

Ekstraksi Opini Berbasis Fitur untuk *Review* Produk Berbahasa Indonesia dengan Algoritma FP-Growth

Feature-based Opinion Extraction for Indonesian Product Reviews using FP-Growth Algorithm

Tetya Arum Dhahyani¹
(tetyaadhahyani@gmail.com)

Eko Darwiyanto²
(ekodarwiyanto@telkomuniversity.ac.id)

Arie Ardiyanti Suryani³
(ardiyanti@telkomuniversity.ac.id)

School of Computing - Universitas Telkom
Jl. Telekomunikasi, Dayeuh Kolot Bandung 40257 Indonesia

ABSTRAK

Seiring dengan berkembangnya teknologi dalam dunia Internet, *e-commerce* menjadi hal yang tidak asing lagi. Salah satu bentuk pemanfaatan *e-commerce* adalah penggunaan *web site* sebagai sarana jual beli. Beberapa *web site* yang memungkinkan pelanggannya untuk berbelanja secara *online* juga memberikan keleluasaan kepada pelanggannya untuk turut aktif menuliskan *review* tentang produk yang dibelinya. Konsekuensinya adalah peningkatan jumlah *review* produk. Setiap hari bisa saja ratusan *review* baru muncul. Hal ini akan berpotensi membuat pelanggan mengalami kesulitan untuk membaca *review* yang dapat membantunya untuk mengambil keputusan membeli produk. Untuk itu, diperlukan sebuah sistem yang dapat memberikan kemudahan bagi pelanggan untuk membaca *review* yang tersedia di Internet. Solusi yang ditawarkan adalah *feature-based opinion mining*.

Sistem ini akan mencari fitur produk dari kalimat opini. Kemudian, kalimat opini yang mengandung fitur yang telah diidentifikasi tersebut akan diklasifikasikan menjadi opini yang berorientasi positif atau negatif. *Data set* yang digunakan berasal dari *review* di www.tabloidpulsa.co.id. Hasil penelitian menunjukkan bahwa proses ekstraksi fitur dengan Algoritma FP-Growth menghasilkan nilai *precision* yang sangat rendah dengan rata-rata nilai sebesar 0,103. Penyebab rendahnya nilai *precision* terutama karena hasil ekstraksi fitur yang terdiri dari lebih satu kata tidak memperhatikan urutan kata dalam kalimat. Sedangkan proses orientasi opini sudah cukup tinggi dengan rata-rata akurasi sebesar 71,9%.

Kata Kunci : *opinion mining*, *FP-Growth*, fitur produk, sentimen analisis, *SentiWordNet*

ABSTRACT

Along with the development of technology of Internet, e-commerce becomes familiar. One form of utilization of e-commerce is the use of a web as a means of buying and selling. Some web that allows customers to shop online also provides the flexibility to customers to participate write a review about the product bought. The consequence is an increase in the number of product reviews. Every day could be hundreds of new review appears. This condition will make customer difficult to read the reviews that can help him to make a decision to buy a product. For that, we need a system that can help customers read the reviews on the Internet easier. The solution offered is a feature-based opinion mining.

This system will look for product features from opinion sentence. Then, opinion sentences containing features that have been identified will be classified into positive or negative orientation. The data used is derived from a review in www.tabloidpulsa.co.id. The results showed that the feature extraction using FP-Growth algorithm produces very low precision value with an average value about 0.103. The low value of precision mainly because of the extracted features that consisting of more than one word did not pay attention to the order of words in a sentence. While the orientation opinion process is already fairly high with an accuracy average about 71.9%.

Keywords : *opinion mining*, *FP-Growth*, product features, sentiment analysis, *SentiWordNet*

BAB 1 PENDAHULUAN

Seiring dengan berkembangnya teknologi dalam dunia *Internet*, *electronic commerce* (*e-commerce*) menjadi hal yang tidak asing lagi. Salah satu bentuk pemanfaatan *e-commerce* adalah penggunaan *web site* sebagai sarana jual beli. Beberapa *web site* yang memungkinkan pelanggannya untuk berbelanja secara *online* juga memberikan keleluasaan kepada pelanggannya untuk turut aktif menuliskan *review* tentang produk yang dibelinya. Konsekuensinya adalah peningkatan jumlah *review* produk. Setiap hari bisa

saja ratusan *review* baru muncul. Hal ini akan berpotensi membuat pelanggan mengalami kesulitan untuk membaca *review* yang dapat membantunya untuk mengambil keputusan membeli produk.

