

PERINGKASAN MULTI-DOKUMEN MENGGUNAKAN KOMBINASI FITUR BERDASARKAN CENTROID DAN KEYWORD

Narandha Arya Ranggianto¹⁾, Diana Purwitasari²⁾, Chastine Fatichah³⁾, and Rizka Wakhidatus Sholikhah⁴⁾

^{1, 2), 3), 4)} Departemen Teknik Informatika, Institut Teknologi Sepuluh Nopember, Indonesia
e-mail: narandharanggi@gmail.com¹⁾, diana@if.its.ac.id²⁾, chastine@if.its.ac.id³⁾, wakhidatus@if.its.ac.id⁴⁾

ABSTRAK

Peringkasan teks pada multi-dokumen memerlukan pemilihan kalimat penting lebih kompleks dibandingkan satu dokumen karena terdapat informasi berbeda-beda yang mengakibatkan kontradiksi dan redundansi informasi. Proses pemilihan kalimat penting dapat dilakukan menggunakan skoring kalimat yang mempertimbangkan informasi utamanya. Kombinasi fitur dilakukan untuk proses skoring kalimat sehingga kalimat yang bernilai tinggi menjadi kandidat peringkasan. Pendekatan centroid memberikan keunggulan dalam mendapatkan informasi utama. Namun, pada pendekatan centroid masih terbatas pada informasi yang dekat dengan titik pusat saja. Penambahan fitur posisi memberikan peningkatan informasi terhadap kepentingan sebuah kalimat, tetapi fitur posisi hanya berfokus pada posisi utama. Oleh karena itu, peneliti menggunakan fitur keyword sebagai kontribusi penelitian yang mampu memberikan informasi tambahan kata-kata penting berupa N-gram pada sebuah dokumen. Pada penelitian ini, fitur centroid, posisi, dan keyword digabungkan untuk proses skoring yang dapat memberikan peningkatan performa untuk multi-dokumen data berita maupun ulasan. Hasil uji coba menunjukkan bahwa penambahan fitur keyword menghasilkan nilai tertinggi untuk data berita DUC2004 ROUGE-1 35.44, ROUGE-2 7.64, ROUGE-L 37.02, dan BERTScore 84.22. Sedangkan data ulasan Amazon didapatkan dengan hasil ROUGE-1 32.24, ROUGE-2 6.14, ROUGE-L 34.77, dan BERTScore 85.75. Nilai ROUGE dan BERTScore mengungguli model unsupervised lainnya.

Kata Kunci: peringkasan teks, multi-dokumen, centroid, posisi, keyword

MULTI-DOCUMENT SUMMARIZATION USING A COMBINATION OF FEATURES BASED ON CENTROID AND KEYWORD

Narandha Arya Ranggianto¹⁾, Diana Purwitasari²⁾, Chastine Fatichah³⁾, and Rizka Wakhidatus Sholikhah⁴⁾

^{1, 2), 3), 4)} Departement of Informatics, Institut Teknologi Sepuluh Nopember, Indonesia
e-mail: narandharanggi@gmail.com¹⁾, diana@if.its.ac.id²⁾, chastine@if.its.ac.id³⁾, wakhidatus@if.its.ac.id⁴⁾

ABSTRACT

Summarizing text in multi-documents requires choosing important sentences which are more complex than in one document because there is different information which results in contradictions and redundancy of information. The process of selecting important sentences can be done by scoring sentences that consider the main information. The combination of features is carried out for the process of scoring sentences so that sentences with high scores become candidates for summary. The centroid approach provides an advantage in obtaining key information. However, the centroid approach is still limited to information close to the center point. The addition of positional features provides increased information on the importance of a sentence, but positional features only focus on the main position. Therefore, researchers use the keyword feature as a research contribution that can provide additional information on important words in the form of N-grams in a document. In this study, the centroid, position, and keyword features were combined for a scoring process which can provide increased performance for multi-document news data and reviews. The test results show that the addition of keyword features produces the highest value for news data DUC2004 ROUGE-1 of 35.44, ROUGE-2 of 7.64, ROUGE-L of 37.02, and BERTScore of 84.22. While the Amazon review data was obtained with ROUGE-1 of 32.24, ROUGE-2 of 6.14, ROUGE-L of 34.77, and BERTScore of 85.75. The ROUGE and BERTScore values outperform the other unsupervised models.

Keywords: Text summarization, multi-document, centroid, position, keyword

I. PENDAHULUAN

Peringkasan teks membantu mendapatkan informasi penting secara cepat dari banyaknya informasi teks yang saat ini terus berkembang. Secara umum, proses peringkasan teks harus memiliki karakteristik dalam mendapatkan informasi penting dan mengurangi waktu baca pengguna [1]. Peringkasan dapat digunakan tidak hanya pada teks berita saja, melainkan teks ulasan. Pada teks berita dan teks ulasan terdiri dari beberapa sumber dokumen untuk diringkas yang disebut multi-dokumen [2]. Proses peringkasan pada multi-dokumen

memerlukan proses yang lebih kompleks dibandingkan satu dokumen karena dokumen berasal dari konten yang memiliki informasi berbeda-beda terhadap suatu topik sehingga menimbulkan kontradiksi. Selain itu, informasi yang berasal dari banyaknya sumber akan menghasilkan redundansi informasi [3]. Hal ini menjadikan peringkasan multi-dokumen rentan terhadap pemilihan informasi pentingnya.

Pada penelitian sebelumnya peringkasan multi-dokumen dilakukan berdasarkan graf menggunakan metode Textrank, dimana setiap kalimat direpresentasikan sebagai simpul (*node*) dan hubungan antar kalimat direpresentasikan sebagai tepi (*edge*) untuk tingkatan kepentingan kalimat [4]. Metode ini dapat memahami hubungan antar kalimat namun tidak dapat menjangkau seluruh informasi kalimat yang ada karena terbatas pada *node* yang saling terhubung. Penelitian lain menggunakan pendekatan *centroid* untuk peringkasan multi-dokumen yang dapat menjangkau topik yang beragam [5]. Namun penggunaan satu fitur berdasarkan *centroid* memberikan performa ringkasan yang menurun karena informasi kalimat menjadi redundan. Penelitian selanjutnya menggunakan fitur *centroid* yang ditambahkan dengan fitur posisi kalimat. Hasil akhir dari kombinasi fitur ditotal untuk mendapatkan skoring kalimat akhir yang digunakan sebagai kandidat peringkasan [6]. Penggunaan kombinasi fitur tersebut menghasilkan peningkatan performa dibandingkan fitur *centroid* saja dan Lexrank yang merupakan metode berbasis graf. Hal ini menjadikan kombinasi yang berbasis fitur *centroid* dan fitur lainnya menghasilkan performa yang lebih meningkat.

Pada kombinasi fitur yang berbasis *centroid* telah dilakukan oleh [7], [8]. Dari penelitian tersebut pendekatan fitur *centroid* hanya digabungkan dengan fitur posisi. Fitur posisi memberikan informasi bahwa posisi di awal kalimat memiliki nilai yang lebih tinggi dibandingkan kalimat yang jauh dari posisi pertama [9]. Fitur posisi masih berfokus pada kalimat di posisi pertama dibandingkan posisi lainnya. Sebagai contohnya, terdapat 10 kalimat dalam sebuah dokumen, skor yang dihasilkan untuk posisi kalimat yaitu 0.63, 0.5, ..., 0.5 [10]. Berdasarkan hal tersebut, fitur posisi memberikan informasi yang bias untuk data *sparse* seperti data ulasan. Hal ini dikarenakan data ulasan memiliki topik yang bermacam-macam dan kalimat utama tidak hanya berada di posisi awal saja. Hal tersebut menjadikan skoring kalimat bernilai sama dan menyebabkan kehilangan informasi penting [11]. Selain fitur posisi, terdapat fitur *keyword* yang dapat memberikan informasi yang penting berdasarkan *N*-gram kata. Penelitian sebelumnya menyebutkan bahwa fitur *keyword* dapat memberikan informasi terhadap dokumen yang bias karena dapat menangkap kata-kata penting yang menunjukkan topik utama pada sebuah dokumen [12]. Oleh karena itu, fitur *keyword* menjadi fitur yang penting pada proses peringkasan berbasis skoring.

