

## Analisis Opini Terhadap Fitur *Smartphone* Pada Ulasan *Website* Berbahasa Indonesia

Doni Setyawan\*<sup>1</sup>, Edi Winarko<sup>2</sup>

<sup>1</sup>Program Studi S2 Ilmu Komputer FMIPA UGM

<sup>2</sup>Departemen Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: \*[donnisoft@gmail.com](mailto:donnisoft@gmail.com), [ewinarko@ugm.ac.id](mailto:ewinarko@ugm.ac.id)

### Abstrak

Melalui toko online konsumen dapat memberikan opini mengenai suatu produk yang dijual, salah satu produk laris yang diminati adalah *smartphone*. Opini-opini konsumen ini ternyata merupakan sumber data yang bisa dimanfaatkan untuk mengetahui kelebihan dan kekurangan suatu produk berdasarkan pengalaman sebenarnya dari pengguna, oleh karena itu agar dapat mendayagunakan sumber data ini maka diperlukan sistem yang otomatis melakukan penambangan dan peringkasan opini terhadap produk *smartphone* pada ulasan opini konsumen. Sistem yang dibangun terdiri lima bagian besar yaitu pengumpulan data, preprocessing review, penambangan fitur, analisis opini, dan visualisasi hasil. Pengumpulan data yaitu mengambil data review dari website target dengan melakukan web scraping, preprocessing yaitu melakukan pembersihan data review. Tahapan penambangan fitur akan menemukan fitur yang terdapat pada kalimat review dengan algoritma apriori untuk menghasilkan frequent itemset, kemudian analisis opininya dengan pendekatan lexicon based, rule bahasa dan score function. Tahapan terakhir menampilkan hasil dalam bentuk grafik. Dari hasil pengujian penambangan fitur diperoleh rata-rata nilai recall adalah 0,63 dan precision 0,72. Hal ini disebabkan ukuran dan kualitas bagus tidaknya review sangatlah berpengaruh. Hasil pengujian akurasi analisis opini menunjukkan nilai yang tinggi yaitu rata-rata akurasinya 81,76%. Teknik yang digunakan terbukti memberikan hasil yang baik yaitu penggunaan kata opini yang sudah terlabeli orientasi opininya, penggunaan rule bahasa dan penerapan score function.

**Kata kunci**— *smartphone*, review, frequent itemset, rule bahasa, analisis opini

### Abstract

Through online stores, consumers can give an opinion of a product, one of the best-selling products is *smartphone*. Their opinions become valuable and can be worthwhile to know the advantages or disadvantages of products based on the user's experience. Therefore, in order to utilize the data of customers' opinions, it is necessary to create a system that automatically performs mining and summarizing opinions on *smartphone* product. The system consist of five parts: data collection, preprocessing review, feature mining, analysis of opinions and then visualize the results. Data collection is taking data reviews website using web scraping, preprocessing review is for cleaning data reviews. Feature mining stage will find features in the reviews with apriori algorithm to produce frequent item set, then analyze the opinion using lexicon based, language rule and score function. The result will be shown in graphical form. From the testing of feature mining obtained average recall score at 0.63 and precision at 0.72. It depends on good or bad quality of reviews. The results of testing accuracy opinion analysis shows high value with accuracy 81.76 %. The technique showed good results with opinion data which is labeled, using language rule and the implementation of score function.

**Keywords**— *smartphone*, review, frequent itemset, linguistic rule, opinion analysis

## 1. PENDAHULUAN

Jumlah pengguna internet di Indonesia pada tahun 2015 adalah 88 juta jiwa, dimana 85% nya mengakses internet menggunakan telepon seluler yaitu 74,8 juta jiwa dan 11 % melakukan aktivitas jual beli *online* yaitu 9,68 juta jiwa[1]. Produk yang dijual melalui toko *online* sangatlah beraneka ragam, salah satu produk yang laris diperjual belikan adalah *smartphone* yang mana di Indonesia jumlah penggunaannya pada akhir tahun 2015 diperkirakan mencapai 55 juta jiwa[2]. Angka ini sangat mungkin akan terus bertambah mengingat tren penggunaan *smartphone* dan jual beli *online* yang terus meningkat. Salah satu cara untuk meningkatkan jumlah dan kepercayaan pengguna yang mengakses toko *online* maka disediakan fasilitas *review* produk, dimana pembeli dapat memberikan opininya mengenai suatu produk dan layanan dari toko *online* tersebut. Opini-opini pembeli ini ternyata merupakan sebuah sumber data yang bisa dimanfaatkan untuk kepentingan tertentu seperti untuk mengetahui kelebihan dan kekurangan suatu produk berdasarkan pengalaman sebenarnya dari pengguna.

*Opinion mining* atau *sentiment analysis* adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Jika diberikan satu himpunan dokumen teks  $D$  yang berisi opini (atau sentimen) mengenai suatu objek, maka penambangan opini bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentari pada setiap dokumen  $d \in D$  dan untuk menentukan apakah komentar tersebut positif, negatif atau netral [3]. Terdapat beberapa teknik dalam melakukan penambangan opini untuk mengklasifikasi apakah opini tersebut positif, negatif atau netral diantaranya *naïve bayesian classification*, *support vector machine* dan *lexicon based*.