Untuk itu, diperlukan sebuah sistem yang dapat memberikan kemudahan bagi pelanggan untuk membaca *review* yang tersedia di *Internet*. Solusi yang ditawarkan adalah *feature-based opinion mining*. Sistem ini akan mencari fitur produk dari kalimat opini. Kemudian, kalimat opini yang mengandung fitur yang telah diidentifikasi tersebut akan diklasifikasikan menjadi opini

positif atau opini negatif [1]. Seluruh opini yang ada akan ditampilkan berkelompok berdasarkan fitur yang dimiliki oleh masing-masing produk. Proses ekstraksi fitur produk menggunakan *Association Mining*. *Association Mining* digunakan karena kecenderungan *review* mengandung banyak hal yang tidak berkaitan langsung dengan fitur produk. Masing-masing pelanggan memiliki cara yang berbeda dalam menuliskan *review*-nya. Namun, kata-kata yang digunakan bersifat konvergen. Jadi, penggunaan *Association Mining* untuk pencarian *frequent itemset* adalah tepat karena *frequent itemset* cenderung menjadi fitur produk [2].

Sebelumnya telah diadakan penelitian menggunakan *data set* berbahasa Inggris [2]. Sedangkan pada penelitian Tugas Akhir ini akan digunakan *data set* berbahasa Indonesia. Dipilihnya dataset berbahasa Indonesia karena dianggap sistem seperti ini juga dibutuhkan untuk memberikan kemudahan kepada pelanggan yang membaca *review* berbahasa Indonesia. Penggunaan *data set* yang berbeda tentunya akan membutuhkan tahapan *preprocessing* yang berbeda. Penelitian menggunakan *data set* berbahasa Indonesia sudah pernah dilakukan [3]. Namun, terdapat perbedaan untuk algoritma yang digunakan dalam proses ekstraksi fitur produk. Algoritma CBA yang menerapkan Algoritma *Apriori* digunakan pada penelitian [3]. Sedangkan dalam penelitian Tugas Akhir ini akan digunakan Algoritma *FP-Growth*. Algoritma *FP-Growth* dipilih karena merupakan pengembangan dari Algoritma *Apriori*, sehingga kekurangan dari Algoritma *Apriori* diperbaiki oleh Algoritma *FP-Growth* [4]. Algoritma *FP-Growth* efisien dan skalabel untuk pencarian *frequent itemset* yang panjang dan pendek, dan sekitar satu kali lebih cepat dari Algoritma *Apriori* dan juga lebih cepat daripada beberapa metode pencarian *frequent pattern* baru [5]. Algoritma *FP-Growth* menggunakan struktur data *Frequent Pattern (FP) Tree*. *FP-Tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FP-Tree*. Untuk setiap transaksi yang dipetakan, ada kemungkinan beberapa transaksi memiliki item yang sama, sehingga lintasannya memungkinkan saling menimpa. Apabila semakin banyak data transaksi yang memiliki item yang sama, maka proses pemampatan struktur data *FP-Tree* akan semakin efektif.

BAB 2 DASAR TEORI

2.1 Opinion Mining

Opinion mining, atau disebut juga *sentiment analysis* merupakan sebuah cabang penelitian di domain *text mining* yang mulai marak pada tahun 2003. *Opinion mining* adalah penelitian komputasional dari opini, *sentiment*, dan emosi yang diekspresikan secara tekstual. Jika diberikan satu set dokumen teks D yang berisi opini mengenai suatu objek, maka *opinion mining* bertujuan untuk mengekstrak atribut dan komponen dari objek yang telah dikomentari pada

setiap dokumen dan menentukan apakah komentar tersebut bersifat positif atau negatif [8].

