih baik dari dua tahap yang lain seperti pada perhitungan *precission*. Tinggi nya hasil perhitungan ekstraksi fitur pada tahapan *stopword* dengan *lemma* menunjukan bahwa normalisasi dengan *lemma* lebih baik karena pemrosesan bahasa pada *lemma* dengan mengembalikan ke bentuk dasar (*root*) harus sesuai dengan tata yang ada pada kamus bahasa, sementara *stemming* mengembalikan suatu kata ke bentuk dasar dengan tidak memperhatikan atau tidak perlu sesuai dengan tata yang ada pada kamus bahasa. Contoh penggunaan *stemming* dan *lemma* :

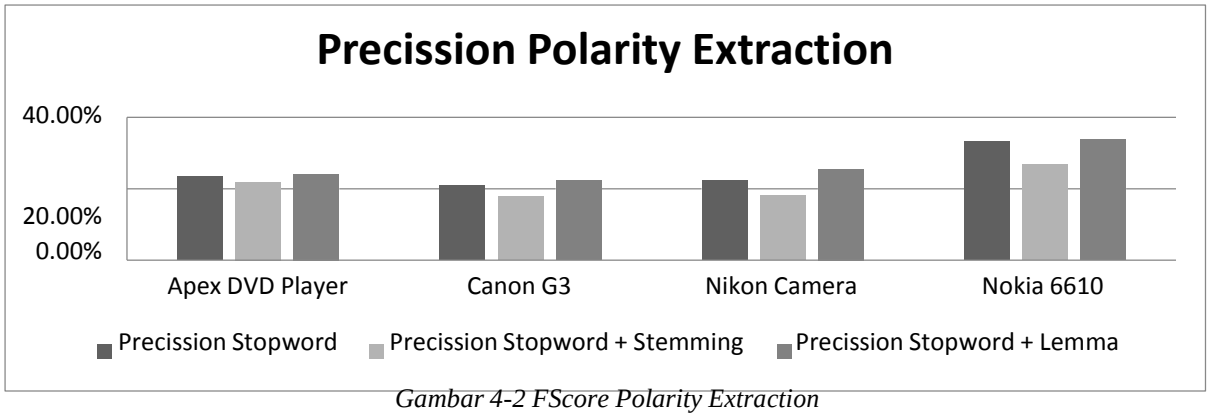
Table 4-0-1 Contoh Stemming dan Lemma

Kata	Stemming	Lemma
Studying ,studied	stud	study
see, saw	s	see

Normalisasi pada *Lemma* memperhatikan kamus bahasa sangat sesuai dan lebih banyak menghasilkan kata yang lebih relevan bagi hasil *preprocess* yang diharapkan sehingga menghasilkan evaluasi perhitungan yang lebih tinggi di bandingkan dengan dua tahapan lainnya.

4.1.2 Analisis Ekstraksi Polaritas

Evaluasi ekstraksi polaritas dilakukan dengan menggunakan perhitungan precission recall antara polaritas hasil frasa dengan menggunakan metode PMI-IR dengan polaritas dataset yang sudah tertera pada penelitian sebelumnya. Evaluasi ini berguna untuk mendapatkan polaritas ektraksi fitur dan opini yang relevan. Berikut adalah hasil evaluasi nilai Fscore yang merupakan rata rata dari nilai *precission recall* ekstraksi polaritas :



Gambar 4-2 FScore Polarity Extraction

Hasil diatas adalah evaluasi nilai rata rata *precission* dan *recall* dari ekstraksi polaritas. Evaluasi polaritas ini merupakan kelanjutan dari evaluasi fitur pada frasa hasil perhitungan PMI-IR dengan hasil polaritas dari dataset sebelumnya. Pada ekstraksi polaritas, fitur yang terambil dari hasil evaluasi fitur sebelumnya akan di evaluasi polaritas frasa fitur dan opini yang dihasilkan. Sama seperti ekstraksi fitur sebelumnya , tahapan dengan preprocess stopword dan lemma masih lebih tinggi dari dua perhitungan yang lain. Perhitungan polaritas dengan menggunakan PMI-IR bisa berubah ubah tergantung dari jumlah *hits* yang dihasilkan pada setiap pencarian kedekatan dengan *near operator*, Karena pada pencarian dengan menggunakan sistem yang online akan mengalami proses *update* seiring dengan perkembangan zaman dan perkembangan opini terhadap fitur dalam dunia maya. Hal ini mempengaruhi hasil dari evaluasi polaritas. Sebagai contoh *precission recall* ekstraksi polaritas pada tanggal 6 Juni 2015 berbeda dengan hasil pada tanggal 10 Juni 2015.

