

PERINGKASAN OPINI UNTUK REVIEW PRODUK MENGGUNAKAN FORMULASI INTEGER LINEAR PROGRAMMING

OPINION SUMMARIZATION FOR PRODUCT REVIEW USING INTEGER LINEAR PROGRAMMING FORMULATION

Tiara Nabila¹, Warih Maharani², M.Syahrul Mubarak³

Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

¹tiaranabilaa@gmail.com, ²wmaharani@gmail.com, ³msyahrulmubarak@gmail.com

Abstrak

Pada era digital, *website* untuk jual beli atau *e-commerce* sudah sangat marak digunakan karena lebih efisien dibandingkan berbelanja langsung di toko *offline*. *Customer* yang membeli barang di *website* biasanya dapat memberikan *review* terhadap produk yang sudah mereka beli. Opini yang diberikan akan menjadi acuan bagi *customer* berikutnya, tetapi opini yang terdapat pada *website* biasanya sangat banyak sehingga dibutuhkan peringkasan agar memudahkan *customer* dalam menentukan pilihan produk mereka. Pada penelitian ini akan dibahas tentang peringkasan opini atau *opinion summarization*. Peringkasan dilakukan dengan terlebih dahulu melakukan ekstraksi menggunakan *pattern matching* lalu akan diperoleh orientasi kalimat opini dengan *SentiWordNet*. Selain itu dibutuhkan dua nilai parameter, yaitu *content score* dan *coherence score* sehingga dapat dilakukan peringkasan pada *review* produk dengan menggunakan formulasi *Integer Linear Programming (ILP)* guna membuat ringkasan yang lebih koheren antara kalimat satu dan lainnya.

Kata kunci : *opinion summarization, integer linear programming (ILP), SentiWordNet*

Abstract

In digital era, website or e-commerce for transaction already widespread, because it is more efficient than shopping at offline stores. Customer who purchased goods on websites can usually provide a review of products they purchased. The opinions given will be the reference for the next customer, but the opinions contained on the website usually too much to be manually handled, and this it needs compaction in order to facilitate the customer in determining the selection of their products. This research discusses opinion summarization. First, summarization is performed by using pattern matching for feature extraction that is obtained opinion orientation using SentiWordNet. Two score parameter, which are content score and coherence score are required in order for summarization can performed well on product reviews by using Integer Linear Programming (ILP) formulation..

Keywords : opinion summarization, integer linear programming (ILP), SentiWordNet

1. Pendahuluan

E-commerce saat ini sudah menjadi tren di seluruh penjuru masyarakat dunia, terlebih lagi teknologi informasi sudah semakin canggih dan sangat mendukung. Maraknya *e-commerce* memberikan efek terhadap seluruh area bisnis, karena sangat membantu dalam proses bisnis terutama untuk berinteraksi dengan customer [1]. Untuk mengetahui kualitas suatu produk pada *website e-commerce*, para konsumen akan memberikan *review* untuk setiap produk yang sudah mereka beli sebelumnya, hal ini bertujuan untuk memberikan informasi kepada calon konsumen yang akan membeli produk tersebut ataupun feedback bagi penjual produk. *Review* yang diberikan oleh konsumen biasanya terdiri dari opini-opini yang positif ataupun negatif, serta *review* yang berjumlah banyak dapat membingungkan calon konsumen untuk memperoleh kesimpulan informasi kualitas barang tersebut. Oleh karena itu, untuk memudahkan dalam memberikan informasi bagi konsumen dalam memilih produk yang sesuai dengan kebutuhan konsumen, maka dilakukan peringkasan kalimat *review* dari suatu produk.

Ryan McDonald [2] melakukan penelitian tentang peringkasan *review* produk menggunakan Integer Linear Programming (ILP) secara global sehingga hanya menggunakan prinsip ILP yang umum yaitu memaksimalkan suatu fungsi objektif berdasarkan beberapa constraint yang ada. Karena model peringkasan yang sudah ada sebelumnya hanya membahas secara global dan tidak memperhatikan koherensi dari hasil peringkasan, sehingga pada penelitian tugas akhir ini digunakan dua parameter tambahan yaitu *content score* dan *coherence score* guna membuat hasil ringkasan yang lebih koheren antara kalimat satu dan lainnya.