Salah satu level dalam *opinion mining* adalah *feature-based opinion mining*. Pada analisis sentimen, umumnya target opini tidak diperhatikan, sehingga sentimen diukur secara global dalam suatu opini. Apabila analisis sentimen melibatkan target opini, maka disebut dengan istilah *feature-based opinion mining*. Ada dua proses dalam *feature-based opinion mining*, yaitu *feature extraction* dan *opinion orientation identification*. *Feature extraction* adalah identifikasi fitur objek yang dikomentari. *Opinion orientation identification* menentukan apakah opini pada fitur bersifat positif atau negatif.

2.2 Association Rule Mining

Association rule mining adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan, sehingga analisis asosiasi juga sering disebut dengan istilah *market basket analysis* [9].

2.3 Algoritma Frequent Pattern (FP)-Growth

FP-Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Algoritma *FP-Growth* tidak melakukan *generate candidate* karena *FP-Growth* menggunakan konsep pembangunan *tree* dalam pencarian *frequent itemsets*. Hal tersebut yang menyebabkan Algoritma *FP-Growth* lebih cepat dari Algoritma *Apriori* [11].

Karakteristik algoritma *FP-Growth* adalah menggunakan struktur data *tree* yang disebut dengan *FP-Tree*. Pembuatan *FP-Tree* dapat dilihat pada Gambar 2-1 berikut.

Input: A transaction database DB and a minimum support threshold ξ .

Output: FP-tree, the frequent-pattern tree of DB .

Method: The FP-tree is constructed as follows.

1. Scan the transaction database DB once. Collect F , the set of frequent items, and the support of each frequent item. Sort F in support-descending order as $FList$, the list of frequent items.
2. Create the root of an FP-tree, T , and label it as "null". For each transaction $Trans$ in DB do the following.

Select the frequent items in $Trans$ and sort them according to the order of $FList$. Let the sorted frequent-item list in $Trans$ be $[p | P]$, where p is the first element and P is the remaining list. Call $insert_tree([p | P], T)$.

The function $insert_tree([p | P], T)$ is performed as follows. If T has a child N such that $N.item-name = p.item-name$, then increment N 's count by 1; else create a new node N , with its count initialized to 1, its parent link linked to T , and its node-link linked to the nodes with the same *item-name* via the node-link structure. If P is nonempty, call $insert_tree(P, N)$ recursively.

Gambar 2-1: Algoritma Pembuatan FP-Tree

Metode *FP-Growth* dapat dibagi menjadi 3 tahapan utama:

- (1) tahap pembangkitan *conditional pattern base*,
- (2) tahap pembangkitan *conditional FP-Tree*, dan

(3) tahap pencarian *frequent itemset*.

Ketiga tahapan tersebut dapat dilihat pada ilustrasi algoritma dalam Gambar 2-2 di bawah.

```

Procedure FP-growth(Tree, a) {
  (01) if Tree contains a single prefix path then // Mining single prefix-path FP-tree {
    (02) let P be the single prefix-path part of Tree;
    (03) let Q be the multipath part with the top branching node replaced by a null root;
    (04) for each combination (denoted as β) of the nodes in the path P do
    (05) generate pattern β ∪ a with support = minimum support of nodes in β;
    (06) let freq pattern set(P) be the set of patterns so generated;
  }
  (07) else let Q be Tree;
  (08) for each item ai in Q do // Mining multipath FP-tree
    (09) generate pattern β = ai ∪ a with support = ai .support;
    (10) construct β's conditional pattern-base and then β's conditional FP-tree Tree β;
    (11) if Tree β ≠ ∅ then
      (12) call FP-growth(Tree β , β);
    (13) let freq pattern set(Q) be the set of patterns so generated;
  }
  (14) return(freq pattern set(P) ∪ freq pattern set(Q) ∪ (freq pattern set(P) × freq pattern set(Q)))
}
    
```

Gambar 2-2: Algoritma FP-Growth

Dengan menggunakan *FP-Tree*, algoritma *FP-Growth* dapat langsung mengekstrak *frequent Itemset* dari *FP-Tree*. Pencarian *itemset* yang *frequent* dengan

menggunakan algoritma *FP-Growth* akan dilakukan dengan cara membangkitkan *FP-Tree*.

2.4 Part-of-Speech (POS) Tagging

POS Tagging adalah proses menandai bagian part-of-speech atau kata dengan kategori gramatikal seperti kata benda, kata kerja, kata sifat, frase, dan lain-lainnya ke kelas-kelas yang sudah didefinisikan ke dalam sebuah teks.