Pada penelitian sebelumnya, fitur *keyword* telah dilakukan pada peringkasan multi-dokumen dengan menggunakan TF-IDF. Penelitian tersebut masih berbasis sintaktik yang hanya mengandalkan kemunculan sebuah kata pada dokumen [13]. Oleh karena itu, pada penelitian ini berfokus pada penggabungan fitur *centroid* yang berbasis semantik dan fitur posisi serta fitur *keyword* yang berbasis sintaktik. Penambahan fitur *keyword* dilakukan dengan menghitung banyaknya kemunculan *keyword* berdasarkan *N*-gram kata penting dalam sebuah dokumen [14]. Kalimat yang terdiri dari banyak *keyword* akan menghasilkan nilai yang lebih tinggi. Proses perhitungan nilai *keyword* berdasarkan ekstraksi *keyword* menggunakan model YAKE berbasis *unsupervised* dibandingkan menggunakan TF-IDF. YAKE memiliki kelebihan untuk mengekstraksi *keyword* dalam dokumennya sendiri tanpa membutuhkan kamus dari luar seperti perhitungan IDF dan korpus lainnya seperti NER, POS Tagger, maupun stopwords. Ekstraksi *keyword* YAKE juga menghasilkan performa yang lebih baik dibandingkan TF-IDF [15]. Kombinasi antar fitur *centroid*, posisi, dan *keyword* dapat memberikan skoring kalimat dengan lebih informatif dan relevan terhadap hasil peringkasan. Model digunakan untuk data berita maupun data ulasan untuk mendapatkan hasil yang terbaik dari kedua domain tersebut.

II. TINJAUAN PUSTAKA

A. Teks Peringkasan Multi-dokumen

Peringkasan multi-dokumen dapat dilakukan dengan beberapa model yaitu graf, *centroid*, dan kombinasi fitur kalimat. Pada model graf, penelitian sebelumnya menggunakan model Textrank, dimana model tersebut berbasis graf yang biasa digunakan untuk peringkasan. Textrank sendiri memiliki konsep algoritma Pagerank untuk menghitung peringkat setiap kalimat. Pada penelitian tersebut model Textrank digabungkan dengan *Maximal Marginal Relevance* (MMR) untuk mengurangi redundansi kalimat [4]. Namun, pada penelitian tersebut tidak dapat mengatasi masalah terhadap kalimat yang tidak formal pada dokumen yang menyebabkan pengurangan nilai ROUGE sebagai evaluasi ringkasan. Sedangkan model *centroid* memiliki konsep dimana kalimat yang penting berada dekat dengan dengan titik pusat dalam dokumen. Metode *centroid* memiliki kelebihan dalam dapat menangkap informasi sebuah kalimat dengan banyak topik dibandingkan graf karena dapat memahami sebuah konteks menggunakan *embedding* kalimat [16].

Selanjutnya pada kombinasi fitur berbasis *centroid* dilakukan oleh [8], dimana peneliti menggunakan fitur berdasarkan *centroid*: fitur relevansi dan fitur keterbaruan informasi/novelty serta fitur posisi. Kedua fitur utama

diadaptasi dari penelitian sebelumnya dengan menggunakan model *autoencoder* [10]. Kemudian dari fitur-fitur tersebut dikembangkan menjadi pendekatan *centroid*. Kombinasi dari kedua fitur utama tersebut (*centroid* dan posisi) memberikan hasil yang lebih baik dibandingkan Textrank dan Textteaser karena pada kombinasi tersebut menggunakan nilai semantik sebuah kalimat. Pendekatan *centroid* dilakukan dengan mengubah kalimat menjadi bentuk vektor berupa *embedding* kalimat dari model *pre-trained* dibanding menggunakan TF-IDF. Penelitian tersebut membandingkan seluruh *embedding* kalimat yang ada seperti model BERT, ELMO, USE-Transformer, serta USIF.

B. *Embedding* Kalimat

Pada proses peringkasan yang dilakukan pada penelitian sebelumnya menggunakan representasi kalimat TF-IDF. Proses tersebut memilih kalimat-kalimat penting dengan perhitungan kata-kata yang muncul dalam sebuah dokumen (TF). Selain itu, kata yang muncul merupakan kata unik atau kata yang tidak muncul pada dokumen lainnya memberikan nilai yang lebih tinggi (IDF) [6]. Namun, TF-IDF tidak dapat memperhatikan urutan dari kata dan tidak memiliki fungsi semantik sehingga kalimat tidak dapat memahami konteks dari sebuah kalimat. Pada penelitian lain melakukan proses *clustering* yang berbasis *centroid* dengan menggunakan model *embedding* kata Word2Vec. Model yang digunakan pada penelitian tersebut yaitu Word2Vec yang memiliki kelemahan dalam menghadapi *out-of-vocabulary* sehingga terdapat kata yang diabaikan karena *embedding* kata tidak berada pada kamus pada pelatihan [5]. Penelitian lainnya menggunakan model *pre-trained* yaitu *Bidirectional Encoder Representations from Transformers* (BERT). Model BERT memiliki kelebihan untuk mengenali hubungan kontekstual kata-kata berbasis *attention* dibandingkan Word2Vec berbasis *skip-gram* atau *Continuous Bag of Words* (CBOW). Model BERT juga dilatih pada data yang lebih besar sebanyak 3300 juta kata sehingga dapat menjangkau lebih banyak kata [17]. Proses *embedding* pada BERT diawali dengan penambahan token [CLS] dan [SEP] pada akhir kalimat. Setiap token akan dihitung menjadi *embedding* dengan proses *self-attention*. Setiap token akan dihitung rata-rata *embedding* sehingga mendapatkan hasil *embedding* kalimat.

C. Fitur *Centroid*

Pada fitur *centroid* digunakan pada penelitian sebelumnya berdasarkan TF-IDF untuk merepresentasikan sebuah kata pada peringkasan. Penelitian tersebut dilakukan dengan 3 tahapan utama yaitu mendeteksi sebuah topik, membangun klaster dari topik-topik tersebut, dan memilih kalimat yang dekat dengan *centroid* sebagai kandidat peringkasan [18]. Selain menggunakan TF-IDF, fitur *centroid* dilakukan berdasarkan *embedding* kata. Model *embedding* kata menggunakan Word2Vec yang digunakan untuk mencari *centroid* pada peringkasan teks untuk memahami sebuah konteks pada kalimat. Pembangunan fitur *centroid* dilakukan berdasarkan perhitungan *similarity*, *clustering*, pemilihan kalimat berdasarkan kalimat yang terdekat dengan *centroid*, dan pengurutan kalimat ringkasan berdasarkan posisi di dalam dokumen awal [19]. Pada penelitian lain *embedding* kata digunakan untuk proses pencarian *centroid* terhadap kata-kata penting pada dokumen. Penelitian tersebut diawali dengan ekstraksi kata penting pada sebuah dokumen. Kalimat akan dihitung berdasarkan kata penting yang telah diekstrak tersebut. Nilai yang lebih tinggi terhadap kata penting digunakan sebagai kandidat peringkasan. Kedekatan kata dalam setiap kalimat dan kata penting dihitung menggunakan *Word Mover's Distance* (WMD). Hasil dari model tersebut memberikan performa peringkasan yang lebih baik dibandingkan Textrank [20]. Penelitian tersebut memberikan kelemahan dalam pemahaman semantik antar kalimat karena perhitungan kedekatan berdasarkan kata yang sensitif terhadap variasi kata dalam sebuah kalimat. Proses perhitungan *centroid* dapat dilakukan menggunakan *embedding* kalimat yang digunakan untuk menghitung kedekatan kalimat terhadap informasi utama menggunakan *cosine similarity* yang disebut dengan fitur relevansi serta fitur keterbaruan informasi/*novelty* untuk mengurangi informasi yang redundan [8]. Kedua fitur tersebut digunakan pada penelitian ini dikarenakan memiliki keunggulan untuk mendapatkan relevansi serta keterbaruan informasi yang dapat memahami konteks seluruh kalimat.