Beberapa penelitian sebelumnya menggunakan *naive bayes* dan seleksi fitur kata untuk mengklasifikasi teks berita dan abstrak akademis yang menghasilkan nilai *accuracy* 91% pada dokumen berita dan 82% pada dokumen abstrak[4]. Penelitian lain menggunakan metode *support vector machine* dan *semantic orientation* untuk menganalisis sentimen terhadap kalimat berbahasa Arab yang menghasilkan nilai *accuracy* 80%, *recall* 0.809, *precision* 0.805 dan *F-measure* 0.806[5]. Pendekatan yang lain adalah menggunakan *pattern knowledge* untuk melakukan ekstraksi fitur produk dan kata opini menggunakan *pattern knowledge* pada *review* konsumen berbahasa Inggris[6]. Penggunaan daftar kata opini atau *lexicon* berbahasa Indonesia dan *rule* bahasa juga dapat digunakan untuk menganalisis sentimen terhadap *movie review*[7]. Dalam penelitian ini digunakan metode *association mining*, *feature pruning*, *lexicon* berbahasa Indonesia, *score function* dan *rule* bahasa untuk menambang fitur *smartphone* pada ulasan berbahasa Indonesia dan menganalisis opini fitur apakah positif, negatif atau netral dan membandingkan nilai orientasi antar fitur pada *smartphone* yang berbeda merek.

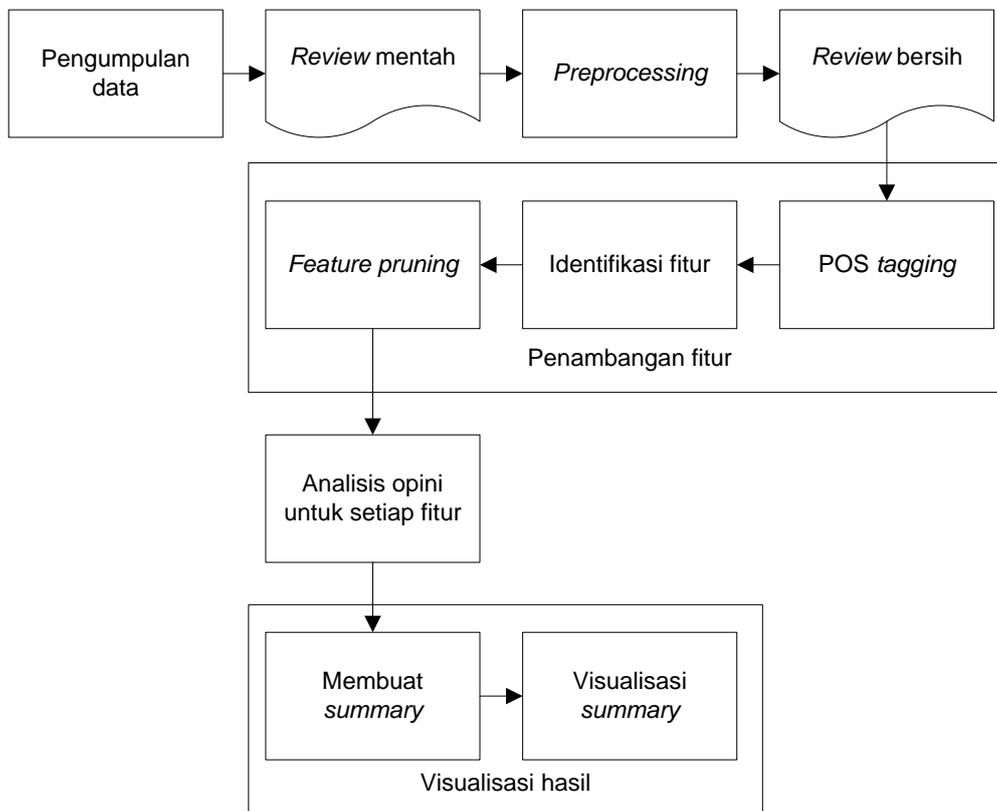
## 2. METODE PENELITIAN

### 2.1 Arsitektur Sistem

Sistem yang akan dibuat merupakan sistem yang dapat melakukan proses penambangan opini terhadap produk *smartphone* dengan cara menemukan fitur dari *smartphone* yang dikomentari oleh *user* pada *review* yang telah di-*crawling* dan di-*preprocessing* sebelumnya, kemudian menganalisis opini yang diekspresikan terhadap fitur tersebut apakah berorientasi positif, negatif atau netral. Tahapan terakhir adalah memvisualkan hasilnya dalam bentuk grafik yang mudah dipahami oleh pengguna sebagaimana tersaji pada Gambar 1.

Pada tahapan pertama sistem akan mengumpulkan data dari *website* target dengan cara melakukan *web scraping*, yaitu mengambil judul dan isi *review* yang menghasilkan *review* mentah. Tahapan selanjutnya adalah *preprocessing* terhadap *review* mentah dengan melakukan *case folding* yaitu mengubah semua huruf yang ada pada *review* menjadi huruf kecil, penggantian kata-kata tidak baku dan penyeragaman kata yang menjadi fitur, kemudian

menyimpan hasilnya ke dalam *review* bersih. Selanjutnya dilakukan *POS tagging* terhadap kata-kata di dalam kalimat, yaitu pemberian label jenis kelas kata (kata benda, kata sambung, kata sifat, dsb), kemudian tiap-tiap kalimat disimpan dalam tabel baru beserta label jenis katanya. Dari tiap-tiap kalimat beserta label jenis katanya akan diambil kata-kata yang berlabel kata benda saja sehingga menghasilkan himpunan kata benda untuk setiap kalimat (*itemset*). Himpunan kata benda inilah yang akan menjadi data transaksi untuk proses selanjutnya.