Proses lainnya yang dilakukan untuk melakukan peringkasan teks dalam penelitian ini yaitu preprocessing data berupa *stopword removal, lemmatization* dan *pos tagging* yang berfungsi untuk membersihkan data agar siap untuk diolah. Selanjutnya proses ekstraksi menggunakan *pattern matching* dengan daftar *rules* yang ada untuk mendapatkan kata-kata fitur opini yang terdapat pada kalimat dan proses akhir sebelum dilakukan peringkasan adalah melakukan klasifikasi kalimat untuk menentukan orientasi kalimat berdasarkan hasil

ekstraksi menggunakan SentiWordNet yaitu berupa kamus yang berisi kata-kata sentiment yang sudah memiliki skor setiap kata nya.

2. Dasar Teori

2.1 Stopword Removal

Stopword removal yaitu proses untuk penghilangan atau pemilihan istilah yang tidak memiliki makna penting pada suatu kalimat berdasarkan daftar kata stopwords yang digunakan[3]. Tujuan dari stopwords removal ini yaitu agar memudahkan pengambilan dari fitur dan memudahkan untuk proses dari ekstraksi.

2.2 Lemmatization

Lemmatization adalah teknik normalisasi yang lain untuk mengubah suatu bentuk kata biasa ke bentuk dasar kata tersebut, . Fungsi dari *lemmatization* hampir sama dengan proses *stemming* yang bertujuan agar penggunaan kata atau morfologi kata yang ada digunakan dengan benar. Teknik ini sudah banyak digunakan pada information retrieval[4].

2.3 Part Of Speech (POS) Tagging

Penggunaan *Part Of Speech* (POS) sangat bermanfaat dalam proses *information retrieval*, karena proses ini memberikan informasi tentang kata dan ketetanggaannya. *Part of Speech* akan melakukan identifikasi terhadap setiap kata lalu akan memberikan tag yang ditempatkan pada akhir dari setiap kata yang sudah diidentifikasi [5]. Pemberian tag yang digunakan yaitu sesuai dengan library dari Stanford POS Tagging.

2.4 Pattern Matching

Pattern matching adalah sebuah *subtask* dari proses discovery pattern, pola yang digunakan bisa pola yang umum atau lebih spesifik [6]. Pola yang digunakan pada tugas akhir ini berupa pola linguistic, dimana pola akan dicocokkan dengan hasil dari POS Tagging yang ada lalu akan menghasilkan output berupa kata *Noun* atau *Noun Phrase*. Berikut adalah pola yang digunakan dalam penelitian ini :

Tabel 2-1 Daftar Rule[7]

No	First Word	Second Word	Third Word
1	JJ	NN/NNS	-
2	RB/RBR/RBS	JJ	not NN nor NNS
3	JJ	JJ	not NN nor NNS
4	NN/NNS	JJ	not NN nor NNS
5	RB/RBR/RBS	VB/VBD/VBN/VBG	-

2.5 SentiWordNet

SentiWordNet adalah kumpulan *lexical database* atau kamus yang merupakan pengembangan dari WordNet. Dimana setiap synset yang terdapat di dalam WordNet akan dikelompokkan ke dalam tiga jenis skor sentimen yaitu netral, positif dan negatif. SentiWordNet merupakan tools online yang open source sehingga dapat digunakan secara bebas untuk berbagai kepentingan penelitian[8].

Perhitungan setiap *term* nya dihitung sebagai berikut :

$$\text{SynsetScore} = \text{PosScore} - \text{NegScore} \quad (2.1)$$

Setelah mendapatkan *SynsetScore*, system akan mengambil seluruh term yang ada sesuai dengan peringkat dan tipe *term*, contoh:

term -> {*synsetscore of term#1, synsetscore of term#2, ... , synsetscore of term#n*}

Seluruh kata yang ada akan dihitung rata-rata nya menurut peringkat kata tersebut sebagai sentiment score menggunakan perhitungan sebagai berikut :

$$\text{Score} = \frac{1}{2} * \text{first} + \frac{1}{3} * \text{second} + \frac{1}{4} * \text{third} + \dots + \frac{1}{n+1} * n \quad (2.2)$$

$$\text{Sum} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n+1} \quad (2.3)$$

$$\text{Sentiment Score} = \frac{\text{Score}}{\text{Sum}} \quad (2.4)$$

2.6 Integer Linear Programming

Integer Linear Programming (ILP) adalah suatu masalah pembatasan optimasi dimana *cost function* dan *constraints* berada dalam satu set variabel integer. ILP yang modern dapat menyelesaikan masalah optimasi yang besar dengan waktu yang cukup singkat. Tujuan dari ILP ini adalah untuk mengatur variabel indikator untuk memaksimalkan subjek *payoff* dari beberapa set *constraints* agar menjamin validitas dari solusi yang dihasilkan[2]. Fungsi objektif yang digunakan pada penelitian ini adalah :

$$\mathcal{L}(S) = \sum_{e_i \in S} w_i e_i + \sum_{a_{i,j} \in S} c_{ij} a_{ij} \tag{2.5}$$