D. Fitur Posisi

Fitur posisi merupakan fitur yang memberikan informasi terhadap posisi kalimat dari dokumen. Fitur posisi ini memiliki konsep bahwa kalimat penting berada posisi utama pada sebuah dokumen. Hal ini ditunjukkan pada hasil perhitungan fitur posisi pada penelitian sebelumnya yang memberikan nilai yang lebih tinggi dibandingkan kalimat yang jauh dari posisi utama [21]. Fitur posisi ini memiliki konsep yang lebih sederhana untuk peringkasan teks dibandingkan fitur judul, fitur kemunculan angka, dan fitur kemunculan kata benda dalam sebuah dokumen. Apabila kemunculan angka atau kata benda perlu dilakukan proses pengenalan terhadap jenis kata tersebut yang biasanya menggunakan *POS Tagging* [14]. Kesederhanaan fitur ini memiliki kelemahan dimana fitur ini menghasilkan informasi yang bias karena hanya bereferensi pada posisi dokumen dibandingkan fitur lainnya seperti fitur topik dan graf [22]. Hal ini menjadikan fitur hanya memberikan penambahan informasi yang kecil karena tidak selalu kalimat pada dokumen memiliki kalimat penting yang beraturan pada posisi pertama, contohnya data ulasan yang dapat ditemukan di awal, tengah, maupun akhir sebuah dokumen. Namun, pada penelitian lain

menyebutkan penghilangan fitur posisi juga akan mempengaruhi performa pada peringkasan [9]. Oleh karena itu, kondisi terbaik untuk peringkasan teks tetap menggunakan fitur posisi sehingga pada penelitian ini tetap menggunakan fitur posisi untuk menjaga performa peringkasan.

E. Fitur *Keyword*

Penggunaan fitur *keyword* akan memudahkan pemodelan peringkasan memahami teks yang bersifat ambigu [14]. Oleh karena itu, fitur *keyword* menjadikan poin yang bernilai untuk proses penangkapan informasi pada peringkasan multi-dokumen. Penelitian lain menggunakan fitur *keyword* dan fitur kata lainnya seperti kemunculan kata dalam judul dan jumlah kata benda yang sesuai. Peneliti tersebut menggunakan pendekatan TF-IDF untuk melakukan ekstraksi *keyword* berupa kata benda dalam kalimat, dimana kata tersebut akan menunjukkan nilai kepentingan dari sebuah kalimat. Apabila kata benda tersebut muncul dalam sebuah kalimat dan bersifat unik maka menghasilkan nilai yang tinggi dibandingkan kalimat lainnya [13]. Selain TF-IDF, terdapat model ekstraksi *keyword* YAKE yang memiliki performa lebih baik dibandingkan TF-IDF. YAKE merupakan metode *keyword* berbasis statistik dimana memiliki memiliki proses pemrosesan teks, ekstraksi fitur, pembobotan individu, dan pembangunan *keyword*. Fitur yang digunakan pada *keyword* YAKE adalah *casings*, *term position*, *term frequency*, *term relatedness*, dan *term different sentence* [15]. Model ini mendapatkan nilai tertinggi untuk *precision*, *recall*, dan *F1-Score* dibandingkan dengan Textrank, TF-IDF, Singlerank, dan RAKE untuk ekstraksi *keyword*.

III. METODOLOGI PENELITIAN

A. *Pre-processing*

Secara umum, *pre-processing* merupakan proses untuk membersihkan teks dari karakter yang tidak bermakna sehingga data siap untuk diproses pada model peringkasan. *Pre-processing* pada model peringkasan multi-dokumen terdiri dari klaster D dokumen yang berisi i dokumen sehingga dinotasikan $D = \{d_i\}$. Tahapan pertama setiap dokumen d_i dari klaster D akan dipisahkan menjadi sebuah kalimat-kalimat atau tokenisasi kalimat berdasarkan tanda titik (.). Setelah proses tokenisasi, setiap kalimat akan dibersihkan dengan proses penghapusan karakter dan *casefolding*.

Penghapusan karakter dilakukan untuk menghilangkan karakter spesial/symbol, tanda baca kecuali tanda titik (.) karena digunakan sebagai tokenisasi, spasi/symbol/tanda baca yang lebih dari satu, XML/HTML tag, dan URL. Proses *casefolding* merupakan perubahan kata menjadi huruf kecil yang bertujuan untuk menghilangkan perbedaan makna antara huruf besar maupun huruf kecil pada model peringkasan. Setiap tahapan *pre-processing* dapat dilihat pada Tabel I, dimana tabel tersebut merupakan contoh hasil *pre-processing* pada data Amazon. Contoh yang digunakan sebanyak 2 dokumen ulasan dalam satu klaster D . Pada hasil tokenisasi terdapat 2 kalimat untuk d_1 dan 2 kalimat untuk d_2 . Pada penghapusan karakter, dokumen d_1 terdapat spasi yang berlebihan diantara kata “thin” dan “but” sehingga perlu dihapus, sedangkan pada contoh d_2 dilakukan penghapusan tanda seru yang berlebihan (!!). Selanjutnya proses *casefolding* untuk mengubah bentuk kecil sebuah kata. Keseluruhan hasil *pre-processing* ditunjukkan pada Tabel I.

B. Model Kombinasi Fitur

Gambar 1 merupakan model kombinasi fitur yang digunakan untuk proses peringkasan data ulasan maupun data berita. Secara keseluruhan pada gambar tersebut menunjukkan proses ringkasan berdasarkan kombinasi fitur untuk satu klaster dokumen, dimana pada satu klaster terdiri dari kumpulan dokumen. Setiap dokumen akan dilakukan *pre-processing*, dimana tahapan awal untuk *pre-processing* dilakukan proses tokenisasi yang berfungsi untuk memecah dokumen menjadi kalimat-kalimat. Setelah proses tokenisasi maka terkumpul kalimat-kalimat yang ditunjukkan pada Tabel I. Setiap kalimat akan dihitung skornya berdasarkan kombinasi fitur yang ditunjukkan garis putus-putus pada Gambar 1. Kombinasi fitur tersebut terdiri dari fitur relevansi, fitur keterbaruan informasi/*novelty*, fitur posisi, dan fitur *keyword*. Pada fitur relevansi dan fitur keterbaruan informasi/*novelty* merupakan uraian dari fitur *centroid* dimana menggunakan nilai *similarity* terhadap nilai *centroid* yang didapatkan dari total *embedding* dari seluruh kalimat pada klaster D . Fitur posisi menggunakan informasi indeks posisi dari dokumen ke- i . Pada contohnya kalimat “my son is 3 and this fits him perfectly” memiliki indeks posisi ke-1 yang bereferensi pada d_1 sedangkan kalimat “he loved it” memiliki indeks posisi ke-2 yang bereferensi pada d_2 . Sedangkan fitur *keyword* diawali dengan proses ekstraksi *keyword* $N - gram$ menggunakan model YAKE berdasarkan kalimat untuk setiap dokumen ke- i . Setelah mendapatkan ekstraksi *keyword* maka kalimat yang terdiri dari unsur *keyword* berupa $N - gram$ akan memiliki total bobot dari *keyword* yang didapatkan dari model YAKE. Setelah semua fitur didapatkan, maka proses terakhir yaitu perhitungan nilai akhir dengan menggunakan kombinasi *linier*. Skor tertinggi akan dijadikan sebagai kalimat ringkasan. Hasil ringkasan akan dievaluasi dengan membandingkan *ground truth* dari ringkasan.

C. Fitur Relevansi

Setiap kalimat $s_{t,i}$, dimana t merupakan indeks kalimat dan i merupakan indeks dokumen yang akan dilanjutkan pada proses skoring menggunakan kombinasi fitur. Pada fitur relevansi dan fitur keterbaruan informasi/*novelty*, membutuhkan nilai *centroid* untuk setiap dokumen $s_{t,i}$. Pada proses ini, kalimat akan diubah menjadi *embedding* kalimat dari BERT yang memiliki ukuran dimensi 768. Fitur relevansi didapatkan dari kedekatan antara *embedding* kalimat dengan *centroid* dari kalimat-kalimat yang berada di klaster D yang dihitung dengan (1). Pada penelitian ini mengadaptasi persamaan fitur relevansi dari [8], apabila kalimat dekat dengan *centroid* maka kalimat memiliki informasi yang dekat dengan topik utama. Kedekatan antara kalimat dengan *centroid* menggunakan nilai *cosine similarity* yang dinotasikan sebagai sim .