Gambar 1 Gambaran umum sistem[8]

Berikutnya akan dilakukan identifikasi fitur menggunakan algoritma apriori untuk menghasilkan *frequent itemset* dari sekumpulan himpunan kata benda tersebut. *Frequent itemset* yang dihasilkan disebut sebagai *candidate feature*. Dari *candidate feature* akan dilakukan *feature pruning*, yaitu menghapus kata-kata yang dianggap bukan fitur dari *smartphone*, hasil akhir setelah dilakukan *feature pruning* itulah yang akan dijadikan sebagai fitur. Tahapan berikutnya adalah menganalisis orientasi opini pada setiap fitur apakah positif, negatif atau netral. Teknik yang digunakan adalah dengan *lexicon based* yang berupa daftar kata opini yang sudah disesuaikan untuk domain *smartphone*, serangkaian *rule* bahasa yang telah didefinisikan, antara lain *but clause rule* dimana apabila dalam suatu kalimat ditemukan kata "tetapi" maka orientasi opini akan dihitung tersendiri pada klausa kalimat yang memuat kata "tetapi" tersebut, *negation rule* yang membuat orientasi kata opini menjadi kebalikannya apabila di sekitar kata opini tersebut ditemukan kata-kata negasi. Penerapan *score function* yang mempertimbangkan jarak antara kata opini dengan fitur yang terdapat dalam sebuah kalimat dengan mengabaikan kata "tidak" dalam perhitungan jaraknya.

Proses terakhir adalah membuat *summary* dan visualisasi hasil dengan penggambaran grafik dimana untuk tiap-tiap fiturnya akan ditampilkan grafik opini positif, negatif dan netral disertai dengan jumlah *review* yang memberikan opini tersebut. Informasi lain yang ditampilkan adalah kalimat-kalimat dari *review* yang memuat fitur tersebut dan nilai perbandingan antar fitur dari *smartphone* merek lain.

### 2.1.1 Perancangan pengumpulan data

Data utama yang dibutuhkan dalam sistem ini adalah *review* pengguna mengenai suatu produk *smartphone*. *Review* ini diperoleh dengan cara melakukan *web scraping* pada *website* target, yaitu mengekstrak judul dan isi *review* dari halaman yang diunduh.

### 2.1.2 Preprocessing review

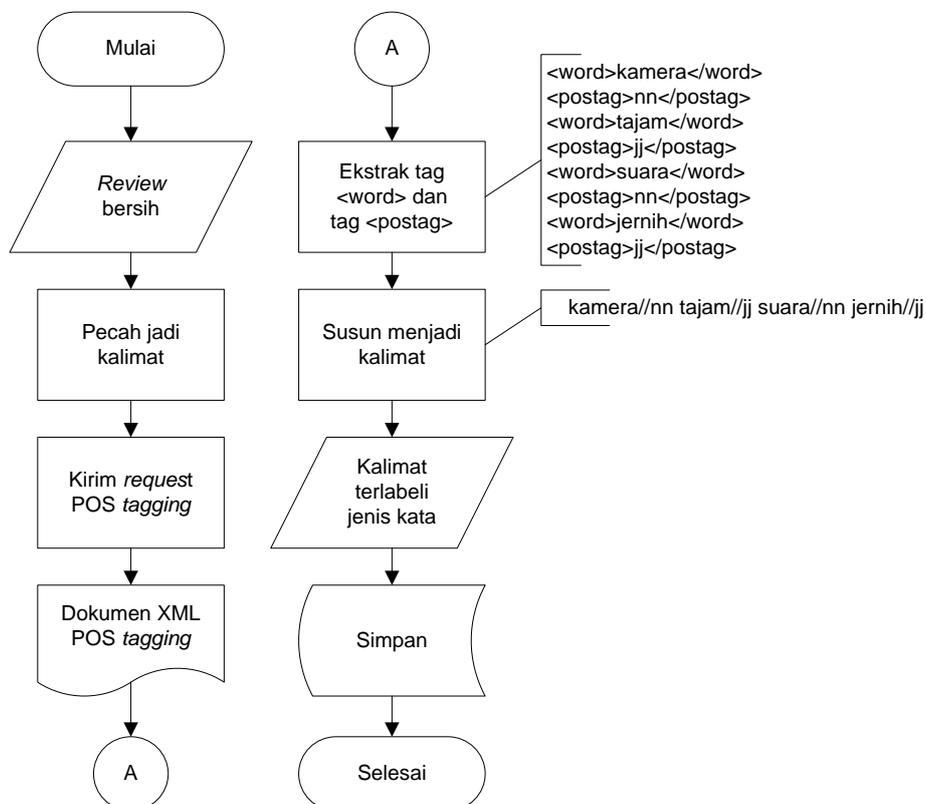
*Review* yang sudah diambil dari *website* target masih berupa data mentah, kemudian dilakukan *preprocessing* dengan melakukan *case folding*, penggantian kata-kata tidak baku dan penyeragaman kata yang menjadi fitur. *Case folding* yaitu mengubah semua huruf yang ada pada isi *review* menjadi huruf kecil yang bertujuan untuk menormalisasi ukuran teks sehingga dapat diimplementasikan pada operasi-operasi teks. Proses selanjutnya adalah penggantian kata prokem (kata tidak baku), misalnya kata 'jos' dan 'mantaf' diganti menjadi 'mantap', contoh lainnya adalah kata-kata yang penggunaannya sering disingkat seperti 'tdk' diganti menjadi 'tidak'. Tahapan yang terakhir adalah penyeragaman kata yang menjadi fitur dari *smartphone* menjadi sebuah kata saja, misalnya kata 'baterai' di dalam *review* sering ditulis dengan bentuk 'batt', 'batere', 'btre' dan 'bateray'.