Table 4-0-2 Perbedaan Ekstraksi Polaritas Nikon Coolpix

Polarity June 6th 2015		Polarity June 10th 2015	
Precission	Recall	Precission	Recall
18.56%	22.35%	19.19%	23.26%
15.61%	18.06%	15.59%	18.11%
20.01%	23.89%	20.92%	25.52%

4.2 Summary Generation

Hasil yang di dapat dari penelitian *opinion extraction* dengan menggunakan pendekatan *pattern rule* adalah hasil ektraksi fitur dan ekstraksi polaritas yang sebelumnya telah melalui evaluasi *precission recall*, hasil tersebut kemudian digabungkan dengan kalimat sebenarnya untuk mendapatkan informasi yang relevan. Hasil dari penelitian tugas akhir ini adalah sebagai berikut :

Table 4-0-3 Summary Nikon Coolpix

Feature and Opinion Extraction	Feature and Opinion Dataset	Feature and Opinion Extraction Sentence
battery life[+]	battery life[+]	battery life[+],##battery life is excellent .
great price [+]	price[+2]	price[+],##great price for all the features .
Compact control[+]	size[+],control[+]	control[+],##it is very compact but the controls are so well designed that they 're still easy to use .

Hasil diatas merupakan salah satu contoh hasil dari ekstraksi fitur dan polaritas opini dari hasil pendekatan *pattern rule* dengan tahapan *stopword* dengan *lemma*, dimana *Feature* and *Opinion Extraction* merupakan hasil ekstraksi kandidat fitur dan opini dengan menggunakan pendekatan *pattern rule* dan PMI-IR sementara *Feature* and Opinion *Dataset* adalah hasil dari label fitur dan polaritas opini dari dataset yang sudah ada sebelumnya. Feature and *Opinion Extraction sentence* adalah penyatuan dari fitur dan polaritas opini dari hasil ekstraksi dengan kalimat sebenarnya. Selanjutnya proses summary adalah dengan mengklasifikasikan fitur dan opini beserta kalimat ke dalam orientasi positif dan negatif untuk memudahkan proses review kalimat, berikut adalah contoh hasil klasifikasi orientasi kalimat positif dan negatif :

Table 4-4 Contoh Hasil Klasifikasi kalimat

Kalimat Positif	Kalimat Negatif
camera[+]##this is a wonderful camera .	auto mode[-]##. this camera keeps on autofocussing in auto mode with a buzzing sound which can 't be stopped .
battery life[+]##battery life is ok .	customer service[-]##it is a good camera in terms of the function and quality , but take your chance with it because nikon absolutely sucks when it comes to customer service .
macro[+]##they take excellent macro shots as well .	lcd[-]##the only things i have found that i havent liked is that the lcd is hard to read in daylight but everyone elses is too .
price[+]##great price for all the features .	viewfinder[-]##2 the camera is so small that when you attach some lenses i have the 19mm wide-angle -- wc-e 68 , the optical viewfinder is partially obscured .

4.3 Grouping Feature

Grouping merupakan metode peringkasan hasil ekstraksi dengan cara mencari sinonim dari fitur, kemudian mengelompokan fitur tersebut ke daam satu kelas yang telah ditentukan sebelumnya untuk membantu mempermudah *review* hasil ekstraksi. Hasil dari grouping fitur adalah seperti contoh pada *grouping* hasil fitur pada *Canon Coolpix* berikut ini：

Input Class : Picture
1. [macro mode, picture]##the macro mode is exceptional , the pictures are very clear and you can take the pictures with the lens unbelievably close the subject . 2.[picture, delay]##otherwise , it takes very good pictures ; shutter delay is n't so bad either . 3.[picture]##after nearly 800 pictures i have found that this nikon takes incredible pictures . 4.[picture]##one neat thing - i have taken some pictures in what i thought would be impossible lighting conditions pitch black rooms - no problem for the camera - rooms looked like they had ample lighting . 5.[camera, picture]##i highly recommend this camera to anyone looking for a good digital camera that takes great pictures yet does n't take weeks to figure out how to operate . 6.[picture]##it takes great pictures . 7.[picture quality, movie]##the picture quality is amazing and you can connect it to your tv and could make silent movies that way if you wanted to . 8.[image]##this camera was affordable , very easy to learn , and produces spectacular images .