POS Tagging merupakan proses penting pada opinion mining, khususnya pada tahap preprocessing. Inputan dari POS Tagging berupa string atau kumpulan kata dalam sebuah NLP (Natural Language Processing) dan sebuah list dari part-of-speech. Outputnya berupa

sebuah POS tag untuk setiap kata yang akan digunakan pada proses berikutnya.

2.5 SentiWordNet Bahasa Indonesia

SentiWordNet Bahasa Indonesia merupakan pengembangan dari SentiWordNet, merupakan kamus leksikal untuk opini atau sentimen Bahasa Indonesia. SentiWordNet membobotkan setiap sentimen dengan 2 nilai, yaitu nilai positif dan nilai negatif. Jika dijumlahkan, nilai positif dan negatif setiap sentimen (*synset*) bernilai antara 0 sampai dengan 1. Dimana nilai ini menjadi salah satu bobot dalam pembobotan

final dalam klasifikasi sentimen positif atau negatif [9].

2.6 Evaluasi Performansi

Evaluasi performansi yang akan digunakan, yaitu precision dan recall. Dalam proses evaluasi performansi untuk klasifikasi, ada empat istilah yang harus diketahui, yaitu:

- (2) true negative, yaitu kelas yang diprediksi salah dan hasilnya salah;
- (3) false positive, yaitu kelas yang diprediksi benar, tetapi hasilnya salah; dan
- (4) false negative, yaitu kelas yang diprediksi salah, tetapi hasilnya benar.

Precision

Precision adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen yang ditemukan oleh sistem.

$$Precision = \frac{TP}{TP + FP}$$

Recall

Recall adalah rasio jumlah dokumen relevan yang ditemukan kembali dengan total jumlah dokumen dalam kumpulan dokumen yang dianggap relevan.

$$Recall = \frac{TP}{TP + FN}$$

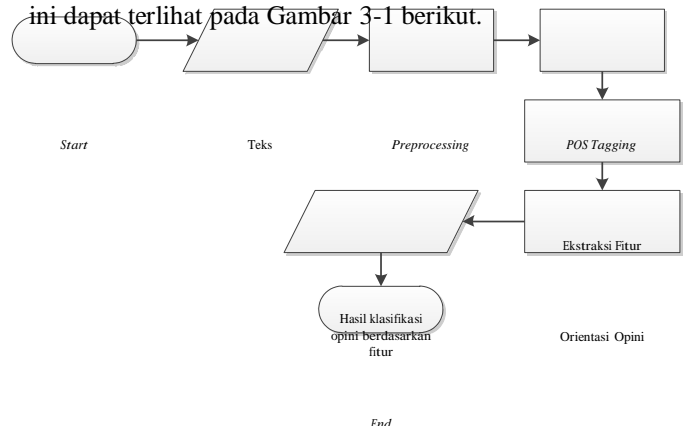
Accuracy

Accuracy adalah rasio dari jumlah ketepatan prediksi tiap kelas terhadap jumlah total semua prediksi yang diklasifikasikan ke dalam kelas-kelas tersebut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

BAB 3 PERANCANGAN SISTEM

Hal yang pertama kali dilakukan dalam sistem adalah pengumpulan *review* sebagai *data set* yang digunakan dalam penelitian. Kemudian, data tersebut akan mengalami proses *preprocessing* dan *POS Tagging*. Dari hasil *POS Tagging*, akan dibentuk *transaction file* yang digunakan sebagai *input* proses ekstraksi fitur produk. Setelah fitur produk teridentifikasi, akan dilakukan proses klasifikasi opini berbasis fitur. Urutan proses dalam sistem ini dapat terlihat pada Gambar 3-1 berikut.



Gambar 3-1: Gambaran umum sistem Tahapan utama dalam sistem ini terdiri dari.

- (1) true positive, yaitu kelas yang diprediksi benar dan hasilnya benar;

- (1) Sistem menerima input berupa *data set* yang berisi review pelanggan mengenai produk elektronik ponsel.
Data set terdiri dari 3 jenis merk ponsel.
- (2) Data melalui tahapan *preprocessing*, diantaranya: *case folding*, *cleaning*, *tokenization*, dan *stop words removal*.
- (3) Data yang sudah melalui tahapan *preprocessing* selanjutnya akan di-*tag* untuk setiap tokennya. Token yang merupakan kata benda maupun frase kata benda dalam satu *review* akan dijadikan *transaction file* sebagai *input* proses ekstraksi fitur produk.