Parameter yang digunakan yaitu content score dan coherence score.
 Untuk mencari nilai *content score* menggunakan rumus berikut :

$$\text{Content Score} = \frac{\sum e_i}{\text{jumlah } e \text{ yang digunakan}} \tag{2.6}$$

Bobot dari masing-masing aspek(e) diambil dari score *SentiWordNet*. Setelah mendapatkan *content score* dapat dilakukan perhitungan mencari *coherence score*.

Untuk mencari nilai *coherence score* terdapat beberapa langkah :

Langkah pertama mencari nilai *local coherence*($c_{(i,j)}$) :

$$c_{(i,j)} = \frac{\text{content}(n) + \text{content}(n)}{2} \tag{2.7}$$

		content 1	content 2	...	content (n)
		S1	S2	...	S(n)
content 1	S1	$\frac{\text{content 1} + \text{content 1}}{2}$	$\frac{\text{content 1} + \text{content 2}}{2}$...	$\frac{\text{content 1} + \text{content}(n)}{2}$
content 2	S2	$\frac{\text{content 2} + \text{content 1}}{2}$	$\frac{\text{content 2} + \text{content 2}}{2}$...	$\frac{\text{content 2} + \text{content}(n)}{2}$
....
content(n)	S(n)	$\frac{\text{content}(n) + \text{content 1}}{2}$	$\frac{\text{content}(n) + \text{content 2}}{2}$...	$\frac{\text{content}(n) + \text{content}(n)}{2}$

Langkah kedua mencari nilai keterkaitan antar kalimat($a_{(i,j)}$) :

$$a_{(i,j)} = \frac{\text{content}(n) + \text{content}(n)}{2} \tag{2.8}$$

content 1	content 2	...	content (n)
S1	S2	...	S(n)
$\frac{\text{content 1} + \text{content 2}}{2}$	$\frac{\text{content 2} + \text{content}(n)}{2}$...	$\frac{\text{content}(n) + 0}{2}$

Langkah ketiga lakukan penjumlahan antara local coherence dan nilai keterkaitan :

$$X_{(n)} = \frac{c_{(i,j)} + a_{(i,j)}}{2} \tag{2.9}$$

S1	S2	...	S(n)
$\frac{c_{(1,1)} + a_{(1,2)}}{2}$	$\frac{c_{(1,2)} + a_{(2,3)}}{2}$...	$\frac{c_{(i,j)} + a_{(i,j)}}{2}$
$\frac{c_{(1,2)} + a_{(1,2)}}{2}$	$\frac{c_{(2,2)} + a_{(2,3)}}{2}$...	$\frac{c_{(i,j)} + a_{(i,j)}}{2}$
...
$\frac{c_{(i,j)} + a_{(i,j)}}{2}$	$\frac{c_{(i,j)} + a_{(i,j)}}{2}$...	$\frac{c_{(i,j)} + a_{(i,j)}}{2}$

Langkah keempat mencari nilai coherence score ($C_{(i,j)}$) :

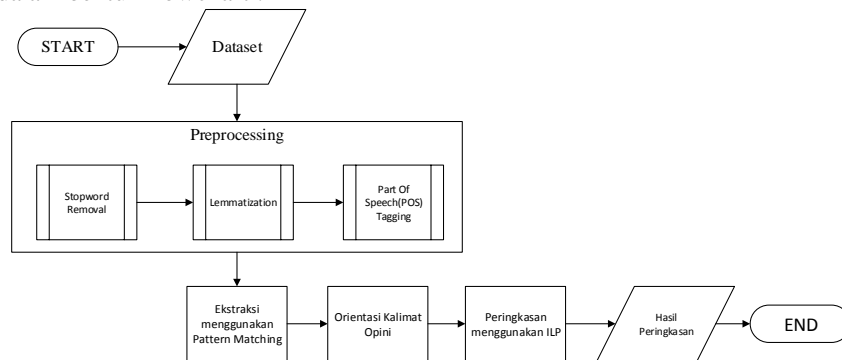
$$C_{(i,j)} = \frac{\text{Content Score}}{X_{(n)}} \tag{2.10}$$

$$C_{(i,j)} = \frac{\sum C_{(i,j)}}{\text{jumlah kalimat}} \tag{2.11}$$

3. Perancangan Sistem

3.1 Gambaran Umum Sistem

Sistem yang akan dibuat untuk penelitian tugas akhir ini yaitu suatu sistem yang dapat melakukan peringkasan opini menggunakan formulasi *integer linear programming*. Dengan melewati tiga tahapan utama yaitu, preprocessing, ekstraksi, klasifikasi dan akhirnya baru bisa dilakukan peringkasan. Berikut gambaran umum sistem yang dibuat dalam bentuk flowchart :