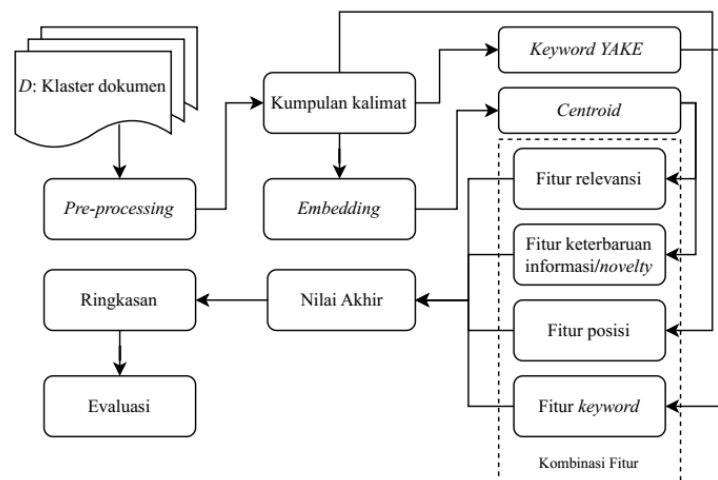
$$relevansi(s_{t,i}) = sim(s_{t,i}, D_{centroid}) \tag{1}$$

D. Fitur Keterbaruan Informasi/*Novelty*

Fitur keterbaruan informasi/*novelty* berfungsi untuk mendapatkan informasi baru dan mengurangi redundansi dari ringkasan. Fitur ini dihitung berdasarkan rumus dari [8] yaitu menggunakan nilai *cosine similarity* yang dinotasikan sebagai sim untuk menghitung kalimat $s_{t,i}$ dengan seluruh kalimat pada dokumen yang dinotasikan sebagai $s_{k,i}$, dimana $1 \leq k \leq |D|, k \neq t$, k merupakan indeks kalimat selain kalimat $s_{t,i}$, dan $|D|$ merupakan banyaknya kalimat yang ada di klaster D . Kemudian maksimal nilai yang didapatkan dari kesamaan $sim(s_{t,i}, s_{k,i})$ akan dibandingkan dengan *threshold*. Kondisi pertama menyatakan bahwa nilai keterbaruan informasi/*novelty* tinggi yaitu bernilai 1, apabila maksimal $sim(s_{t,i}, s_{k,i})$ dibawah *threshold*. Kondisi kedua menyatakan bahwa $sim(s_{t,i}, s_{k,i})$ lebih dari *threshold*, maka akan dilanjutkan pada perhitungan nilai relevansi. Indeks kalimat untuk hasil dari $sim(s_{t,i}, s_{k,i})$ tertinggi dinotasikan $s_{j,i}$, dimana j merupakan indeks kalimat dengan nilai $sim(s_{t,i}, s_{k,i})$ tertinggi. Jika nilai relevansi dari kalimat $s_{t,i}$ lebih dari $s_{j,i}$ maka nilai keterbaruan informasi/*novelty* akan tinggi. Kondisi ketiga ketika tidak memenuhi kondisi sebelumnya maka dihitung dengan persamaan $1 - \max(sim(s_{t,i}, s_{k,i}))$. Nilai *novelty* secara keseluruhan ditunjukkan pada persamaan (2), dimana $s_{j,i}$ merupakan hasil dari *argmax* yang bertujuan untuk mendapatkan indeks kalimat dari nilai maksimal $sim(s_{t,i}, s_{k,i})$ dan τ_n merupakan *threshold* untuk nilai *novelty*, dimana menggunakan nilai 0.95 [8].

TABEL I
CONTOH HASIL *PRE-PROCESSING* PADA DATA AMAZON

<i>Pre-processing</i>	Data Asli	Hasil <i>Pre-processing</i>
Tokenisasi	d_1 : My son is 3 and this fits him perfectly. The hat is thin, but completes the outfit.	My son is 3 and this fits him perfectly. The hat is thin, but completes the outfit.
	d_2 : I ordered this for my 3 yr old for Halloween. He loved it!!	I ordered this for my 3 yr old for Halloween. He loved it!!
Penghapusan karakter	My son is 3 and this fits him perfectly.	My son is 3 and this fits him perfectly
	The hat is thin, but completes the outfit.	The hat is thin but completes the outfit.
	I ordered this for my 3 yr old for Halloween. He loved it!!	I ordered this for my 3 yr old for Halloween He loved it
Casefolding	My son is 3 and this fits him perfectly	my son is 3 and this fits him perfectly
	The hat is thin but completes the outfit.	the hat is thin but completes the outfit
	I ordered this for my 3 yr old for Halloween	i ordered this for my 3 yr old for halloween
	He loved it	he loved it



Gambar 1. Model Kombinasi Fitur untuk Peringkasan Multi-dokumen

$$novelty(s_{t,i}) = \begin{cases} 1, & \text{jika } \max (sim(s_{t,i}, s_{k,i})) < \tau_n, 1 \leq k \leq |D|, k \neq t \\ 1, & \text{jika } \max (sim(s_{t,i}, s_{k,i})) > \tau_n \text{ dan } relevansi(s_{t,i}) > relevansi(s_{j,i}), j = \\ & argmax(sim(s_{t,i}, s_{k,i})) \\ 1 - \max (sim(s_{t,i}, s_{k,i})), & \text{jika tidak} \end{cases} \quad (2)$$

E. Fitur Posisi

Fitur posisi kalimat dari sebuah dokumen memiliki konsep bahwa kalimat pertama merupakan kalimat yang paling penting. Nilai akan lebih kecil apabila kalimat berada jauh dengan posisi pertama. Rumus untuk nilai posisi diadaptasi dari [10] dimana $p(s_{t,i}, d_i)$ merupakan posisi kalimat $s_{t,i}$ pada sebuah dokumen d_i , dimana d_i merupakan dokumen untuk kalimat $s_{t,i}$ dan $|d_i|$ merupakan jumlah kalimat pada dokumen ke- i yang ditunjukkan pada persamaan (3). Rentang nilai posisi yaitu [0.5, 1].

$$posisi(s_{t,i}) = \max (0.5, \exp \left(\frac{-p(s_{t,i}, d_i)}{\sqrt[3]{|d_i|}} \right)) \quad (3)$$

F. Fitur Keyword

Pada fitur *keyword*, proses diawali dengan ekstraksi *keyword* dari hasil metode YAKE. Proses metode YAKE berbasis kalimat untuk setiap kluster D yang menghasilkan beberapa *keyword* yang dinotasikan sebagai *keyword* $k_{c,i}$ dan bobot *keyword* $v_{c,i}$, dimana c merupakan indeks *keyword* dari dokumen i . *Keyword* $k_{c,i}$ merupakan kumpulan $N - gram$ *keyword* untuk dokumen d_i . Setiap kalimat $s_{t,i}$ akan dicek berupa $N - gram$, dimana $N - gram$ yang dimaksud adalah kumpulan kata yang berjumlah $N = 1, N = 2$, dan $N = 3$ yang dinotasikan sebagai $N - gram(s_{t,i})$. Apabila $N - gram(s_{t,i})$ sama dengan $N - gram$ *keyword* $k_{c,i}$, maka kalimat tersebut memiliki bobot $v_{c,i}$. Notasi matematika untuk kondisi diatas yaitu $V(s_{t,i}) = \{v_{c,i} \mid (k_{c,i}, v_{c,i}) \in YAKE(d_i)\}$, dimana $V(s_{t,i})$ merupakan himpunan bobot dari YAKE yang digunakan sebagai skoring kalimat $s_{t,i}$. Perhitungan akhir untuk mendapatkan fitur *keyword* ditunjukkan pada persamaan (4).

$$keyword(s_{t,i}) = \sum_{V(s_{t,i})=\{v_{c,i} \mid (k_{c,i}, v_{c,i}) \in YAKE(d_i)\}} V(s_{t,i}) \quad (4)$$

G. Nilai Akhir

Setelah mendapatkan fitur relevansi, fitur keterbaruan informasi/*novelty*, fitur posisi, dan fitur *keyword*, maka keempat fitur tersebut akan dihitung untuk nilai akhir menggunakan kombinasi *linier*. Setiap nilai akan dikalikan dengan bobot dimana α untuk fitur relevansi, β untuk fitur keterbaruan informasi/*novelty*, γ untuk fitur posisi, dan δ untuk fitur *keyword* dengan ketentuan $\alpha + \beta + \gamma + \delta = 1$. Persamaan untuk menghitung nilai akhir kalimat ditunjukkan pada persamaan (5). Setiap kalimat memiliki nilai akhir sehingga kalimat dengan nilai tertinggi akan digunakan sebagai kalimat ringkasan sejumlah kalimat yang telah ditentukan.

$$nilai_akhir(s_{t,i}) = \alpha \times relevansi(s_{t,i}) + \beta \times novelty(s_{t,i}) + \gamma \times posisi(s_{t,i}) + \delta \times keyword(s_{t,i}) \quad (5)$$