- (4) Proses ekstraksi fitur produk menggunakan Algoritma FP-Growth. Tidak semua hasil fitur yang terekstrak dengan Algoritma FP-Growth tersebut relevan, maka akan dilakukan proses *pruning* untuk menghapus fitur yang dianggap tidak relevan.
- (5) Kalimat-kalimat *review* akan ditentukan polaritasnya menggunakan bantuan *SentiWordNet* Bahasa Indonesia.
- (6) Kalimat opini dalam *review* akan dikelompokkan berdasarkan fitur produk untuk masing-masing merk ponsel untuk mempermudah calon pembeli dalam membaca *review* yang ada.

BAB 4 PENGUJIAN DAN ANALISIS

4.1 Data Set

Data set yang digunakan berupa *file .sql* yang merupakan *review* opini berbahasa Indonesia dari produk elektronik ponsel. Data tersebut diambil dari situs www.tabloidpulsa.co.id. Berikut adalah rincian *data set* yang digunakan.

Tabel 4 Error! No text of specified style in document.-1: Data set review produk

Nama data set	Rincian data
Nokia Lumia	130 <i>review</i>
Samsung Galaxy	193 <i>review</i>
Sony Xperia	329 <i>review</i>

Kalimat opini yang digunakan pada *data set* memiliki beberapa aturan, yaitu.

- (1) Kalimat opini yang eksplisit menyebutkan fitur produk yang dikomentari oleh pelanggan.
- (2) Kalimat opini memiliki ejaan yang sesuai dengan kamus leksikal *SentiWordNet* Bahasa Indonesia.

Kalimat opini tidak mengandung singkatan dan kata tidak baku.

4.2 Analisis Hasil Pengujian

Pengujian yang dilakukan pada sistem bertujuan untuk mengetahui tingkat keberhasilan sistem dalam melakukan keseluruhan prosesnya. Hasil yang akan diuji adalah ekstraksi fitur dan orientasi opini.

Tabel 4-2: Recall dan Precision ekstraksi fitur produk

Nama Produk	Recall	Precision
Nokia Lumia	0.6341463415	0.1477272727
Samsung Galaxy	0.5090909091	0.0332147094
Sony Xperia	0.2597402597	0.1290322581

Dari Tabel 4-2 dapat dilihat bahwa nilai *recall* untuk ekstraksi fitur pada data Nokia sudah cukup, berada pada kisaran angka 0,6. Sedangkan pada Samsung dan Sony, nilai *recall* dapat dikatakan rendah, kisaran angka 0,2 hingga 0,5. Kemudian untuk nilai *precision* dalam proses ekstraksi fitur terbilang sangat rendah untuk semua data, berada

pada kisaran angka 0,0 hingga 0,1.

Besarnya jumlah fitur set yang terekstrak oleh sistem menyebabkan nilai *recall* menjadi tinggi karena fitur set tersebut sudah mengandung sebagian besar fitur yang dianotasi manual. Sebaliknya untuk nilai *precision*, nilai *precision* akan semakin rendah apabila jumlah fitur set yang terekstrak oleh sistem semakin besar. Nilai *precision* yang rendah mengindikasikan bahwa jumlah fitur yang relevan lebih kecil dibandingkan dengan fitur yang tidak relevan.

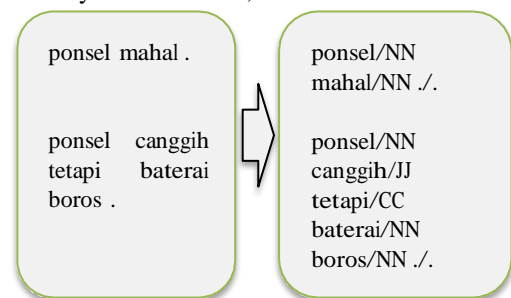
Jika dilakukan metode pengujian yang sama dengan penelitian [3], maka nilai *precision* yang dihasilkan Tugas Akhir ini dengan *data set* Nokia Lumia (Data A) adalah 0,221, Nokia Lumia (Data B) adalah 0,046. Sedangkan untuk sistem [3] berturut-turut menghasilkan nilai 0,75; 0,8. Hasil ekstraksi fitur dalam Tugas Akhir ini masih sangat rendah, namun fitur yang dihasilkan oleh sistem dapat dikatakan relevan untuk produk yang diekstrak.