### 2.1.3 Penambahan fitur

Pada tahapan ini sistem akan melakukan penambahan fitur suatu produk *smartphone* dari sekumpulan *review* dalam data bersih sebelumnya. Terdapat 3 tahapan utama dalam proses penambahan fitur yaitu POS *tagging*, identifikasi fitur, dan *feature pruning*.

#### 2.1.3.1 POS Tagging

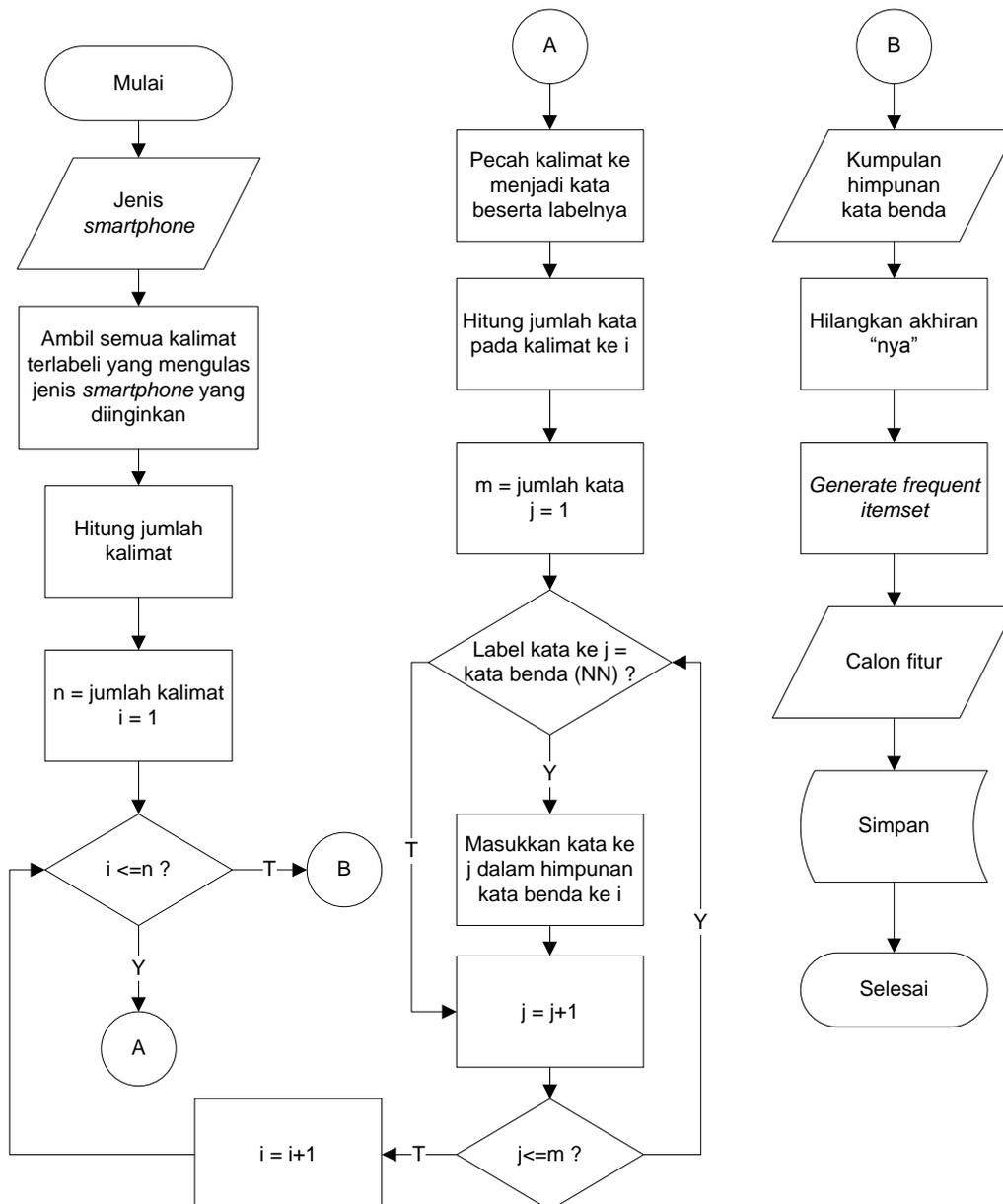
Tahapan POS *tagging* akan memberi label jenis kata pada setiap kata di dalam kalimat apakah termasuk kata benda, kata kerja, kata sifat, kata keterangan, dsb. Proses POS *tagging* disajikan pada Gambar 2.



Gambar 2 Flow chart POS tagging

### 2. 1.3.2 Identifikasi fitur

Untuk melakukan identifikasi fitur diperlukan beberapa tahapan sebagaimana yang tersaji pada Gambar 3. Pertama-tama seluruh *review* dari suatu produk *smartphone* yang telah melalui tahapan *POS tagging* sebelumnya, untuk setiap kalimatnya diambil kata yang labelnya kata benda saja, sehingga setiap kalimat akan menghasilkan himpunan kata benda. Kumpulan dari seluruh himpunan kata benda ini nantinya akan menjadi data awal untuk proses identifikasi fitur. Dari sekumpulan himpunan kata benda ini dilakukan penghapusan imbuhan 'nya' pada akhir kata yang bertujuan untuk mencari bentuk kata dasar dari suatu kata, misalnya kata "kameranya" menjadi "kamera".