H. Evaluasi

Evaluasi yang digunakan untuk mengukur model peringkasan yaitu *Recall-Oriented Understudy for Gisting Evaluation* berupa ROUGE-1, ROUGE-2, dan ROUGE-L sesuai dengan penelitian [8]. ROUGE-1 dan ROUGE-2 mengukur *unigram* dan *bigram* yang *overlap* antara ringkasan prediksi dan *ground truth* atau referensi ringkasan. Sedangkan ROUGE-L merupakan pengukuran *longest common subsequence*, dimana pengukuran berdasarkan rangkaian kata terpanjang yang sama antara ringkasan prediksi dengan *ground truth*. Secara matematis nilai ROUGE ditunjukkan pada (6), dimana N merupakan panjang $N - gram$, $jumlah_{overlap}(N - gram)$ merupakan jumlah maksimum dari $N - gram$ yang muncul pada ringkasan prediksi dan *ground truth*, $jumlah(N - gram)$ merupakan jumlah $N - gram$ pada ringkasan sebagai referensi. ROUGE diukur berdasarkan nilai *F1-Score* yang merupakan rata-rata harmonik antara *recall* dan *precision*. Semakin tinggi nilai ROUGE maka semakin banyak kata yang sesuai dengan *ground truth*. Nilai ROUGE memiliki rentang nilai 0-1. Namun, nilai tersebut akan dikonversi menjadi persentase untuk memudahkan dalam interpretasi sehingga menjadi 0%-100%. Selain nilai ROUGE, pada penelitian ini menggunakan tambahan evaluasi yaitu BERTScore yang biasa

digunakan sebagai pengukuran pembangunan kalimat [23]. BERTScore merupakan evaluasi berbasis kontekstual dari *embedding* kalimat menggunakan model BERT. Evaluasi ini menggunakan metode *greedy* untuk mencari kedekatan *cosine similarity* antara ringkasan prediksi dengan *ground truth*.

$$ROUGE - N = \frac{\sum_{S \in Summ_{ref}} \sum_{N-gram \in S} jumlah_{overlap}(N - gram)}{\sum_{S \in Summ_{ref}} \sum_{N-gram \in S} jumlah(N - gram)} \quad (6)$$

IV. HASIL DAN PEMBAHASAN

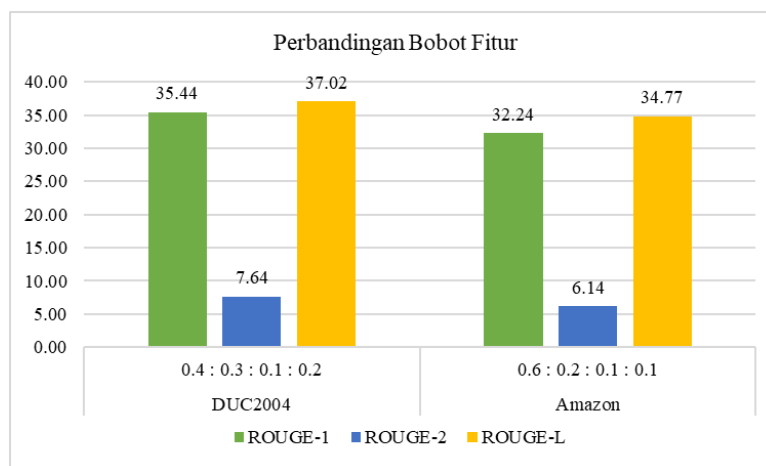
A. Dataset

Pada penelitian ini, dataset yang digunakan terdiri 2 jenis yaitu DUC2004 dan Amazon yang digunakan sebagai peringkasan multi-dokumen. DUC 2004 merupakan data teks berita yang telah digunakan oleh [8], sedangkan Amazon merupakan dataset untuk ulasan dari penelitian [24]. DUC 2004 dibuat oleh NIST yang biasa digunakan untuk mengevaluasi teks ringkasan. Tabel II merupakan deskripsi dari data DUC2004 dan Amazon. Pada data DUC2004 Task 2 terdiri dari 50 kluster, dimana setiap kluster terdiri dari 10 dokumen sehingga total dokumen yang ada yaitu 500. Jumlah *ground truth* untuk DUC2004 yaitu 4 yang dibuat oleh pakar yang berbeda. Sedangkan pada data Amazon merupakan data ulasan pada suatu produk yang terdiri dari kategori: elektronik, baju/sepatu/perhiasan, rumah/dapur, dan kesehatan. Jumlah data terdiri dari 60 produk dimana setiap produk memiliki jumlah 8 ulasan sehingga total data untuk Amazon adalah 480. Jumlah *ground truth* pada data Amazon yaitu 3 yang dilakukan oleh pakar yang berbeda.

B. Uji Coba Pembobotan Fitur

Ketika melakukan perhitungan nilai akhir, setiap fitur akan dikalikan dengan bobot masing-masing dimana α untuk fitur relevansi, β untuk fitur keterbaruan informasi/*novelty*, γ untuk fitur posisi, dan δ untuk fitur *keyword* yang ditunjukkan pada persamaan (5). Jumlah kalimat yang digunakan sebanyak 5 kalimat berdasarkan penelitian sebelumnya [8]. Pada penelitian ini melakukan pengujian terhadap bobot untuk mendapatkan nilai ROUGE-1, ROUGE-2, dan ROUGE-L yang terbaik. Proses pembobotan dilakukan menggunakan konsep dari [8] yaitu nilai α memiliki nilai lebih tinggi dibandingkan β , γ , dan δ . Sedangkan nilai β lebih besar dibandingkan γ dan δ . Jumlah α , β , γ , dan δ akan bernilai 1 sehingga pada nilai α dan β memiliki rentang nilai [0, 1] dengan nilai kenaikan konstan 0.1. Sedangkan untuk nilai γ dan δ menggunakan nilai yang lebih rendah dengan rentang nilai [0.05, 0.2] dengan kenaikan konstan 0.05. Dari hasil uji coba didapatkan 23 kombinasi untuk mendapatkan nilai yang optimal dalam pembobotan.

Pada Gambar 2 ditunjukkan hasil perbandingan bobot fitur untuk DUC2004 (bagian kiri) dan Amazon (bagian kanan). Berdasarkan gambar tersebut, nilai perbandingan bobot fitur untuk DUC2004 memiliki nilai yang lebih tinggi dibandingkan nilai dari Amazon. Selain itu, pada Gambar 2 menunjukkan hasil bobot optimal yang berbeda dari setiap data. Pada data DUC2004 bobot optimal didapatkan dengan perbandingan $\alpha = 0.4$, $\beta = 0.3$, $\gamma = 0.1$, dan $\delta = 0.2$. Hal ini menunjukkan bahwa fitur *keyword* memiliki kontribusi yang lebih tinggi dibandingkan dengan fitur posisi. Sedangkan pada data Amazon nilai optimal didapatkan dari kombinasi $\alpha = 0.6$, $\beta = 0.2$, $\gamma = 0.1$, dan $\delta = 0.1$. Kombinasi tersebut menunjukkan bahwa bobot untuk fitur posisi dan fitur *keyword* memiliki kontribusi yang sama.



Gambar 2. Perbandingan Bobot Fitur untuk DUC2004 dan Amazon

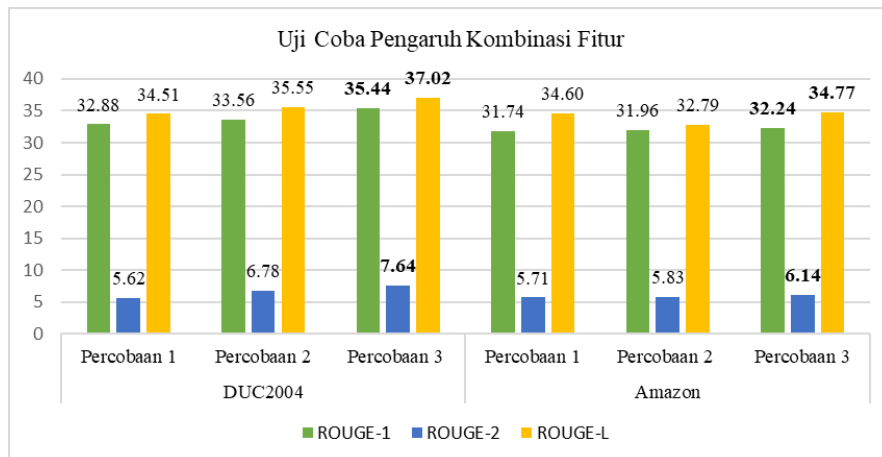
TABEL II

DESKRIPSI DATA DUC2004 DAN AMAZON

Data	Domain	Klaster	Dokumen	Jumlah <i>Ground Truth</i>
DUC2004	Berita	50	500	4
Amazon	Ulasan	60	480	3

C. Uji Coba Pengaruh Fitur *Keyword*

Uji coba pertama dilakukan untuk mengevaluasi performa pada keempat fitur yang digunakan pada penelitian ini. Tujuan dari uji coba ini adalah melihat pengaruh kombinasi fitur yang digunakan. Skenario yang dilakukan yaitu: i) percobaan 1, kombinasi fitur relevansi dan keterbaruan informasi/*novelty*, ii) percobaan 2, kombinasi fitur relevansi, fitur keterbaruan informasi/*novelty*, dan fitur posisi, iii) percobaan 3, kombinasi fitur relevansi, fitur keterbaruan informasi/*novelty*, fitur posisi, dan fitur *keyword*. Uji coba dilakukan untuk keseluruhan data DUC2004 dan Amazon.