Gambar 3-1 Gambaran umum sistem

3.2 Preprocessing

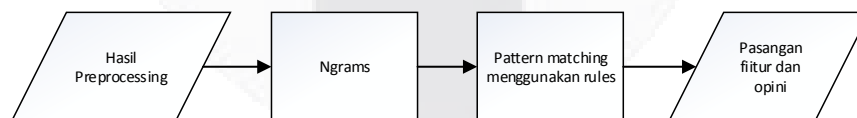
Tahapan awal yang dilakukan di dalam sistem ini yaitu *preprocessing* data, dimana inputan dataset awal akan diolah menjadi data yang berkualitas untuk selanjutnya diproses ke tahapan yang lain. Pada tahapan *preprocessing* ini dibagi kembali menjadi beberapa proses, yaitu :

1. Dataset yang berupa file .txt pertama akan melalui tahapan *stopword removal* untuk menghilangkan kata-kata yang dianggap tidak relevan dengan membandingkan kalimat di dalam dataset dengan daftar stopwords yang digunakan
2. Tahapan preprocessing selanjutnya adalah *lemmatization* dimana kata yang terdapat pada kalimat akan diubah ke bentuk kata dasarnya dengan menggunakan library Stanford.
3. Setelah proses *stopword removal* dan *lemmatization*, proses terakhir adalah pemberian tag, yaitu proses POS Tagging. Proses ini sangat diperlukan karena mendukung untuk proses ekstraksi. Pada proses ini juga digunakan library Stanford POS Tagging. Setelah semua proses preprocessing dilakukan maka data sudah siap diproses ke tahapan selanjutnya.

3.3 Ekstraksi menggunakan pattern matching

Tahapan selanjutnya yaitu ekstraksi kalimat opini, tujuannya yaitu untuk mencari pasangan fitur dan opini yang terdapat pada kalimat opini yang sudah mengalami proses *preprocessing*. Pada proses ekstraksi digunakan rule pada tabel 2.2 untuk mendapatkan pasangan fitur dan opini.

Berikut adalah detail dari setiap tahapan untuk proses ekstraksi fitur dan opini menggunakan pattern matching :



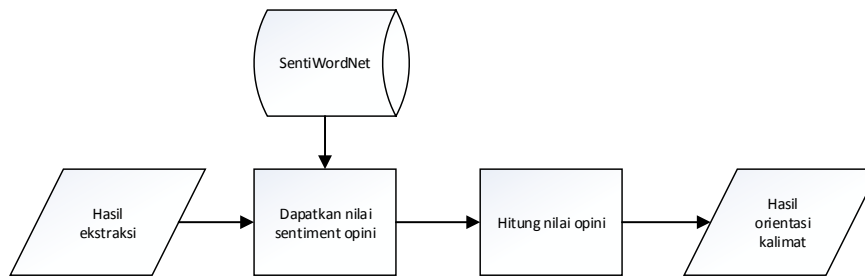
Gambar 3-2 Tahapan proses ekstraksi

1. Data input yang digunakan telah mengalami proses *preprocessing*.
2. Dengan menggunakan n-grams, dilakukan pemotongan sebanyak dua hingga tiga kata yaitu bi-gram dan tri-gram. Ini dilakukan karena rule yang digunakan mencocokkan dua hingga tiga kata.
3. Hasil dari bi-gram dan tri-gram kemudian akan dicocokkan terhadap rule yang digunakan.
4. Jika kata-kata tersebut cocok dengan rule yang digunakan, maka kata tersebut akan diekstrak sebagai pasangan fitur dan opini.