Gambar 3 Flow chart identifikasi fitur

### 2. 1.3.2 Feature pruning

*Feature pruning* adalah proses membuang kata-kata dalam calon fitur (*candidate feature*) yang diperkirakan bukan fitur. Teknik yang digunakan dalam *feature pruning* adalah

dengan menggunakan daftar *stop word* yang dibuat dengan menggunakan sampel data dari beberapa *review*.

#### 2. 1.4 Analisis opini untuk tiap fitur

Pada tahapan ini akan dilakukan analisis opini untuk setiap fitur yang terpilih dari hasil proses penambangan fitur. Analisis opini akan menentukan apakah opini tersebut berorientasi positif, negatif, atau netral. Untuk melakukan analisis opini menggunakan bantuan daftar *lexicon word* yaitu kata-kata sifat dan kerja yang telah diberi label positif dan negatif, beberapa *rule* bahasa yang telah didefinisikan yaitu *but clause rule*, *negation rule* dan penerapan *score function* yang mempertimbangkan jarak antara kata opini dengan fitur yang terdapat dalam sebuah kalimat, dimana jarak yang dekat berbanding lurus dengan besar dari skor nilainya. Cara perhitungannya *score function* adalah sebagai berikut, untuk setiap kalimat yang memuat satu atau lebih fitur, kata-kata yang berlabel JJ (kata sifat) dan VB (kata kerja) akan diambil, kata sifat dan kata kerja ini akan menjadi kata opini (*opinion word*). Selanjutnya untuk setiap fitur  $f$  yang teridentifikasi di dalam kalimat tersebut akan dihitung skor orientasinya dengan melakukan pengecekan kata-kata opini dalam kalimat tersebut terhadap database kata opini (*opinion word list*) yang telah diberi label orientasi positif dan negatif. Kata opini berorientasi positif akan diberi skor orientasi +1, sedangkan yang berorientasi negatif akan diberi skor orientasi -1. Kemudian keseluruhan skor orientasi dari kata-kata sifat dan kerja dalam kalimat tersebut dijumlahkan menggunakan *score function* seperti yang disajikan pada persamaan (1) [9]:

$$score(f) = \sum_{w_i: w_i \in S \wedge w_i \in V} \frac{w_i \cdot SO}{dis(w_i, f)} \quad (1)$$

$w_i$  adalah kata opini yang diambil dari kalimat,  $V$  adalah *database* kata opini yang telah diberi orientasi positif dan negatif,  $S$  adalah kalimat yang di dalamnya terdapat fitur  $f$ , dan  $dis(w_i, f)$  adalah jarak antara opini  $w_i$  dan fitur  $f$  di dalam kalimat  $S$ , sedangkan  $w_i \cdot SO$  adalah nilai orientasi awal  $w_i$  yang terdapat pada  $V$  yaitu 1 untuk positif, -1 untuk negatif dan 0 untuk netral. Kedekatan jarak antara  $w_i$  dan  $f$  akan menghasilkan bobot yang besar, sedangkan jarak yang jauh akan memberikan bobot yang semakin kecil. Hal ini adalah sesuatu yang logis pada sebuah kalimat, dimana suatu kata opini yang letaknya berdekatan dengan fitur maka kata tersebut memang memberikan opini untuk fitur tersebut. Untuk kata opini yang jaraknya jauh dengan fitur maka kata tersebut dimungkinkan tidak memberikan opini untuk fitur tersebut. Untuk algoritma analisis opini secara keseluruhan disajikan pada Gambar 4.

#### 2. 1.5 Visualisasi hasil analisa

Tahapan selanjutnya setelah analisis opini fitur terhadap keseluruhan *review* yang ada pada *database*, hasilnya akan divisualisasikan dalam bentuk grafik yang menggambarkan jumlah orientasi opini positif, negatif, dan netral dari fitur *smartphone* dan juga akan ditampilkan *link* ke kalimat yang memuat opini tersebut.

#### 2.2 Implementasi

Spesifikasi *hardware* dan *software* yang digunakan untuk merancang, membuat dan menjalankan sistem ini menggunakan *processor Intel Core 2 Duo @ 2 GHz*, memori 4 GB dengan sistem operasi *Windows 7 Ultimate 32-bit*. Untuk manajemen *database* menggunakan RDBMS MySQL 5.6.21 dan *webserver* menggunakan Apache 2.4.10 (Win32). Pengembangan bagian *backend* menggunakan bahasa pemrograman PHP versi 5.6.3, JavaScript, *library Sastrawi Sentence Detector* untuk memecah *review* menjadi kalimat, *library Apriori-Master* untuk menghasilkan *frequent itemset* dan pemanfaatan *web service* dengan alamat <http://fws.cs.ui.ac.id:80/RESTFulWSSstanfordPOSTagger/POSTagger> dalam proses POS

*tagging*. Sedangkan bagian *frontend* visualisasi selain PHP juga menggunakan *JQuery* 1.10.1, *fancybox* 2.1.5 dan *chart.js* untuk menampilkan grafik orientasi opini. Implementasi visualisasi opini tiap fitur disajikan dalam bentuk *bar chart* sebagaimana ditampilkan pada Gambar 5.