Gambar 3. Pengaruh Kombinasi Fitur untuk DUC2004 dan Amazon

Gambar 3 menunjukkan perbandingan hasil uji coba untuk kombinasi fitur yang digunakan, dimana pada nilai yang ditulis tebal yang menandakan nilai tersebut merupakan nilai tertinggi dibandingkan percobaan lainnya. Pada gambar tersebut terdapat 3 percobaan yang terdiri dari kombinasi fitur relevansi dan keterbaruan informasi/*novelty* (percobaan 1), kombinasi fitur relevansi, fitur keterbaruan informasi/*novelty*, dan fitur posisi (percobaan 2), dan kombinasi fitur relevansi, fitur keterbaruan informasi/*novelty*, fitur posisi, dan fitur *keyword* (percobaan 3). Pada Gambar 3 dapat dilihat bahwa diagram tertinggi didapatkan oleh percobaan 3 yaitu kombinasi fitur relevansi, fitur keterbaruan informasi/*novelty*, fitur posisi, dan fitur *keyword* untuk data DUC2004 maupun Amazon. Selain itu, pada label diagram ditulis tebal yang merupakan nilai tertinggi dibandingkan percobaan lainnya. Sedangkan pada percobaan 2 memiliki peningkatan performa dibandingkan percobaan 1, namun percobaan 2 tidak dapat mengungguli percobaan 3. Nilai ROUGE untuk percobaan 3 untuk data DUC2004 adalah ROUGE-1 35.44, ROUGE-2 7.64, dan ROUGE-L 37.02. Sedangkan pada data Amazon adalah ROUGE-1 32.24, 6.14, dan 34.77. Hal ini menunjukkan bahwa penambahan fitur *keyword* memberikan pengaruh positif terhadap pemodelan kombinasi fitur.

D. Hasil Evaluasi dengan Penelitian Sebelumnya

Pada uji coba ini, peneliti melakukan perbandingan antara model kombinasi fitur *centroid*, fitur posisi, dan fitur *keyword* (*centroid* + posisi + *keyword*) dengan penelitian sebelumnya yaitu model *unsupervised*: Textteaser [25], Textrank [26], dan BERT Ekstraktif [27]. Textrank merupakan model berbasis graf yang melakukan perhitungan berdasarkan nilai kepentingan dari sebuah kalimat. Textteaser merupakan model statistik yang biasa digunakan sebagai ringkasan berita. Kemudian BERT Ekstraktif yang merupakan peringkasan berbasis model *pre-trained* BERT. Ketiga metode diimplementasikan menggunakan *library* Python textteaser¹, summa-textrank², dan bert-extractive-summarizer³.

Tabel III menunjukkan hasil dari perbandingan model dengan metode *unsupervised* yang digunakan oleh penelitian sebelumnya. Nilai pada tabel yang ditulis tebal menunjukkan bahwa nilai yang tertinggi dibandingkan nilai lainnya. Pada tabel tersebut terdiri dari dua data, dimana model kombinasi fitur *centroid*, posisi, dan *keyword* (*centroid* + posisi + *keyword*) dibandingkan dengan model Textteaser, Textrank, dan BERT Ekstraktif. Data pertama merupakan data DUC2004, model tertinggi didapatkan oleh *centroid* + posisi + *keyword* dengan nilai

¹ <https://github.com/IndigoResearch/textteaser>

² <https://github.com/summanlp/textrank>

³ <https://github.com/dmmiller612/bert-extractive-summarizer>

ROUGE-1 35.44, ROUGE-2 7.64, ROUGE-L 37.02, dan BERTScore 84.22. Data kedua yaitu data Amazon, dimana nilai tertinggi didapatkan oleh model *centroid* + posisi + *keyword* dengan nilai ROUGE-1 32.24, ROUGE-2 6.24, ROUGE-L 34.77, dan BERTScore 85.75. Dari kedua data, urutan pertama dihasilkan pada model *centroid* + posisi + *keyword*, sedangkan pada urutan kedua didapatkan oleh model BERT Ekstraktif. Hal tersebut menunjukkan bahwa model *centroid* + posisi + *keyword* memberikan keunggulan hasil ringkasan dilihat dari nilai ROUGE dan BERTScore dibandingkan model Textteaser, Textrank, dan BERT Ekstraktif.

Berdasarkan hasil yang ditunjukkan pada Tabel III, model Textteaser hanya berdasarkan nilai sintaktik yang mengabaikan konteks sebuah kalimat sehingga menghasilkan nilai yang rendah dibandingkan model lainnya. Sedangkan pada model Textrank memiliki kemampuan pemahaman konteks berbasis graf namun tidak memberikan hasil optimal karena menghasilkan konteks yang terbatas. Pada BERT Ekstraktif memberikan hasil yang lebih baik dibandingkan Textteaser dan Textrank karena dapat memahami konteks pada peringkasan menggunakan model *pre-trained*. Namun BERT Ekstraktif tidak masih rendah dibandingkan dengan kombinasi fitur dari *centroid* + posisi + *keyword*. Kombinasi fitur *centroid* + posisi + *keyword* menghasilkan informasi yang baik dimana selain memahami konteks kalimat, model dapat menggunakan informasi posisi untuk mendapatkan kalimat utama, dan *keyword* sebagai informasi kata-kata penting pada kalimat.

TABEL III
PERBANDINGAN HASIL RINGKASAN DENGAN PENELITIAN SEBELUMNYA

Model	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore	Data
Textteaser	31.10	5.78	34.09	83.90	DUC2004
Textrank	33.37	6.78	33.90	83.90	
BERT Ekstraktif	33.71	7.06	34.76	83.92	
<i>Centroid</i> + posisi + <i>keyword</i>	35.44	7.64	37.02	84.22	
Textteaser	30.75	5.49	32.22	85.11	Amazon
Textrank	31.00	5.47	33.91	85.55	
BERT Ekstraktif	31.18	5.48	29.94	85.57	
<i>Centroid</i> + posisi + <i>keyword</i>	32.24	6.14	34.77	85.75	

E. Hasil Peringkasan

Pada bagian ini, peneliti menunjukkan hasil ringkasan yang diperoleh dari model kombinasi fitur *centroid*, fitur posisi, dan fitur *keyword* (*centroid* + posisi + *keyword*). Pada Tabel IV merupakan hasil ringkasan untuk data DUC2004 untuk model *centroid* + posisi + *keyword* dan model BERT Ekstraktif. Kalimat ringkasan yang didapatkan dari masing-masing model berjumlah 5 kalimat. Pemilihan hasil BERT Ekstraktif pada Tabel IV dikarenakan BERT Ekstraktif memiliki nilai tertinggi dibandingkan model *unsupervised* lainnya. Kemudian pada kolom *ground truth* menunjukkan hasil yang diinginkan yang dibuat oleh *annotator* yang berjumlah 3. Hasil model *centroid* + posisi + *keyword* menghasilkan ringkasan yang informatif dibandingkan dengan BERT Ekstraktif. Hal tersebut ditunjukkan pada kalimat berwarna jingga yang berisi tentang “peledakan bom yang terjadi di Jerusalem”. Pembahasan tentang informasi tersebut juga sesuai dengan *ground truth* ke-3. Sedangkan model BERT Ekstraktif tidak dapat menangkap informasi tersebut.