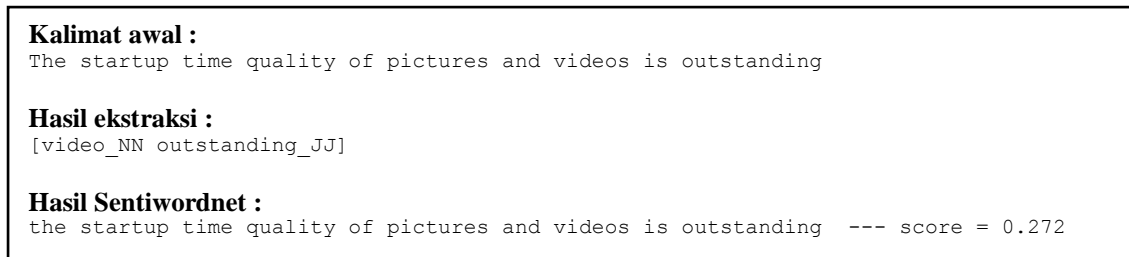
3.4 Orientasi kalimat opini

Setelah mendapatkan hasil dari proses ekstraksi, selanjutnya hasil ekstraksi tersebut dihitung per kata nya dengan perhitungan (rumus sentiwordnet) sesuai dengan daftar nilai yang ada pada kamus sentiwordnet untuk menentukan apakah kalimat tersebut merupakan kalimat negatif atau positif supaya lebih memudahkan proses peringkasan kalimat. Pada penelitian ini, perhitungan hanya dilakukan terhadap kata yang memiliki tag JJ (adjective) saja, jika hasilnya lebih dari 0 maka termasuk kalimat positif dan jika

hasilnya kurang dari 0 maka termasuk kalimat negatif. Berikut adalah detail dari setiap tahapan untuk menentukan orientasi kalimat opini :



Gambar 3-3 Tahapan proses orientasi kalimat opini

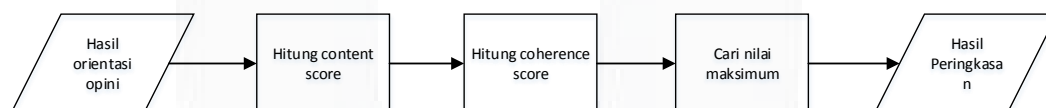


Gambar 3-4 Hasil Proses Ekstraksi

Dilihat dari contoh diatas, hasil ekstraksi dari kalimat tersebut yaitu *video_NN outstanding_JJ* sehingga kita hanya menghitung nilai dari kata *outstanding*. Sesuai dengan nilai di dalam kamus SentiwordNet score yang dihasilkan yaitu 0,272 sehingga orientasi kalimat tersebut adalah positif.

3.5 Peringkasan menggunakan ILP

Setelah mendapatkan hasil orientasi kalimat opini dilakukan peringkasan menggunakan formulasi integer linear programming dengan memaksimalkan fungsi (2.1). Hasil orientasi kalimat opini sudah diklasifikasikan antara kalimat positif dan negatif sehingga lebih memudahkan proses perhitungan ILP. Berikut adalah detail tahapan untuk mendapatkan hasil peringkasan menggunakan formulasi ILP :



Gambar 3-5 Tahapan proses peringkasan

Contoh kalimat yang digunakan :

1. a digic iipowered imageprocessing system enables the sd500 to snap a limitless stream of 7megapixel photos at a respectable clip its startup time is tops in its class and it delivers decent photos when compared to its competition
2. if you want something resembling a real and professional review before plunking down your 500 00 i suggest going to dpreview or dcresource and reading the reviews there
3. i thought given all the 10 reviews about this camera that i would buy it in spite of being a bit pricey
4. the camera was sitting on my home desk waiting for the new digital card and the weekend for the first round of serious shooting
5. the startup time quality of pictures and videos is outstanding

List fitur yang digunakan :

- e1 = digital card
- e2 = video outstanding
- e3 = time top class
- e4 = bit pricey
- e5 = snap limitless stream

- e6 = photo respectable clip
- e7 = decent photo
- e8 = professional review
- e9 = real professional

Tabel Error! No text of specified style in document.-1 Content Score dan Coherence Score