Tinggi rendahnya nilai *recall* dan *precision* dari sistem ekstraksi fitur dapat dipengaruhi oleh beberapa faktor, diantaranya: karakteristik *review* dan hasil proses POS-Tagging dari *review*. Selain itu, rendahnya nilai *precision* juga disebabkan oleh hasil ekstraksi fitur yang terdiri dari lebih satu kata munculnya tidak sesuai dengan urutan kata dalam *review*. Algoritma FP-Growth menghasilkan *frequent* fitur dengan urutan berdasarkan nilai *minimum support* dari yang tertinggi ke terendah.

Tabel 4-3 : Persebaran fitur produk

Nama Data Set	Hasil Ekstraksi Fitur
Nokia Lumia	103 fitur
Samsung Galaxy	348 fitur
Sony Xperia	133 fitur

Berdasarkan Tabel 4-3 di atas, Samsung mempunyai jumlah fitur yang paling banyak. Hal ini dikarenakan karakteristik *review* pada produk Samsung Galaxy cenderung menggunakan kalimat panjang yang mengomentari fitur yang variatif. Selain karakteristik *review*, hasil POS-Tagging juga mempengaruhi sistem dalam mengambil kata benda atau frase kata benda yang akan diekstrak menjadi fitur set. Contoh hasil POS-Tagging yang tidak sesuai dapat dilihat pada Gambar 4-1 Gambar 4-1 di bawah ini. Kata “mahal” dan “boros” adalah kata sifat, seharusnya bernilai “JJ”, bukan “NN”.



Gambar 4-1: Hasil POS-Tagging yang tidak sesuai

Tabel 4-4: Hasil Compactness Pruning

Nama Produk	Compactness Pruning	
	Recall	Precision
Nokia Lumia	0,6341463415	0,2063492063
Samsung Galaxy	0,5090909091	0,0691358025
Sony Xperia	0,2597402597	0,1333333333

Tabel 4-5: Hasil Redudancy Pruning

Nama Produk	Redudancy Pruning	
	Recall	Precision
Nokia Lumia	0,4634146341	0,1844660194
Samsung Galaxy	0,2909090909	0,0459770115
Sony Xperia	0,2597402597	0,1503759398

Dari Tabel 4-3 dan Tabel 4-4 di bawah, dapat disimpulkan bahwa penggunaan metode *pruning* dapat berpengaruh dalam menentukan fitur produk yang relevan. Sistem mengekstrak fitur set yang relevan dengan hasil anotasi fitur manual apabila nilai *precision* semakin tinggi.

Pada Tabel 4-5, dapat terlihat bahwa *redudancy pruning* dapat menurunkan nilai *recall*. Cara kerja *redudancy pruning* adalah menghapus kandidat fitur produk yang terdiri dari satu kata. Kecenderungan fitur produk yang terdiri dari satu kata lebih banyak daripada fitur produk yang terdiri dari lebih satu kata. Maka dari itu, nilai *recall* akan menurun karena jumlah fitur set berkurang.

Pada Tabel 4-4 dapat terlihat bahwa *compactness pruning* menaikkan nilai *precision* dan *redudancy pruning* cenderung menurunkan nilai *precision*. *Compactness pruning* menaikkan nilai *precision* karena cara kerjanya berkebalikan dengan *compactness pruning*. Kecenderungan fitur produk yang terdiri lebih dari satu kata lebih sedikit jika dibandingkan dengan fitur produk yang terdiri dari satu kata. Sehingga, kemungkinan kandidat fitur yang dihapus setelah proses *compactness pruning* tidak akan mempengaruhi jumlah fitur yang relevan. Apabila jumlah fitur yang relevan tetap atau bertambah dan jumlah fitur set semakin berkurang, maka nilai *precision* akan naik. Sedangkan untuk *redudancy pruning*, nilai *precision* akan menurun karena jumlah fitur yang relevan akan tetap atau cenderung berkurang.