Hasil Peringkasan untuk Amazon ditunjukkan pada Tabel V, dimana peneliti juga melakukan perbandingan model *Centroid* + posisi + *keyword* dengan BERT Ekstraktif sama seperti Tabel IV untuk data DUC2004. Kalimat ringkasan yang didapatkan dari masing-masing model berjumlah 5 kalimat. Pada Tabel V menunjukkan kolom *ground truth* sebagai hasil ringkasan yang diinginkan, dimana terdapat jumlah *ground truth* sebanyak 3. Warna jingga pada Tabel V menunjukkan informasi yang dihasilkan pada model *centroid* + posisi + *keyword* dimana tidak ditemukan pada model BERT Ekstraktif. Hal tersebut menunjukkan ringkasan yang lebih informatif. Informasi tersebut menyebutkan bahwa “*It may fit tight on guys with large thighs*” yang menjelaskan bahwa pengguna menilai bahwa produk setelan cocok untuk orang yang memiliki bentuk badan yang besar. Hal ini sesuai dengan *ground truth* yang memiliki informasi yang sama terhadap hasil prediksi oleh model *centroid* + posisi + *keyword* yang ditunjukkan pada *ground truth* ke-2.

TABEL IV
HASIL RINGKASAN DUC2004 DENGAN PENELITIAN SEBELUMNYA

BERT Ekstraktif	<i>Centroid</i> + posisi + <i>keyword</i>	<i>Ground Truth</i>
<ul style="list-style-type: none"> israel's cabinet announced within hours of a market bombing friday that it will put off a vote indefinitely on whether to ratify the wye river accord until palestinians 	<ul style="list-style-type: none"> a car rigged with explosives blew up friday morning in jerusalem's mahane yehuda market packed with israelis shopping for the jewish sabbath, killing two people 	On Nov. 6, 1998 a suicide auto-bomb attack on a Jerusalem market killed two Palestinian "martyrs" and wounded 21 Israelis. Israel's cabinet immediately suspended consideration of the Wye peace agreement and the Prime Minis-

<p>crack down further on terrorism.</p> <ul style="list-style-type: none"> • in response to the attack, the israeli cabinet suspended ratification of the wye agreement until there `is verification that the palestinian authority is indeed fighting terrorism.' • prime minister benjamin netanyahu had held up the work in recent months in an apparent attempt to not disrupt negotiations with the palestinians. • the radical group islamic jihad claimed responsibility saturday for the market bombing and vowed more attacks to try to block the new peace accord. • itaf alayan, a supporter of the group who runs the islamic purity nursery in bethlehem, said palestinian officers had raided the school and confiscated videotapes and educational cassettes before locking it up. 	<p>and wounding 21.</p> <ul style="list-style-type: none"> • the islamic militant group hamas, which has tried to stop the peace agreement, claimed responsibility, police said. • israel's cabinet announced within hours of a market bombing friday that it will put off a vote indefinitely on whether to ratify the wye river accord until palestinians crack down further on terrorism. • the radical group islamic jihad claimed responsibility saturday for the suicide bombing of a crowded jerusalem market and promised more attacks to try to block the new peace accord. • groundbreaking at the site led to angry palestinian protests and a breakdown of peace talks in march 1997, netanyahu's cabinet delayed action on the latest israeli-palestinian accord following friday's suicide bombing at a jerusalem market, and his remarks about building on har homa may be seen as a provocation by the palestinians at a politically sensitive moment. 	<p>ter vowed to expand Jewish settlement in the Arab sector of Jerusalem. Islamic Jihad claimed responsibility for the bombing and identified the Wye accord as its target. Israel called for outlawing Islamic Jihad and Hamas while the Palestinians accused Israel as using the bombing as a pretext for delaying implementation of Wye. The "martyrs", 21 and 24, were both alumni of Israeli jails.</p> <p>After a bombing in a Jerusalem market Fri., the Israeli Cabinet postponed indefinitely its vote on the Wye River peace accord. At first, Hamas claimed responsibility for attack in which 2 suicide bombers were killed and 24 people were hurt. On Sat., Islamic Holy War (Islamic Jihad) took credit for attack, vowing more to block the accord. Israel would not debate or vote on the accord until Palestinians took steps to stop terrorism and outlaw military wings of radical groups. Israel intends to continue building homes in Jerusalem, including in Arab sector. One of the Palestinian suicide bombers had spent much of his teen years in Israeli prisons.</p> <p>A car bomb exploded prematurely near a busy Jerusalem market killing the two suicide bombers and injuring 21 Israelis. One of the Islamic Jihad "martyrs" had said he was "going to Paradise". The Israeli cabinet was meeting at the time of the explosion and they put off ratification of the Wye River "land for security" accord. The Islamic Jihad promises more attacks in hope of derauling the accord. The US expects, at least hopes, for ratification. PM Netanyahu vows to continue Jewish building in Jerusalem and Israel demanded that radical Islamic groups be outlawed. The Palestinian Authority condemned the attack and claimed to have already made some arrests</p> <p>The Wye River accord has not been implemented. As the Israeli cabinet was considering the agreement, Islamic Jihad militants exploded a car bomb in nearby Mahane Yehuda market. The cabinet suspended ratification of the agreement, demanding the Palestinian Authority take steps against terrorism. Further, after the bombing, Israeli Prime Minister Netanyahu announced the resumption of construction of a new settlement, Har Homa, in a traditionally Arab area east of Jerusalem. Israel also demands that Arafat outlaw the military wings of Islamic Jihad and Hamas. The attack injured 24 Israelis, but only the two assailants, Sughayer and Tahayneh,</p>
---	--	--

were killed.

TABEL V
HASIL RINGKASAN AMAZON DENGAN PENELITIAN SEBELUMNYA

BERT Ekstraktif	Centroid + posisi + keyword	Ground Truth
<ul style="list-style-type: none"> although the quality of construction seems adequate the sizing isnt accurate. the suit seemed ok in quality however it was the wrong size. it is always a gamble buying online even following the sizing chart. i bought a size small so that i would not have to worry about them falling off but i should have bought my correct size as they fit true to size. the only reason for the 1star reduction is the seams around the legs are pretty stiff and tight. 	<ul style="list-style-type: none"> although the quality of construction seems adequate the sizing isnt accurate. the size on the waist was correct but if you have large thighs like i do or should i say muscular thighs it will be a tight fit. i ordered a size 34 which is the size i normally wear and the suit not only pinches around my legs but also is almost uncomfortably tight around the waist. i wear a 32 and i read some of the reviews and yes it does have a nice feel and fit but does run small with that being said i ordered a 34 and it was a nice fit and feel but i think a 36 would have been a better fit i bought a size small so that i would not have to worry about them falling off but i should have bought my correct size as they fit true to size. 	The suit fits ok on me, it is good around the waist but tight in the legs/thighs. I will probably return this and move up a size, it's easier to use the drawstring on the waist than deal with a swimsuit strangling my thighs. Consider a larger size than normal if you have large thighs.
		This suit is a good fit for me. It may fit tight on guys with large thighs. I like the feel of this suit and the style is good for anyone swimming as a workout or doing any real training. It may run a bit small though it does seem to stretch at the seams with use and wear.
		The suit tends to run a bit small as the fitting in the waist was adequate but it was tight in the thighs. After heavy use the seams tend to loosen up and even loose the elasticity over time. Overall it's a nice suit, has a nice feel, and is nicely constructed.

V. KESIMPULAN

Peneliti membangun model kombinasi fitur yang terdiri dari fitur *centroid*, fitur posisi, dan fitur *keyword*. Berdasarkan kelemahan fitur *centroid* untuk mengatasi informasi yang sama dan fitur posisi yang hanya berfokus pada posisi utama, maka peneliti berkontribusi untuk menambahkan fitur *keyword* untuk kombinasi fitur peringkasan multi-dokumen. Pembobotan fitur yang optimal ketika fitur *keyword* memiliki bobot yang lebih tinggi dibandingkan fitur posisi atau fitur *keyword* memiliki bobot yang sama dengan fitur posisi. Berdasarkan pengujian yang dilakukan pada data berita DUC2004 didapatkan hasil ROUGE-1 35.44, ROUGE-2 7.64, ROUGE-L 37.02, dan BERTScore 84.22. Sedangkan data ulasan Amazon didapatkan dengan hasil ROUGE-1 32.24, ROUGE-2 6.14, ROUGE-L 34.77, dan BERTScore 85.75. Nilai tersebut merupakan nilai tertinggi dibandingkan model *unsupervised* lainnya yang telah dilakukan pada penelitian sebelumnya yaitu Textrank, Textteaser, dan BERT Ekstraktif. Oleh karena itu, penggunaan fitur *keyword* memberikan informasi terhadap kata penting berupa *N – gram* yang menghasilkan ringkasan lebih informatif untuk data berita maupun data ulasan.