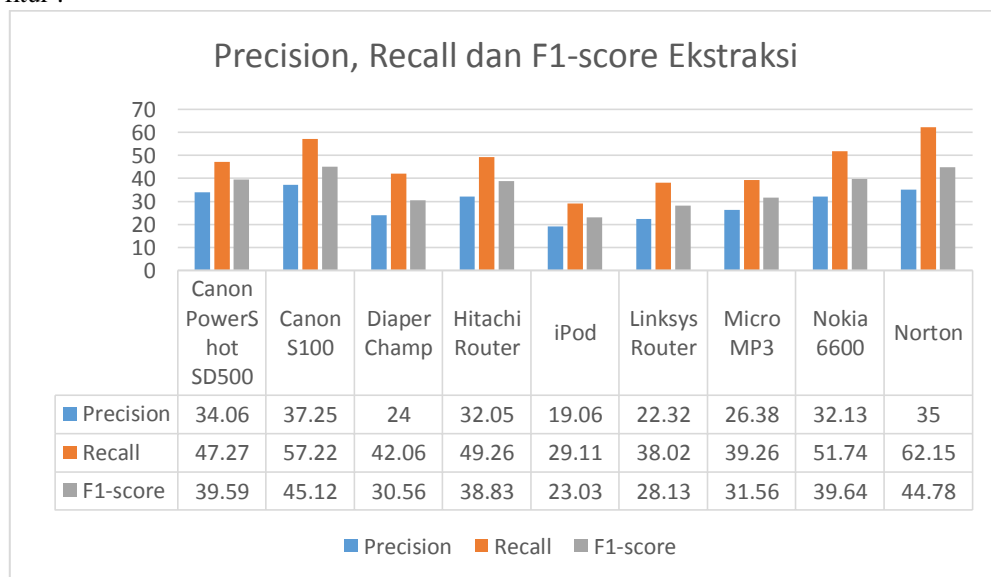
Sentence		e1	e2	e3	e4	e5	e6	e7	e8	e9	Content Score	Coherence Score
	s1	0	0	1	0	1	1	1	1	0	0	0.058
s2	0	0	0	0	0	0	0	0	1	1	0	0.006
s3	0	0	0	1	0	0	0	0	0	0	0.003	0.023
s4	1	0	0	0	0	0	0	0	0	0	0.049	0.080
s5	0	1	0	0	0	0	0	0	0	0	0.002	0.009

Data yang digunakan adalah data kalimat opini yang sudah diklasifikasikan berdasarkan kalimat positif dan negatif. Proses ILP yang pertama yaitu mencari *content score*. *Content score* setiap kalimat dihitung berdasarkan bobot w_i yang diperoleh dari *SentiWordNet* dan jumlah aspek yang terdapat pada kalimat tersebut. Setelah itu mencari nilai *coherence score*, nilai ini dapat dihitung setelah kita sudah menghitung *content score*. Jika *content score* dan *coherence score* masing-masing kalimat sudah diperoleh, maka dapat dilihat kalimat mana yang memiliki total *content score* dan *coherence score* yang paling maksimal. Kalimat yang memiliki total dengan nilai maksimal akan terpilih sebagai ringkasan kalimat. Pada contoh kasus ini, kalimat yang memiliki nilai maksimal yaitu kalimat 1.

4. Pembahasan

4.1 Analisis Hasil Ekstraksi Fitur

Analisis dilakukan terhadap hasil fitur produk yang terekstraksi menggunakan pattern matching dengan beberapa rules yang sudah ditentukan. Berikut adalah hasil evaluasi precision, recall dan f1-score pada ekstraksi fitur :

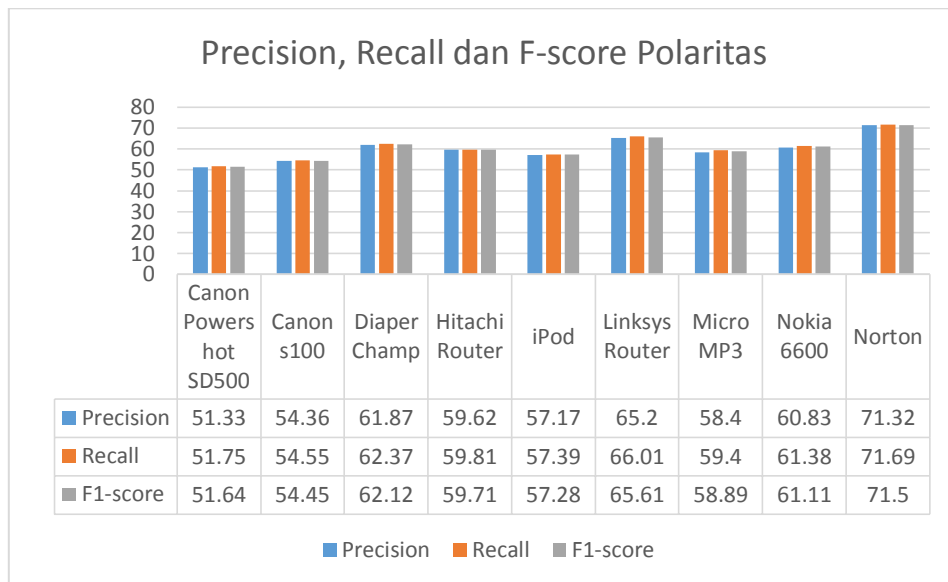


Gambar 4-1 Evaluasi Ekstraksi Fitur

Berdasarkan gambar 4-1 hasil *precision*, *recall* dan *F-score* berada pada angka yang sama. Sehingga dapat ditunjukkan berdasarkan hasil tersebut, tingkat kedekatan antara hasil prediksi dari sistem dengan nilai kebenarannya.