Berikut ini contoh perbandingan hasil ekstraksi fitur yang tanpa *pruning* dan dengan *pruning*. Tabel 4-6 berisi contoh hasil ekstraksi fitur tanpa *pruning*. Tabel 4-7 berisi contoh hasil ekstraksi fitur dengan *compactness pruning* dan Tabel 4-8 berisi contoh hasil ekstraksi fitur dengan *redudancy pruning*.

Tabel 4-6: Hasil ekstraksi fitur tanpa pruning

Tanpa Pruning

video
transisi
ponsel transisi
efek ponsel transisi
efek transisi
transfer
bluetooth transfer
bluetooth cara transfer
bluetooth data transfer
bluetooth sesama transfer
cara transfer

Tabel 4-7: Hasil ekstraksi fitur dengan Compactness Pruning

dengan Compactness Pruning
video
transisi
efek transisi
transfer
bluetooth transfer
bluetooth cara transfer
bluetooth data transfer
bluetooth sesama transfer
cara transfer

Tabel 4-8: Hasil ekstraksi fitur dengan Redudancy Pruning

dengan Redudancy Pruning
video
efek transisi
bluetooth transfer
bluetooth cara transfer
bluetooth data transfer
bluetooth sesama transfer
cara transfer

Dengan menggunakan metode *association mining* menggunakan Algoritma FP-Growth sudah dapat menentukan *frequent itemsets* yang merupakan fitur produk, namun tidak semua fitur produk yang dihasilkan merupakan fitur produk yang asli. Sehingga, diperlukan proses *pruning* untuk menyeleksi fitur yang relevan. Hasil *pruning* tergantung dari batasan nilai minimum yang diberikan. Jika batasan nilai semakin tinggi, jumlah fitur set yang dihasilkan akan semakin berkurang. Dalam Tugas Akhir ini, batasan nilai yang digunakan sesuai dengan penelitian [2].

Tabel 4-9: Hasil Akurasi Orientasi Opini

Nama Produk	Accuracy
Nokia Lumia	72,87 %
Samsung Galaxy	71,35 %
Sony Xperia	71,47 %

Dari Tabel 4-9 dapat dilihat bahwa nilai akurasi untuk orientasi opini pada semua jenis *data set* terbilang cukup tinggi, yaitu berada pada kisaran angka 71%-73%. Hal ini menggambarkan bahwa

sistem sudah dapat memprediksi orientasi opini dengan benar.

Proses penentuan orientasi opini dilakukan dengan menggunakan SentiWordNet. Kelengkapan dan ketepatan *synset* dalam SentiWordNet yang digunakan ikut mempengaruhi ketepatan prediksi orientasi opini. Sebagai contoh, kata “baterai” jika dilihat pada SentiWordNet yang digunakan dalam penelitian ini akan menghasilkan nilai $pos=n$. Hal ini sesuai dengan hasil *POS-Tagging* yang juga menyatakan bahwa kata “baterai” menghasilkan $tag=NN \rightarrow pos=n$. Namun, kata baterai tersebut memiliki bobot negatif. Seharusnya hanya kata yang memiliki nilai $pos=a$ yang dapat memiliki bobot positif atau negatif. Hasil klasifikasi opini yang mengandung kata “baterai” menjadi salah karena definisinya dalam SentiWordNet tidak tepat. Hasil *synset* “baterai” dapat dilihat pada Tabel 4-10.

Tabel 4-10: Bobot kata “baterai” dalam SentiWordNet

<i>Synset</i>	<i>POS</i>	<i>Positive Score</i>	<i>Negative Score</i>
baterai	n	0	0,25
baterai	n	0	0
baterai	n	0	0
baterai	n	0	0
baterai	n	0	0
baterai	n	0	0
Rata-rata Skor		0	0,0357142857