Gambar 4-3 Grouping feature Extraction Canon Coolpix

Pada gambar diatas merupakan contoh dari hasil *Grouping* fitur. Fitur di dalam tanda kurung siku “[ ]” merupakan sinonim dari kelas *picture* yang di inputkan sebelumnya.

5 Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan dari hasil proses implementasi, pengujian, dan analisis maka dapat ditarik kesimpulan berikut.

1. Pendekatan *Pattern Rule* dapat diterapkan pada proses ekstraksi fitur dan polaritas opini.
2. *Pointwise Mutual Information (PMI-IR)* dalam Proses ekstraksi polaritas dari suatu frasa fitur dan opini dapat diterapkan dengan cara menghitung *hits* asosiasi antara frasa dengan kata yang di jadikan polaritas kata positif dan negatif yaitu *excellent* dan *poor*
3. Tahapan normalisasi kata *preprocess* dengan *stopword* dan lemma lebih baik untuk menghasilkan fitur dan polaritas opini yang relevan di bandingkan dengan tahapan dengan hanya proses *stopword* dan tahapan dengan proses *stopword* dan *stemming*. Process stopword dan lemma pada nokia 6610 menjadi dataset dengan nilai fscore atau rata rata precission recall tertinggi yaitu 35.40% untuk rata rata precission recall ekstraksi fitur dan 31.39% untuk ekstraksi polaritas opini.
4. Normalisasi kata dengan Lemma di nilai lebih baik karena kata yang mengalami proses normalisasi disesuaikan dengan kamus besar bahasa.
5. Klasifikasi polaritas fitur dan opini dalam kalimat dapat dilakukan dengan cara memisahkan kalimat yang mengandur fitur positif dan kalimat yang mengandung fitur dengan polaritas negatif

5.2 Saran

Saran yang dapat diajukan untuk penelitian lebih lanjut mengenai topik ini adalah:

1. Pengembangan pola pengetahuan dengan *pattern rule* dapat di lakukan dengan menambahkan penelitian pencarian pattern yang baru sebai pola ekstraksi.
2. Proses ekstraksi yang dilengkapi dengan pendeteksi kata ganti fitur/*correference* dalam kalimat
3. Mengganti *API search* dengan *NEAR operator* dengan transaksi *searching* yang tidak terbatas memungkinkan dalam mengekstrak kalimat dengan jumlah ekstraksi pattern fitur dan opini yang lebih banyak
4. Pembersihan data dengan *preprocess* tambahan dibutuhkan untuk mendapatkan hasil yang relevan.

Daftar Pustaka :

[1] M. Hu and B. Liu, "Mining Opinion Feature in Customer Reviews," p. 6.

[2] bharati and Remegeri, "Data Mining techniques and Application," p. 1, 2006.

[3] D. Osimo and F. Mureddu, "Research Challange on Opinion Mining and Sentiment Analysis," p. 2.

[5] L. Dybekjaer, H. Hemsen and W. Minker, "Evaluation of Text and Speech System," p. 99, 2007.

[7] Pyle, Loshin and Redman, "Data Mining," p. 47, 2014.

[8] A. Gelbukh and E. Morales, "Advances in Artificial Intelligence: 7th Mexican International Conference," *MICAI 2008*, p. 125, 2008.

[9] M. Bramer, "Principles of Data Mining," p. 242, 2007.

[10] J. Plison, N. Lavrac and D. Mladenic, "A Rule based Approach to Word Lemmatization," p. 1.

[11] D. P. Turney, "Thumbs Up or Thumbs Down ? Semantic Orientation Applieed to Unsupervised Classification Classification of Reviews," p. 2, 2002.

[12] A. G. Miller, R. Beckwith, C. Fellbaum, D. Gross and k. Miller, "Introduction to WordNet: An On-line Lexical Database," p. 2, 1993.