4.2 Analisis Hasil Identifikasi Polaritas Kalimat Opini

Evaluasi hasil polaritas opini dilakukan dengan menggunakan perhitungan precision, recall dan f-score dengan menggunakan perhitungan dari rumus Sentiwordnet dengan polaritas yang terdapat pada dataset dari penelitian sebelumnya. Berikut adalah hasil evaluasi precision, recall dan f-score :



Gambar 4-2 Evaluasi Hasil Polaritas Menggunakan SentiWordNet

Berdasarkan gambar diatas, hasil precision, recall dan f-score dari seluruh dataset berada pada angka yang cenderung sama. Sehingga dapat ditunjukkan berdasarkan hasil tersebut, tingkat kedekatan antara hasil prediksi dari sistem dengan nilai kebenarannya.

4.3 Analisis Parameter Content Score dan Coherence Score

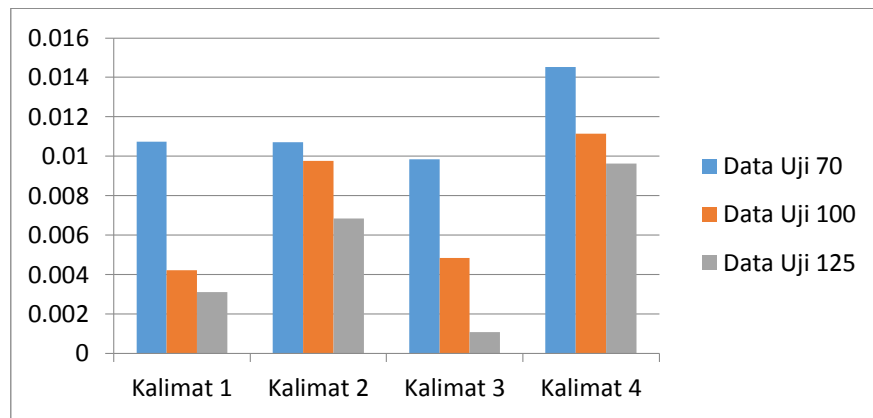
Pada penelitian ini peringkasan dilakukan menggunakan formulasi ILP, dan yang menjadi faktor utama nya yaitu dua parameter yang terdapat pada persamaan (2.1) yaitu *content score* dan *coherence score*. Maka skenario pengujian dilakukan pada analisis pengaruh dari parameter *content score* dan *coherence score* terhadap hasil peringkasan.

a. Pengaruh Parameter *Content Score*

Seperti yang sudah dipaparkan pada sub bab 3.2.5 terdapat perhitungan *content score* untuk menentukan hasil peringkasan. Skenario pengujian dilakukan terhadap empat data uji yang masing-masing berjumlah 70, 100 dan 125 kalimat untuk melakukan analisis nilai content score dari masing-masing data yang jumlahnya berbeda tetapi memiliki beberapa kalimat yang sama.

Tabel 4-1 Contoh Kalimat yang Digunakan

No	Kalimat
1	<i>Powerup is quick and you can even set your favorite image to display for a second or so as it turns on</i>
2	<i>I have been fairly impressed with its ability to capture decent details in low light situations such as theaters inside a bus at night ect</i>
3	<i>Also as someone who at least knows a little bit about the technical work of taking a photo i really miss having manual controls</i>
4	<i>I have especially enjoyed the 640x480 mode while not a replacement for a dedicated video camera its quality easily surpasses that of my old hi8 camera and is perfect for quick clips of things that still images just dont describe max 8 mins at 640 30fps on a one gig card</i>



Gambar Error! No text of specified style in document.-1 Hasil Pengujian Content Score terhadap Tiga Data Uji

Dapat dilihat grafik hasil data uji, nilai *content score* dari setiap kalimat mengalami penurunan. Semakin banyak jumlah kalimat yang digunakan maka nilai *content score* akan semakin menurun karena perhitungan *content score* dipengaruhi juga oleh jumlah fitur yang digunakan pada setiap perhitungan.

b. Pengaruh Parameter Coherence Score

Pengujian untuk *coherence score* dilakukan pada dataset Canon Powershot SD500. Berikut hasil *coherence score* dari pengujian diambil hanya lima kalimat tertinggi dan lima kalimat terendah pada *result* positif.