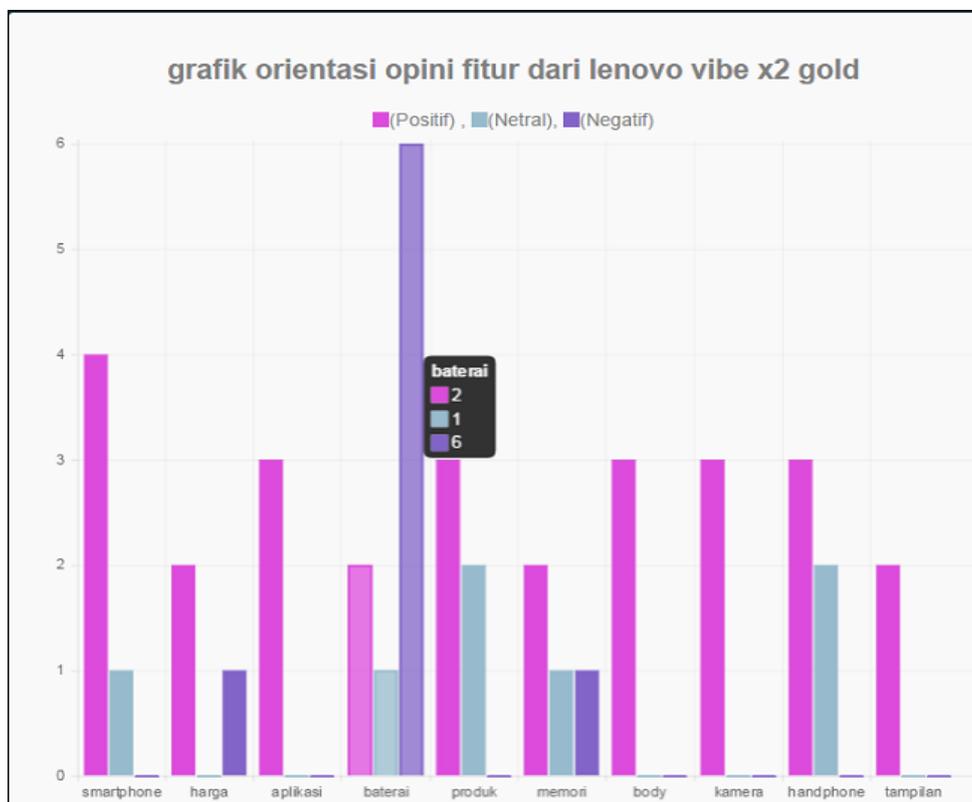
```

Algorithm Opinion Orientation( )
for each detected feature  $f_i$  of a smartphone product do
  for each sentence  $S_j$  in the review database do
    if  $f_i$  appear in  $S_j$  do
      orientation = 0
      if feature  $f_i$  is in the "but" clause then
        orientation = apply the "but" clause rule
      else remove "but" clause from  $S_j$  if it exist
      for each opinion word  $ow$  in  $S_j$  do
        orientation += wordOrientation( $ow, f_i, S_j$ )
      end for
    end if
  else // there's no feature  $f_i$  in  $S_j$ 
    continue
  end if

  if orientation > 0 then
     $f_i$  orientation in  $S_j = 1$  (positif)
  else if orientation < 0 then
     $f_i$  orientation in  $S_j = -1$  (negatif)
  else
     $f_i$  orientation in  $S_j = 0$  (neutral)
  end if
end for
end for

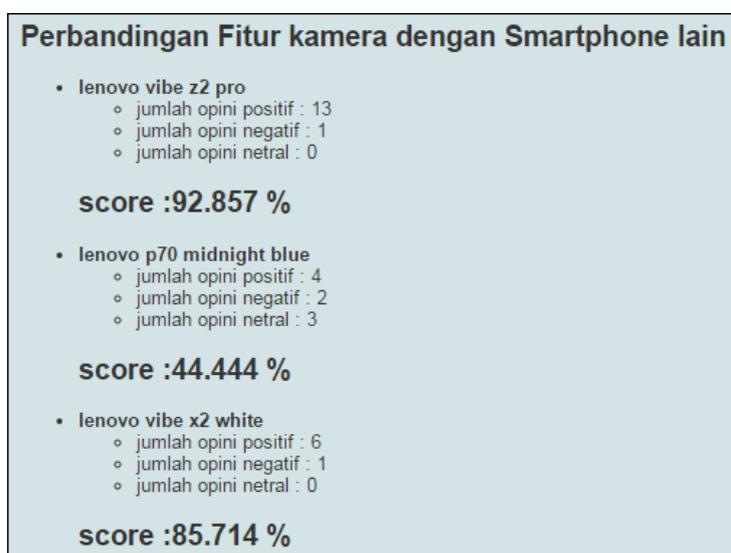
```

Gambar 4 Algoritma analisis opini



Gambar 5 Bar chart opini tiap fitur

Gambar 6 menyajikan nilai perbandingan fitur kamera antar merek *smartphone* yang berbeda, dengan begitu konsumen dapat mengetahui yang terbaik diantara beberapa merek *smartphone*.



Gambar 6 Nilai perbandingan fitur antar merek *smartphone*

### 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini digunakan *review* dari *website* [www.bhinneka.com](http://www.bhinneka.com) untuk melakukan pengujian terhadap sistem yang meliputi pengujian *running time preprocessing*, penambahan fitur dan analisis opini.