BAB 5 KESIMPULAN DAN SARAN

5.1 Kesimpulan

- (1) Berdasarkan hasil pengujian dan analisa yang telah dilakukan pada pembahasan sebelumnya, maka dapat diambil beberapa kesimpulan sebagai berikut Ekstraksi fitur dengan menggunakan Metode *Association Mining*, yaitu Algoritma *FP-Growth*, sangat berpengaruh dalam menentukan suatu fitur produk yang *frequent*.
- (2) Karakteristik *review* dan hasil *POS-Tagging* juga berpengaruh terhadap proses ekstraksi fitur. *Review* yang berupa kalimat panjang akan sulit diidentifikasi fitur produknya karena mengomentari fitur yang variatif. Semakin variatif fitur, kecenderungan fitur tersebut muncul di *review* semakin sedikit sehingga tidak memenuhi batasan nilai *minimum support*. Hal ini akan berpengaruh pada rendahnya nilai *precision* ekstraksi fitur produk. Dari urutan karakteristik *review* dengan kalimat panjang ke pendek, dapat dilihat nilai *precision* untuk Samsung Galaxy, Sony Xperia, dan Nokia Lumia berada di kisaran 0,033; 0,129; dan 0,147.

- (3) *Compactness pruning* dapat menaikkan nilai rata-rata *precision* dari 0,103 menjadi 0,136.
- (4) Nilai rata-rata *accuracy* untuk proses orientasi opini dari *data set* Nokia Lumia, Samsung Galaxy, dan Sony Xperia adalah 71,9%. Orientasi opini dipengaruhi oleh kelengkapan dan ketepatan definisi *synset* pada SentiwordNet. Semakin banyak kata dalam suatu opini yang tidak tercantum pada Sentiwordnet mengakibatkan banyak kata tidak dapat dibobotkan dengan tepat, sehingga hasil klasifikasi orientasi opini menjadi tidak tepat.

5.2 Saran

Saran yang ingin disampaikan untuk perbaikan Tugas Akhir ini, yaitu.

- (1) Proses *pruning* yang dapat secara otomatis dilakukan oleh program.
- (2) Analisis sentimen terhadap kalimat yang fitur produknya tidak disebutkan secara langsung (kalimat implisit).
- (3) Penggunaan *POS Tagging* Bahasa Indonesia dengan *accuracy* yang lebih baik.
- (4) Penggunaan SentiWordNet Bahasa Indonesia yang lebih lengkap.

REFERENSI

- [1] Bing Liu, "Sentiment analysis and subjectivity," in *andbook of Natural Language Processing, Second Edition*. Taylor and Francis Group, Boca., 2010.
- [2] Minqing Hu and Bing Liu, "Mining Opinion Features in Customer Reviews," *AAAI*, pp. 755-760, July 2004.
- [3] Galih Arisona, Warih Maharani, and Siti Saadah, "Opinion Summarization Terhadap Fitur Produk Elektronik Menggunakan Metode Association Mining," 2014.
- [4] Cornelia Györödi, Robert Györödi, and Stefan Holban, "a Comparative Study of Association Rules Mining Algorithms".
- [5] Jiawei Han, Jian Pei, and Yiwen Yin, "MiningFrequent Patterns without Candidate Generation: A Frequent-Pattern Tree Aproach," in *SIGMOD '00 Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, New York, NY, USA, 2000, pp. 1-12.
- [6] Efraim Turban, Jay E. Aronson, and Ting-Peng Liang, *Decision Support and Business Intelligence Systems.*: Prentice Hall, 2005.
- [7] Usama Fayyad, Gregory Piatetsky-shapiro, and Smyth Padhraic, "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, vol. 17, pp. 37-54, 1996.
- [8] Bing Liu, "Sentiment Analysis: A Multi-Faceted Problem," *Invited contribution to IEEE Intelligent Systems*, 2010.
- [9] Ayu Komalasari, Warih Maharani, and Adiwijaya

- , "Klasifikasi Sentimen pada Review Produk Gadget Menggunakan Graph Based Word Sense Disambiguation," Dec. 2013.
- [10] Michael J.A. Barry and Gordon S. Linnof, *Data Mining Techniques For Marketing Sales Second Edition.*: Wiley Publishing, 2004.
- [11] Erwin, "Analisis Market Basket Dengan Algoritma Apriori dan FP-Growth," *Jurnal Generic*.
- [12] David Samuel, "Penerapan Stuktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset".
- [13] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval.*: Cambridge University Press, 2008.