Tabel 4-2 Lima Kalimat dengan Coherence Score Tertinggi

Kalimat	Coherence Score
<i>as any digital camera owner knows is that when you go to take a picture of a moving object to expect some blurring but with some preliminary testing we had someone move there hands back and forth under the lense and another person take the picture and suprisingly a crystal clear image apeared on the decently sized screen</i>	0.0310
<i>digital zoom is clear too say you used all 3x optical zoom try to do digital zoom then just try not to shake the camera at all and you'll still get a clear good quality picture</i>	0.0309
<i>the view finder shows crystal clear pictures that will look even better when you get them to the computer</i>	0.0260
<i>you can set the optical zoom before shooting and than while shooting video adjust a digital zoom in and out however due to the cameras light weight camera shake is a problem to contend with</i>	0.0178
<i>the disadvantages of this camera is it has been having some trouble with its autofocus on a few occasional pictures especially quick candid shots where you start up and shot as quick as you can</i>	0.0176

Berdasarkan pemahaman nya, *coherence score* pada suatu kalimat berarti nilai keterkaitan antara kalimat satu dengan kalimat lainnya. Hal ini dapat dibuktikan dari hasil pada tabel 4.5 dimana lima kalimat tersebut saling berkaitan satu sama lain, berkaitan disini artinya kalimat tersebut masih membicarakan topik yang sama atau mengandung aspek yang sama.

Tabel 4-3 Lima Kalimat dengan Coherence Score Terendah

Kalimat	Coherence Score
<i>i would imagine this is a problem with any camera of a compact nature</i>	0.00135
<i>and the video quality is amazing</i>	0.00129
<i>this is my third digital elph and it is definitley the best ever</i>	0.00113
<i>the camera was sitting on my home desk waiting for the new digital card and the weekend for the first round of serious shooting</i>	0.00102
<i>but if you want a pocket camera to use when you cant take your digital slr with you that will capture quality images this is your choice of camera</i>	0.00094749

Sedangkan tabel diatas menunjukkan lima kalimat dengan nilai coherence score terendah, dapat dilihat tidak ada keterkaitan sama sekali antara kalimat satu dengan kalimat lainnya. Tidak ada aspek atau topik yang sama didalam kalimat tersebut.

5. Kesimpulan

Berdasarkan analisis terhadap pengujian yang telah dilakukan pada penelitian tugas akhir ini, maka dapat ditarik kesimpulan sebagai berikut :

- a) Hasil performansi pada tahapan ekstraksi menggunakan *pattern matching* berada pada angka sekitar 19% sampai 62% terhadap *dataset* yang digunakan.
- b) Hasil performansi penentuan orientasi opini menggunakan *SentiWordNet* berada pada angka sekitar 51% sampai 71%.
- c) Nilai *content score* mengalami penurunan berdasarkan data uji yang digunakan. Semakin banyak data yang digunakan maka semakin kecil nilai *content score* yang dihasilkan.
- d) Nilai *coherence score* akan menghasilkan kalimat yang saling berkaitan satu sama lain dilihat berdasarkan aspek serta topik yang sama.

6. Daftar Pustaka

- [1] Gangeshwer, D, "E-Commerce or Internet Marketing: A Business Review from Indian Context". International Journal of u- and e-services, Science and Technology, 2013.
- [2] McDonald, Ryan. "A Study of Global Inference Algorithms in Multi-document Summarization". In Proc. of the 29th European Conference on Information Retrieval 2007.
- [3] C. D. Manning, P. Raghavan, H. Schutze, "Introduction to Information Retrieval". 2008.
- [4] V. Balakrishnan and E. L. Yemoh, "Stemming and Lemmatization: A Comparison of Retrieval Performances". Lecture Notes on Software Engineering Vol. 2, No. 3, 2014
- [5] D. Jurafsky and J. H. Martin, "Speech and Language Processing", 2015
- [6] G. Navarro, "Pattern Matching" Department of Computer Science, University of Chile
- [7] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002.
- [8] A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining", Proceedings of the 5th Conference on Language Resources and Evaluation (LREC), 2006