#### 3.1 Hasil pengujian *running time*

Pada pengujian ini setiap produk akan dihitung berapa detik waktu yang diperlukan dalam proses *preprocessing*. Tujuan dari pengujian ini adalah untuk mengetahui rata-rata waktu *preprocessing* sebuah *review*. Tabel 1 menyajikan hasil *running time preprocessing*.

Tabel 1 *Running time preprocessing*

No	<i>Smartphone</i>	Jumlah <i>review</i>	<i>Preprocessing</i> (detik)	Rata-rata <i>preprocessing</i> tiap <i>review</i>
1	Lenovo Vibe Z2 Pro	30	0,928	0,031
2	Lenovo P70 - Midnight Blue	25	0,914	0,036
3	Lenovo Vibe X2 Gold	21	0,622	0,029
4	Lenovo Vibe X2 White	15	0,678	0,042
5	Lenovo A6000 Plus 16GB	8	0,27	0,034
6	Xiaomi Mi 4i LTE – White	10	0,271	0,027
7	Himax pure s black caviar	14	0,498	0,036
8	Samsung Galaxy Note 5 - Gold	4	0,152	0,038
9	ADVAN Star Note S5L Coffee	3	0,135	0,045
10	SMARTFREN Andromax T	5	0,186	0,037

Dari Tabel 1 dapat dilihat bahwa rata-rata waktu *preprocessing* untuk sebuah *review* bervariasi, hal ini karena panjang pendeknya sebuah *review* tidaklah sama antara satu dengan yang lain. Rata-rata waktu *preprocessing* sebuah *review* dapat disimpulkan termasuk cepat yaitu masih dibawah angka 0,05 detik.

### 3.1 Hasil pengujian penambangan fitur

Pada pengujian ini akan mengukur kinerja penambangan fitur yang dilakukan oleh sistem terhadap 5 produk *smartphone* menggunakan *precision* dan *recall* sebagaimana disajikan pada Tabel 2. *Precision* diperoleh dengan membagi antara jumlah fitur benar yang dihasilkan oleh sistem dengan jumlah fitur yang dihasilkan oleh sistem baik benar maupun salah. *Recall* diperoleh dengan membagi antara jumlah fitur benar yang dihasilkan oleh sistem dengan jumlah fitur yang terdapat dalam koleksi *review*.

Tabel 2 *Recall* dan *precision* penambangan fitur *smartphone*

No	Smartphone	Jumlah review	Jumlah fitur di dalam review (manual)	Jumlah fitur dihasilkan sistem		Recall	Precision
				Total	Benar		
1	BOLT 4G Powerphone	11	8	7	5	0,625	0,71
2	SAMSUNG Galaxy Grand	9	12	11	8	0,67	0,72
3	APPLE iPhone 5S 16GB Gold	6	9	7	5	0,55	0,71
4	MICROSOFT Lumia 430 Orange	10	22	22	16	0,73	0,73
5	NOKIA Lumia 520 Black	8	12	11	7	0,58	0,64

### 3.1 Hasil pengujian analisis opini fitur

Pengujian ini akan menguji akurasi hasil analisis opini dari masing-masing fitur pada *smartphone* dengan *confusion matrix*, yaitu matrik yang berisi hasil klasifikasi data yang terprediksi benar dan salah sebagaimana yang tersaji pada Tabel 3 dan rumus akurasinya pada persamaan 2 [10].

Tabel 3 *Confusion matrix*

		Predicted	
		Negative	Positive
Actual	Negative	A	B
	Positive	C	D

$$Accuracy = \frac{a+d}{a+b+c+d} \quad (2)$$

Dari Tabel 3 nilai a menyatakan jumlah prediksi negatif yang sebenarnya negatif, b menyatakan jumlah prediksi positif yang sebenarnya negatif, c menyatakan jumlah prediksi negatif yang sebenarnya positif dan d menyatakan jumlah prediksi positif yang sebenarnya positif. Tabel 4 menyajikan hasil akurasi analisis orientasi opini pada setiap fitur dari 5 buah produk *smartphone*. Hasil akurasi cukup tinggi yaitu rata-ratanya 81,76 % sehingga teknik yang digunakan untuk melakukan orientasi opini sangat menjanjikan yaitu penggunaan kata opini yang sudah diberi label orientasinya, *but clause rule* yang akan mengecek kalimat apakah mengandung kata “tapi” atau “tetapi”, *negation rule* mengecek apakah sebelum fitur terdapat kata negasi dan *score function* yang mempertimbangkan jarak antara kata opini dan fitur di dalam sebuah kalimat.

Tabel 4 Akurasi analisis opini fitur

Smartphone	Jumlah prediksi (P) dan jumlah aktual (A)									Akurasi (%)
	P + A +	P + A -	P + A N	P - A +	P - A -	P - A N	P N A +	P N A -	P N A N	
BOLT 4G Powerphone IVO	14	1	0	1	0	0	1	0	1	83,33
SAMSUNG Galaxy Grand I9082 White	24	2	0	0	2	1	2	1	2	82,35
APPLE iPhone 5S 16GB Gold - White	10	0	2	2	0	0	0	0	4	77,78
MICROSOFT Lumia 430 Orange	55	1	1	2	4	0	9	0	8	83,75
NOKIA Lumia 520 Black	23	0	0	2	5	0	4	1	3	81,58

Keterangan Tabel 4:

- P +, A + = Jumlah prediksi positif yang aktualnya positif
- P +, A - = Jumlah prediksi positif yang aktualnya negatif
- P +, A N = Jumlah prediksi positif yang aktualnya netral
- P -, A + = Jumlah prediksi negatif yang aktualnya positif
- P -, A - = Jumlah prediksi negatif yang aktualnya negatif
- P -, A N = Jumlah prediksi negatif yang aktualnya netral
- P N, A + = Jumlah prediksi netral yang aktualnya positif
- P N, A - = Jumlah prediksi netral yang aktualnya negatif
- P N, A N = Jumlah prediksi netral yang aktualnya netral

#### 4. KESIMPULAN

Dari penelitian yang telah dilakukan terdapat beberapa kesimpulan yaitu :

1. Dari hasil pengujian *running time preprocessing* dapat disimpulkan bahwa semakin banyak jumlah *review* pada sebuah *smartphone*, maka semakin lama waktu yang diperlukan untuk *preprocessing*, tetapi untuk rata-rata waktu *preprocessing* sebuah *review* termasuk cepat yaitu 0.035 detik. Untuk *running time* analisis opini fitur *smartphone* sangat dipengaruhi dengan jumlah fitur dan jumlah kalimat yang termuat di dalam *review* untuk produk tersebut, yaitu rata-rata waktu tercepat adalah 2.444 detik dan terlama 5.989 detik.

2. Dari hasil pengujian penambahan fitur diperoleh rata-rata nilai *recall* adalah 0.63 dan *precision* 0.72. Hal ini disebabkan ukuran dan kualitas bagus tidaknya *review* sangat berpengaruh, *review* yang memuat rata-rata 7 buah kalimat dan pada setiap kalimatnya mengulas minimal 2 buah fitur akan memberikan hasil *recall* dan *precision* yang tinggi. Sebaliknya *review* yang didalamnya hanya berisi beberapa kalimat dan setiap kalimat rata-rata hanya mengulas sebuah fitur akan memberikan hasil *recall* dan *precision* yang rendah.
3. Hasil pengujian akurasi analisis opini pada setiap fiturnya menunjukkan angka yang bagus yaitu rata-rata akurasinya 81.76%. Teknik yang digunakan terbukti memberikan hasil yang baik yaitu penggunaan kata opini yang sudah terlabeli orientasi opininya, penggunaan *rule* bahasa dan penerapan *score function*.
4. Hasil pengujian menunjukkan penggunaan *score function* yang mempertimbangkan jarak kata opini dengan fitur pada suatu kalimat menghasilkan akurasi yang lebih tinggi daripada penggunaan *score function* yang tidak mempertimbangkan jarak kata opini dan fitur. Hal ini membuktikan bahwa keberadaan kata opini dalam suatu kalimat yang tidak berdekatan dengan fitur adalah tidak terkait dengan fitur tersebut.

## 5. SARAN

Pengembangan sistem ini masih memiliki keterbatasan yang dapat dijadikan referensi untuk penelitian selanjutnya, sehingga dapat disarankan beberapa hal yaitu :

1. Perlu penggabungan dan sistem seleksi *review* dari beberapa *website*, sehingga dapat diperoleh jumlah *review* yang lebih banyak dan lebih layak untuk digunakan sebagai data yang memungkinkan untuk menghasilkan nilai *recall* dan *precision* yang lebih baik dalam proses penambahan fitur.
2. Perlu perhitungan analisis opini fitur pada level kalimat yang lebih detail, karena dalam sebuah kalimat dapat saja memuat beberapa fitur.
3. Perlu dikembangkan korpus kata opini yang dapat digunakan untuk domain tertentu.

## DAFTAR PUSTAKA

- [1] Berliyanto, 2015, Profil Pengguna Internet di Indonesia Tahun 2015, <http://blog.idkeyword.com/profil-pengguna-internet-di-indonesia-tahun-2015/>, diakses 3 Desember 2015
- [2] Prabancono, 2015, Wow Pengguna Smartphone di Indonesia Capai 55 Juta Orang, <http://www.solopos.com/2015/09/20/pengguna-smartphone-wow-pengguna-smartphone-di-indonesia-capai-55-juta-orang-644446>, 20 September 2015, diakses 3 Desember 2015..
- [3] Liu, B., 2010, *Handbook of Natural Language Processing, chapter Sentiment Analysis and Analysis*, 2nd Edition.
- [4] Hamzah, A., 2012, *Klasifikasi Teks Dengan Naive Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita dan Akademis*, ISSN:1979-9111, vol. 3, 2012.
- [5] Shoukry, A. M., 2013, *Arabic sentence level sentiment analysis*, Dissertation, Department of Computer Science and Engineering, The American University in Cairo.
- [6] Htay, S. dan Lynn, K.T., *Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews*, The Scientific World Journal, 2013.

- 
- [7] Komansilan, E., 2012, *Penambangan Opini Pada Situs Review Film Berbahasa Indonesia*, Tesis, Jurusan Ilmu-ilmu Komputer FMIPA UGM, Yogyakarta.
- [8] Hu, M., dan Liu, B. 2004. Mining and Summarizing Customer Reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Seattle, Washington, USA, Agustus
- [9] Ding, X., Liu, B. dan Yu, P. 2008. A holistic lexicon-based approach to opinion mining. *In Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008)*.
- [10] Kohavi, R., dan Provost, F., 1998, *On Applied Research in Machine Learning*. In *Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process*, Columbia University, New York, Volume 30.