Pada penelitian ini, selain penambahan fitur untuk meningkatkan hasil ringkasan, proses tokenisasi kalimat juga termasuk hal yang penting. Proses tokenisasi yang hanya menggunakan tanda baca titik (.) menghasilkan tokenisasi yang tidak sempurna atau menghasilkan *noise*. Hal ini dapat menghasilkan kalimat yang tidak memiliki makna. Oleh sebab itu, pada penelitian selanjutnya, mampu melakukan proses tokenisasi kalimat dengan baik. Salah satu cara tokenisasi kalimat dapat menggunakan proses segmentasi. Metode segmentasi kalimat dapat dieksplor menggunakan aspek dan opini untuk mengurangi kalimat yang tidak memiliki makna. Sedangkan pada data berita dapat dilakukan eksplorasi *Name Entity Recognition* (NER) yang dapat mengidentifikasi entitas yang relevan.

DAFTAR PUSTAKA

- [1] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, p. 113679, 2021, doi: 10.1016/j.eswa.2020.113679.
- [2] C. Ma, W. E. Zhang, M. Guo, H. Wang, and Q. Z. Sheng, "Multi-document Summarization via Deep Learning Techniques: A Survey," *ACM Comput. Surv.*, Apr. 2022, doi: 10.1145/3529754.
- [3] Z. Liang, J. Du, and C. Li, "Abstractive social media text summarization using selective reinforced Seq2Seq attention model," *Neurocomputing*, vol. 410, pp. 432–440, Oct. 2020, doi: 10.1016/j.neucom.2020.04.137.
- [4] D. Gunawan, S. H. Harahap, and R. Fadillah Rahmat, "Multi-document Summarization by using TextRank and Maximal Marginal Relevance for Text in Bahasa Indonesia," *Proceeding - 2019 Int. Conf. ICT Smart Soc. Innov. Transform. Toward. Smart Reg. ICISS 2019*, pp. 1–5, 2019, doi: 10.1109/ICISS48059.2019.8969785.
- [5] A. Khan *et al.*, "Sentence Embedding Based Semantic Clustering Approach for Discussion Thread Summarization," *Complexity*, vol. 2020, pp. 1–11, Aug. 2020, doi: 10.1155/2020/4750871.
- [6] H. C. Manh, H. Le Thanh, and T. L. Minh, "Extractive Multi-document Summarization using K-means, Centroid-based Method, MMR, and Sentence

- Position,” in *Proceedings of the Tenth International Symposium on Information and Communication Technology - SoICT 2019*, 2019, pp. 29–35, doi: 10.1145/3368926.3369688.
- [7] J. Lovinger, I. Valova, and C. Clough, “Gist: general integrated summarization of text and reviews,” *Soft Comput.*, vol. 23, no. 5, pp. 1589–1601, Mar. 2019, doi: 10.1007/s00500-017-2882-2.
- [8] S. Lamsiyah, A. El Mahdaouy, B. Espinasse, and S. El Alaoui Ouatik, “An unsupervised method for extractive multi-document summarization based on centroid approach and sentence embeddings,” *Expert Syst. Appl.*, vol. 167, no. September 2020, p. 114152, 2021, doi: 10.1016/j.eswa.2020.114152.
- [9] A. Joshi, E. Fidalgo, E. Alegre, and R. Alaiz-Rodriguez, “RankSum—An unsupervised extractive text summarization based on rank fusion,” *Expert Syst. Appl.*, vol. 200, p. 116846, Aug. 2022, doi: 10.1016/j.eswa.2022.116846.
- [10] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles, “SummCoder: An unsupervised framework for extractive text summarization based on deep auto-encoders,” *Expert Syst. Appl.*, vol. 129, pp. 200–215, Sep. 2019, doi: 10.1016/j.eswa.2019.03.045.
- [11] G. A. M. Mendoza, Y. Ledeneva, and R. A. García-Hernández, “Determining the importance of sentence position for automatic text summarization,” *J. Intell. Fuzzy Syst.*, vol. 39, no. 2, pp. 2421–2431, Aug. 2020, doi: 10.3233/JIFS-179902.
- [12] H. Li, J. Zhu, J. Zhang, C. Zong, and X. He, “Keywords-Guided Abstractive Sentence Summarization,” *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 05, pp. 8196–8203, Apr. 2020, doi: 10.1609/aaai.v34i05.6333.
- [13] D. Patel, S. Shah, and H. Chhinkaniwala, “Fuzzy logic based multi document summarization with improved sentence scoring and redundancy removal technique,” *Expert Syst. Appl.*, vol. 134, pp. 167–177, Nov. 2019, doi: 10.1016/j.eswa.2019.05.045.
- [14] A. P. Widyassari et al., “Review of automatic text summarization techniques & methods,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1029–1046, Apr. 2022, doi: 10.1016/j.jksuci.2020.05.006.
- [15] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, “YAKE! Keyword extraction from single documents using multiple local features,” *Inf. Sci. (Nijl.)*, vol. 509, pp. 257–289, Jan. 2020, doi: 10.1016/j.ins.2019.09.013.
- [16] N. Franciscus, J. Wang, and B. Stantic, “Mining Summary of Short Text with Centroid Similarity Distance,” 2019, pp. 447–461.
- [17] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. M1m, pp. 4171–4186, 2019.
- [18] W. Guan, I. Smetannikov, and M. Tianxing, “Survey on Automatic Text Summarization and Transformer Models Applicability,” in *2020 International Conference on Control, Robotics and Intelligent System*, Oct. 2020, pp. 176–184, doi: 10.1145/3437802.3437832.
- [19] Z. Cai, N. Lin, C. Ma, and S. Jiang, “Indonesian Automatic Text Summarization Based on A New Clustering Method in Sentence Level,” in *Proceedings of the 2019 International Conference on Big Data Engineering*, Jun. 2019, pp. 30–35, doi: 10.1145/3341620.3341626.
- [20] D. Seitkali and R. Musabayev, “Using Centroid Keywords and Word Mover’s Distance for Single Document Extractive Summarization,” in *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval*, Jun. 2019, pp. 149–152, doi: 10.1145/3342827.3342852.
- [21] F. B. Goularte, S. M. Nassar, R. Fileto, and H. Saggion, “A text summarization method based on fuzzy rules and applicable to automated assessment,” *Expert Syst. Appl.*, vol. 115, pp. 264–275, Jan. 2019, doi: 10.1016/j.eswa.2018.07.047.
- [22] S. Gong, Z. Zhu, J. Qi, C. Tong, Q. Lu, and W. Wu, “Improving extractive document summarization with sentence centrality,” *PLoS One*, vol. 17, no. 7, p. e0268278, Jul. 2022, doi: 10.1371/journal.pone.0268278.
- [23] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, “BERTScore: Evaluating Text Generation with BERT,” Apr. 2019, [Online]. Available: <http://arxiv.org/abs/1904.09675>.
- [24] A. Bražinskas, M. Lapata, and I. Titov, “Unsupervised Opinion Summarization as Copycat-Review Generation,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 5151–5169, doi: 10.18653/v1/2020.acl-main.461.
- [25] W. Widodo, M. Nugraheni, and I. P. Sari, “A comparative review of extractive text summarization in Indonesian language,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 3, p. 032041, Mar. 2021, doi: 10.1088/1757-899X/1098/3/032041.
- [26] C. Liu, Y. Che, and R. Duan, “Research and Improvement of TextRank Algorithm Adding Degree Adverbs,” *J. Phys. Conf. Ser.*, vol. 2005, no. 1, p. 012058, Aug. 2021, doi: 10.1088/1742-6596/2005/1/012058.
- [27] S. Lamsiyah, A. El Mahdaouy, S. E. A. Ouatik, and B. Espinasse, “Unsupervised extractive multi-document summarization method based on transfer learning from BERT multi-task fine-tuning,” *J. Inf. Sci.*, vol. 49, no. 1, pp. 164–182, Feb. 2023, doi: 10.1177/0165551521990616.